Recent horizontal intron transfer to a chloroplast genome

Elena V. Sheveleva and Richard B. Hallick*

Department of Biochemistry and Molecular Biophysics, The University of Arizona, 1041 East Lowell Street, Tucson, AZ 85721-0088, USA

Received September 22, 2003; Revised and Accepted December 18, 2003

ABSTRACT

Evidence is presented for the recent, horizontal transfer of a self-splicing, homing group II intron from a cyanobacteria to the chloroplast genome of Euglena myxocylindracea. The psbA gene of E.myxocylindracea was found to contain a single 2566 nt group II intron with a gene in domain 4 for a 575 amino acid maturase. The predicted secondary structure and tertiary interactions of the group II intron, as well as the derived maturase primary sequence, most closely resemble the homing intron of the cyanobacterium Calothrix and the rnl introns of Porphyra purpurea mitochondria, while being only distantly related to all other Euglena plastid introns and maturases. All main functional domains of the intron-encoded proteins of known homing introns are conserved, including reverse transcriptase domains 1–7, the zinc finger domain and domain X. The close relationship with cyanobacterial introns was confirmed by phylogenetic analysis. Both the full-length psbA intron and a Δ-maturase variant self-splice in vitro in two independent assays. The psbA intron is the first example of a self-splicing chloroplast group II intron from any organism. These results support the conclusion that the psbA intron is the result of a recent horizontal transfer into the E.myxocylindracea chloroplast genome from a cyanobacterial donor and should prompt a reconsideration of horizontal transfer mechanisms to account for the origin of other chloroplast genetic elements.

INTRODUCTION

Group II introns are found in prokaryotic mRNA and in tRNA, mRNA and rRNA of organelles in fungi, plants and protists. Group II introns are ribozymes. They have a conserved secondary structure, first proposed by Michel as a central core with six radiating helical domains, designated d1–d6 (1; see review in 2). The catalytic center is formed by d1 and d5. Several tertiary interactions are involved in the stabilization of the catalytic core, including EBS1/IBS1, EBS2/IBS2, EBS3/IBS3, α/α’, β/β’, ε/ε’, δ/δ’ and γ/γ’. In addition to Watson–Crick pairings, an 11 nt RNA motif ζ (AUGG ... CCUAA) that is conserved in d1 of group II self-splicing introns interacts with the domain V hairpin terminal loops with a GNRA consensus sequence ζ’. Similarly, in the η/η’ interaction a GNRA hairpin loop that caps d6 interacts with another 11 nt motif within d2 (see review in 3).

Euglenoid chloroplast group II introns have the same core secondary structure as other group II introns (4,5), confirmed by comparative analysis of cognate introns from two or more euglenoid species (6,7). Euglenoid group II introns splice by the same two-step mechanism as other group II introns (8). Nevertheless, these introns comprise a class distinct from their counterparts in other organelles and prokaryotes (see review in 9). Many are considerably shorter (252–671 nt) than typical group II introns and lack some expected tertiary contacts. More variability occurs in the catalytic d5 domain than in more typical group II introns. Plastid genomes of all 17 Euglena spp. examined to date also contain a unique class of chloroplast introns designated group III (9), a streamlined derivative of group II introns. Group III introns have a narrow size range of 73–119 nt, lack the group II intron-like d2–d5 domains and splice via a lariat intermediate utilizing the unpaired adenosine (A*) within d6 (10). The 5’-region of some group III introns is somewhat similar to domain ID3 of group II introns.

A small subset of group II introns, none of which are from chloroplast genomes, have ‘homing’ activity, the ability to move horizontally into an intron-less allele of the same genome or into a new genome. Subsequent intragenomic homing could result in vertical evolution among direct descendants of the ancestral genome. A detailed understanding of group II intron homing has come from studies of the group II intron a12 of yeast mitochondria and L1.Ltr B of Lactococcus lactis (11,12). The L1.Ltr B intron RNA reverse splices directly into the sense strand of a double-stranded DNA target site, while the intron-encoded reverse transcriptase/maturase cleaves the antisense strand and uses it as primer for reverse transcription of the inserted intron RNA (13–15). The ability of group II introns to spread into ectopic sites has also recently gained additional experimental support (16) and is an attractive model for the introduction of splicing introns into the eukaryotic nucleus (16,17). Intron homing was likely important for the proliferation of group II introns into the genomes of bacteria and the organelle genomes of lower eukaryotes and plants. There is little evolutionary evidence for
this process, as most group II introns in eukaryotes have lost homing activity and an intron-encoded protein. The phylogenetic relationships and evolutionary history of group II introns with reverse transcriptase genes have been reported (3,5,18,19).

Although chloroplasts may have acquired group II introns by horizontal transfer from a bacterial origin, none of the current models envision recent evolutionary transfer into a chloroplast genome. Furthermore, chloroplast intron-encoded proteins (IEPs) are all evolutionarily distinct from their counterparts in mobile introns and are thought to be the result of vertical evolution. Plastid genomes originated by endosymbiosis of a cyanobacterium. After the discovery of introns in cyanobacteria, these photosynthetic prokaryotes were considered the source of introns in chloroplast genomes (20–22). More important, horizontal gene transfer into chloroplast genomes from other organisms was believed to be unlikely (23–27). The acquisition of new introns and insertion into new sites during vertical evolution could be explained by functional intragenomic mobility (homing) of group I and II introns (28). There are contradictions to vertical chloroplast evolution, such as the absence of chloroplast introns in some algal chloroplast genomes considered the predecessors of plant chloroplast genomes (29–31). In this study we present evidence for recent transfer of a homing intron from a cyanobacterium to the chloroplast genome of *Euglena myxocylindracea*. This result may exemplify a more generalized ability of the chloroplast genomes of some species to acquire other types of genetic information within plasmids, viruses and transposable elements via horizontal DNA transfer.

**MATERIALS AND METHODS**

**Euglena cultures and nucleic acid extraction**

*Euglena myxocylindracea* (UTEX 1989) was obtained from the University of Texas Culture Collection. Total nucleic acid (TNA) extracts were prepared as previously described (32) either directly from cultures obtained from the UTEX culture collection grown on solid slants or following additional collection grown on solid slants or following additional

Intron and intron-encoded maturase sequences were isolated from *E. myxocylindracea* by PCR amplification from primers targeted to the *psbA* gene. A sample of 0.05–0.2 µg/ml of nucleic acid extract (TNA) was amplified from the synthetic oligonucleotides P1 and P2. Primer P1 corresponds to coordinates 559–581 (5′-GGACCTTACCAGTTAATTGT-ATG) and primer P2 corresponds to coordinates 3598–3621 (5′-AAGAAGAAAATGTAAGAACGGAGA) of the *E. myxocylindracea* *psbA–psbK* operons (accession no. AY290861). Exon–intron boundaries were determined by RT–PCR followed by DNA sequencing with the same primers. Superscript I (Invitrogen Inc.) was used for reverse transcription.

**Recombinant plasmids**

Plasmid PEYC 2009 contains 226 nt from the 5′-exon, the 2566 nt group II intron containing *mat4* and the first 271 nt of the 3′-exon of *psbA* from *E. myxocylindracea*. The plastid sequence was cloned distal to the phage T3 promoter in pBSKS+ as an stl1–stl2 insert. PEYC 2010 contains the same 5′-exon, 173 nt from the 3′-exon and *mat4* with a large internal deletion leaving only 62 nt, also cloned behind the phage T3 promoter in pBSKS+.

**In vitro transcription**

**In vitro** radiolabeled transcripts were synthesized from 1 µg linearized plasmids using transcription buffer, 3 µCi/µl [α-32P]UTP (300 Ci/mmol), 100 µM unlabeled UTP, 500 µM each of the other NTPs and MAXIscript T3 RNA polymerase (Ambion Inc.). Transcripts were gel purified on a denaturing 4% polyacrylamide gel. RNA was excised and soaked overnight at 4°C in 500 mM NH4Ac, 10 mM MgCl2, 0.1 mM EDTA, 0.1% SDS. To obtain non-radiolabeled transcripts MEGAscript T3 (Ambion Inc.) was used. For in *vitro* splicing reactions both labeled and non-labeled transcripts were ethanol precipitated and then dissolved in buffer containing 40 mM MOPS, 1 mM EDTA, 1 mM dithiothreitol, 0.5 U/µl RNase inhibitor. Prior to splicing, 200 nM RNA was renatured by heating to 90°C for 1 min and then immediately diluted into assay buffer at 30°C (33).

**Self-splicing reactions in vitro**

Transcripts were incubated for the indicated times at 42°C in 40 mM MOPS, 10 mM MgCl2, 1 mM dithiothreitol, 1 µl RNase inhibitor and 500 mM (NH4)2SO4. Splicing products of radioactive transcripts were analyzed on a denaturing 4% polyacrylamide gel, followed by autoradiography. Splicing products of non-radioactive transcripts were gel purified and analyzed by RT–PCR followed by DNA sequencing.

**Computer analysis**

Sequences similar to *mat4* were identified by BLAST searches (34). Open reading frame analysis and determination of putative amino acid sequences were done with the computer program DNAStrider. Nucleotide and protein sequence alignments were carried out using the PILEUP and Clustal X programs (Genetics Computer Group Sequence Analysis Package, version 8.0, Madison, WI). RNA folding was done mostly manually and using the mFold program (35).

**Phylogenetic analysis**

The amino acid sequences of IEPs were aligned by PILEUP and Clustal X. Phylogenetic trees on 19 amino acid sequences were constructed using the neighbor joining and parsimony analysis default settings of the program PAUP and two amino acid sequences were taken as an outgroup (36). Bootstrap values were from 1000 re-samplings. The following representatives were selected for analysis (gene name or locus tag is followed by species name, with accession no. in parentheses). Chloroplast genome: *mat4*, *E. myxocylindracea* (AY290861); unnamed, *Scenedesmus obliquus* (P19593). Cyanobacteria: ORF2, *Calothrix* sp. (CAA50529); alr8560, *Nostoc* sp. PCC 7120 (BAB77479); thr0620, *Thermosynechococcus elongatus* BP-1 (NP_681409); gil0177, *Gloeobacter violaceus* (NP_923123); tery0355, *Trichodesmium erythraeum* (NP_00071211); all5206, *Nostoc* sp. PCC 7120 (NP_489246). Other bacteria: *Avin1823*, *Azotobacter vinelandii* (ZP_00090141); pX01-07, *Bacillus anthracis*
RESULTS

Identification of a novel group II intron in the *E. myxocylindracea* psbA gene

The psbA gene of *Euglena gracilis* was previously reported to contain four ‘euglenoid’ type group II introns. During a survey of several additional *Euglena* spp. to determine if the *E. gracilis* psbA introns are conserved, a single intron was discovered in the chloroplast psbA gene of *E. myxocylindracea* (Fig. 1). The intron insertion site, unique to *E. myxocylindracea*, is located within exon 3 in the corresponding *E. gracilis* psbA gene. Six other *Euglena* spp. surveyed lacked both the four *E. gracilis* psbA introns and the novel *E. myxocylindracea* intron. The intron and flanking coding regions were amplified by PCR, cloned and sequenced on both strands (accession no. AY290861). The new intron is 2566 nt in length. Encoded in the intron is an ORF of 1728 nt that would encode a protein of 575 amino acids.

Structure of the intron

The *E. myxocylindracea* intron has typical group II intron secondary structure with six helical domains (dI–dVI) radiating from a central core (Fig. 2). The intron-encoded protein gene, located within dIV as expected, is designated *mat4*, being the fourth example of a *Euglena* chloroplast intron-encoded protein. All expected tertiary interactions between introns and exons (EBS1/IBS1, EBS2/IBS2 and EBS3/IBS3) and within the intron (α/α′, β/β′, ε/ε′, δ/δ′, γ/γ′, ζ/ζ′ and η/η′) were identified (Fig. 2). The intron belongs to subgroup IIb (37), with a putative δ–δ′ pairing in dI and the EBS3 base A paired with the first nucleotide of the 3′-exon. The new group II intron is not similar to any other known *Euglena* chloroplast ‘euglenoid-type’ group II introns. Rather, the predicted secondary structure model and tertiary interactions conform very closely to the *Cal.xI* intron of *Calothrix* (22) and *rnl* introns 1 and 2 of *P. purpurea* mitochondrial DNA (38).

The intron-encoded protein resembles cyanobacterial reverse transcriptases

The deduced amino acid sequence of *mat4*, when compared to known proteins by BLAST analysis (34), was found to have E values in the range 2 × e−91 to e−126 for alignments with group II intron-encoded and free standing reverse transcriptases from cyanobacteria, as well as intron-encoded proteins of the mitochondrial gene of the red alga *P. purpurea*. The highest similarity of the *E. myxocylindracea* intron protein, with 43% identity and 59–60% positives over the entire coding region, is to a reverse transcriptase homolog from *Nostoc* sp. PCC 7120 which is encoded by the *alr8560* gene (accession no. NC_003273) and to a group II intron-encoded protein from *Calothrix* (accession no. S40013). The *Nostoc* ORF (45989–47791) is in d4 of a putative group II intron (45422–47910). After cyanobacteria the highest similarity to *mat4* was with reverse transcriptase homologs ORF544 and ORF546 from mitochondria of *P. purpurea* (35–37% identity and 55% positives). Alignment scores between *mat4* and the previously described *mat1–mat3* loci of *Euglena* chloroplasts and the various *matK* loci of plant chloroplasts were very low compared to the cyanobacterial reverse transcriptases, indicative of only a distant relationship to previously described intron-encoded proteins of chloroplasts.

It is noteworthy that both the intron structure and the intron-encoded protein gene are most similar to cyanobacterial and red algal mitochondrial introns. This would be expected if the intron resulted from horizontal transfer of a retromorphing cyanobacterial intron.

The *mat4* encoded protein

An amino acid sequence alignment of the *mat4* IEP with IEPs from *P. purpurea* and *Calothrix* is shown in Figure 3. All of the main functional domains from *P. purpurea* and *Calothrix* are conserved in *E. myxocylindracea*. These are the reverse transcriptase, a Zn finger domain and domain X. Within the reverse transcriptase domain are seven conserved amino acid motifs (RT1–RT7) characteristic of retrotransposons and other reverse transcriptases. The *mat4* IEP sequence aligns with the cyanobacterial and mitochondrial sequences in the Zn finger domain, domain X and RT1–RT7 as well as they align with each other. Additionally, the reverse transcriptase has an additional domain RT0 which is characteristic of IEPs. One functional motif in RT5 has the sequence YADD typical of bacterial IEPs in retrohoming introns. Mutations in either Asp residue cause the impairment of Mg2+ binding at the reverse transcriptase active site and elimination of reverse transcriptase activity (11,39,40). A YADD motif is not present in other maturases of euglenoids nor in maturases of other chloroplast genomes.

The distance between domains RT4 and RT5 is slightly longer in *E. myxocylindracea* compared to the other three species. Domain X is a putative RNA-binding domain associated with maturase activity (41). The Zn finger domain is located distal to domain X of the protein and is associated with intron mobility. It can be divided into N- and C-terminal...
regions. The N-terminal region is required for reverse splicing into double-stranded DNA and DNA unwinding, while the C-terminal region containing the conserved Zn finger-like and endonuclease motifs is required for endonuclease activity (42,43). The Zn finger domain is present in \textit{mat4}, but absent in all previously investigated maturases from euglenoids. Thus the \textit{mat4} IEP is most closely related to cyanobacterial IEPs and only distantly related to chloroplast IEPs.

\textit{mat4} from \textit{E.myxocylindracea} could be expressed and isolated from \textit{E.coli} BL21(DE3) (data not shown), although apparently in an inactive form. This is the first matrase from euglenoids which we have been able to express in \textit{E.coli} cells, perhaps a consequence of a recent prokaryotic origin.

\textbf{Self-splicing \textit{in vitro} of the intron containing \textit{mat4}}

The ability of the \textit{psbA} group II intron and a $\Delta$-IEP variant to undergo self-splicing \textit{in vitro} was investigated. \textit{In vitro} RNA transcripts were incubated at 42°C in 0.5 M ammonium sulfate for 0–60 min. The resulting RNA products were amplified by RT–PCR and resolved by agarose gel electrophoresis (Fig. 4A). A product corresponding in size to spliced exons is detected following a 10–60 min incubation. Lowering the temperature or ionic strength or substituting ammonium chloride all resulted in decreased spliced product formation. The RT–PCR products of the $\Delta$-IEP \textit{psbA} pre-mRNA are shown in Figure 4B. The amplified product from the spliced \textit{psbA} mRNA is the major species after either 20 or 60 min of incubation. The products labeled E5′–E3′ were confirmed to represent correctly spliced \textit{psbA} mRNA exons by cloning and DNA sequence analysis (data not shown). The \textit{in vitro} self-splicing reaction was repeated with 32P-labeled \textit{psbA} pre-mRNA (Fig. 4C). Products corresponding to both the spliced exons and the excised intron lariat were obtained. The excised intron lariat is evident after 20 and 60 min incubation. Thus,
by two different assays, the *E. myxocylindracea* psbA intron was shown to self-splice in vitro. This is the first example of a self-splicing group II intron from any chloroplast genome.

**Phylogenetic analysis of mat4**

To investigate the evolutionary relationship between *mat4* and related intron-encoded proteins and reverse transcriptase proteins, a phylogenetic analysis was performed on proteins of cyanobacterial, bacterial, mitochondrial and chloroplast origin using neighbor joining and parsimony analysis (36). Sequences for evolutionary comparisons were chosen from among the 50 highest BLAST similarity scores. Multiple sequences from the same species and sequences which did not have all domains present in *mat4*, with the exception of *Nostoc* sp., were omitted. The only chloroplast genome sequence represented, other than *mat4*, was a reverse transcriptase from *S. obliquus*. Other euglenoid maturases and maturases from chloroplast genomes of charophytes and plants were also too...
divergent for this analysis. Some reverse transcriptases included are in the chloroplast-type reverse transcriptase group defined by Zimmerly et al. (19). Two bacterial reverse transcriptases of mitochondrial type were chosen as an outgroup. The resulting 19 sequences were from cyanobacteria, proteobacteria and mitochondrial genomes. The results of the evolutionary analysis are shown in Figure 5. Both mat4 and the reverse transcriptase from the mitochondrial genome of P. purpurea are deeply rooted in the cyanobacterial clade. One of two bacterial clades includes representative sequences from cyanobacteria as well as the mitochondria genome of brown algae. The most closely related chloroplast sequence to mat4 is a reverse transcriptase from S. obliquus, which is the most distantly related representative of a clade with a mitochondrial protein from Pylaiella and several bacteria sequences. From this phylogenetic analysis, mat4 is more closely related to cyanobacteria sequences than reverse transcriptase from the green algae S. obliquus or any other plastid maturase.

DISCUSSION

The structural similarities of the psbA intron RNA, the in vitro self-splicing activity and the phylogenetic position of the IEP within a cyanobacterial clade support the conclusion that the psbA intron is the result of a recent horizontal transfer into the E. myxocylindracea chloroplast genome from a cyanobacterial donor. Previous models for the evolution of chloroplast

![Figure 4](https://example.com/figure4.png)

**Figure 4.** In vitro splicing. Splicing products of non-radioactive transcripts were purified and a RT±PCR was carried out: (A) plasmid PEYC2009; (B) plasmid PEYC2010. Splicing products of radioactive transcripts were analyzed on a denaturing 4% polyacrylamide gel, which was dried, put on X-ray film and developed the next day (C). M, marker λ DNA±BstEII; C, control.

![Figure 5](https://example.com/figure5.png)

**Figure 5.** Phylogenetic trees on 21 amino acid sequences including two outgroup sequences were constructed using the distance (neighbor joining, NJ) method of the program PAUP. Bootstrap values from 1000 re-samplings are given at the nodes of the tree. Branch lengths are proportional to the expected mean number of substitutions per site along the branch, as quantified by the scale bar. Parsimony analysis had similar branching as in the NJ tree (data not shown). chl., chloroplast genome; mit., mitochondrial genome. Bacteria include bacteria other than cyanobacteria.
genetic elements have emphasized a common origin during the primary endosymbiotic event that gave rise to chloroplast genomes. The absence of introns in chloroplast genomes of the early branching algal species, but their reappearance in green algal chloroplast genomes and their descendants, including the vascular plants, has been explained by loss of introns in the chloroplast genomes of early branching species (29–31). The variation in intron content among plastids from different sources was attributed to vertical transfer subsequent to primary endosymbiosis. There is now evidence for an alternative pathway of chloroplast genome evolution. Introns in chloroplasts may have originated either from the primary endosymbiont or due to secondary horizontal transfer(s) from a cyanobacterial donor. Given the variety in chloroplast intron location and structure, horizontal transfer may have occurred many times in different plastid lineages, most notably in the euglenoid protists.

_Euglena_ plastid genomes are unique with respect to chloroplast introns, due to the large numbers, structural diversity, multitude of IEPs and high degree of degeneracy. The discovery of a new _psbA_ intron with strictly cyanobacterial features suggests an attractive explanation for intron history in euglenoid plastids. Multiple horizontal intron transfer events from cyanobacteria into plastid genomes in this lineage may have given rise to different populations of chloroplast introns. Following acquisition of a foreign intron, intragenomic mobility could account for the spread of introns to new target sites, including into existing introns, accounting for the formation of a variety of classes of twintrons (8). One cyanobacterial intron may have evolved into group III introns, while other cyanobacterial introns evolved into different varieties of group II introns. Support for this model lies in the variety of _Euglena_ chloroplast IEPs (‘maturases’) described to date, including _mat1_ specific for a group III twintron of _psbC_, _mat2_ specific for a group II intron of _psbC_, _mat3_ within a group II intron of _psbD_, _mat4_ described in this report and _mat5_ within a group II intron of _psbA_ in _Euglena spirigryra_ (E.Sheveleva and R.B.Hallick, unpublished observations). Several factors may have influenced the evolution of introns in plastid genomes. An early event may be the loss of intragenomic mobility, caused by mutation of domains in the IEPs. Additional streamlining of both the IEPs and the RNA is likely the consequence of host factors assuming part of the splicing reaction. The ultimate loss of IEPs in most plastid introns reflects a major role of the host in catalyzing the splicing reactions. Another evolutionary factor is the type of gene where introns are inserted, as evolution of introns follows the evolution of the genes in which they are located. Group III introns are more common in non-photosynthetic genes, whereas group II introns are more common in photosynthesis-related genes (9). Finally, different classes of plastid introns must also reflect the evolutionary time since a horizontal transfer event. The most deeply rooted introns, such as the group III twintron with the _IEP mat1_ in _psbC_ of many euglenoid species, were presumably acquired very early compared to the recent arrival of the cyanobacterial-like _psbA_ intron of this report.

The formation of algal chloroplast genomes occurred more than 1.2 billion years ago. Primary plastids of red and green algae have spread laterally among distantly related organisms through the process of secondary endosymbiosis, in which a heterotrophic eukaryote, including many euglenoids, retains the photosynthetic apparatus of algae (44). Green algae are the closest predecessors of euglenoids. The possibility of horizontal gene transfer into chloroplast genomes has been discussed by others, but either not proven or considered unlikely (23–27). There is precedent for the concept of horizontal intron transfer between organisms. Horizontal intron and gene transfer are well known in bacteria (reviewed in 19,45). Likewise, the appearance of cyanobacterial-like introns in the mitochondrial genome of red and brown alga are presumably results of horizontal gene transfer (19,38,46).

The discovery of horizontal intron transfer in _Euglena_ should prompt a reconsideration of horizontal transfer of introns and other genetic elements in the evolutionary history of chloroplast genomes. The differences in intron and gene content among chloroplast genomes of different species could reflect both differential gene loss from the primary endosymbiont and also horizontal transfer of new genes to some plastid lineages. Many contemporary plastid genomes may be an evolutionary mosaic of genes from the primary endosymbiotic event that gave rise to all plastid genomes and other genes acquired through secondary horizontal transfer events from many different sources.

**ACKNOWLEDGEMENTS**

We would like to thank Jane Dugas for technical assistance with the manuscript. We would also like to thank Dr W. Birky for assistance with the phylogenetic analysis. This work was supported by NIH grant GM35665.

**REFERENCES**


