RNA (1995), *1*:1079–1088. Cambridge University Press. Printed in the USA.
Copyright © 1995 RNA Society.

The *Euglena gracilis* intron-encoded *mat2* locus is interrupted by three additional group II introns

LIQUN ZHANG,[1] KRISTIN P. JENKINS,[2] ERHARD STUTZ,[3] and RICHARD B. HALLICK[1]

[1] Department of Biochemistry, University of Arizona, Tucson, Arizona 85721, USA
[2] Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA
[3] Laboratoire de Biochimie Végétale, Université de Neuchâtel, Chantemerle 18, CH-2000 Neuchâtel, Switzerland

ABSTRACT

The 4,144 nt *Euglena gracilis* chloroplast *psbC* intron 2 has been characterized as a single, *cis*-spliced 593 nt group II intron interrupted by an open reading frame of 758 codons in the loop region of domain IV. The 2,277 nt coding region of orf 758 is interrupted by two additional group II introns of 369 nt and 352 nt. Another 553 nt group II intron is located in the 5′ untranslated leader region of orf 758. Because the *psbC* intron 2 orf encodes a maturase-like protein that has reverse transcriptase domains and a domain X characteristic of group II intron-encoded proteins, the locus has been designated *mat2*. The *psbC* intron 2 is the first member of a new category of twintron, characterized by introns within a gene within another intron. A potential role of *psbC* intron 2 as a ''founder'' intron involved in the spread of introns to new sites in the plastid genome of the Euglenophyceae is discussed.

Keywords: *Euglena*; group II introns; reverse transcriptase; twintron

INTRODUCTION

The *Euglena gracilis* chloroplast genome contains many novel introns unlike any introns found in other genomes. Introns account for up to 39% of the *E. gracilis* chloroplast genome. Among the 155 introns identified to date, 74 are group II introns ranging in size from 277 to 671 nt (Hallick et al., 1993). *Euglena* chloroplast group II introns conform to the core secondary structure model proposed by Michel et al. (1989). However, only the catalytic domains V and VI are well conserved in *Euglena* chloroplast introns, whereas domains I–IV are often abbreviated. The *Euglena* chloroplast genome also contains 15 twintrons (Copertino & Hallick, 1993), or introns-within-introns. They include simple twintrons, where one intron is inserted into another intron (Copertino & Hallick, 1991; Copertino et al., 1991, 1992), and complex twintrons in which the external intron is interrupted by more than one intron (Hong & Hallick, 1994a, 1994b) or even by another complex twintron (Drager & Hallick, 1993). The internal intron is usually inserted in a functional domain of the external intron so that the internal intron must be removed to restore the splicing ability of the external intron (Copertino & Hallick, 1993).

Splicing of group II introns in vivo is thought to be mediated by protein factors, including intron-encoded maturases (Michel et al., 1989; Lambowitz & Perlman, 1990; Saldanha et al., 1993). Group II intron-encoded maturases often contain a reverse transcriptase domain. The reverse transcriptase activity has been shown to contribute to intron mobility during evolution (Moran et al., 1995). A potential maturase has been identified in intron 4 of *psbC*, a gene coding for the CP43 chlorophyll-binding protein of photosystem II. The putative maturase gene is located in the internal intron of this group III twintron (Copertino et al., 1994). A gene encoded in *psbC* intron 2 may be a second intron maturase (Mohr et al., 1993).

The 4,144 nt-long *psbC* intron 2 is the largest intron in the *E. gracilis* chloroplast genome (Hallick et al., 1993). Here we report that it is a group II intron containing a 758 codon open reading frame (orf), which has been designated *mat2*. *mat2* encodes a maturase-like protein that is interrupted by three internal group II introns (Fig. 1). This is the first example of an intron-encoded orf containing introns.

Reprint requests to: Richard B. Hallick, University of Arizona, Department of Biochemistry, BioSciences West 524, P.O. Box 210088, 1041 E. Lowell St., Tucson Arizona 85721-0088, USA; e-mail: hallick@arizona.edu.

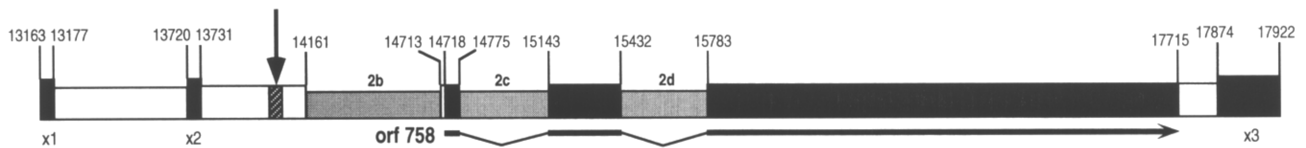


FIGURE 1. Schematic description of the overall structure of *psbC* intron 2. *psbC* exons 1, 2, and 3 (x1, x2, x3) are shown as solid black boxes. *psbC* intron 1 and external intron 2a regions are shown as open boxes. Orf 758 coding regions are shown as dark shaded boxes, and the three internal introns are shown as light shaded boxes. Coordinates of introns 2b, 2c, and 2d are shown above. An arrow indicates the potential ribosomal binding site of orf 758.

RESULTS

Identification of partially spliced *psbC* intron 2 pre-mRNAs by reverse transcriptase-PCR (RT-PCR) analysis

The *E. gracilis* chloroplast *psbC* intron 2 is 4,144 nt long. The size range for group II introns in *E. gracilis* chloroplasts is 277–671 nt. Present within the intron are several putative group II intron 5′-splice sites and 3′ domains V and VI, along with potential exons of an intron-encoded protein.

To test the hypothesis that *psbC* intron 2 is a twintron with an internally encoded protein, a series of oligonucleotide primers were designed to amplify *psbC* intron 2 RNA processing intermediates. Locations of the cDNA and PCR primers used in the analysis are shown in Figure 2A.

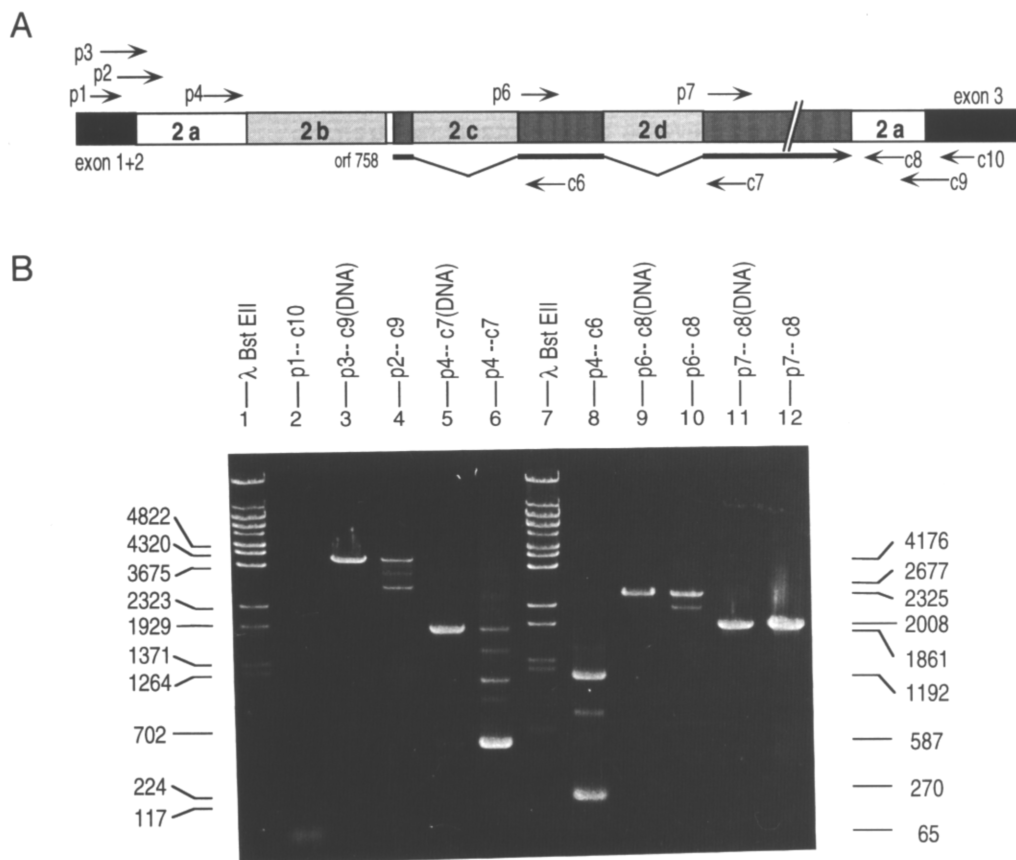


FIGURE 2. cDNA-PCR analysis of *psbC* pre-mRNAs. **A:** Structure of the *psbC* intron 2 is shown. Black boxes represent exons 1, 2, and 3; white boxes are external intron 2a; light shaded boxes represent internal introns (2b, 2c, 2d); and dark shaded boxes represent orf 758. Positions of the primers used for cDNA synthesis and PCR amplification are shown as arrows in the map of *psbC* intron 2. p1–p7 are PCR primers, c6–c12 are cDNA primers. **B:** PCR amplification of *psbC* partially spliced mRNAs. Amplification products were fractionated on an agarose gel. PCR and cDNA primers used for each amplification are shown above lanes 2–6 and 8–12. Control PCR with *E. gracilis* chloroplast total nucleic acid is labeled "DNA" above lanes 2, 5, 9, and 11. λ *Bst*E II molecular weight markers are shown in lanes 1 and 7, and the sizes (in bp) are shown on the left of the gel. Sizes of some PCR products are shown on the right.

Two control reactions were used to define the sizes of the fully spliced and fully unspliced RNAs. RT-PCR with primers specific to *psbC* exons 1–2 and exon 3 (Fig. 2B, lane 2, oligonucleotide p1 spans spliced exons 1 and 2) resulted in a single product of 65 bp, corresponding to spliced exons 1–3. Chloroplast total nucleic acid was PCR amplified with primers specific to the 5′ and 3′ exon–intron boundary sequences (Fig. 2B, lane 3). The resulting 4,176 bp product is the size of the unprocessed intron 2 pre-mRNA.

To test for partially spliced intron 2 pre-mRNAs, chloroplast RNA was RT-PCR amplified with primers spanning exon 2–intron 2 and intron 2–exon 3 boundary sequences (Fig. 2B, lane 4). A PCR product of about 4,200 bp, which comigrates with the control chloroplast total nucleic acid amplification product, and amplified cDNAs from four partially spliced pre-mRNAs of 3,900 bp, 3,500 bp, 3,300 bp, and 3,000 bp, were obtained. The signal from the 3,300 bp RT-PCR product is faint. The 3,000–3,900 bp PCR products are evidence of at least three introns, of a total length of about 1,200 bp, internal to *psbC* intron 2.

Three group II introns are located in domain IV of *psbC* intron 2

A preliminary secondary structure model for an external group II intron interrupted in domain IV by internal introns and an orf was developed (Figs. 2A, 4A). To test this model, primers for RT-PCR reactions were designed to complement either the putative external intron domain IV (p4, Fig. 2A) or the coding region of the putative intron-encoded polypeptide (p6, p7, c6, and c7, Fig. 2A).

RT-PCR with primers p4 and c7 resulted in a product from fully unprocessed precursor (1,861 bp) that comigrates with the product of a total nucleic acid control template, and products of four splicing intermediates (1,509 bp, 1,140 bp, 956 bp, and 587 bp) (Fig. 2B, lanes 5, 6). These products correspond to the four intermediates obtained by RT-PCR with primers p3 and c9. Therefore, all the internal splicing events occur between p4 and c7 (coordinates 13979–15839).

RT-PCR with primers p4 and c6 resulted in four products of 1,192 bp, 823 bp, 639 bp, and 270 bp (Fig. 2B, lane 8). The 1,192 bp species was identified by sequencing as the fully unprocessed precursor (data not shown). Therefore, one of the three internal splicing events must occur 3′ of primer c6 and at least two splicing events must occur within the region flanked by p4 and c6 (coordinates 13979–15170).

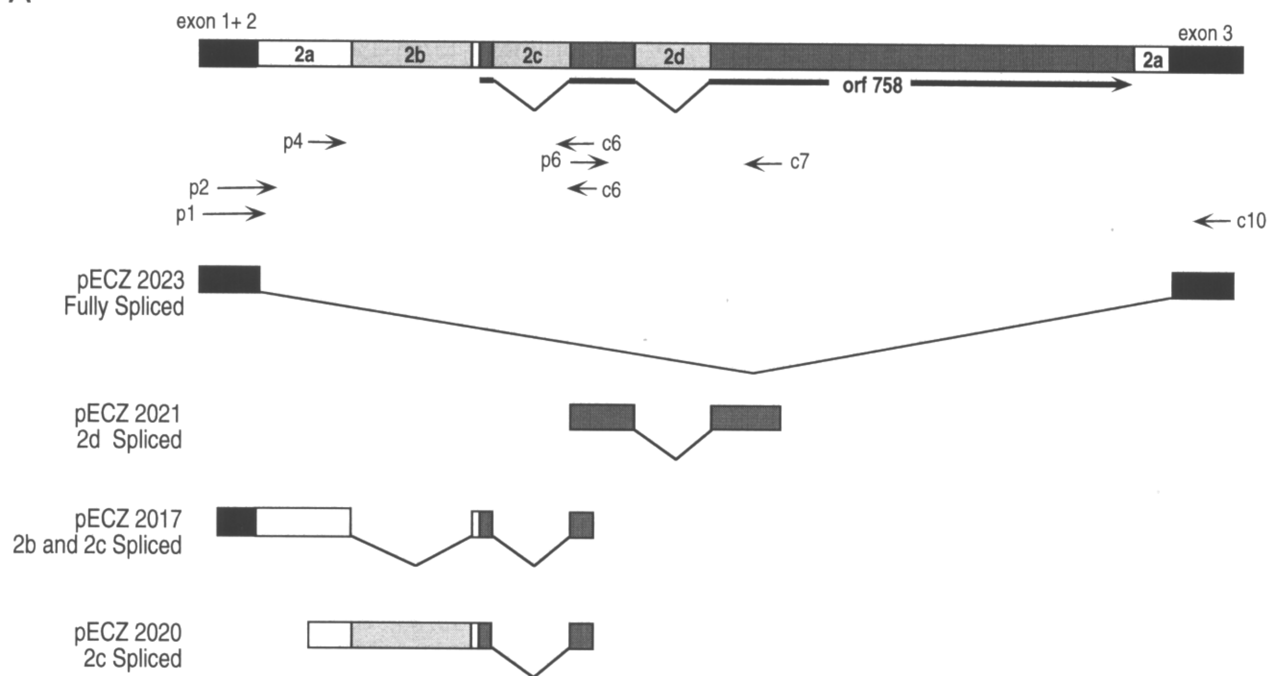
RT-PCR amplification with primers p6 and c8 resulted in a 2,677 bp product corresponding to unprocessed precursor and one partially spliced intermediate of approximately 2,300 bp (Fig. 2B, lane 10). The only RT-PCR product with primers p7 and c8 is the unprocessed precursor (Fig. 2B, lane 12). Because p6 and p7 are the compliment of c6 and c7, respectively, there is only one internal processing event between c6 and c7 (coordinates 15148–15839). We therefore conclude that *psbC* intron 2 contains at least three internal introns and all the splicing events occur within the region flanked by p4 and c7 (Fig. 2A).

Characterization of internal 5′ and 3′ splice sites within *psbC* intron 2

In order to locate the insertion sites of the internal introns, products from PCR reactions with primers p2-c6, p4-c6, p6-c7 (PCR data not shown), and p1-c10 were cloned and sequenced. Schematic diagrams of the cDNAs are shown in Figure 3A. Representative sequencing data showing the splice sites of internal introns (designated 2b, 2c, and 2d) and the external intron 2a are displayed in Figure 3B. Intron 2b is 553 nt (coordinates 14161–14713), intron 2c is 369 nt (coordinates 14775–15143), and intron 2d is 352 nt (coordinates 15432–15783). Splice sites for introns 2a and 2d are unambiguous. Splice sites for introns 2b and 2c could be shifted −1 nt and yield the same cDNA sequence.

Intron secondary structure models

The size of the three introns internal to *psbC* intron 2 are typical of *Euglena* chloroplast group II introns (Copertino & Hallick, 1993; Hallick et al., 1993). Models for the secondary structures of introns 2a, 2b, 2c, and 2d are shown in Figure 4. These intron structures conform to the model proposed by Michel for group II intron secondary structure (Michel et al., 1989) and are similar to other *Euglena* chloroplast group II introns (Hong & Hallick, 1994b). Each model includes six helical domains (I–VI) radiating from a central core. Domains V and VI are most diagnostic of *Euglena* group II introns. Domain V is a potential catalytic domain, with the conserved pairing 5′-(A/U)AGCU .... (A/G)GUUU-3′ at the base of the lower stem, and a 5′-(A/U)G-3′ bulge. Each domain VI contains the adenosine residue involved in lariat formation at position −7 or −8 from the 3′-splice site (Michel et al., 1989; Lambowitz & Belfort, 1993). In addition to the helical domains, several potential tertiary interactions believed to be involved in splicing are suggested, including the exon binding site I (EBSI) in the subdomain Diii loop of domain I for pairing with IBS1 of the 5′-exon; EBSII in the subdomain Di bulge of domain I for potential pairing with IBSII of the 5′-exon (Michel & Jacquier, 1987); the γ-γ' interaction involved in recognizing the 3′ splice site (Michel & Jacquier, 1987; Jacquier & Michel, 1990); the ϵ-ϵ' interaction involved in locating the 5′ splice site (Jacquier & Michel, 1990); and the α-α' interaction between domains IB and domain ID (Harris-Kerr et al., 1993). Note that the γ-γ', ϵ-ϵ', and α-α' interactions are not



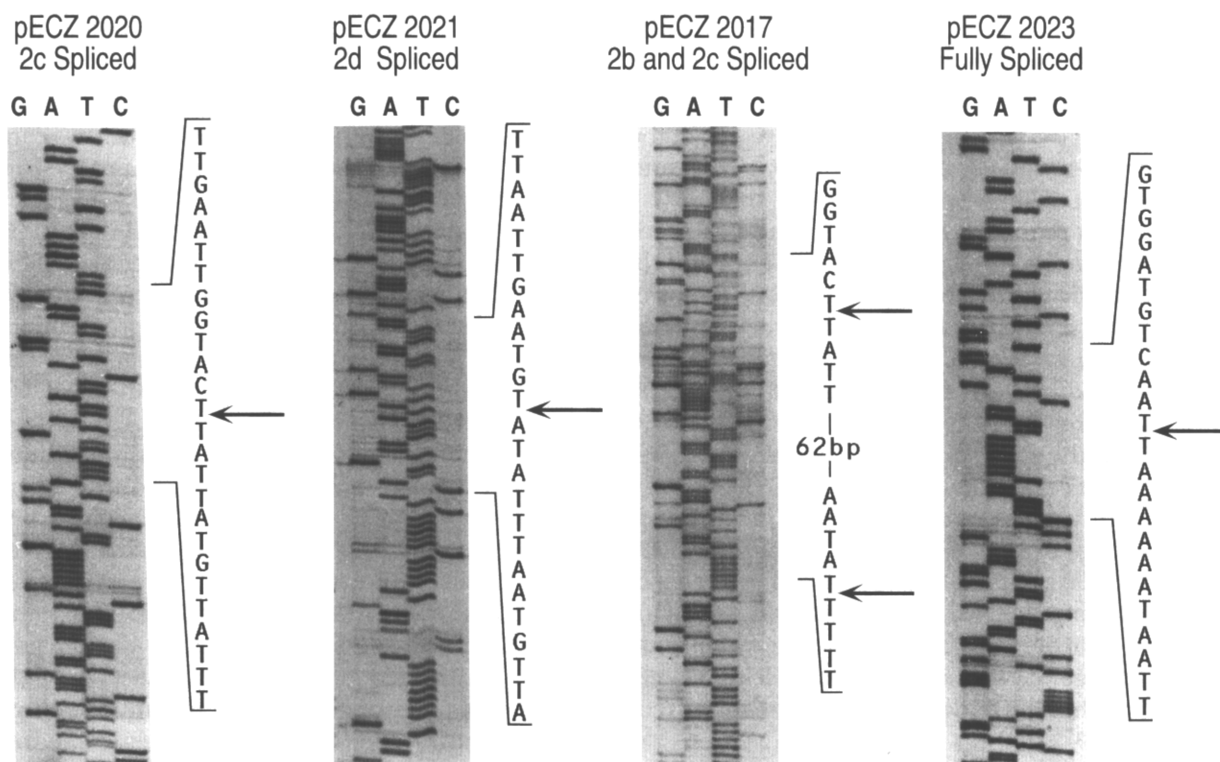


FIGURE 3. Characterization of three internal introns by cDNA cloning and sequencing. **A:** Schematic diagram of the four PCR fragments cloned and sequenced. Names of the plasmids containing the PCR fragments are listed at right, with a description of the cDNA. pECZ2020 is obtained from PCR p4-c6, pECZ2021 is from PCR p6-c7, pECA2017 is from PCR p2-c7, and pECZ2023 is from PCR p1-c10. **B:** Sequencing data showing the splice junctions of introns (from left to right) 2c, 2d, 2b, and 2c, and 2a. Splice sites are indicated by arrows. External intron 2a or *psbC* exon sequence flanking each intron is shown to the right of each panel.

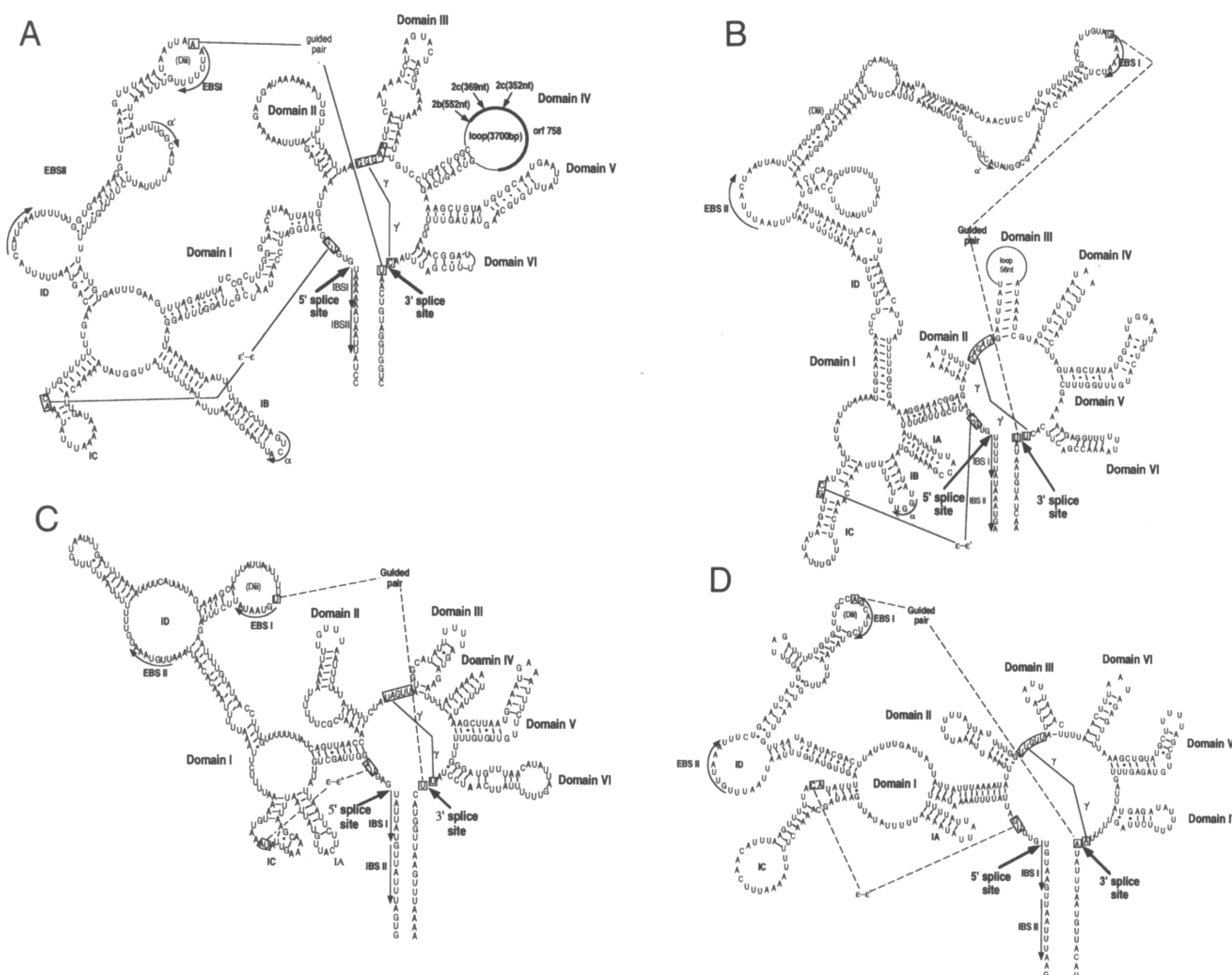


FIGURE 4. Secondary structure model for *psbC* intron 2 external intron 2a and internal introns 2b, 2c, and 2d. The secondary structures of *psbC* intron 2a (**A**), 2b (**B**), 2c (**C**), and 2d (**D**) are based on the model proposed by Michel et al. (1989) and on comparative analysis of other *Euglena* group II introns (Hong & Hallick, 1994b). Orf 758 is located in the loop of domain IV of intron 2a. Insertion sites of the three internal introns are all within the loop of domain IV of intron 2a and are indicated by arrows. The 5′ and 3′ splice sites of each intron are indicated by thin arrows. IBS(I/II)–EBS(I/II) pairing regions are indicated by thin arrows. Possible nucleotides in the guided pair, γ-γ' and ϵ-ϵ' interacting regions are enclosed in boxes and hypothetical tertiary interactions are shown as dotted or solid lines. The α-α' interactions are indicated by thin arrows.

well-conserved in these introns. As is typical of *E. gracilis* chloroplast group II introns, secondary structures and tertiary interactions may be abbreviated when compared to self-splicing group II introns from other genomes (Copertino & Hallick, 1993).

The helical stem of domain IV of external intron 2a, the largest of the four introns, has more extensive pairing than in most other *Euglena* group II introns. All three internal introns, as well as the orf, are within domain IV of intron 2a.

psbC intron 2 splicing pathway is partially ordered

If splicing of the three internal introns, 2b, 2c, and 2d, were unordered, PCR amplification of cDNA using primer pairs p2-c9 or p4-c7 should result in eight products. However, only five products were detected (Fig. 2, lanes 4, 6). Splicing intermediates lacking either intron 2c or 2d alone are similar in size, and would not be resolved. Similarly, intermediates lacking both intron 2c and 2b or 2d and 2b would not be resolved either. These could account for two of the missing products. Also missing was an intermediate corresponding in size to a pre-mRNA with intron 2b excised and 2c and 2d retained. During DNA sequence analysis of partially excised intermediates, pre-mRNAs with 2d excised, 2d and 2c excised, 2d and 2b excised, and all three introns excised were identified. No products with 2d retained and 2b excised were found. To directly test if removal of intron 2d precedes excision of intron 2b, cDNA was synthesized using primer c12, which crosses the 3′ splice junction of intron 2d (Fig. 5A), and was PCR amplified

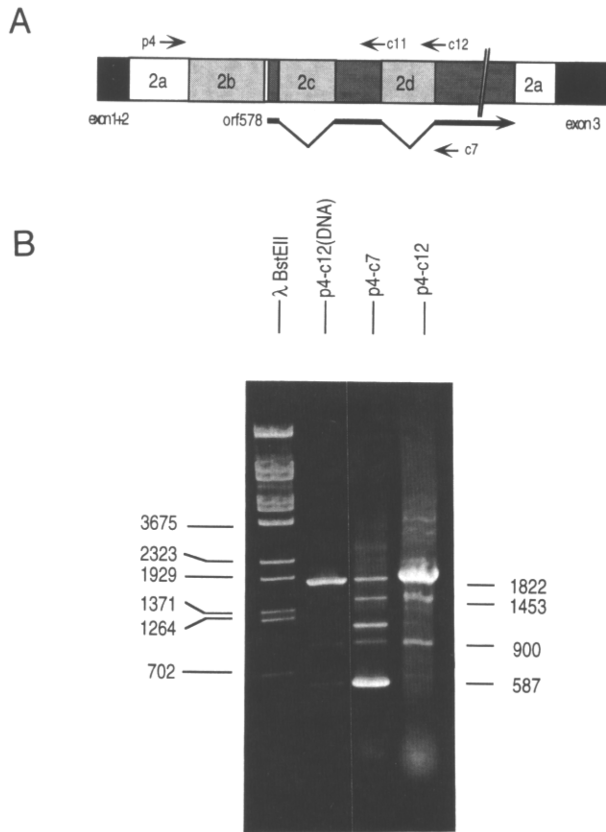


FIGURE 5. **A:** Dominant processing pathway of the three internal introns. Primer c12 crosses the 3′ splice junction of intron 2d. Primer c11 crosses 5′ splice junction of intron 2d. **B:** Control PCR with *E. gracilis* chloroplast total nucleic acid is labeled "DNA." λ *Bst*E II molecular weight markers are shown in the left lane and the sizes (in bp) are shown on the left of the gel. Sizes of some PCR products are shown on the right. Location of primers used for cDNA synthesis and PCR amplification are shown on the *psbC* map in A.

with primers c12 and p4 (Fig. 5). All products amplified in this reaction must contain 2d. Products corresponding to the unexcised mRNA (1,822 nt), intron 2c excised (1,453 nt), and introns 2b and 2c excised (900 nt) were obtained. No intermediate with only 2b excised was detected. In addition, relative to the splicing intermediates, much more of the unspliced mRNA was observed than in previous reactions. A partially ordered RNA splicing pathway for the three internal introns, consistent with the intermediates detected by RT-PCR and DNA sequence analysis is shown in Figure 6. Introns 2c and 2d are preferentially spliced prior to 2b. The splicing of 2b prior to both 2c and 2d was not observed.

Internal introns of *psbC* interrupt an orf encoding a putative maturase

Our initial description of the DNA sequence of the *psbC* operon included a 4,143 nt *psbC* intron 2 with orfs of 177 and 241 codons (Hallick et al., 1993). This sequence has since been corrected to 4,144 nt (EMBL Accession X70810 Release 37, correction at position 17874). In the revised sequence, orfs 177 and 241 are part of a continuous orf of 635 amino acids. After splicing of introns 2b, 2c, and 2d, the orf is extended to 758 amino acids. Intron 2b is located between a putative ribosome binding site and a methionine codon at the beginning of orf 758 (Fig. 1). Intron 2c is present 57 nt into orf 758 between the codons for amino acids 19 and 20. Intron 2d is located 287 nt into orf 758 and splits the codon for amino acid 95. *psbC* intron 2 is the first example of an intron-encoded gene being interrupted by introns.

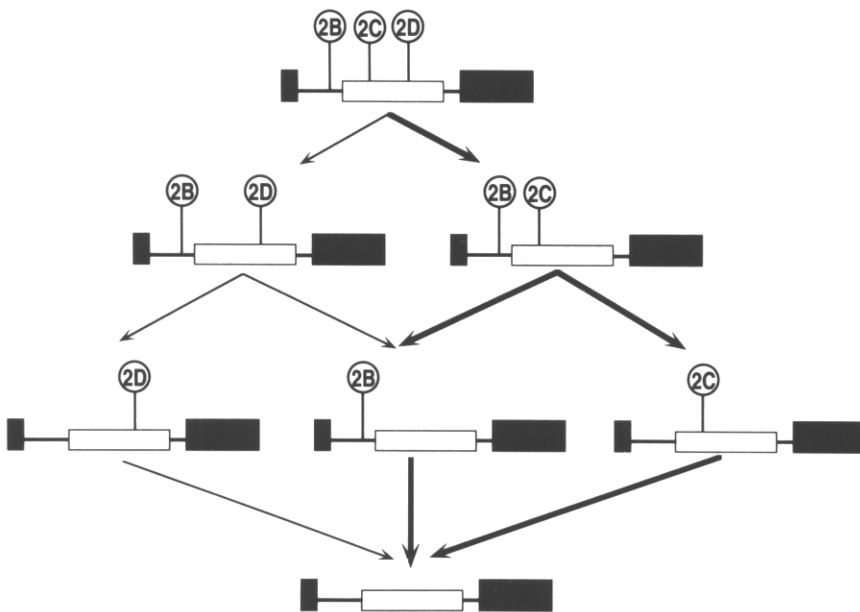


FIGURE 6. Predominant pathway of *psbC* intron 2 mRNA processing. A model of the predominant processing pathway of the three internal *psbC* introns are shown. Black boxes represent exons 2 and 3; solid lines represent intron 2a. White open box represents orf 758. Internal introns 2b, 2c, and 2d are circled. Thick arrows label the preferential pathways.

Orf 758 would encode a protein with a predicted size of 93 kDa and is composed of 25% aromatic amino acids. The predicted overall charge of orf 758 is +81 with a pI of 10.69. High positive charge and aromatic amino acid content may be indicative of RNA-binding activity. In fact, orf 758 has previously been identified as maturase-like based on the presence of an X domain (Mohr et al., 1993). Many maturase-like proteins also include reverse transcriptase domains I–VII, where domains V–VII are usually the best conserved. To search for domains V–VII in orf 758, the amino acid sequence was aligned visually with a chloroplast group II intron maturase (*matK*) consensus sequence (Fig. 7A) and a general reverse transcriptase consensus (Fig. 7B) (Xiong & Eickbush, 1990). In both comparisons, characteristic motifs of YVRY in domain V and FLG in domain VII could be aligned. In addition, the alignments could be extended to many positions where a class (such as hydrophobic) as opposed to a particular amino acid was specified in the consensus. Because orf 758 is located within domain IV of a group II intron, the canonical position of intron-encoded maturases, and contains maturase-like domain X and reverse transcriptase-like domains V–VII, it has been designated *mat2*. Reverse transcriptase domains I–IV, domain Z, and zinc-finger domains could not be identified in *mat2*.

DISCUSSION

A new category of twintron

psbC intron 2 is a group II intron (2a) encoding the 758-amino acid *mat2* locus. *mat2* is interrupted by three additional group II introns. Two of the internal introns (2c, 2d) are within the coding region of *mat2* . The third internal intron (2b) is located between a potential ribosome binding site and the start codon for *mat2*. The potential ribosome binding site of 5′-GUAGU, centered at −18 upstream of *mat2* ATG codon 1 is complementary in 4/5 positions to the 3′-CCUCA-5′ 3′-end of *Euglena* chloroplast 16S rRNA. Intercistronic introns have been reported previously (Barkan, 1988; Stevenson et al., 1991), however, this is the first example of introns within an intron-encoded maturase, and the first example of a group II intron in a 5′-nontranslated leader sequence (Fig. 1). As a complex twintron with an internal gene, *psbC* intron 2 represents a new type of twintron.

A partially ordered *psbC* intron 2 splicing pathway

The partially ordered *psbC* intron 2 internal introns splicing pathway, with preferential initial splicing of intron 2d, was an unanticipated result. Internal group II introns in other *Euglena* chloroplast genes normally splice via independent, unordered events (Koller et al., 1985). Intron 2d may interfere with the normal folding of introns 2b and 2c, or keep splicing factors from contacting the introns. It is possible that only 2d is correctly folded and exposed to splicing factors in the unspliced pre-mRNA. Splicing of 2d may allow a shift in the tertiary structure, allowing 2b and 2c to splice in an unordered sequence.

Excised, intact 4,144 nt *psbC* intron 2 mRNA was not detectable by northern hybridization with intron-specific probes (data not shown). This is a preliminary indication that the external intron might not excise prior to splicing of internal introns. Unspliced domain IV, at 3,707 nt, is possibly too large to fold properly and may interfere with the folding of the entire intron.

mat2 may be required for splicing

Why are internal introns 2b, 2c, and 2d maintained? It has been proposed that the internal introns of twintrons maintain their ability to splice as a species evolves because they interrupt functional domains of the exter-

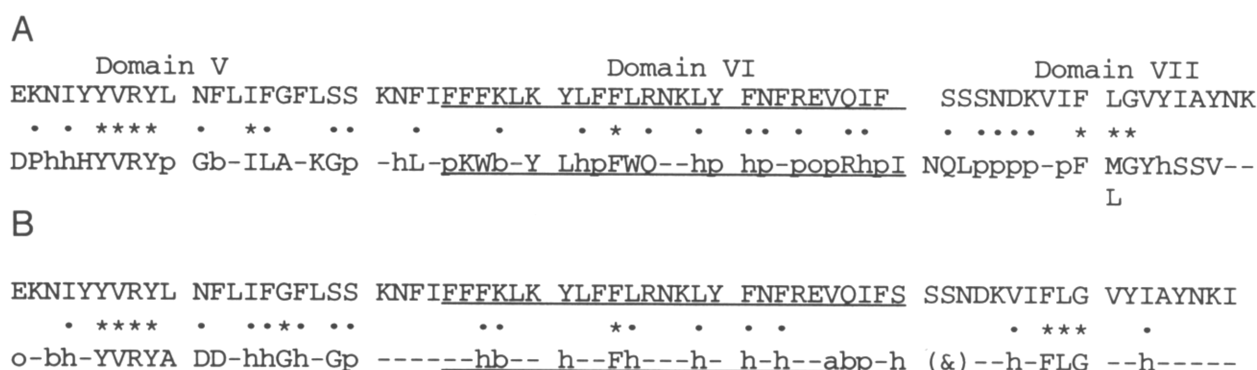


FIGURE 7. Comparison of the reverse transcriptase domains V–VII from orf 758 (top line) with the chloroplast consensus (**A**, bottom line) and general consensus (**B**, bottom line) are shown. Domain VI is underlined. A variable sized gap of 4–14 amino acids occurs in the general consensus between domains VI and VII (&). A gap of two amino acids occurs between domains VI and VII in the orf 758 alignment in A. Dot, similar amino acids; asterisk, identical amino acids; dash, positions with no consensus amino acid. Classes of amino acids are as follows: h, hydrophobic; p, polar; b, basic; a, acidic; o, aromatic.

nal intron. If the internal intron lost the ability to splice, so would the external intron, and, as a consequence, the host gene would not be expressed (Copertino & Hallick, 1993). It is unlikely that internal introns 2b, 2c, and 2d interrupt any essential RNA structure. In the secondary structure model of *psbC* intron 2a, the three internal introns are located in the loop of domain IV. No RNA structural elements required for intron excision have been identified in domain IV (Michel et al., 1989; Copertino & Hallick, 1993; Lambowitz & Belfort, 1993). Experiments also showed that deletion of domain IV did not block group II self-splicing in vitro (Hebbar et al., 1992; Koch et al., 1992). *Trans*-splicing of introns segmented within domain IV occurs in chloroplasts in vivo (Bonen, 1993). Therefore, the three internal introns in domain IV may not be sufficient to block the excision of intron 2a at the RNA level.

Alternatively, besides the additive effect on intron size, the internal introns 2b, 2c, and 2d could be viewed as interrupting the coding region of an essential protein. Splicing of intron 2b, 2c, and 2d may be a prerequisite for expression of *mat2*, a possible intron splicing factor.

Possible functions of *mat2*

psbC intron 2 is one of three *E. gracilis* chloroplast introns encoding internal maturase-like polypeptides. Intron-encoded genes for maturase-like proteins are known in group II introns of both mitochondria and chloroplasts. Some of these proteins have been shown to mediate splicing or potentially contribute to intron mobility (Carignani et al., 1983; Moran et al., 1995). The group II intron-encoded maturase-like proteins comprise a subgroup of retroelements (Xiong & Eickbush, 1988; Doolittle et al., 1989) and are characterized by the presence of one or more domains found in retroelements including Z, reverse transcriptase, X, and zinc finger (Mohr et al., 1993). The X domain is a potential RNA-binding domain, which has been found in all group II intron-encoded maturases or maturase-like proteins. The zinc finger domain may be a remnant of a retroviral endonuclease and recently has been shown not to be necessary for maturase activity (Zimmerly et al., 1995). Only the X and reverse transcriptase domains have been shown to be required for maturase activity (Moran et al., 1994, 1995). Significantly, an X domain and reverse transcriptase domains V–VII have been identified in *mat2*.

The *E. gracilis mat2* protein diverges significantly from other group II intron-encoded maturase-like proteins. Only reverse transcriptase domains V–VII have been identified, and the X domain also differs from the consensus. There are also significant differences between the mitochondrial and chloroplast group II intron-encoded maturase-like proteins. For example, the mitochondrial reverse transcriptase domain V consensus is YVRYADD, and the YADD motif is required for activity in retroviral reverse transcriptases. In chloroplasts, the consensus is YVRY. Loss or modification of domains may be due to differences in substrates, or changes in function. The term "maturase activity" has not been defined rigorously, but is generally used to describe any enhancement or enablement of splicing activity in maturase-dependent introns. It is possible that maturase-like proteins serve different roles in different organisms.

As a potential maturase, *mat2* is a prime candidate for playing a role in splicing one or more *Euglena* chloroplast group II introns. Known group II intron-encoded maturases act specifically on the intron that encodes them or closely related introns (Carignani et al., 1983, 1986). However, of the 155-plus *Euglena* chloroplast introns, only 3 encode maturase-like proteins. It is possible that one result of divergence from consensus domains is more general maturase activity. Functional assays should provide additional information regarding the activity and specificity of *mat2*.

Evolution of *mat2*

Many intriguing questions about the origin and spread of group II and group III introns in genus *Euglena* remain. Our working hypothesis is that, at the DNA level, the chloroplast genome of genus *Euglena* arose from an intron-less common ancestor with the chloroplast genome of other photosynthetic eukaryotes, including land plants (Thompson et al., 1995). In this model, also suggested by others (Lambowitz & Belfort, 1993), ancestral introns were mobile, retrotransposable genetic elements that invaded the genome from another organism, relying in part on internally encoded enzyme activities for mobility. Possibly a single "founder" intron arrived via intergenomic (and interspecies) transfer, and was subsequently propagated intragenomically, either into other genes, or into introns or intragenomic spacers. Introns of contemporary *Euglena* species have subsequently lost mobility, but retained the ability to be spliced. To explore some of these questions, an analysis of *mat2* in evolutionarily diverse species of the genus *Euglena* has been initiated. The first significant result is that an intron homologue encoding *mat2* is present in a similar *psbC* group II intron of *E. viridis* (L. Zhang & R.B. Hallick, unpubl. obs.). *E. viridis* lacks the majority of group II and group III introns of *E. gracilis*, but does contain a homologue of *E. gracilis psbC* intron 2.

MATERIALS AND METHODS

RNA isolation

E. gracilis chloroplasts were isolated and RNA was purified from isolated chloroplasts as described previously (Hallick et al., 1982).

cDNA synthesis

Primers were synthesized by the Midland Certified Reagent Company or the University of Arizona Biotechnology Center. Seven synthetic deoxynucleotide primers complementary to the RNA-like strand were used for cDNA synthesis. cDNA primer c10, (5′-GCAAATCCAGTTGATTCCTG-3′, coordinates 17895–17914, EMBL accession 70810) is complementary to the RNA-like sequence of *psbC* exon 3; primer c9 (5′-CCACC TACAGTTAGTTAAATCG-3′, coordinates 17866–17887) is complementary to the RNA-like sequence at the *psbC* intron 2a–exon 3 junction; primer c8 (5′-GCACATACAGCTTTCTG ACTGAC-3′, coordinates 17802–17824) is complementary to the RNA-like strand at the 3′ end of *psbC* intron 2a and includes the hypothetical 3′ stem of domain IV of the external intron 2a; primer c6 (5′-CCAACCACATTTTAAACTTAACC-3′, coordinates 15148–15170) and primer c7 (5′-GGACACTGCT TACAAAAAAATGG-3′, coordinates 15817–15839) are complementary to the RNA-like sequence at the 5′ end of orf 758.

In order to determine the splicing pathway, primer c11 (5′-ATAATCACACACATTCAATTAAATTCATGC-3′, coordinates 15412–15441) and primer c12 (5′-GATGTAACATTAA ATATTAAAAACTAAG-3′, coordinates 15773–15800) were designed to cross the junction of the intron 2d 5′ and 3′ splice sites, respectively.

cDNA was synthesized using 200 ng of cDNA primer and 10 μg of purified total chloroplast RNA as template. Reactions were conducted as described previously (Copertino & Hallick, 1991).

PCR analysis, cloning, and sequencing

Oligos for PCR amplification are as follows: primer p1 (5′-GT GGAAACGCTCTTTAATAAAAAAT-3′, coordinates 13163–13176 + 13720–13730) is RNA-like and covers *psbC* exons 1 and 2; primer p2 (5′-CGCTCTTTAATAAAAAATGTGTGGC-3′, coordinates 13169–13176 + 13720–13737) contains *psbC* exons 1 and 2 and crosses the exon 2–intron 2a junction; primer p3 (5′-TAATAAAAAATGTGTGGCATGG-3′, coordinates 13720–13741) is located at the *psbC* exon 2–intron 2 junction; primer p4 (5′-GTTTAGATTTATCCGCTTTGGG-3′, coordinates 13979–14000) and primer p5 (5′-GGTAAAAATTAATTGTCCT GACTGGC-3′, coordinates 14087–14112) are located at the 5′ end of intron 2. P5 corresponds to the hypothetical 5′ stem of domain IV of external intron 2a. Primer p6 (5′-GGTTAAG TTTAAAATGTGGTTGG-3′, coordinates 15148–15170) is the reverse complement of c6. Primer p7 (5′-CCATTTTTTTGTAA GCAGTGTCC-3′, coordinates 15817–15839) is the reverse complement of c7.

Various combinations of PCR and cDNA primers were used to amplify synthetic cDNA and, as a control, *E. gracilis* chloroplast total nucleic acid. All reactions were incubated at 80 °C for 3 min and 3 units of Taq polymerase (Perkin-Elmer) were added to each reaction. Amplification cycles consisted of 94 °C for 1 min, 50 °C for 2 min, and 72 °C for 3 min for 25 cycles. PCR products were gel purified and cloned into the ddT-tailed *Eco*R V site of *Bluescript* KS(−) vector(Stratagene) (Kovalic et al., 1991). Plasmids containing inserts were sequenced using the Sequenase kit (USB).

Computer analysis of orf 758

Multiple alignments were performed with PILEUP and PRETTY from the GCG Wisconsin Package (Genetics Computer Group, 1994). Chloroplast *matK* protein sequences from *Pinus contorta* (P24685), *P. thunbergii* (Q00866), *Oryza sativa* (P12175), *Secale cereale* (JN0302), *Nicotiana tabacum* (P12176), *Solanum tuberosum* (P32088), *Saxifragia integrifolia* (P36436), *Pisum sativuum* (S08056), *Sinapis alba* (P09364), and *Hordeum vulgaris* (S28765) were aligned. In addition, the Swissprot, GenBank, and EMBL databases were searched for homologous proteins according to Smith and Waterman on Genquest (Devereux et al., 1984).

Received October 26, 1995; returned for revision November 17, 1995; revised manuscript received November 28, 1995

REFERENCES

Barkan A. 1988. Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *EMBO J 7*:2637–2644.

Bonen L. 1993. *Trans*-splicing of pre-mRNA in plants, animals, and protists. *FASEB J 7*:40–46.

Carignani G, Groudinsky O, Frezza D, Schiavon E, Bergantino E, Slonimski PP. 1983. An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell 35*:733–742.

Carignani G, Netter P, Bergantino E, Robineau S. 1986. Expression of the mitochondrial split gene coding for cytochrome oxidase subunit I in *S. cerevisiae*: RNA splicing pathway. *Curr Genet 11*: 55–63.

Copertino DW, Christopher DA, Hallick RB. 1991. A mixed group II/group III twintron in the *Euglena gracilis* chloroplast ribosomal protein S3 gene: Evidence for intron insertion during gene evolution. *Nucleic Acids Res 19*:6491–6497.

Copertino DW, Hall ET, Van Hook FW, Jenkins KP, Hallick RB. 1994. A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Res 22*:1029–1036.

Copertino DW, Hallick RB. 1991. Group II twintron: An intron within an intron in a chloroplast cytochrome *b*-559 gene. *EMBO J 10*: 433–442.

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.

Copertino DW, Shigeoka S, Hallick RB. 1992. Twintrons: Introns-within-introns in the chloroplast genes of *Euglena gracilis*. *EMBO J 11*:5041–5050.

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, Johnson MS, McClure MA. 1989. Origins and evolutionary relationships of retroviruses. *Quart Rev Biol 64*:1–30.

Drager RG, Hallick RB. 1993. A complex twintron is excised as four individual introns. *Nucleic Acids Res 21*:2389–2394.

Genetics Computer Group I. 1994. *Program manual for the Wisconsin package, version 8 (September)*. 575 Science Drive, Madison, Wisconsin 53711, USA.

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.

Hallick RB, Richards OC, Gray PW. 1982. Isolation of intact, superhelical chloroplast DNA from *Euglena gracilis*. In: M. Edelmann R, Hallick B, Chua NH, eds. *Methods in Chloroplast Molecular Biology*. New York: Elsevier Biomedical. pp 281–294.

Harris-Kerr CL, Zhang M, Peebles CL. 1993. The phylogenetically predicted base-pairing interaction between α and α' is required for group II splicing in vitro. *Proc Natl Acad Sci USA 90*:10658–10662.

Hebbar SK, Belcher SM, Perlman PS. 1992. A maturase-encoded group II intron of yeast mitochondria self-splices in vitro. *Nucleic Acids Res 20*:1747–1754.

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

Hong L, Hallick RB. 1994b. A group III intron is formed from domains of two individual group II introns. *Genes & Dev 8*:1589–1599.

Jacquier A, Michel F. 1990. Base-pairing interactions involving the 5′ and 3′-terminal nucleotides of group II self-splicing introns. *J Mol Biol 213*:437–447.

Koch JL, Boulanger SC, Dib-Hajj SD, Hebbar SK, Perlman PS. 1992. Group II intron deleted for multiple substructures retain self-splicing activity. *Mol Cell Biol 12*:1950–1958.

Koller B, Clarke J, Delius H. 1985. The structure of precursor mRNAs and of excised intron RNAs in chloroplasts of *Euglena gracilis*. *EMBO J 4*:2445–2450.

Kovalic D, Kwak J, Weisblum B. 1991. General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Res 19*:4560.

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

Lambowitz AM, Perlman PS. 1990. Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem Sci 15*:440–444.

Michel F, Jacquier A. 1987. Long-range intron–exon and intron–intron pairings involved in self-splicing of class II catalytic introns. *Cold Spring Harbor Symp Quant Biol 52*:201–212.

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 PS, 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.

Moran JV, Mecklenberg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P. 1994. Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: Initial definition of the maturase domain of the group II intron aI2. *Nucleic Acids Res 22*:2057–2064.

Moran JV, Zimmerly S, Eskes R, Kennell JC, Lambowitz AM, Butow RA, Perlman PS. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol 15*:2828–2838.

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

Stevenson JK, Drager RG, Copertino DW, Christopher DA, Jenkins KP, Yepiz-Plascencia G, Hallick RB. 1991. Intercistronic group III introns in polycistronic ribosomal protein operons of chloroplasts. *Mol Gen Genet 228*:183–192.

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.

Xiong Y, Eickbush T. 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. *Mol Cell Biol 8*:114–123.

Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J 9*:3353–3362.

Zimmerly S, Guo H, Perlman PS, Lambowitz AM. 1995. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell 82*:545–554.