Mol Gen Genet (1995) 246:128–131

SHORT COMMUNICATION

Uwe-G. Maier · Stefan A. Rensing · Gabor L. Igloi
Martina Maerz

Twintrons are not unique to the *Euglena* chloroplast genome: structure and evolution of a plastome *cpn60* gene from a cryptomonad

Received: 15 February 1994 / Accepted: 30 June 1994

Abstract Introns within introns (twintrons) are known only from the *Euglena* chloroplast genome. Twintrons are group II or III introns, into which another group II or III intron has been transposed. In this paper we describe a non-*Euglena* twintron structure within a plastid-encoded chaperone gene (*cpn60*) of the cryptomonad alga *Pyrenomonas salina*. In addition, the evolutionary relationships between members of the Cpn60 protein family are determined. Our findings permit the inclusion of cryptomonad plastomes in phylogenetic studies of intron evolution and present further evidence for the origin of modern plastids from a cyanobacterial ancestor.

Key words Cryptomonads · Intron · Twintron · Cpn60

Introduction

Cryptomonads are unicellular algae whose plastids are surrounded by four membranes (see Gibbs 1990). Plastids of this type originated from a eukaryotic endosymbiont ("second hand", McFadden 1993; Douglas et al. 1991; Maier et al. 1991; Gray 1993). Phylogenetic analyses using rRNA sequences, as well as comparisons of the gene order and gene content of red algal and cryptomonad plastid genomes have demonstrated that red algae and the endocytobiont of cryptomonads have a common ancestor (e.g. Valentin et al. 1993; Douglas 1992). With a single exception (in a red alga, Bernard et al. 1992), both types of plastomes are thought to be intron-free (Douglas 1992; Reith and Munholland 1993). In order to obtain more information about the similarities between red algae and cryptomonads, we searched for plastome *cpn60* genes in the cryptomonad *Pyrenomonas salina*. By hybridization analysis we localized two different *cpn60* genes in the plastid genome. The cloning and sequencing of one of these genes led to the identification of the first split gene in a cryptomonad plastid genome. In this gene, two exons are separated by a 1510-nucleotide intervening sequence.

Communicated by R. G. Herrmann

U.-G. Maier (✉) · S. A. Rensing · M Maerz
Institute of Biology II, Cell Biology; Schaenzlestr.l, 79104 Freiburg, Germany

G. L. Igloi
Institute of Biology III, Schaenzlestr. 1, 79104 Freiburg, Germany

This paper is dedicated to Prof. Dr. Peter Sitte on the occasion of his 65th birthday

Materials and methods

Cloning, sequencing

Whole-cell DNA was fractionated by CsCl/Hoechst 33258 gradients as described earlier (Maerz et al. 1992). Plastid DNA was restricted with *Sac*I and cloned into pBluescript. Manual sequencing was done according to Sambrook et al. (1989), while automated sequencing was done according to Igloi and Schiefermayr (1993).

Gene identification

The initial analysis of the sequence information was carried out by a database search using the BLAST algorithm (Altschul et al. 1990). Detailed studies were done with the program packages PC/GENE (version 6.60, IntelliGenetics) and HUSAR (Heidelberg Unix Sequence Analysis Resource, a service at the German Cancer Research Center). The consensus sequence was calculated using the program CONSENSE V1.1 (available from S. A. Rensing).

RNA secondary structure

Initial calculation of the RNA secondary structure was done with the GCG program FOLDRNA (Genetics Computer Group 1991). Manual improvements were performed later by comparison with existing secondary structure models (Michel et al. 1989).

Phylogenetic analysis. Cpn60 sequences were retrieved from the SWISS-Prot release 27 (Bairoch 1993) by using the semantic information retrieval system IRX (available from HUSAR). A subset of 30 mitochondrial, plastidal and prokaryotic sequences was selected for phylogenetic analysis. Short partial sequences were omitted. Sequence positions which were reported as ambiguous in the database were substituted by an X. The transit signal sequences of the nuclear-encoded organellar proteins were identified on the basis of the sequence reports and removed. The amino acid sequence of the thermophilic factor 55 from *Sulfolobus shibatae* (TF55), which is a member of the TCP (T-complex protein) chaperone family, was selected as the outgroup. This sequence shares about 40% amino acid similarity with the Cpn60 sequences, whereas the Cpn60 sequences share about 60–80% (up to 100% in the case of closely related species) similar positions among themselves.

The alignment was calculated using CLUSTAL V (Higgins et al. 1992) with a PAM250 matrix. After manual improvement, the alignment served as the data set for the reconstruction of the phylogenetic tree. Tree construction was performed using PHYLIP 3.5c [available from ftp-site evolution.genetics.washington.edu; see Felsenstein (1989) for version 3.2]. Two different methods (PAM-based distance-matrix calculation/neighbour-joining and protein parsimony) were used as well as bootstrap resampling (Felenstein 1985) to check the reliability of the topologies found.

Results and discussion

Intron structure in a cryptomonad *cpn60* gene

The intron in the *cpn60* plastome gene from the cryptomonad *Pyrenomonas salina* is 1510 nucleotides long and has the conserved boundaires of a group II intron, 5′-GUGUG and AC, respectively (Fig. 1). Furthermore, as shown in Fig. 2, the intervening sequence can be folded into a six-helix domain radiating from a central core and harboring several highly conserved regions characteristic of group II introns (e.g. the base composition of domains V and VI; Michel et al. 1989), which are important for correct splicing (Jaquier 1990). However, in several respects the cryptomonad *cpn60* intron structure seems to diverge from that generally found for group II sequences. Several of the criteria used to divide group II introns into subgroups IIA and IIB (Michel et al. 1989) are missing. For example, the exon binding site, which is part of a possible stem structure, is not localized in the D3 terminal loop (Michel et al. 1989). Furthermore, domain II, which is flanked by a polyA stretch, cannot be stably folded. However, expecially in the case of domain II, an alternative secondary structure is possible (see below), suggesting that a novel processing mechanism may be employed, which remains to be determined. Domain IV has unique features; translation of the nucleotide sequence of domain IV indicates the presence of an open reading frame (ORF439) with a striking similarity to a recently published ORF from a cyanobacterial group II intron (Ferat and Michel 1993). This cryptomonad maturase-like ORF439 consists of a reverse transcriptase-like domain and the so called domain X (Mohr et al. 1993). A zinc-finger domain, which is found in some ORFs of other group II introns, is missing. As shown in Fig. 1, the reverse transcriptase-like domain is homologous to others, whereas domain X is, in part, similar to the corresponding domain of some *Euglena* ORFs.

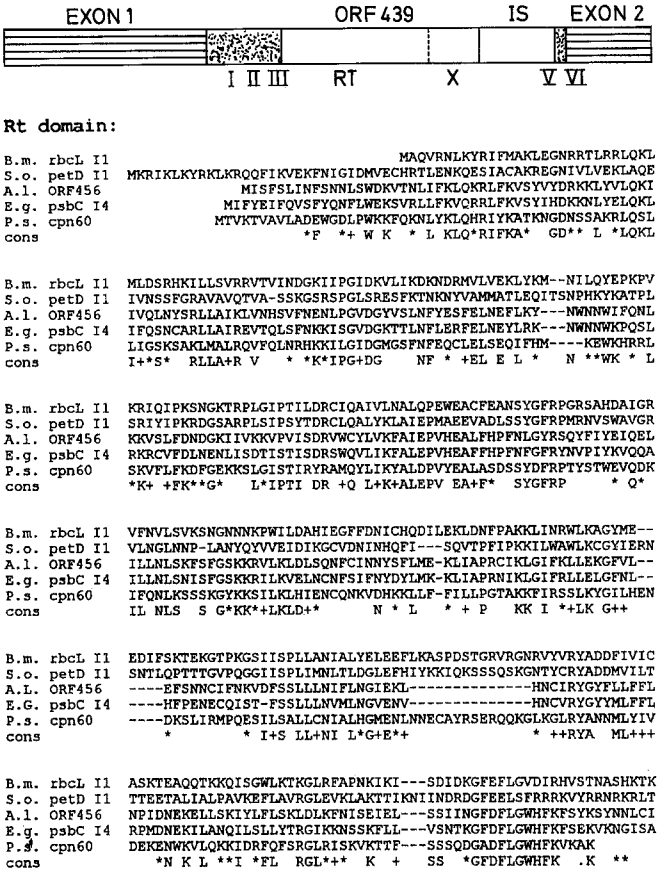


Fig. 1 Physical map of a plastome-encoded *cpn60* gene of *Pyrenomonas salina* and alignment of subdomains of ORF439. The exon-intron structure of *cpn60* is shown (*upper panel*). The intron comprises subdomains I–VI of a group II intron. Subdomain IV (not marked in the Figure) contains an ORF coding for 439 amino acids (counting the homologous methionine as the first amino acid) as well as an insertion (IS) of 502 nucleotides. *Bar* represents 200 bp. In the lower part an alignment is shown of the reverse transcriptase-like domain of ORF439. Symbols in the consensus sequence are given if a minimum of 60% of the sequences are similar at the respective positions. *Asterisk*, hydrophilic amino acid (S, R, H, G, K, Q, N, D, E); +, hydrophobic amino acid (C, F, Y, I, L, M, W, V). B.m., *Bryopsis maxima*; S.o., *Scenedesmus obliquus*; A.l., *Astasia longa*; E.g., *Euglena gracilis*; P.s., *Pyrenomonas salina*.

The most surprising feature of the *cpn60* intron is an insertion of 502 nucleotides in domain IV, upstream of ORF439; its putative secondary structure is shown in Fig. 3. This insertion cannot be classified as a typical group I, II, or III intron, but the putative secondary structure has a striking similarity to that of domain 1 of group II introns. This suggests that the transposition of a group II intron into another group II intron (Lambowitz and Belfort 1993) gave rise to an incomplete insertion of the internal intron. The structure of

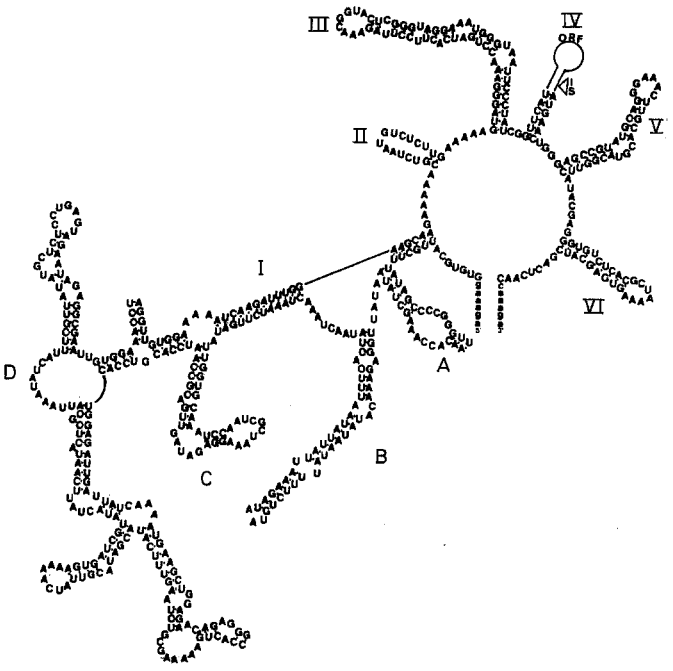


Fig. 2 Possibe secondary structure of the *cpn60* intron of *Pyrenomonas salina*. IS, insertion; ORF, open reading frame

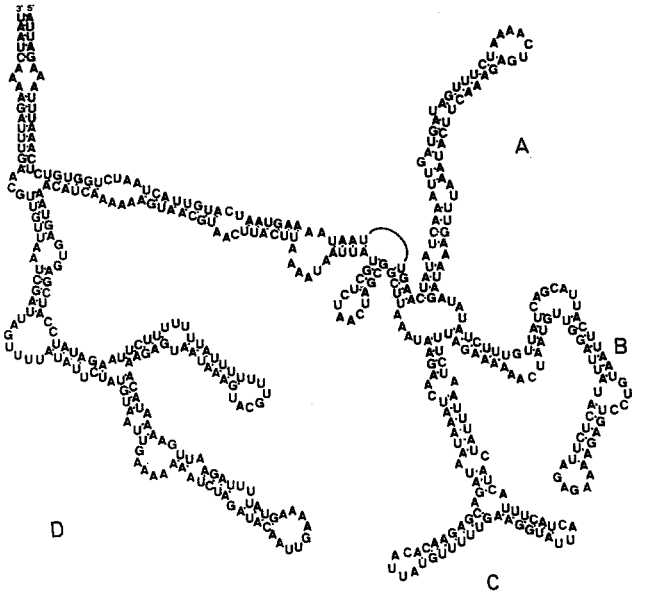
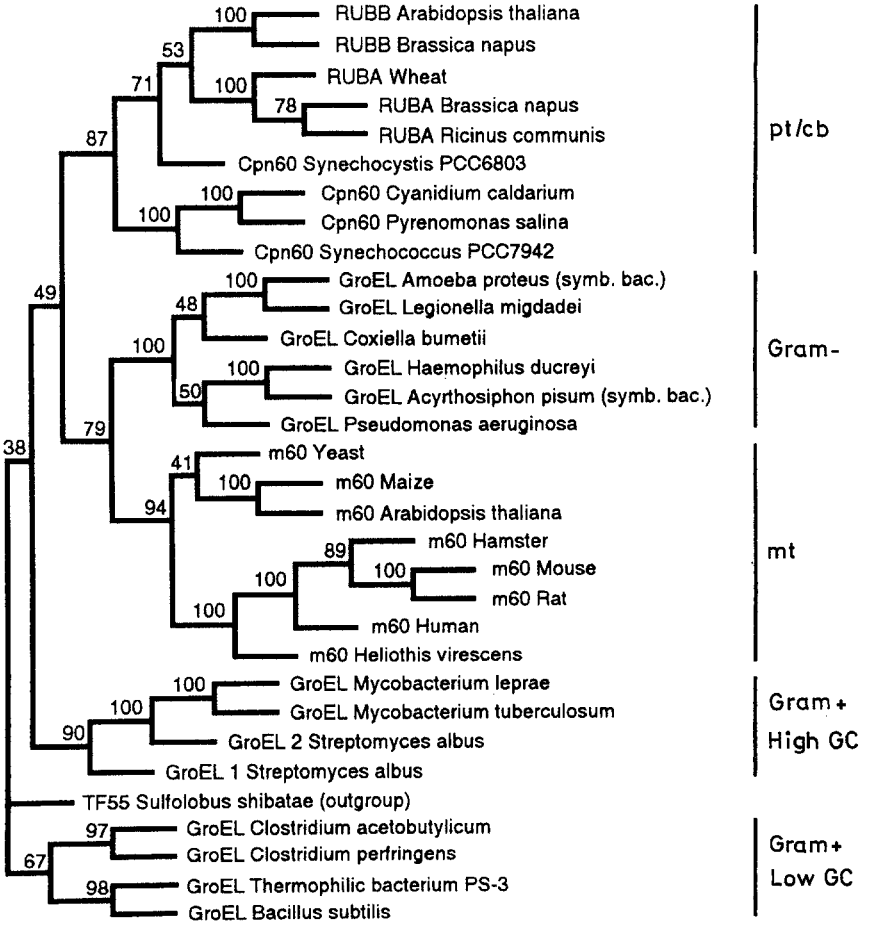


Fig. 3 Possible secondary structure of an insertion in domain IV of the *cpm60* intron.

Fig. 4 Consensus phylogenetic tree of the Cpn60 protein family, constructed by the neighbour-joining method. The numbers at the nodes show how often the group to the right was found in 100 bootstrap samples and therefore reflect the stability of the tree topology. The tree was rooted with the TF55 sequence from *Sulfolobus shibatae*. The topologies of two trees, calculated by different methods (see Materials and methods) were virtually identical. The same sequences clustered together, only minor differences of the cluster order were observed. pt, plastid; cb, cyanobacteria; mt, mitochondrion; RUBA, RUBISCO small subunit binding protein alpha subunit; RUBB, RUBISCO small subunit binding protein beta subunit; m60, mitochondrial Hsp60



the insertion has the sequence 5′-AAGAGA in the external loop of subdomain B, which is complementary to the sequence 5′-UCUCUU in subdomain II of the group II intron, suggesting base pairing between the two introns and possible stabilization of the subdomain II structure.

Phylogenetic implications

The overall structure of the *cpn60* intron is in accordance with the existence of an intron within an intron. Our results thus demonstrate not only that introns within introns ("twintrons"; Copertino and Hallick 1991) are not restricted to the *Euglena* chloroplast genome, but also that cryptomonad plastid genomes are not intron-free.

The analysis of the ORF439 shows it to be homologous to a recently discovered cyanobacterial ORF, which is further supporting evidence for the hypothesis that cyanobacteria are the ancestors of modern plastids (e.g. Gray 1993). The same conclusion may be drawn from the phylogenetic tree constructed on the basis of chaperone sequences. As expected, the coding region of *cpn60* is highly similar to homologous sequences of the hsp 60 family. The phylogenetic tree is divided into five main clusters (Fig. 4); three of them consisting of prokaryotic sequences, whereas the remaining sequences comprise of plastidal as well as of cyanobacterial sequences in one case, and mitochondrial sequences in the other. Thus, the tree is consistent with the endosymbiont theory. The proximity of the *Pyrenomonas salina* sequence to the *Cyanidium caldarium* (Rhodophyta) sequence (Maid et al. 1992) supports the results derived from previoius rDNA analyses (Douglas et al. 1991; Maier et al. 1991), which led to the conclusion that the endosymbiont of cryptomonads was rhodophyte-like.

In summary, our studies have uncovered the first intron in a cryptomonad plastome. The analysis of its structure and of the coding region of the host *cpn60* gene shows several interesting features. (i) The intron is a twintron. (ii) The predicted secondary structure of the insertion may be stabilized by base pairing parts of the group II intron, implying the use of a novel splicing mechanism. (iii) The phylogenetic status of an intron-encoded ORF, as well as of the Cpn60 sequence, reflects the origin of plastids from a cyanobacterial ancestor. Moreover, the fact that cryptomonad plastomes are not intron-free permits the inclusion of this plastome type into phylogenetic studies of intron evolution.

Acknowledgements The authors would like to thank Elfriede Schiefermayr and Klara Ronai for excellent technical assistance. This work is supported by the Deutsche Forschungsgemeinschaft.

References

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

Bairoch A (1993) The PROSITE of sites and patterns in proteins, its current status. Nucleic Acids Res 21:3097–3103

Bernard C, Thomas JC, Mazel D, Mousseau A, Castets AM, Tandeau de Marsac N, Dubacq JP (1992) Characterization of the genes encoding phycoerythrin in the red alga *Rhodella violacea*: evidence for a splitting of the *rpe*B gene by an intron. Proc Natl Acad Sci USA 89:9564–9568

Copertino DW, Hallick RB (1991) Group II twintron: an intron within an intron in a chloroplast cytochrome *b559* gene. EMBO J 10:433–442

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

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

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

Felenstein J (1989) PHYLIP-Phylogeny Interference Package (Version 3.2). Cladistics 5:164–166

Ferat J-L, Michel F (1993) Group II self-splicing introns in bacteria. Nature 364:358–361 Genetics Computer Group (1991) Program Manual for the GCG Package, version 7. 575 Science Drive, Madison, WI 537114, USA

Gibbs SP (1990) The evolution of algal chloroplasts. In: Wiessner W, Robinson DG, Starr RC (eds) Experimental phycology I. Springer-Verlag, Heidelberg, pp 145–157

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

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

Igloi GL, Schiefermayr E (1993) Enzymatic addition of fluorescein- or biotin-riboUTP to oligonucleotides results in primers suitable for DNA sequencing and PCR. BioTechnology 15:486–497

Jacquier A. (1990) Self-splicing group II and nuclear pre-mRNA introns: how similar are they? Trends Biochem Sci 15:351–354

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622

Maerz M, Wolters J, Hofmann CJB, Sitte P, Maier U-G (1992) Plastid DNA from the cryptomonad *Pyrenomonas salina* (Cryptophyceae): physical map, genes, and evolutionary implications. Curr Genet 22:479–482

Maid U, Steinmüller R, Zetsche K (1992) Structure and expression of a plastid-encoded *gro*EL homologous heat-shock gene in a thermophilic red alga. Curr Genet 21:521–525

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

McFadden GI (1993) Second-hand chloroplasts: evolution of cryptomonad algae. Adv Botan Res 19:189–230

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns – a review. Gene 82:5–30

Mohr G, Perlman S, Lambowitz AM (1993) Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucleic Acids Res 21:4991–4997

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning; a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Valentin K, Cattolico RA, Zetsche K (1993) Phylogenetic origin of plastids. In: Lewin RA (ed) Origins of Plastids. Chapman and Hall, London pp 193–222