Curr Genet (1994) 26: 79–86

Current Genetics

Cloning and characterization of the nuclear gene encoding plastid glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracilaria verrucosa*

Yi-Hong Zhou, Mark A. Ragan

Institute for Marine Biosciences, National Research Council of Canada, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1, Canada

Received: 2 December 1993 / 22 December 1993

Abstract. The single-copy nuclear gene (*GapA*), encoding the plastid-localized glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of the marine red alga *Gracilaria verrucosa*, has been cloned and sequenced. The *GapA* transcriptional initiation site was located 49 bp upstream of the start codon, and a putative TATA box was found 54 bp farther upstream. A spliceosomal intron was identified in the transit-peptide-encoding region in a position very similar to intron 1 of *GapA* and *GapB* of higher plants; no introns occur in the region encoding the mature protein. These observations provisionally suggest that both red algae and higher plants descend from a single ancestral photosynthetic eukaryote, i.e. that a single endosymbiotic event gave rise to red algal and higher-plant plastids.

Key words: Glyceraldehyde-3-phosphate dehydrogenase gene – *Gracilaria verrucosa* – Endosymbiotic origin of plastids – Molecular evolution

Introduction

Two kinds of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) isoenzymes exist in investigated algae and higher plants. NAD^+-GAPDH (EC 1.2.1.12) is a cytosolic homotetramer active in glycolysis, while in higher plants the plastid-localized $NADP^+$-GAPDH (EC 1.2.1.13) is heterotetrameric (A_2B_2) and is involved in photosynthetic assimilation of carbon dioxide via the Calvin cycle (Harris and Waters 1976; Cerff 1982). Nuclear genes (*GapA*, *GapB*, *GapC*) encode not only the cytosolic (GAPC) but also the two types of plastid-localized (GAPA, GAPB) subunits. Sequences of *GapA* and *GapB* clearly exhibit a closer relationship to the GAPDH genes of eubacteria (Martin and Cerff 1986; Shih et al. 1986; Brinkmann et al. 1987, 1989; Smith 1989; Doolitle et al. 1990), and more specifically cyanobacteria (Martin et al. 1993), than to the *GapC* genes of even the same organism. This has been interpreted to mean that the ancestor(s) of these genes entered the eukaryotic lineage via one or more cyanobacterium-like endosymbiont(s), with subsequent transfer to the nucleus of the host (Shih et al. 1986; Brinkmann et al. 1987). Genes for the A and B subunits appear to have arisen within the plant lineage by duplication (Brinkmann et al. 1989; Morden et al. 1992) before the diversification of angiosperms (Cerff and Kloppstech 1982; Crane et al. 1989; Li et al. 1989; Martin et al. 1989 a, b) but subsequent to the divergence of red algae (Zhou and Ragan 1993).

It is now widely accepted that plastids are descendants of one or more lineages of free-living photosynthetic prokaryotes that successfully established endosymbioses within eukaryotes (Schimper 1883; Gray 1992). Mereschkowsky (1905) was the first to propose that the prokaryotic endosymbiont was cyanobacterial, a view that has received extensive support from recent investigations of gene arrangements within gene clusters and from phylogenetic analyses of rRNA and other molecular sequences (reviewed by Douglas 1993; Reith and Munholland 1993). Mereschkowsky (1905 page 602; 1910) was also the first to suggest that plastids of red, brown, and green algae arose from different groups of cyanobacteria through separate endosymbioses. Although this multiple-origins view was once widely held (Schiff and Epstein 1965; Sagan 1967; Raven 1970; Taylor 1974), most molecular data now appear to indicate that plastids emerged from a single point within the cyanobacteria (Morden et al. 1992; Douglas 1993), presumably, although not necessarily, in a single endosymbiotic event.

How can we decide between single and multiple origins for plastids? Given the possibilities of lineage-specific effects arising from genetic context (e.g. Liaud et al. 1990), gene conversion and related mechanisms (Scherer and Davis 1980; Dover 1982), and differential nucleotide-substitution rates (Lockhart et al. 1992 a, b), sequence-based phylogenetic analyses may ultimately prove insufficient to resolve the matter, particularly for genes suspected of arising from closely-related cyanobacterial lineages. The discovery and characterization of homologous presence/absence characters may hold more promise.

Correspondence to: M. A. Ragan

Characters present in specific groups of modern plastids and in some, but not all, prokaryotic lineages (e.g. chlorophyll *b* in higher-plant chloroplasts and prochlorophytes: Lewin and Cheng 1989) would argue for multiple origins (Gray and Doolittle 1982; Gray 1992), whereas synapomorphic characters present in different plastid-gene lineages but absent from prokaryotes (e.g. clustering of *rpo/rps2/atp* genes in red algal and higher-plant plastids: Reith and Munholland 1993) would indicate a monophyletic origin from a common photosynthetic eukaryotic ancestor for those organisms.

To our knowledge, however, no unambiguous examples of the latter type of synapomorphy have been described; the *rpo/rps2/atp* data cited above require additional hypotheses of selective gene transfer to the nucleus, while the arrangement of these genes in the *Chlamydomonas reinhardtii* chloroplast genome is anomalous (Reith and Munholland 1993). Clearer examples might most productively be sought among genes which entered the eukaryotic lineage via the genome of the endosymbiont, and were subsequently transferred to the host nucleus and modified, for example, by rearrangement, addition of 5′ or 3′ extensions or intracellular-sorting signal sequences, or the invasion of introns.

Herein we report the cloning and genomic organization of *GapA* from the marine red alga *Gracilaria verrucosa*, and describe the occurrence of a GT-AG (putatively splice-osomal) intron within its transit-peptide-encoding region in a position very similar to that of intron 1 in the GAPDH genes of pea and maize. The well-known lack of sequence conservation among transit peptides (von Heijne et al. 1989; von Heijne and Nishikawa 1991) ultimately precludes an unambiguous decision on whether these red algal and higher-plant introns are homologous; however, similarities in the *GapA* transit peptides and intron positions provisionally suggest that red algal and higher-plant plastids (hence red algae and higher plants) have descended from a common photosynthetic eukaryotic ancestor.

Materials and methods

Cloning, sequencing and characterization of GapA *from* Gracilaria. *G. verrucosa* (Hudson) Papenfuss was collected near Oslo, Norway, and cultured at the IMB Aquaculture Research Station, Sandy Cove, Halifax County, Nova Scotia. Nuclear DNA was isolated as reported (Zhou and Ragan 1993). A genomic library (1.1×10^6 plaques before amplification) was constructed by partially digesting *G. verrucosa* DNA (10 μg) with *Sau*3A I, sizing (15–24 kb), cloning the fragments using the Lambda GEM-11 *Xho*I half-site arms cloning system, and packaging with the Packagene lambda DNA system (Promega). *Escherichia coli* LE392 was used as host strain. An aliquot (3.4×10^5 pfu) was subsequently amplified to 5.0×10^{11} recombinants. A *G. verrucosa* GAPA cDNA fragment from clone G8 (an 0.4-kb *Eco*RI fragment encoding the central part of the mature protein, Val-39 through Glu-170; Zhou and Ragan 1993) was random-prime-labeled with α-^{32}P and used as a probe to isolate genomic clones. Hybridization conditions were the same as reported earlier for screening the *G. verrucosa* cDNA library (Zhou and Ragan 1993). A 3.0-kb *Pst*I/*Xho*I fragment including the entire *GapA* gene sequence was subcloned into pBluescript SK(+), and a 3.1-kb *Eco*RI fragment including the *N*-terminal portion of the gene (through Glu-170) was subcloned into pUC18 (Fig. 1). DNA sequencing was carried out on

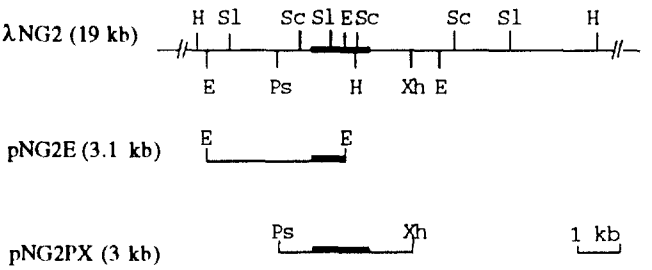

Fig. 1. Restriction map of the *G. verrucosa GapA* genomic clone, and depiction of plasmid subclones. *E*, *Eco*RI; *H*, *Hin*dIII; *Ps*, *Pst*I; *Sl*, *Sal*I; *Sc*, *Sac*I; *Xh*, *Xho*I

double-stranded plasmid templates using synthetic oligonucleotide primers. The transcriptional initiation site was determined by 5′-primer extension (Sambrook et al. 1989) using *G. verrucosa* polyA RNA (Zhou and Ragan 1993), M-MLV reverse transcriptase, and a synthetic 25-bp primer (5′-dGACGAAAGAGCAAGCAGGGAAACTC-3′) complementary to the region immediately upstream of the start codon.

Sequence alignment, data analysis and nomenclature. Deduced transit-peptide sequences were aligned using CLUSTAL V (Higgins et al. 1992) and Multalign (Corpet 1988) with a range of substitution parameters and gap penalties; alignments reported herein were generated with CLUSTAL V using the 250 PAM parameters (Dayhoff et al. 1979), a gap-opening penalty of 15 and a gap-extension penalty of 10, with minor subsequent visual adjustment only at the immediate 5′ end. The definition of conservative amino-acid substitutions follows Dayhoff et al. (1979). Unrooted phylogenetic trees were inferred for the various alignments of transit-peptide sequences using PROTPARS in PHYLIP 3.4 (Felsenstein 1989) and were bootstrapped (100 replicates). Nomenclature of GAPDH proteins (GAPA, GAPB, GAPC), cDNAs and genes (*GapA*, *GapB*, *GapC*) follows the recommendations of the ISPMB Commission on Plant Gene Nomenclature (see Martin et al. 1993).

Results and discussion

Cloning and characterization of GapA *from G. verrucosa*

Three *GapA* clones were isolated from about 1.8×10^5 recombinants in our amplified *G. verrucosa* genomic library. Restriction-endonuclease analyses and Southern hybridizations (data not shown) confirmed that the same gene was present in all three clones, as expected based on our earlier demonstration (Zhou and Ragan 1993) that a single copy exists in the genome of *G. verrucosa*. Clone NG2 was found to contain the complete gene sequence, while the other two clones are truncated at either the 5′ (clone NG1) or 3′ (clone NG5) end.

Comparison of the *G. verrucosa GapA* gene sequence with the corresponding cDNA sequence (Zhou and Ragan 1993) demonstrated that no introns interrupt the region that encodes the mature protein. This was unexpected, as highly conservative and putatively ancient introns occur in the mature-protein-encoding region of both *GapA* and *GapC* in investigated animals and higher plants (Quigley et al. 1988; Liaud et al. 1990). The absence of introns in the mature-peptide-encoding region of *G. verrucosa GapA*, together with the apparent ancestry of its transfer

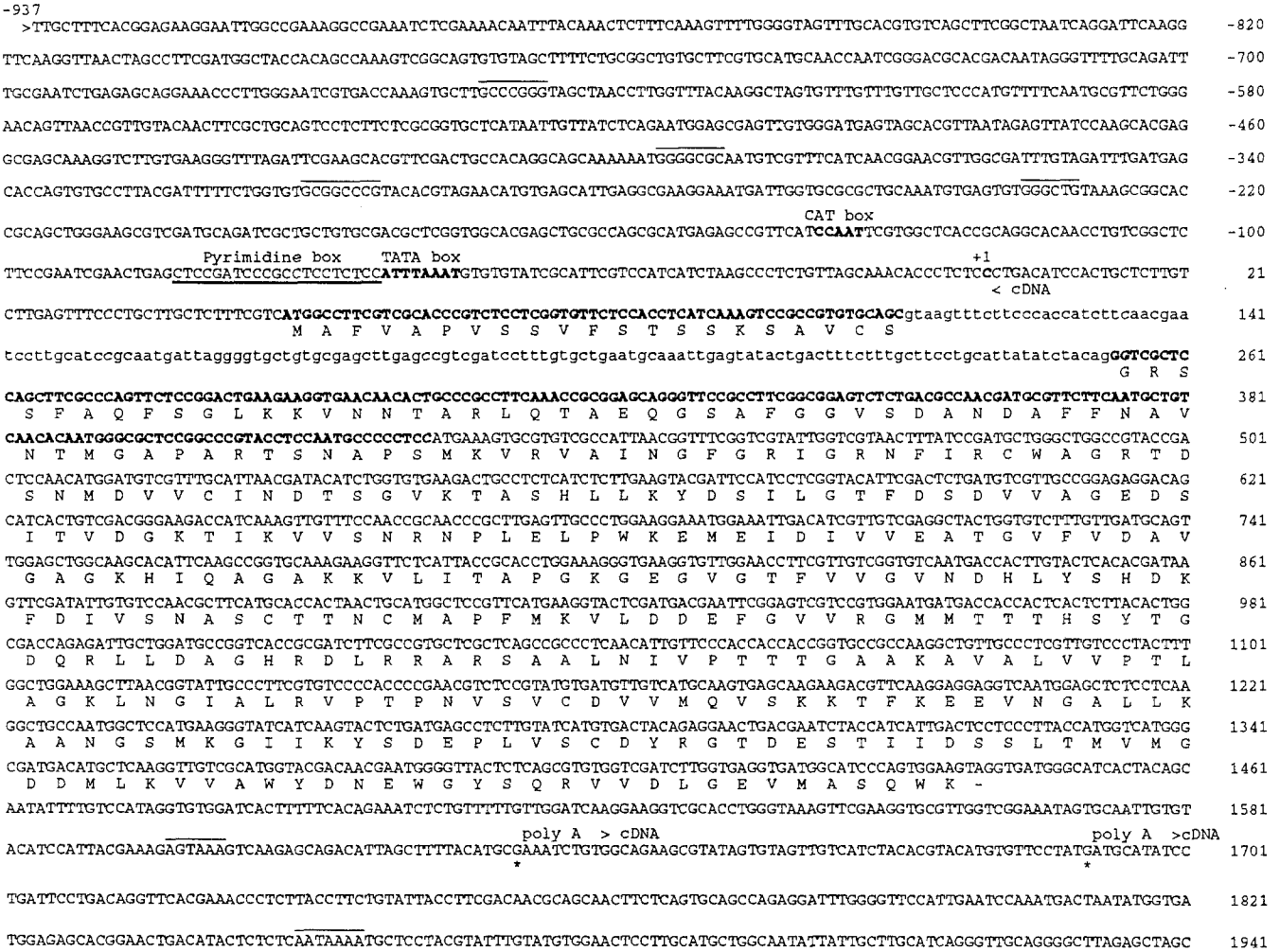

Fig. 2. Nucleotide sequence of the *G. verrucosa GapA* gene with upstream and downstream nontranscribed regions, and the derived GAPA translation product. The transit-peptide-encoding region is shown in *boldface*. Numbering makes reference to the transcription initiation site (+1). Upstream, the putative TATA and CAT elements are shown in *boldface*, the pyrimidine box is *underlined*, and GC motifs are *overlined*. Downstream from the mature protein-encoding region, functional polyadenylation sites are marked with *asterisks*, and the functional polyadenylation signal AGTAAA and the apparently nonfunctional signal region AATAAAA are *overlined*. The nucleotide sequence of *G. verrucosa gapA* has been deposited in GenBank under the accession number L22011

to the nucleus (before the divergence of the red algal and green algal-higher plant lineages: Zhou and Ragan 1993), caution against ready acceptance of the proposal (Liaud et al. 1990) that the degree of intron loss is directly related to the time GAPDH genes have resided in the plastid genome.

A 141-bp intron (Fig. 2, lower case) does, however, occur in the transit-peptide-encoding region of *G. verrucosa GapA*. Like intron 1 in the *GapA* and *GapB* genes of higher plants (Liaud et al. 1990), it is of the GT-AG (putatively spliceosomal) type, with neighboring nucleotides conforming (Table 1) to higher-plant consensus splice-junction sequences (Brown et al. 1986).

Promoter structure and 3′-nontranscribed region of G. verrucosa GapA

The major transcription initiation site (presumptive cap site) was located 49 bp upstream of the start codon by primer extension (data not shown). The initiator element, CTCTCCCTGA (the initiation site is shown in boldface), differs from those known from mammals (PyPyC**A**PyPyPyPyPy: Corden et al. 1980), *Drosophila* (PyATCA**GT**TPy: Hultmark et al. 1986), and *Saccharomyces cerevisiae* (PuPu**Py**PuPu: Hahn et al. 1985). As *G. verrucosa GapA* is the first protein-coding sequence published from a red algal nuclear genomic clone, we cannot propose a consensus initiator element at this time.

To characterize the *GapA* promoter further, the noncoding region extending 937 bp upstream of the initiation site was sequenced. The only significant uninterrupted AT element in this region consists of eight nucleotides, ATTTAAAT, and is located –61 to –54 bp from the transcription initiation site (Fig. 2). Although somewhat farther upstream than expected (Martinez et al. 1989), this is likely to be the RNA polymerase-binding region, similar to the canonical TATA box, as DNA double-strand unwinding would be facilitated in this region; the ATCTAA motif at –26 bp is a less appealing candidate. Immediately up-

Table 1. Exon-intron splice junctions of plants and *G. verrucosa*

A. Comparison of nucleotide sequences at the 5′ exon – intron splice junction of plants and *G. verrucosa*

Position[a]		−3	−2	−1	:	+1	+2	+3	+4	+5	+6
	A	33.9	55.4	10.1		0	0	69.0	55.4	16.7	23.2
168 plant	G	19.1	10.7	72.6		100	0	13.7	6.0	65.5	10.1
introns (%N)	C	33.3	10.7	10.7		0	0.6	7.8	19.0	7.7	17.9
	T[b]	13.7	23.2	6.6		0	99.4	9.5	19.6	10.1	48.8
Plant consensus[c]		A/C	A	G		G	T	A	A	G	T
G. verrucosa *GapA* intron		A	G	C		G	T	A	A	G	T

B. Comparison of nucleotide sequences at the 3′ exon – intron splice junction of plants and *G. verrucosa*

Position[a]		−15	−14	−13	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	:	1
	A	19.2	17.3	12.0	29.3	20.3	20.5	20.3	24.7	19.8	25.7	12.0	19.8	4.8	100	0		13.8
167 plant	G	11.4	−15.0	−15.0	16.2	14.4	19.9	19.2	21.7	23.9	18.6	10.8	49.1	1.8	0	100		61.6
introns (%N)	C	22.1	15.0	20.3	11.4	13.2	12.6	18.6	15.1	15.0	20.0	9.6	13.8	65.9	0	0		13.8
	T[b]	47.3	52.7	52.7	43.1	52.1	47.0	41.9	38.5	41.3	43.7	67.6	17.3	27.5	0	0		10.8
Plant consensus[c]		T	T	T	T/Pu[d]	T	T	T/Pu	T/Pu	T/Pu	T/Pu	T	G	C	A	G		G
G. verrucosa *GapA* intron		G	C	A	T	T	A	T	A	T	C	T	A	C	A	G		G

[a] Position are numbered from the splice site (:)
[b] Sequences are given in DNA form
[c] Plant consensus from analysis of 168 plant intron sequences (Brown et al. 1986)
[d] At positions designated T/Pu, T is the most abundant nucleotide but the combined %G and %A is equivalent to or greater than the %T

stream from the putative TATA box is a 21-base pyrimidine-rich (86% C+T) region similar to pyrimidine box I characterized in the maize *GapC*, maize alcohol dehydrogenase (*Adh*), and pea *Adh* genes (Martinez et al. 1989). A CCAAT element occurs at −133 to −137 bp from the initiation site, and GC-rich regions, six to eight bp in length (Dynan and Tjian 1985), extend upstream from positions −231, −302, −387, and −645 (Fig. 2).

We reported earlier (Zhou and Ragan 1993) that *G. verrucosa* GAPA cDNAs exhibit two polyadenylation sites. For both sites (a total of seven cDNA clones), the only likely polyadenylation signal is the AGTAAA at +1 598 to +1 603 (Fig. 2). Sequencing the genomic clone revealed an AATAAAA motif (+1 851 to +1 857), containing the canonical polyadenylation signal, located downstream from all characterized polyA addition sites. We have no evidence that this AATAAAA motif functions in polyadenylation. Efficient expression of rabbit β-globin mRNAs requires GU- and U-rich elements (Gil and Proudfoot 1987) 3′ to the primary polyadenylation signal; no such elements occur in the 84 bp downstream from the *G. verrucosa GapA* AATAAAA (Fig. 2).

Origin of transit peptides and common ancestry of plastids in red algae and higher plants

Martin et al. (1993) recently provided evidence that nuclear-encoded chloroplast-localized GAPDHs of higher plants arose from endosymbiotic cyanobacteria. Our own phylogenetic analyses (Zhou and Ragan 1993) demonstrate a specific common ancestry for the gene regions encoding mature plastid-localized GAPDHs in *G. verrucosa* and higher plants. These results are consistent with two alternative histories: either red algae and higher plants have descended from a common photosynthetic eukaryotic ancestor, or red algal and higher-plant plastids arose via independent endosymbioses of closely related cyanobacteria.

The intron in the *G. verrucosa GapA* gene (Fig. 2) occurs at a position very similar to that of intron 1 in transit sequences of the corresponding nuclear genes in pea (Liaud et al. 1990) and maize (Quigley et al. 1988) (Fig. 3 and below). This observation could help us decide between the two above alternative histories if the evolutionary origin of chloroplast transit peptides can be identified, and if the homology of red algal and higher-plant *GapA* transit peptides can be established.

After reviewing the structures of nearly 100 chloroplast transit peptides, Keegstra et al. (1989) concluded that "...sequence similarities generally exist among transit peptides of the same precursor derived from different plant species. In contrast, few similarities are found among different precursors, even when the precursors are derived from the same plant species." This presumably means that, at least for some genes, transit peptide-encoding regions can be stably associated with a gene lineage over evolutionarily long periods of time. For *GapA/GapB*, Liaud et al. (1990) date these associations back to the progenitor of higher-plant chloroplasts.

Liaud et al. (1990) assumed that this progenitor was a eubacterium, and concluded that both the GAPA/GAPB transit peptide and the intron in the corresponding DNA are thus of eubacterial origin. However, no known eubacterial GAPDH gene (Branlant and Branlant 1985; Conway et al. 1987; Branlant et al. 1989), including those of the cyanobacterium *Anabaena variabilis* ATCC 29 413 (Martin et al. 1993), exhibits an upstream extension suggestive of being an ancestral transit-peptide-encoding region. The same is true of a cyanobacterial gene encoding the small

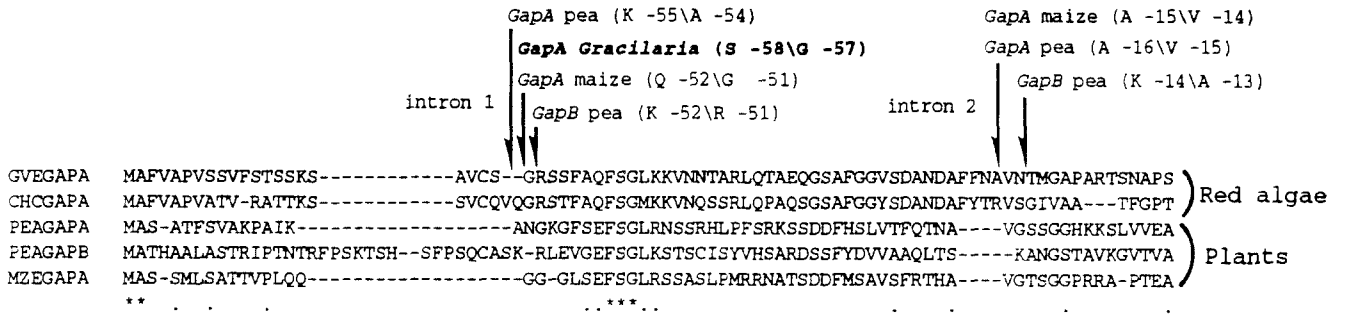

Fig. 3. Alignment (CLUSTAL V: Higgins et al. 1992) of selected transit peptides: GVEGAPA, *G. verrucosa* GAPA; CHCGAPA, *Chondrus crispus* GAPA; PEAGAPA, pea GAPA; PEAGAPB, pea GAPB; MZEGAPA, maize GAPA (see text for references). Positions of introns 1 and 2 in the corresponding genes are shown above. Conserved positions are marked below the sequences: *asterisk*, identity; *dot*, conservative change (Dayhoff et al. 1979)

subunit of chloroplastic ribulose-1,5-bisphosphate carboxylase, which in green algae and higher plants is nuclear-encoded and contains an amino-terminal transit peptide (Nierzwicki-Bauer et al. 1984). Nor is any modern eubacterium known to contain spliceosomal introns. Given these observations, it seems more defensible to hypothesize that transit-peptide-encoding regions were instead recruited from the genome of the eukaryotic host at an early stage in the establishment of the endosymbiosis(es) which led to plastids, became covalently incorporated into individual genes by exon shuffling (Rogers 1985), and thereafter were stabilized and retained by selection. An analogous scenario has been proposed for mitochondrially targeted proteins (Cavalier-Smith 1987). If this scenario is valid for GAPDH genes, demonstration of homology between the transit-peptide-encoding regions of plastid GAPDHs from *G. verrucosa* and higher plants would make it highly likely that red algae and higher plants descend from a common photosynthetic eukaryotic ancestor.

Homology of genes is typically inferred by demonstrating that sequences of the exons are significantly more similar than could be expected from convergent evolution or chance; the identical positioning of introns increases the confidence we can place in such inferences. Application of both of these criteria requires sequences to be aligned relatively unambiguously. Unfortunately, despite the generalization of Keegstra et al. (1989) cited above, no strongly conserved sequence features can be identified even within a single type of transit peptide (von Heijne et al. 1989; von Heijne and Nishikawa 1991). For GAPDH transit peptides, sequence conservation is poor not only between red algae and higher plants (Zhou and Ragan 1993) but even within florideophycidean red algae. For *GapA* from *G. verrucosa* (Zhou and Ragan 1993) and *Chondrus crispus* (Liaud et al. unpublished; GenBank accession X73035), optimal pairwise alignment reveals 92.0% and 80.7% identities at the amino-acid and nucleotide levels respectively for the mature-protein regions, but only 56.2% and 62.0% respectively for the transit-peptide-encoding regions, indicating that many of the mutations accepted in the former region are silent whereas those in the latter regions often result in amino-acid substitutions. With selective pressures relaxed to this extent even within florideophytes, unambiguous alignment of red algal with plant GAPDH transit peptides is impossible.

Indeed, simple comparisons based on percent similarity do not reveal any specific relationship between red algal and higher-plant GAPDH transit peptides; by such measures, red algal GAPDH transit peptides are not significantly more similar to higher-plant GAPDH transit peptides than to a higher-plant consensus transit-peptide sequence for the small subunit of ribulose 1,5-bisphosphate carboxylase (ss-RuBisCO; Table 2), which also contains a (possibly convergent) (A/S/G)(Q/E)FSG(L/M)(K/R) motif. These identity values fall within the "twilight zone" (Doolittle 1987) where relationships among even longer peptide sequences are typically difficult to prove.

Sequence comparisons based on overall similarity do not, however, adequately take into account character change at individual positions, and it might be thought that position-by-position analysis (e.g., by protein parsimony) could provide adequate evidence of homology. Bootstrapped parsimony analyses based on various alignments (Fig. 3 and others not shown; see Materials and methods) do in fact group red algal GAPDH and higher-plant GAPDH (especially GAPB) transit peptides (data not shown). However, problems arise in attempting to root such trees. Of known transit peptides, only those of ss-RuBisCO are alignable with those of GAPDH; all others (Keegstra et al. 1987) are unalignably divergent from these in both length and sequence. In principle, it would be possible to root such trees on gene duplication (Iwabe et al. 1989), but the red algal ss-RuBisCO gene is plastid-localized (Reith and Munholland 1993) and thus has no transit peptide. For all of these reasons, we cannot adequately demonstrate homology between red algal and higher-plant GAPDH transit peptides, or align them with confidence.

These uncertainties necessarily preclude definitive conclusions about the possible homology between intron 1 in *GapA/GapB* transit-peptide-encoding regions of pea and maize, on one hand, and the similarly positioned intron of *G. verrucosa*, on the other hand. However, four lines of evidence argue that these introns may be homologous; these arguments and the corresponding counterarguments are presented below.

First, these are phase-zero GT-AG introns, as expected in eukaryotes for proteins which have arisen by exon assembly (Rogers 1985). Phase class is the most conservative feature retained by homologous introns (Patthy 1987). As 43.5–53% of nuclear pre-mRNA introns in higher or-

Table 2. Comparisons among transit peptides of GAPDHs and ribulose-1,5-bisphosphate carboxylase small subunits

Comparison	Identical	Identical or similar	Non-conservative	Gapped	Nonconservative plus gapped
Within red algae	57%	81%	14%	5%	19%
Within plant GAPA	30–48	49–63	24–33	6–23	37–51
Within plant GAPB	49–51	77–78	12–13	11	22–23
Plant GAPA vs GAPB	14–23	35–43	23–38	20–38	57–65
Gracilaria to plant GAPA	10–18	27–41	38–44	17–35	59–73
Gracilaria to plant GAPB	14–21	35–42	36–43	22–23	58–65
Plant GAPA to ssRuBPCase	11–28	31–51	26–33	23–43	49–69
Plant GAPB to ssRuBPCase	16–17	35–39	22–29	35–39	61–65
Gracilaria to ssRuBPCase	11–12	28–29	35–37	35	71–72

ganisms occur in phase zero (Finchant 1992), there is about one chance in four that unrelated red algal and higher-plant introns would occur in this phase.

Second, protein alignments (including Fig. 3) inferred using various gap-weighting criteria consisently show these introns to be positioned no more than six nucleotides apart, a disparity elsewhere observed between intron 2 in pea *GapA* and its obviously homologous counterpart in pea *GapB*, and readily attributable to misalignment, lineage-specific insertions or deletions in the poorly conserved transit-peptide-encoding regions, and/or intron sliding (Craik et al. 1983; Patthy 1987). Although sequence divergence precludes confident alignment along much of the length of the transit peptide, the *Gracilaria* and higher-plant *GapA* introns are positioned only four to six codons from the (A/S/G)(Q/E)FSG(L/M)(K/R)-encoding motif, the most conservative in the entire transit peptide. Similar positioning might possibly occur, however, even if the red algal and higher-plant introns had separate origins (Wistow 1993).

Third, preliminary evidence suggests that intron 1 in the transit-peptide-encoding regions of maize *GapA* may be involved in expression of the gene (W. Martin, personal communication), making it conceivable that this intron has been specially conserved by selection. Corresponding experiments have not been conducted with the *Gracilaria GapA* intron.

Fourth, the poor primary-structural conservation (above) and apparent selection for random-coil secondary structure in the plastid transit-peptide (von Heijne and Nishikawa 1991; Zhou and Ragan 1993) render intron targeting (Dujon et al. 1986) unappealing as an alternative explanation for the similar positions of these introns in evolutionarily diverse lineages. But, because this is the first nuclear intron characterized from red algae, we may not be able to recognize proto-splice sites (Dibb and Newman 1989) in red algal nuclear DNA.

Thus there is circumstantial (but not definitive) evidence, from *GapA(/B)* genes, that both the plastid and nuclear lineages of red algae and higher plants arose from the corresponding genomes of a single photosynthetic ancestral eukaryote. Such a conclusion would be consonant with nuclear 18*s* rRNA gene trees (Bhattacharya et al. 1990; Douglas et al. 1991; Bird et al. 1991; Hendriks et al. 1991; Ragan et al. 1994 (in press) that show red algae diverging just before plants, animals and fungi. There is increasing evidence, both ultrastructural and molecular, that plastids of cryptophytes and chromophytes arose via secondary endosymbioses involving red alga-like eukaryotes (Douglas and Turner 1991; Douglas 1993), while the precursor to euglenoid chloroplasts is increasingly thought to have been a eukaryotic green alga (Gibbs 1978; Lefort-Tran 1981; Ragan and Lee 1992). If red and green plastids share a single origin, a single cyanobacterial endosymbiosis may have given rise to plastids in most or all lineages of photosynthetic eukaryotes (Palmer 1993).

Acknowledgements. We thank W. Martin for helpful discussions and for communicating results prior to publication, and M. E. Reith and an anonymous reviewer for critical review of the manuscript. Issued as NRCC no. 34914.

References

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

Bird CJ, Murphy CA, Rice EL, Gutell RR, Ragan MA (1991) Towards an rRNA gene phylogeny of the red algae. J Phycol 27 [suppl]: 9

Branlant C, Oster T, Branlant G (1989) Nucleotide sequence determination of the DNA region coding for *Bacillus stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase and of the flanking DNA regions required for its expression in *Escherichia coli*. Gene 75: 145–155

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

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

Brown JWS, Feix G, Frendewey D (1986) Accurate in vitro splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. EMBO J 5: 2749–2758

Cavalier-Smith T (1987) Intron phylogeny: a new hypothesis. Ann NY Acad Sci 503: 55–71

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

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

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

Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P (1980) Promoter sequences of eukaryotic protein-coding genes. Science 209: 1406–1414

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16: 10881–10890

Craik CS, Rutter WJ, Fletterick R (1983) Splice junctions: association with variation in protein structure. Science 220: 1125–1129

Crane PR, Donoghue MJ, Doyle JA, Friis EM (1989) Angiosperm origins. Nature 342: 131

Dayhoff MO, Schwartz RM, Orcutt BC (1979) In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. Natl Biomed Res Fdn, Silver Spring Md. pp 345–352

Dibb NJ, Newman AJ (1989) Evidence that introns arose at proto-splice sites. EMBO J 8: 2015–2021

Doolittle RF (1987) Of URFs and ORFs. A primer on how to analyze derived amino-acid sequences. University Science Books, Mill Valley, California

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

Douglas S (1993) Chloroplast origins and evolution. In: Bryant DA (ed) The molecular biology of the cyanobacteria. Kluwer, Dordrecht (in press)

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

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

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299: 111–116

Dujon B, Colleaux L, Jacquier A, Michel F, Monteilhet C (1986) Mitochondrial introns as mobile genetic elements: the role of intron-ncoded proteins. In: Wickner RB (ed) Extrachromosomal elements in lower eukaryotes. Plenum Press, New York, pp 5–27

Dynan WS, Tjian R (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316: 774–778

Felsenstein J (1989) PHYLIP – phylogeny inference package (version 3.2).Cladistics 5: 164–166

Finchant GA (1992) Constraints acting on the exon positions of the splice-site sequences and local amino-acid composition of the protein. Hum Mol Genet 1: 259–267

Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can J Bot 56: 2883–2889

Gil A, Proudfoot NJ (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3′ end formation. Cell 49: 399–406

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

Gray MW, Doolittle WF (1982) Has the endosymbiont hypothesis been proven? Microbiol Rev 46: 1–42

Hahn S, Hoar ET, Gurennte L (1985) Each of three "TATA elements" specifies a subset of the transcription initiation sites at the *CYC-1* promoter of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 82: 8562–8566

Harris JI, Waters M (1976) Glyceraldehyde-3-phosphate dehydrogenase. In: Boyer PD (ed) The enzymes, 3rd edn, vol 13. Academic Press, New York, pp 1–49

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

Heijne G von, Steppuhn J, Herrmann RG (1989) Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem 180: 535–545

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

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

Hultmark D, Klemenz R, Gehring WJ (1986) Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22*, Cell 44: 429–438

Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86: 9355–9359

Keegstra K, Olsen LJ, Theg SM (1989) Chloroplastic precursors and their transport across the envelope membranes. Annu Rev Plant Physiol Plant Mol Biol 40: 471–501

Lefort-Tran M (1981) The triple layered organization of the *Euglena* chloroplast envelope (significance and function). Ber Dt Bot Ges 94: 463–476

Lewin RA, Cheng L (1989) *Prochloron*, a microbial enigma. Chapman and Hall, New York

Li W-H, Gouy M, Wolfe KH, Sharp PM (1989) Angiosperm origins. Nature 342: 131–132

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

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD (1992 a) Substitutional bias confounds inference of cyanelle origins from sequence data. J Mol Evol 34: 153–162

Lockhart PJ, Penny D, Hendy MD, Howe CJ, Beanland TJ, Larkum AWD (1992 b) Controversy on chloroplast origins. FEBS Lett 301: 127–131

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 a) Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339: 46–48

Martin W, Gierl A, Saedler H (1989 b) Angiosperm origins (reply to Crane et al. and to Li et al.). Nature 342: 132

Martin W, Brinkmann H, Savona C, Cerff R (1993) 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

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

Mereschowsky C (1905) Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol Centralbl 25: 593–604, 689–691

Mereschowsky C (1910) Theorie der zwei Plamaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. Biol Centralbl 30: 278–288, 289–303, 321–347, 354–367

Morden CW, Delwiche CF, Kuhsel M, Palmer JD (1992) Gene phylogenies and the endosymbiotic origin of plastids. BioSystems 28: 75–90

Nierzwicki-Bauer SA, Curtis SE, Haselkorn R (1984) Cotranscription of genes encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase in the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 81: 5961–5965

Palmer JD (1993) A genetic rainbow of plastids. Nature 364: 762–763

Patthy L (1987) Intron-dependent evolution: preferred types of exons and introns. FEBS Lett 214: 1–7

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

Ragan MA, Lee AR III (1992) Making phylogenetic sense of biochemical and morphological diversity among the protists. In: Dudley EC (ed) The unity of evolutionary biology. Dioscorides Press, Portland, vol I, pp 432–441

Ragan MA, Bird CJ, Rice EL, Gutell RR, Murphy CA, Singh RK (1994) A molecular phylogeny of the marine red algae (Rhodophyta) based on the nuclear small-subunit rRNA gene. Proc Natl Acad Sci USA 91: in press.

Raven PH (1970) A multiple origin for plastids and mitochondria. Science 169: 641–646

Reith M, Munholland J (1993) A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. Plant Cell 5: 465–475

Rogers J (1985) Exon shuffling and intron insertion in serine protease genes. Nature 315: 458–459

Sagan L (1967) On the origin of mitosing cell. J Theor Biol 14: 225–274

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, vol 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Scherer S, Davis RW (1980) Recombination of dispersed repeated DNA sequences in yeast. Science 209: 1380–1384

Schiff JA, Epstein HT (1965) The continuity of the chloroplast in *Euglena*. In: Locke M (ed) Reproduction: molecular, subcellular and cellular. Academic Press, New York, pp 131–189

Schimper AFW (1883) Über die Entwickelung der Chlorophyllkörner und der Farbkörper. Bot Ztg 41: 105–114

Shih M-C, 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

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

Taylor FJR (1974) Implications and extensions of the serial endosymbiosis theory for the origin of eukaryotes. Taxon 23: 229–258

Wistow G (1993) Protein structure and introns. Nature 364: 107–108

Zhou Y-H, Ragan MA (1993) cDNA cloning and characterization of the nuclear gene encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracilaria verrucosa*. Curr Genet 23: 483–489

Communicated by R. W. Lee