Curr Genet 31: 89–95

ORIGINAL PAPER

Michael D. Thompson · Liqun Zhang · Ling Hong
Richard B. Hallick

Two new group-II twintrons in the *Euglena gracilis* chloroplast are absent in basally branching *Euglena* species

Received: 4 March / 11 September 1996

Abstract Studies of the phylogeny and chloroplast intron content of selected *Euglena* species have led to insights in our understanding of the timing of intron acquisition. In the current study, two new twintrons, found in *E. gracilis*, have been characterized by the analysis of partially spliced pre-mRNAs. Intron 1 of *atpE* is a 463-nt group-II intron interrupted by a second group-II intron 320 nt long. Intron 1 of *psbD* is also a group-II twintron with external and internal introns of 635 nt and 463 nt, respectively. The two introns composing the *psbD* twintron, as well as six additional group-II introns found in the *E. gracilis psbD* gene, are not present in several basally branching *Euglena* species, including *E. myxocylindracea*, *E. stellata* and *E. viridis*. The distribution of *psbD* introns in *Euglena* is consistent with a late evolutionary acquisition of group-II introns in this lineage.

Key words *Euglena* · Chloroplast · Twintron · Evolution

Introduction

Studying the phylogeny and chloroplast intron content of selected *Euglena* species is proving useful for understanding the timing of intron acquisitions in this lineage. A phylogenetic approach to studying *Euglena* chloroplast introns has also buttressed our understanding of the structure and polarity of the evolution of group-II and group-III introns (Thompson et al. 1995).

Group-III introns have been identified only in chloroplast genes from members of the *Euglena* and *Astasia* genera. Group-II introns are widely distributed among plant and fungal mitochondria, some lower eukaryotic mitochondria, in the plastids of photosynthetic eukaryotes (Michel et al. 1989), and in prokaryotes (Ferat and Michel 1993). Among 11 completely sequenced plastid genomes, the number of group-II introns ranges from 0 in *Cyanophora paradoxa*, *Porphyra purpurea* and *Odontella sinensis* (Kowallik et al. 1995; Reith and Munholland 1995; Stirewalt et al. 1995) to approximately 20 in plant chloroplasts (Hiratsuka et al. 1989; Maier et al. 1995; Ohyama et al. 1986; Shinozaki et al. 1986; Wakasugi et al. 1994), with the exception, however, of *Euglena gracilis* (Hallick et al. 1993).

The *E. gracilis* plastid genome is the richest known source of group-II introns; 155 group-II and group-III (a streamlined form of group-II) introns are present, nearly ten times the number found in any other plastid genome (Hallick et al. 1993). Among the *E. gracilis* introns, 15 twintrons have been identified including several complex twintrons composed of two or three individual internal introns located within the same external intron (Copertino and Hallick 1993). The excision of all internal introns, prior to that of the external, is essential for complete twintron excision since all known internal introns disrupt a functional domain of the external intron. It is by virtue of this sequential splicing pathway that twintrons were first experimentally characterized through the identification of partially spliced intermediates (Copertino and Hallick 1991). All of the 15 twintrons identified in the plastid genome of *E. gracilis* interrupt protein-coding genes.

Group-II introns are defined by the presence of conserved boundary sequences and secondary structure elements that form six helical domains (Koller et al. 1984; Michel and Jean-Luc 1995; Michel et al. 1989). The secondary structures and tertiary interactions formed by group-II intron sequences mediate intron

M.D. Thompson · L. Zhang · R.B. Hallick (✉)
Department of Biochemistry, University of Arizona,
Tucson, AZ 85721, USA

L. Hong
University of California, Department of Molecular
and Cellular Biology, Berkeley, CA 94720, USA

Communicated by D. R. Wolstenholme

excision, as a lariat, and exon ligation through two trans-esterification reactions. This splicing mechanism is similar to that of nuclear pre-mRNA introns. This similarity has led to the suggestion that group-II introns may be the evolutionary progenitors of nuclear introns (Sharp 1985; Cech 1986; Roger and Doolittle 1993; Saldanha et al. 1993; Sontheimer and Steitz 1993).

Two new twintrons have been characterized by the analysis of partially spliced pre-mRNAs. Intron 1 of *atpE* is a 463-nucleotide group-II intron interrupted in domain VI by a 320-nucleotide group-II intron. Intron 1 of *psbD* is also a group-II twintron, with a 635-nucleotide internal intron inserted into domain V of a 463-nucleotide external intron. The *psbD* gene in *E. gracilis* contains ten introns. The first seven *psbD* introns are not present in several basally branching *Euglena* species. This distribution of *psbD* introns in *Euglena* is consistent with a late evolutionary acquisition of introns in this lineage.

Materials and methods

Euglena strains and nucleic acid isolation. The *Euglena* species, *E. geniculata* var. *terricola* (UTEX 366), *E. myxocylindracea* (UTEX 1989), *E. pisciformis* var. *typica* (UTEX 1604), *E. stellata* (UTEX 372), *E. viridis* (UTEX 85) and *E. gracilis* var. *Z strain* were obtained from the culture collection of algae at the University of Texas at Austin. Total nucleic-acid isolation was described by Thompson et al. (1995).

cDNA synthesis, amplification, cloning and sequencing. cDNAs were synthesized and amplified by the polymerase chain reaction (PCR) as previously described (Copertino and Hallick 1991) and 5 μl of total-nucleic-acid extract was used as a template for cDNA synthesis. cDNAs representing partially or fully spliced *psbD* twintron 1 were synthesized using the oligonucleotide primers C1 and C2 (Table 1), respectively. cDNAs complementary to *psbD* pre-mRNAs were amplified using the oligonucleotide primer P1 (see Table 1). The oligonucleotide primers C3 and C4 (see Table 1) were used for cDNA synthesis of *atpE* pre-mRNA containing partially or fully spliced twintron 1, respectively. cDNAs complementary to *atpE* pre-mRNAs were amplified using the primer P2 (see Table 1). Amplification products were ligated into the ddT-tailed *Eco*RV site of the plasmid vector KS+ (Stratagene and Holton and Graham 1991). The nucleotide sequences of recombinant plasmids were determined by the chain-termination technique (Sanger et al. 1977) and modified for use with Sequenase^{TM} (U.S. Biochemicals). Degenerate oligonucleotide primers P3/C5 or P3/C6 (see Table 1) were used for the amplification of *psbD* gene fragments extending from exons 1–8 or exons 1–3, respectively. This amplification was carried out for 30 cycles using the following parameters: 94°C for 1 min, 43°C for 2 min and 72°C for 3 min.

Results

Twintron locations

Figure 1 shows the locations of the two twintrons described in this report. One is located in the *psbD* gene (position 3931–5028 in GenBank accession #X70810) which encodes the D2 polypeptide of photosystem II (Orsat et al. 1994). The second twintron is located in the *atpE* gene (position 88642–89398) which encodes the ε subunit of chloroplast ATP synthase (Hong and Hallick 1994).

RT-PCR analysis of *psbD* and *atpE* pre-mRNAs

Based on sequence analysis, the *psbD* and *atpE* twintrons were each predicted to be composed of one group-II intron internal to a second group-II intron (Hallick et al. 1993). To experimentally test this prediction, RT-PCR products of *psbD* and *atpE* pre-mRNAs were isolated and analyzed. Synthetic oligonucleotide primers (shown in Fig. 1), were used for cDNA synthesis of *psbD* and *atpE* pre-mRNAs. cDNAs were amplified by PCR and recombinant plasmids containing the amplification products were produced. Figures 2 and 3 show the cDNA sequence data corresponding to partial and complete excision of the *psbD* and *atpE* twintrons, respectively. The recombinant plasmids pEZC1094 (Fig. 2) and pEZC 2015 (Fig. 3) are missing predicted internal intron sequences and therefore represent *psbD* and *atpE* pre-mRNAs, respectively, with excised internal introns. Based on the sequences of pEZC 1094 and pEZC 2015, the 5′ and 3′ splice boundaries of the internal introns from *psbD* intron 1 and *atpE* intron 1, respectively, are confirmed. pEZC1095 and pEZC 2016 (Fig. 2 and 3) contain only exon sequences and therefore represent mRNAs with these twintrons fully excised. The *psbD* internal intron 5′ and 3′ boundaries were confirmed by direct primer-extension RNA sequencing (data not shown).

Table 1 Sequences and coordinates (EMBL #X70810) of primers used for cDNA synthesis and PCR amplification

Oligo Name	Oligo sequence	Location
C1	5′-CCGATACTATTTGGAAAGAATAG-3′	4998–5020
C2	5′-GCAATAATAAAGAATGACCC-3′	5021–5040
C3	5′-CTTTCAGGAATAAGTAAAGTAG-3′	88 636–88 657
C4	5′-CCAAATTCTGCTTCATTGAC-3′	87 753–87 772
C5	5′-CC(G/A)AA(G/A)ATTTG(T/A)GACCAAAACG-3′	8841–8863
C6	5′-CCATAG(A/T)CC(G/T)CCTAGTTG(A/C)(A/C)ACCAACG-3′	5483–5510
P1	5′-GCAATGCTTTGACAGCAGC-3′	3899–3917
P2	5′-ATGACATTAGATGTATC-3′	89 410–89 428
P3	5′-TGGTATAC(T/A)CATGG(T/A)TT(A/G)GC-3′	3860–3879

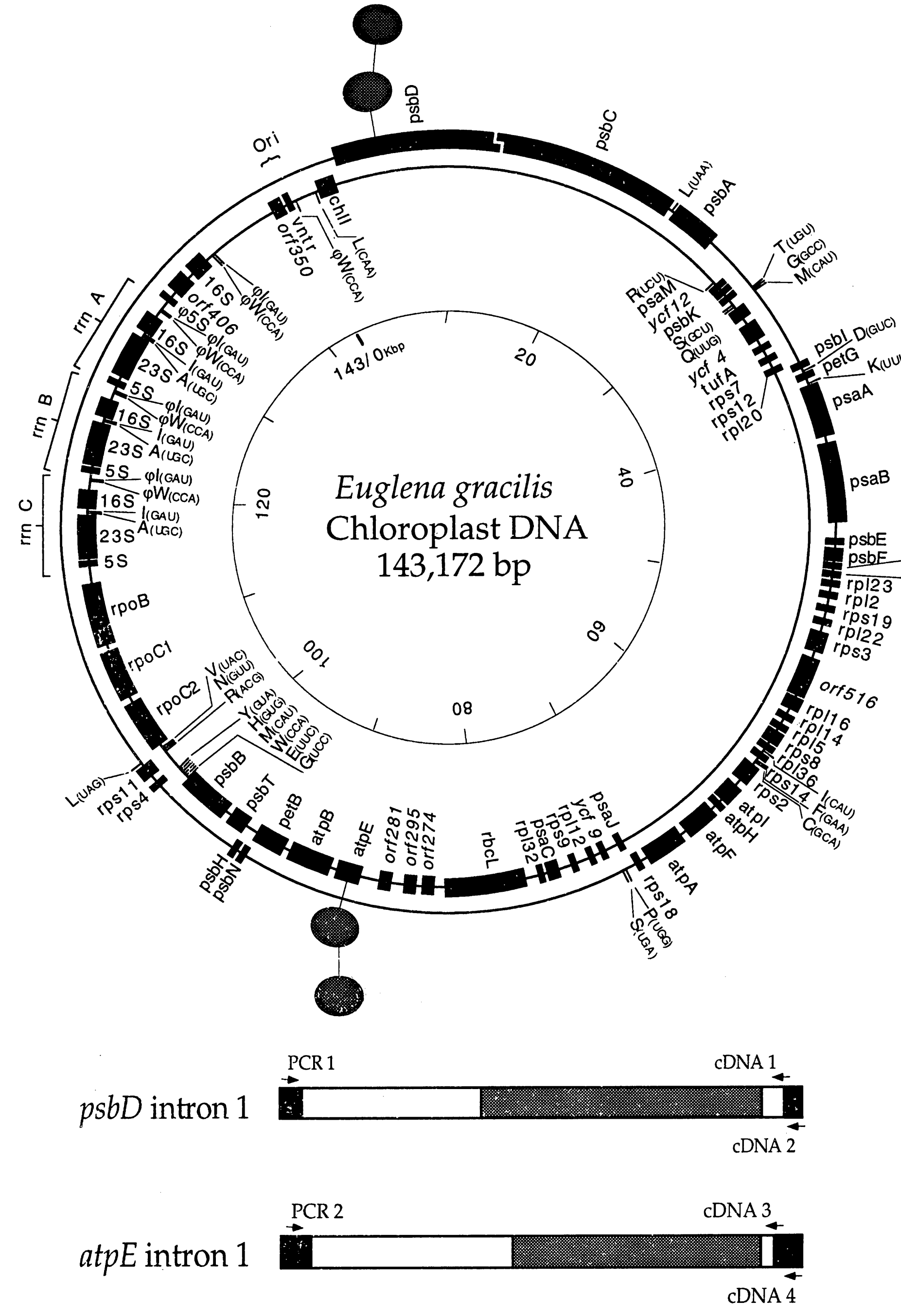

Fig. 1 Locations of two new group-II twintrons in the *E. gracilis* chloroplast genome. The overall organization of the *E. gracilis* chloroplast genome is shown (adapted from Hallick et al. 1993). The locations of two group-II twintrons, within the *psbD* and *atpE* genes are indicated by *shaded lollipops*. One lollipop inside another represents one group-II intron in another. A schematic diagram of each twintron is also shown. *Black boxes* represent exons. *Open boxes* represent external introns. Internal introns are shown as *shaded boxes*. Oligonucleotide primer locations are indicated with *arrows*. The primers PCR1, PCR2, cDNA2 and cDNA4 are located entirely within exon sequences. cDNA1 and cDNA3 span an intron-exon boundary

psbD intron content in selected *Euglena* species

To study the evolutionary history of introns in the *psbD* gene in the genus *Euglena*, the occurrence of *psbD* introns in five different *Euglena* species was examined. The presence of introns was inferred based on the size of PCR-amplified *psbD* gene segments. A schematic diagram of a portion of the *E. gracilis psbD* gene and the data from this experiment are shown in Fig. 4. The completely sequenced *psbD* gene from *E. gracilis* served as a (+) intron control. In *E. gracilis*, the P3/C5-amplified gene segment is 4981 nucleotides long and contains seven different introns (Hallick et al. 1993). The P3/C5-amplified gene segment from *E. geniculata* co-migrated with the P3/C5-amplified gene segment from *E. gracilis*. Based on these data, the *psbD* gene from *E. geniculata* most likely contains introns that are homologous to *psbD* introns 1–7 from *E. gracilis*. The amplified gene segments from *E. viridis*, *E. stellata* and *E. myxocylindracea* co-migrate with one another at approximately 600 nt. Based on the *E. gracilis* gene sequence, the predicted size of an intronless P3/C5-

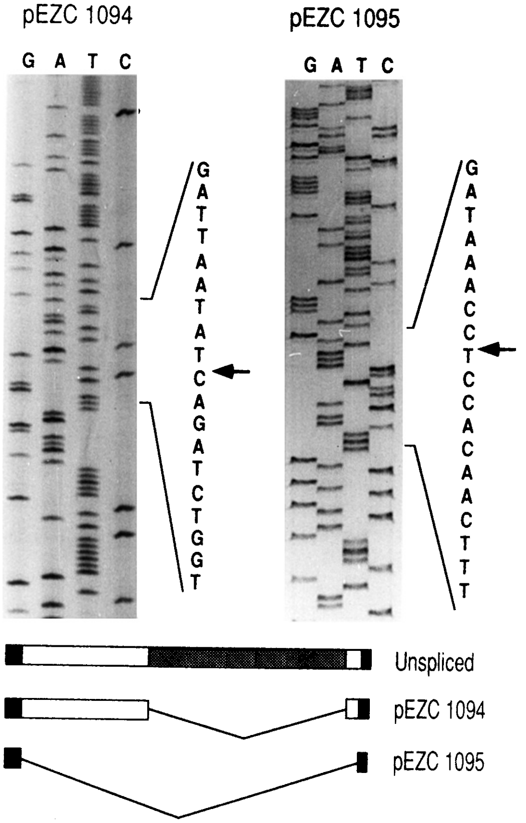

Fig. 2 Sequences of the cDNAs corresponding to partial and complete excision of *psbD* twintron 1. Synthetic oligonucleotide primers were used for RT-PCR cloning of partially and fully spliced *psbD* mRNAs. Nucleotide sequences of representative cDNAs corresponding to partially (pEZC 1094) and fully (pEZC 1095) spliced mRNAs are shown. The expanded sequences represent internal and external intron boundaries. The locations of the internal and the external introns are indicated by *arrows*. Schematic diagrams of the unspliced message and the partially and fully spliced messages corresponding to the sequence ladders are shown

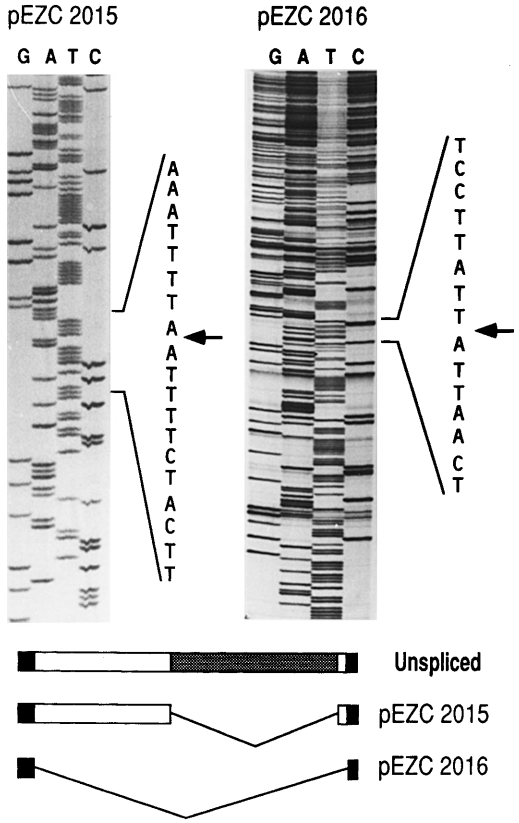

Fig. 3 Sequences of the cDNAs corresponding to partial and complete excision of *atpE* twintron 1. Templates for *atpE* sequences were produced as described in Fig. 2. Nucleotide sequences of cDNAs corresponding to partially (pEZC 2015) and fully (pEZC 2016) spliced mRNAs are shown. The expanded sequences represent internal and external intron boundaries. The locations of the internal and the external introns are indicated by *arrows*. Schematic diagrams of the unspliced message and the partially and fully spliced messages corresponding to the sequence ladders are shown

amplified gene segment is 586 nt. Therefore, the *E. viridis*, *E. stellata* and *E. myxocylindracea psbD* gene segments are most likely intronless.

In *E. pisciformis*, the P3/C5-amplified gene segment is intermediate in size between *E. gracilis* and *E. viridis*. Therefore, *E. pisciformis* most likely contains introns homologous to a subset of the *E. gracilis psbD* introns 1–7. Alternatively, the P3/C5-amplified gene segment from *E. pisciformis* may contain a set of introns that are not homologous with some or any of the *psbD* introns 1–7 from *E. gracilis*.

To determine whether homologs of *E. gracilis psbD* introns 1 (twintron) and 2 are present in *E. pisciformis*, a segment of the *E. pisciformis psbD* gene extending from exon 1 (P3) to exon 3 (C6) was analyzed. The approximately 600-nucleotide P3/C6-amplified gene segment from *E. pisciformis* was cloned and completely sequenced. From a comparison of the *E. pisciformis* gene sequence with the *E. gracilis* coding sequence, a 449-nucleotide group-II intron located between codons 72 and 73 was revealed. Based on this analysis, a single intron is present in the *E. pisciformis psbD* gene segment defined by primers P3 and C6. Homologs of *E. gracilis psbD* introns 1 and 2 are not present in *E. pisciformis*.

Discussion

The *psbD* and *atpE* twintrons are each composed of one group-II intron inside a second group-II intron

The primary sequences of three individual introns comprising the *psbD* twintron and the external *atpE* intron can each be represented in the secondary structural model (Figs. 5 and 6) proposed by Michel

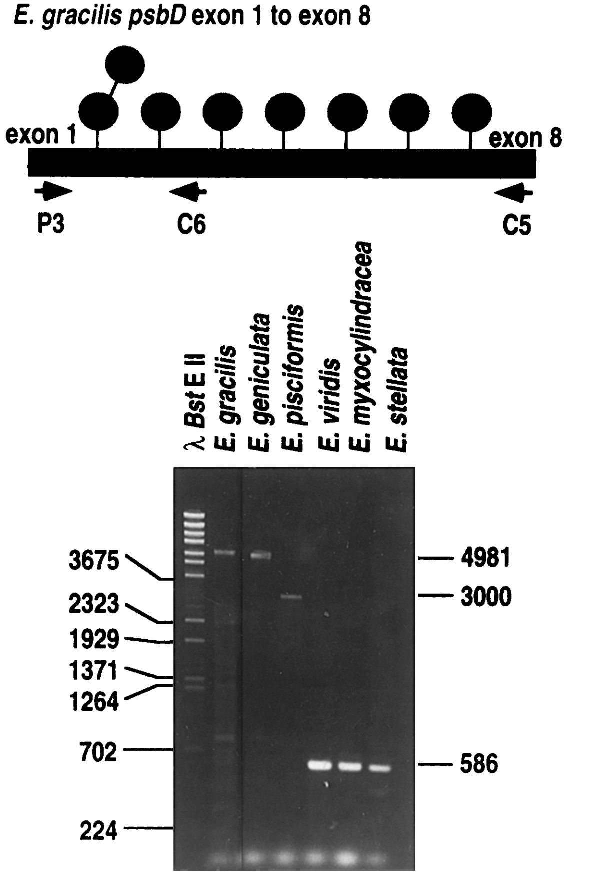

Fig. 4 *psbD* primer locations and PCR data. A schematic diagram of the *E. gracilis psbD* gene is shown. The *black box* represents exons, *filled lollipops* represent introns. The locations of the P3, C5 and C6 primers are indicated by *arrows*. The oligonucleotide primers P3 and C5 were used to amplify a segment of the psbD gene from *E. gracilis*, *E. geniculata*, *E. pisciformis*, *E. viridis*, *E. myxocylindracea* and *E. stellata*. PCR reaction products were separated in an agarose gel and visualized by ethidium-bromide staining

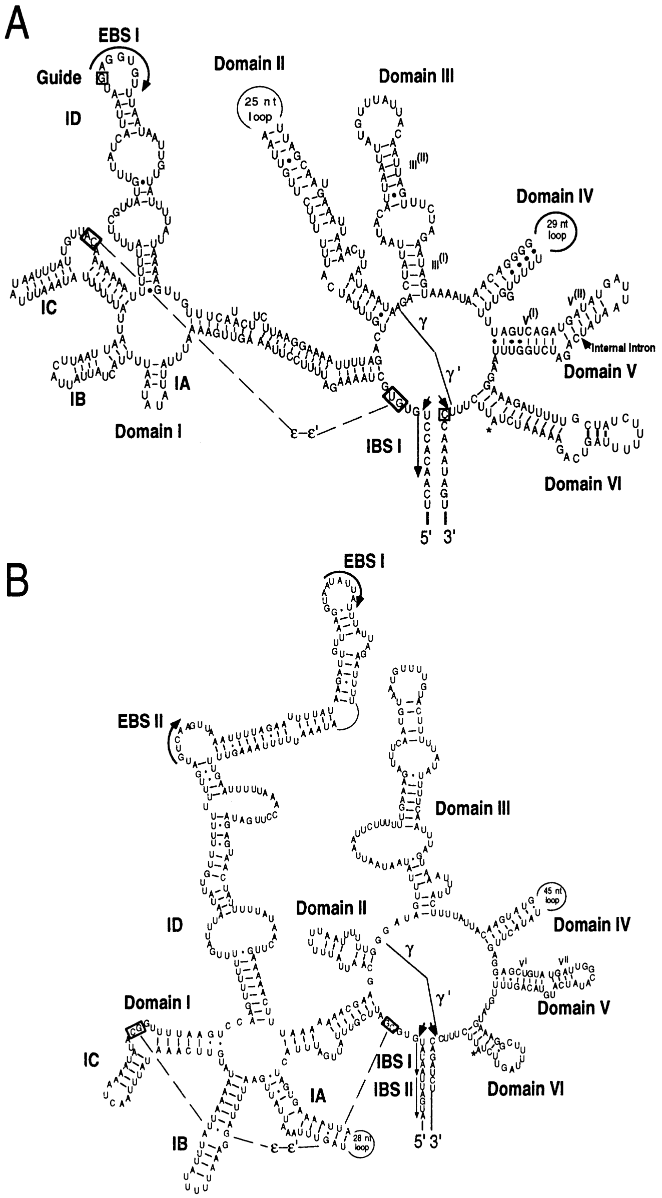

Fig. 5A, B Secondary structure models for the introns of the *psbD* twintron. **A** external intron. **B** internal intron. *Dashed lines and boxed nucleotides* indicate the ε–ε′ tertiary interactions. The γ–γ′ interactions are indicated by *solid lines*. The branch site A is marked with an *asterisk* (*). Nucleotides involved in the EBS1-IBS1 interaction are marked with *curved arrows*; 5′ and 3′ splice sites are indicated by *straight arrows*

et al. (1989). In this model, each intron has six helical domains (I–VI) radiating from a central core flanked by the 5′ and 3′ intron boundary sequences. The 5′-boundary sequences of both *psbD* introns conform to the group-II 5′-GUGYG consensus sequence (Michel and Jean-Luc 1995). Both *atpE* introns have variant nucleotides at intron position +3. Each intron has a domain VI with a bulged A-nucleotide, at position –7 or –8 from the 3′-splice boundary, as the putative site of 2′–5′ branch formation with the intron 5′-end during splicing.

In addition to a conserved secondary structure, each intron model accommodates most of the tertiary interactions previously described for group-II introns, including base pairing between the 5′-exon and intron domain Id (EBS1-IBS1), ε–ε′ pairing between the 5′-splice boundary positions +3 and +4 with intron domain IC, and the γ–γ′ interaction between the 5′-RRGAY region separating intron domains II–III and the last nucleotide of the intron. Some introns also have a second putative exon-intron pairing (IBS2-EBS2) and the "guided pair" interaction involving the 3′-exon.

A partial secondary structure of the internal *atpE* intron is shown in Fig. 6 B. Putative domains III, IV, V and VI that conform to the canonical secondary structure model are identified. Notable features are a bulged A-nucleotide present in domain VI, seven

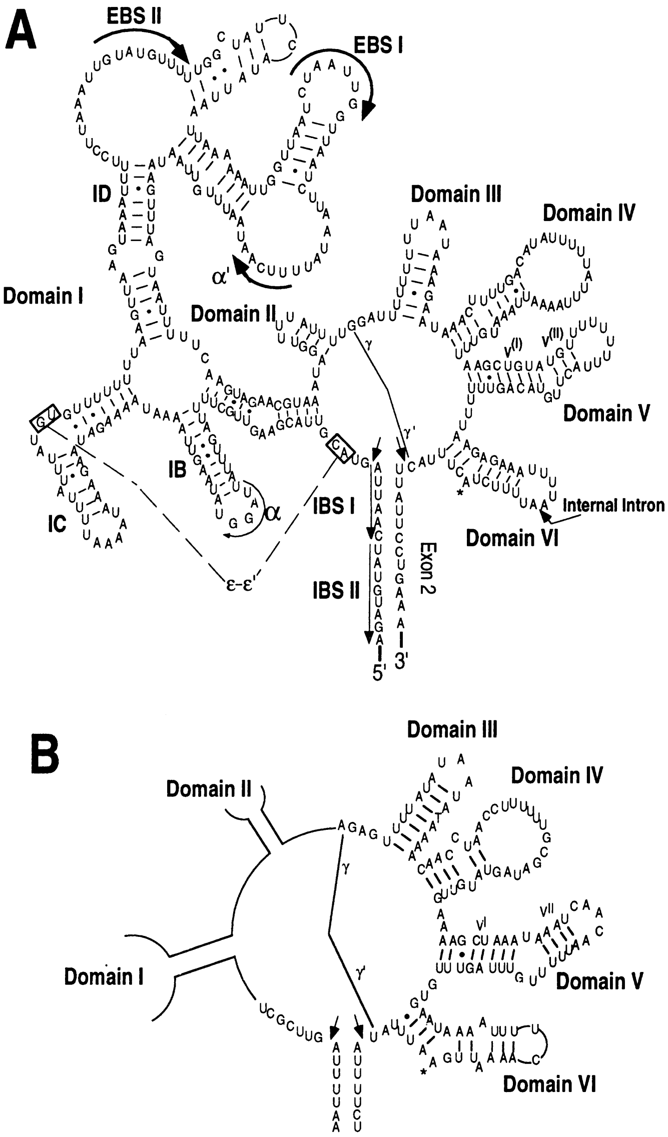

Fig. 6A, B Secondary structure models for the introns of the *atpE* twintron. **A** external intron. **B** internal intron. Notations are as described in Fig. 5

nucleotides from the 3′-splice boundary, and a γ–γ′ tertiary interaction 5′ of the putative domain III. The secondary structures of domains I and II are unclear. Putative domains I and II, including potential EBS1-IBS1, EBS2-IBS2 and ε–ε′, can be found. However, the overall structure of these putative domains does not conform well to the consensus model. The *Euglena* chloroplast group-II introns are generally shorter, have less extensive domain-I base pairing, and less distinct tertiary interactions than their counterparts in fungal mitochondria and plant chloroplasts.

The secondary structures presented in Figs. 5 and 6 are based on the model for group-II introns proposed by Michel et al. (1989). Whether the stem-loop structures shown in Figs. 5 and 6 are biologically active is not known. Support for the structures of domains V and VI, the closure of domain I, and the central core of several *Euglena* group-II introns has been obtained by a comparison of homologous introns from different species of *Euglena* (Thompson et al. 1995; Thompson and Zhang, unpublished). In each case, the central core is well conserved at the primary sequence level among homologous introns. Compensatory base changes maintain the structures of domains IV, V, VI, the closure of domain I, and the position of a putative RRGAY region possibly involved in the γ–γ′ interaction. Homologous intron comparisons of the *psbD* and *atpE* introns presented in this report have not been possible since no homologs to these introns from sufficiently diverse species were identified in the evolutionary study.

Both the *psbD* intron-1 and *atpE* intron-1 internal introns disrupt essential functional domains of their respective external introns. In *psbD*, the external intron is interrupted in domain 5 at a position identical to the site of insertion of the internal intron of the *psbF* twintron. In *atpE*, the external intron is disrupted in domain VI, only 7 nt proximal to the putative branch-A nucleotide. Both intron insertions would be expected to prevent external intron splicing prior to internal intron removal.

Euglena psbD introns: relatively late acquisitions

The occurrence of twintrons is suggestive of a mechanism for intron addition. Copertino and Hallick proposed that twintrons were formed by the insertion of one or more introns into another intron (Copertino and Hallick 1991). One underlying assumption is the paucity of introns early in the evolution of this lineage. In this scenario, introns accumulated as the main line leading to *E. gracilis*, evolved. The near absence of *rbcL* introns in basally branching *Euglena* species compared to nine *rbcL* introns in derived species is supportive of the introns-late scenario.

In order to test the hypothesis that *Euglena psbD* introns are relatively late introductions to the *Euglena* chloroplast lineage, the *psbD* intron content of the *Euglena* species examined in this study was superimposed onto a *Euglena* phylogeny (Fig. 7) (Thompson et al. 1995). The branching order of *E. myxocylindracea*, *E. pisciformis, E. stellata* and *E. viridis* relative to each other is not well supported by this phylogeny. However, the basal positions of *E. myxocylindracea, E. pisciformis*, *E. stellata* and *E. viridis*, relative to *E. gracilis*, and *E. geniculata* are strongly supported (Thompson et al. 1995).

The most parsimonious explanation for the distribution of *psbD* introns is that the *psbD* gene in *Euglena* evolved from an intronless ancestor and introns were added to this gene during its evolution. The *Euglena* species that do not have *psbD* introns (*E. viridis*, *E. myxocylindracea* and *E. stellata*) are all basally

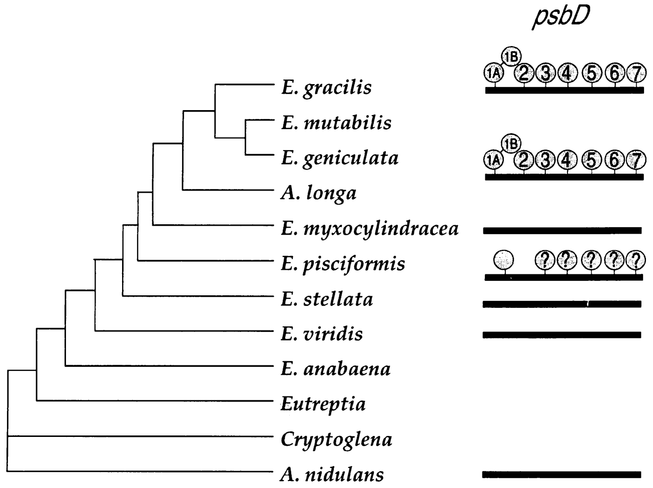

Fig. 7 *psbD* intron distribution in *Euglena.* The phylogenetic tree is based on the *rbcL* coding sequence (Thompson et al. 1995). The content of *psbD* introns from select *Euglena* species is shown on the right. Introns are indicated by *numbered lollipops*

branching. The most derived species (*E. gracilis* and *E. geniculata*) contain at least eight introns, two of which are members of a twintron. *A. longa* is a non-photosynthetic species that has undergone a secondary loss of genes for photosynthetic membranes, including *psbD*.

The specific position of *E. pisciformis* among the basally branching species is not well supported in the current phylogeny. However, based on the introns-late scenario, the presence of *psbD* introns in *E. pisciformis*, whether they are homologous to *E. gracilis psbD* introns or not, is supportive of *E. pisciformis*[1] placement as more derived than *E. viridis*, *E. myxocylindracea* and *E. stellata*. This placement of *E. pisciformis* is also consistent with a phylogeny derived from nuclear 18*s* rDNA sequences (Richard Triemer, personal communication).

References

Cech TR (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell 44: 207–210

Copertino DW, Hallick RB (1991) Group-II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. EMBO J 10: 433–42

Copertino DW, Hallick RB (1993) Group-II and group-III introns of twintrons: potential relationships to nuclear pre-mRNA introns. Trends Biochem Sci 18: 467–471

Ferat J-L, Michel F (1993) Group-II self-splicing introns in bacteria. Nature 364: 358–361

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E (1993) Complete DNA sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21: 3537–3544

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217: 185–94

Holton TA, Graham MW (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res 19: 1156

Hong L, Hallick RB (1994) Gene structure and expression of a novel *Euglena gracilis* chloroplast operon encoding cytochrome b6 and the beta and epsilon subunits of the H(+)-ATP synthase complex. Curr Genet 25: 270–281

Koller B, Gingrich JC, Stiegler GL, Farley MA, Delius H, Hallick RB (1984) Nine introns with conserved boundary sequences in the *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene. Cell 36: 545–553

Kowallik KV, Stoebe B, Schaffran I, Freier U (1995) The chloroplast genome of a chlorophyll a+c-containing alga, *Odontella sinensis*. Plant Mol Biol Rep 13: 226–342

Maier RM, Neckermann K, Igloi GI, Kossel H (1995) Complete sequence of the *maize* chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251: 614–628

Michel F, Jean-Luc F (1995) Structure and activities of group-II introns. Annu Rev Biochem 64: 435–461

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

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organizatoin deduced from complete sequence of liverwort *Marchantia polymorpha*. Nature 322: 572–574

Orsat B, Spielmann A, Marc-Martin S, Lemberger T, Stutz E (1994) Analysis of the 22-kbp-long *psbD-psbC* gene cluster of *Euglena gracilis* chloroplast DNA: evidence for overlapping transcription units undergoing differential processing. Biochim Biophys Acta 1218: 75–81

Reith M, Munholland J (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol biol Rep 13: 333–345

Roger JA, Doolittle WF (1993) Why introns-in-pieces? Nature 364: 289–290

Saldanha R, Mohr G, Belfort M, Lambowitz AM (1993) Group-I and group-II introns. FASEB J 7: 15–24

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

Sharp PA (1985) On the origin of RNA splicing and introns. Cell 42: 397–400

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049

Sontheimer EJ, Steitz JA (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science 262: 1989–1996

Stirewalt V, Michalowski C, Loffelhardt W, Bohnert H, Bryant D (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol Biol Rep 13: 327–332

Thompson MD, Copertino DW, Thompson E, Favreau MR, Hallick RB (1995) Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. Nucleic Acids Res 23: 4745–4752

Wakasugi T, Tsudsuki J, Ito, S, Nakashima K, Tsudsuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91: 9794–9798