Curr Genet (2009) 55:69–79
DOI 10.1007/s00294-008-0225-7

RESEARCH ARTICLE

# The horsetail *Equisetum arvense* mitochondria share two group I introns with the liverwort *Marchantia*, acquired a novel group II intron but lost intron-encoded ORFs

**Dominique Bégu · Alejandro Araya**

Received: 17 November 2008 / Revised: 5 December 2008 / Accepted: 5 December 2008 / Published online: 27 December 2008
© Springer-Verlag 2008

**Abstract** We studied the genomic structure and RNA editing of mitochondrial *cox1*, *cox2*, *cob and atp9* from the horsetail *Equisetum arvense,* a representative of an old fern lineage. Editing of *cox1, cob* and *atp9* mRNAs occur only by C-to-U transitions. No changes were found in *cox2* transcripts constituting one of the rare examples of unedited mitochondrial mRNA in land plants. From three intervening sequences in *cox1, cox1i395* and *cox1i624* are group IB introns homologous to the *Marchantia polymorpha cox1* introns, and *cox1i747* is a group IIA intron different to other introns found in plant mtDNA. The group II intron *cox2i373* is very similar to other introns found in *cox2* from vascular plants. While *cob* and *atp9* have no introns and display the gene structure found in seed plants, various nucleotide substitutions abolish the only potential ORF, a LAGLIDADG endonuclease present in *cox1i395*. Thus, *E. arvense* mitochondria conserve two group I introns from non-vascular plants, probably inherited from a common ancestor with liverworts. Analogous to seed plants, *E. arvense* has no potential mitochondrial splicing factors encoded in these introns. This is the first report concerning the presence of vertically inherited group I introns in vascular plant mitochondria.

**Keywords** Mitochondria · RNA editing · Introns · Pteridophyte · *cox1* · *cox2*

Nucleotide sequence data reported are available in GEN Bank databases under the accession numbers FJ376598 (*cox1*), FJ376799 (*cox2*), FJ376600 (*cob*) and FJ389746 (*atp9*).

Communicated by L. Tomaska.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00294-008-0225-7) contains supplementary material, which is available to authorized users.

D. Bégu (✉) · A. Araya
Laboratoire de Microbiologie Cellulaire et Moléculaire et Pathogénicité (MCMP), UMR5234 CNRS, Université Victor Segalen Bordeaux2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France
e-mail: dominique.begu@reger.u-bordeaux2.fr

A. Araya
e-mail: aaraya@reger.u-bordeaux2.fr

## Introduction

Group I and II introns are present in all land plant mitochondrial genomes so far described (Chaw et al. 2008). Both intron types are characterized by a distinct and conserved RNA secondary structure (Waring et al. 1982; Michel et al. 1982), and differ by their splicing mechanism (Saldanha et al. 1993). Comparative analysis of complete genomes and phylogenetic studies of individual introns revealed that their occurrence in different genes varies by sporadic gain and loss during embryophyte evolution (Knoop 2004; Bonen 2008).

In mitochondria of vascular plants, all the introns belong to group II type except for a group I intron present in *cox1* from Peperomia first described by Vaughn et al. (1995) and more recently identified in a relatively large number of flowering plants (Cho et al. 1998; Sanchez-Puerta et al. 2008). This intron seems to be acquired from independent horizontal transfers during angiosperm evolution. In contrast, non-vascular plants mitochondria present several group I introns. Beside 25 group II introns, 7 group I introns split mitochondrial genes in the liverwort *Marchantia polymorpha* (Oda et al. 1992). Six of them are present in *cox1* considered the main group I intron reservoir, and one is found in *nad5*. The moss *Physcomitrella patens* has two group I introns, the

counterparts of *cox1* and *nad5* introns from *M. polymorpha* (Terasawa et al. 2007). The *nad5* group I intron is found in all liverworts and mosses investigated so far, but is absent in hornworts (Steinhauser et al. 1999; Duff 2006). Thus, it seems that a gradual loss of group I introns occur in non-vascular plants after the divergence between liverwort and mosses. While a small number of mitochondrial group II introns have been vertically transmitted between non-vascular and vascular plants during evolution, such inheritance has not been observed for group I introns (Bonen 2008).

Some ORFs encoded in mitochondrial introns are required for splicing and intron mobility (Dujon et al. 1986; Lambowitz and Zimmerly 2004). Group I and II intron-encoded proteins have different characteristics. Group I introns encode site-specific endonucleases with conserved LAGLIDADG or GIY-YIG motifs, while group II introns encode for proteins with homology to viral reverse transcriptase (RT), referred as maturase-reverse transcriptases (MAT-R). Several ORFs have been found in introns from bryophyta mitochondria where eight group II introns encode MAT-R proteins and two group I introns encode LAGLIDADG endonucleases (Oda et al. 1992). *Physcomitrella* mitochondrial genome (accession number AB251495) presents two maturase-encoding group II introns, which are different to *Marchantia*, while no group I intron-encoded endonucleases are detected. In seed plant, only one ORF related to fungal RT-maturases (*mat-r*), is highly conserved among species in the same group II intron of *nad1* (Wahleithner et al. 1990). This ORF is absent in *E. arvense nad1* (Dombrovska and Qiu 2004). The loss of the coding capacity of introns might be partially compensated by the transfer of maturase genes into the nuclear genome (Mohr and Lambowitz 2003; Nakagawa and Sakurai 2006).

The aim of this work was to investigate the status of selected mitochondrial introns in pteridophytes, using the horsetail *E. arvense* as a model. We were particularly interested in intron-encoded ORFs which are present in six mitochondrial genes from the liverwort *M. polymorpha* (Oda et al. 1992). Among vascular plants, horsetails, living ferns, and seed plants (euphylophytes) are assumed to be the sister group of hornworts and mosses (Groth-Malonek et al. 2005). Liverworts are proposed to be the basal clade to the whole embryophyte (Henrick and Crane 1997; Qiu et al. 2006). Up to date, no pteridophyte mitochondrial genome has been reported, and only partial *cox1* gene sequence is known for four species (Sper-Whitis et al. 1996). Moreover, *cox2* has been described only in the primitive vascular plant *Psilotum nudum* (Sper-Whitis et al. 1994), and neither *cob* nor *atp9* gene sequence is available for that clade. The study of horsetails (equisetopsids) is important since they constitute an early branching group in the moniliformopses which are considered as the closest living relatives to seed plants (Pryer et al. 2001).

We report here the genomic structure of *cox1*, *cox2*, *cob* and *atp9* from the horsetail *E. arvense* to elucidate the gain and loss of coding introns in pteridophyte. Two others genes bearing potential coding introns, *rrn18* and *atp1* have been described (Duff and Nickrent 1997; Wikström and Pryer 2005). By analyzing genomic and cDNA sequences, we verify the actual exon-joining boundaries and determine the RNA editing status of the respective coding regions. We found that *E. arvense* mitochondria have conserved group I introns from non-vascular plants, but similar to seed plants, have lost the coding capacity of introns. This work constitutes the first report concerning the presence of two group I introns in the mitochondrial genome of a vascular plant, probably inherited from a common ancestor with liverworts.

## Materials and methods

### Nucleic acid extraction

*Equisetum arvense* plants were field-collected near Bordeaux (France). Total DNA and RNA were extracted from fresh lateral shoots using the DNeasy Plant Mini Kit and the RNeasy Plant Mini Kit (Qiagen), respectively, following the protocol specified by the manufacturer.

### Genomic and cDNA amplification, cloning and sequencing

Genomic fragments were amplified from total DNA by nested PCR to improve the signal and the specificity. The orthologous primers used are detailed below. Two different thermostable DNA polymerase were used: Advantage 2 Polymerase Mix (Clontech) in current PCR assay and *Tfl* DNA polymerase (Promega) to amplify products longer than 3 kbp. PCR amplification assays contained 10–100 ng of total DNA, 250 µM each dNTP, 10 µM primers and 2.5 units of DNA polymerase in a final volume of 50 µl. PCR reactions were performed essentially as described by the supplier. Hybridization temperature and elongation time were adapted according to the primer melting temperature and the fragment size as follows: for *cox1*, the hybridization temperature was 47 and 50°C for PCR1 and PCR2, respectively, and the elongation step was 5 min; for *cox2*, the hybridization temperature was 50 and 48°C for PCR1 and PCR2, respectively, and 2 min elongation; for *cob*, 40 and 44°C hybridization temperature for PCR1 and PCR2, respectively, and 1 min elongation; for *atp9* hybridization was at 44°C and 1 min elongation. In all cases, the elongation temperature was 68°C.

cDNA synthesis was performed on total RNA with the Access RT-PCR kit (Promega) according to the protocol of the manufacturer, followed by a nested PCR performed as described for genomic PCR, using 1 µl of RT-PCR product. The results presented correspond, at least, to two

independent experiments performed with different nucleic acids preparations.

PCR and RT-PCR products were purified with the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and cloned into the pGEMT-Easy vector (Promega). PCR clones were purified with Wizard plus SV minipreps DNA purification kit (Promega) and sequenced on an ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) using the Big Dye Terminator v1.1 cycle sequencing kit (Applied Biosystems). Data presented here are the result of the analysis of 20 genomic and at least 40 cDNA clones for each gene.

### PCR and RT-PCR primers

Orthologous PCR primers were designed based on conserved 5′ and 3′ exonic sequences of homologous genes from *M. polymorpha* (Oda et al. 1992). The primers chosen cover about 85% of the coding sequence of the *E. arvense* genes compared to the *M. polymorpha* homologous genes. In some cases, the primers used for nested PCR (S2/AS2) were designed from the actual *Equisetum* sequence (*Ea*) after cloning and sequencing the first PCR product (S1/AS1 primers).

*cox1*-S1, GATATAGGTACTCTATATTT; *cox1*-AS1, AGGATTCTGTTCAACCGC; *cox1*-S2, CTATATTTAA TCTTCGGTGC (*E.a*); *cox1*-AS2, AACCGCCGCCCAA GG (*E.a*); *cox2*-S1, CACTCCTATGATGCAAGG; *cox2*-AS1, ATACCCAAGAAACATAATCA; *cox2*-S2, AAGG AATAATTGACTTACATC; *cox2*-AS2, CAAAGAAACA GCTTCTA (*E.a*); *cob*-S1, CATTTGATAGATTATCC; *cob*-AS1, AGAATGGGCGTTAT; *cob*-S2, AGTTATTGG TGGGG; *cob*-AS2, GGCGAAATACAAGAA (*E.a*); atp9-S, AATGGAGCAGGAGC; atp9-AS, TATAAAAATGCC ATCATT.

### Sequence analysis and secondary structure prediction

The search for similar regions was done with the BLASTN 2.2.18+ program (http://www.ncbi.nlm.nih.gov/BLAST/). Nucleotidic sequences were aligned using Gene Jockey II software (Biosoft). Intron secondary structure prediction was obtained with the mfold software (Zuker 2003) (http://www.mfold.bioinfo.rpi.edu/), or by analogy with known homologous introns and adjusted manually according to the canonical group I and group II intron structures described by Michel et al. (1989).

## Results

### Genomic structure of *cox1*, *cox2*, *cob* and *atp9*

PCR amplification using *cox1* orthologous primers generated a product of 4,345 bp length from total DNA preparation, while a RT-PCR analysis showed a cDNA product of 1,451 bp, suggesting the presence of large intronic sequences. In the case of *cox2,* the genomic PCR product is 2,029 bp in length, while the RT-PCR product has only 651 bp indicating the presence of intervening sequences in this gene. The genomic and cDNA amplification products from *cob* produced an identical 1,061-bp band; therefore, we conclude that *cob* is not interrupted by introns. Similar to *cob*, *atp9* presents no introns. In all cases, the RT-PCR amplification products were obtained from specific mRNA, since no amplification products were obtained in the absence of RT which is a clear evidence that the RT-PCR products originated from the respective mRNAs.

To define precisely the position of the putative introns, the genomic and cDNA amplification products of *cox1* and *cox2* were cloned and sequenced. To facilitate the comparison with the homologous *M. polymorpha* introns, thereafter the introns will be named according the nomenclature proposed by Dombrovska and Qiu (2004). *cox1* presents three introns, intron 1 *(cox1i395),* intron 2 *(cox1i624)* and intron 3 (*cox1i747*) which are 1,126, 1,085 and 712 bp in length, respectively. In the case of *cox2*, an intron of 1,378 bp, *cox2i373*, interrupts the ORF. Sequence analysis of cDNA clones showed that the mitochondrial *cob* gene has not intervening sequences in the horsetail *E. arvense*. A diagram of the exon–intron structure of the three genes is depicted in Fig. 1b (for details, see Supplementary material).

### RNA editing of *Equisetum arvense* transcripts

RNA editing is a hallmark of land plant mitochondria characterized mainly by C-to-U, and in some species U-to-C changes. At least 40 independent cDNA clones were sequenced for each gene transcript and compared to the respective genomic sequences.

*cox1* mRNA is extensively edited with 47 C-to-U changes distributed throughout the mature transcripts: 6 changes in exon1, 7 in exon 2, 6 in exon 3 and 28 C-to-U changes in exon 4 (Fig. 1c). A total of 13.7% of exonic C residues are edited, concerning 41 out of 463 codons in *cox1* mRNA, but no silent editing was observed (Supplementary material S1). Thirty-six out of 41 codons are edited either at the first or second position. Four codons are edited (underlined) at two residues: codon 71 (CCU); codon 428 (CCC); codon 434 (CCA) and codon 86 (UCC). Finally, a Pro 242 (CCC) codon is changed to Phe (UUU) by triple editing. RNA editing is quite efficient for *cox1* spliced mRNAs since 27 out of 41 codons were found edited in all cDNA clones, the remaining 14 codons were edited at frequencies from 76 to 98%. Twenty-six out of 50 *cox1* mature mRNAs were fully edited, 22% present only one unedited codon and 16% have two unedited codons.

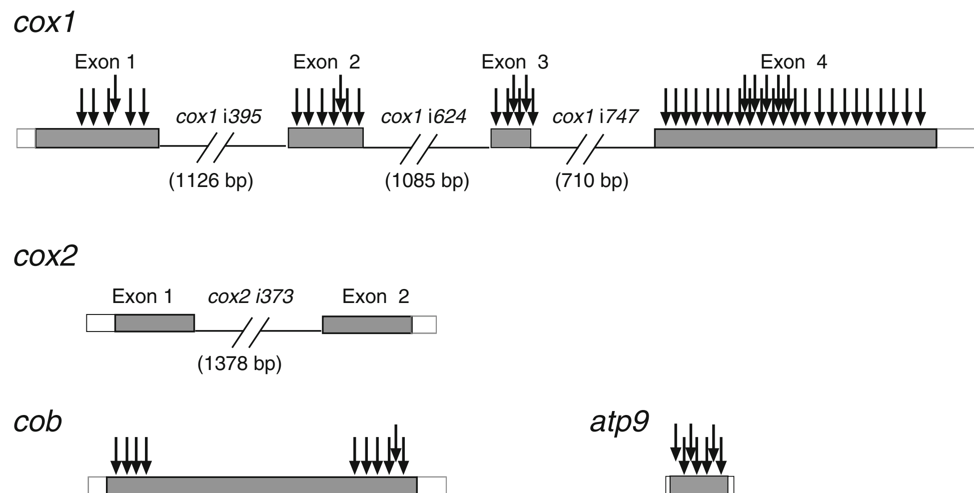



**Fig. 1** Genomic organization of *cox1*, *cox2*, *cob* and *atp9*, and RNA editing of the mature transcripts from the horsetail *Equisetum arvense*. Gene structure was established after sequencing genomic and cDNA PCR products. Exons are represented by *grey rectangles*. Introns are depicted by *split thin lines* connecting exons. Intron names are according to the nomenclature suggested by Dombrovska and Qiu (2004), and intron size is indicated in parentheses. Missing information at exonic ends (*white rectangles*) is estimated to 18 and 30 codons at 5′ and 3′ ends of *cox1*, respectively, 30 and 4 codons at 5′ and 3′ ends of *cox2*, 29 and 36 codons at 5′ and 3′ ends of *cob*, and 7 and 4 codons for *atp9* mRNAs based on comparison with *Marchantia* coding sequences. *Vertical arrows* indicate the localization of edited C residues

The putative COX1 protein encoded in the *E. arvense* mitochondrial genome share 85.8% identity with the *M. polymorpha* and 85.3% with *A. thaliana* homologous protein. However, the protein encoded by the edited mRNA presents 94.6 and 94.2% identity with the *M. polymorpha* and *A. thaliana* homologous counterparts, respectively.

In the case of *cob* transcripts, 10 C-to-U conversions were detected after sequence analysis of 40 cDNA clones compared to the genomic sequence. The editing sites were formed two clusters with four editing sites located at the 5′ end and six sites at the 3′ end of the mRNA (Fig. 1c). RNA editing modifies 9 out of 322 codons (Supplementary material S2). Eight codons are modified at only one residue, three of them are in position 1 in the codon and seven in position 2. Only one, Pro 363 (CCC) codon was edited at two positions. Among the sequenced cDNA clones, 27 out of 40 corresponds to fully edited mRNAs, the others were partially edited presenting one (7/40), two (4/40), three (1/40) or six (1/40) unedited codons. Seven C residues are changed to Us in *atp9* mRNA (Supplementary material S3). Different to *cox1* and *cox2* transcripts, the *atp9* mRNAs were fully edited.

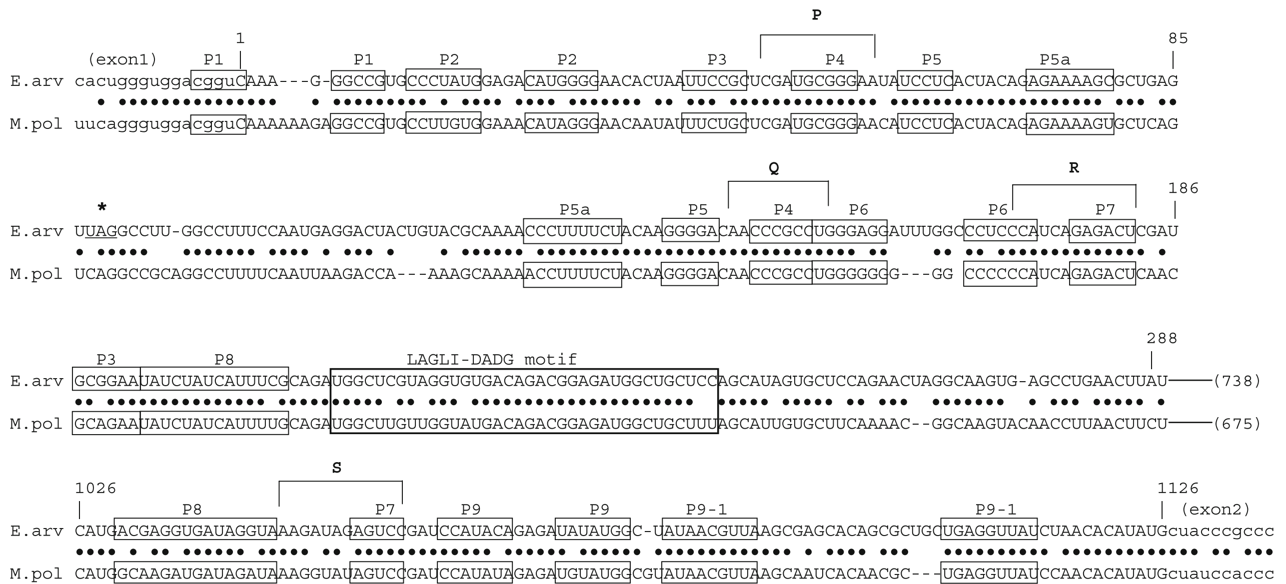
While several C-to-U changes were found in different transcripts, no U-to-C changes were observed in any the cDNA clones sequenced. It is interesting to note that no difference between *cox2* genomic and cDNA nucleotide sequence was found, indicating that *cox2* transcripts are correctly spliced but are not edited in *E. arvense* mitochondria. The sequence identity between the putative COX2 proteins of *Equisetum* and *Marchantia* is about 90%.

## Secondary structure and coding capacity of *Equisetum arvense cox1* and *cox2* introns

The introns interrupting the horsetail *cox1* and *cox2* ORFs were compared to *M. polymorpha* counterparts (Ohta et al. 1993). *E. arvense* intron1 (*cox1i395)* and intron 2 (*cox1i624)* from *cox1* are homologous to intron 4 (*aI4*) and intron 6 (*aI6*) from *M. polymorpha cox1*, respectively (Fig. 2). No sequence similar to *cox1i747* was found in mitochondrial genomes available in data banks.

The mitochondrial intervening sequences have canonical secondary structures with particular regions allowing to define them as type I or type II introns. *cox1i395* and *cox1i624* present highly conserved elements of secondary structure designated P1 through P9 and the conserved consensus sequence elements P, Q, R, and S characteristic of group I introns according to standard representation established by Burke et al. (1987). Both sequences can be folded in a secondary structure presenting a catalytic core made up of two extended helices, the P4–P6 domain formed by the stacking of P5, P4, P6, and P6a helices whereas the P3–P9 domain is formed by the stacking of P8, P3, P7, and P9 (see S4 and S5 of Supplementary material).

*Marchantia polymorpha aI4* encodes for a LAGLIDADG endonuclease located in loop 8, fused to the preceding exon1. *cox1i395,* the *Equisetum* homologue of *aI4,* presents only one LAGLIDADG motif (Fig. 2a). The downstream region in loop 8 differs from the *Marchantia* sequence leading to the loss of the endonuclease reading



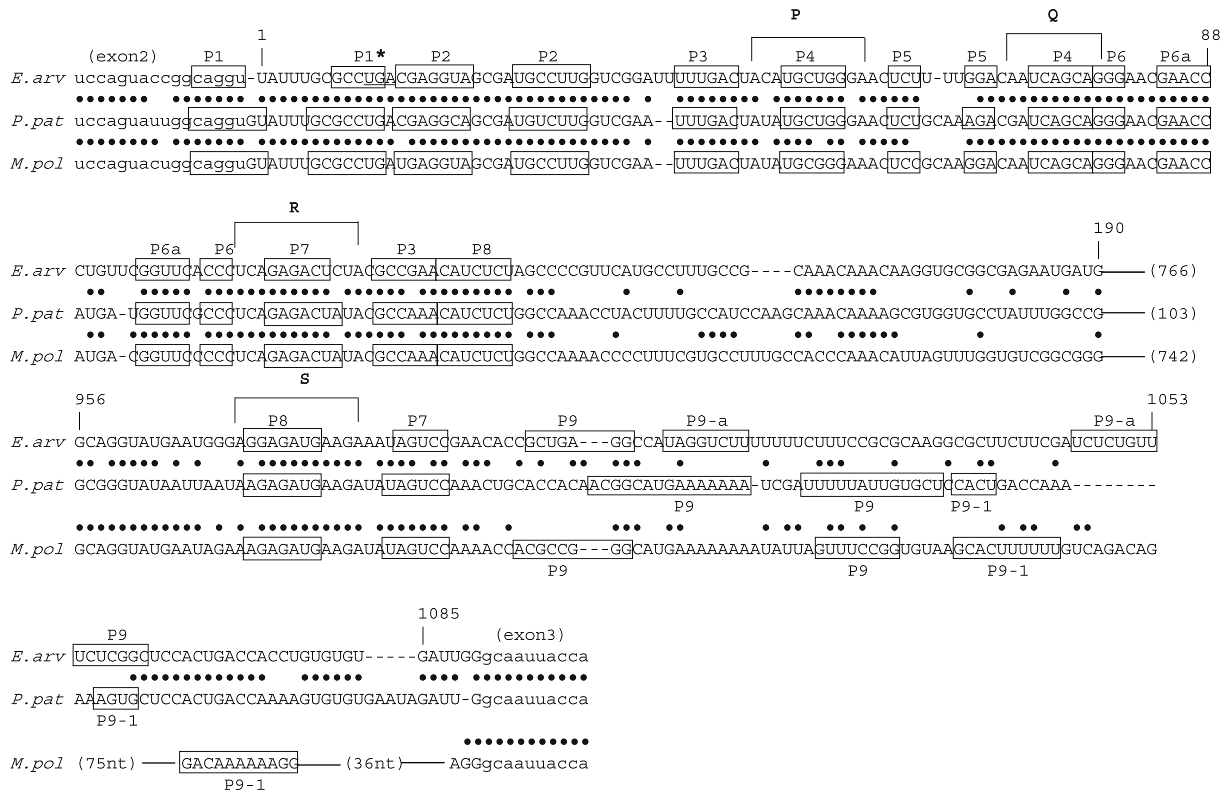




**Fig. 2** Comparison of group I introns from *E. arvense cox1i395* and *cox1i624 with* the *P. patens cox1i624* and *M. polymorpha ai4* and *ai6* homologues. **a** *E. arvense cox1i395* (*E. arv*) is compared to the *Marchantia ai4* intron (*M. pol*). The conserved LAGLIDADG endonuclease motif is indicated. **b** Sequence alignment of *E. arvense cox1i624* with *Physcomitrella* (*P. pat*) *cox1i624* and *Marchantia ai6* introns. Nucleotides not shown in the alignment are signaled in parentheses. Intron sequences are in uppercase and exonic sequence junctions in lowercase letters. *Numbers above alignments* indicate the nucleotide position within each intron in *Equisetum* sequences. Nucleotide identity with *Equisetum* sequence is indicated by *dots*. Motifs P1–P9 (*small rectangles*) involved in base pairing in the secondary structure of group I introns were defined as reported by Ohta et al. (1993). *Brackets* indicate conserved core sequences P, Q, R and S. The premature stop codon in the putative encoded-ORF is *underlined and signaled by an asterisk*

frame. In contrast, *Equisetum cox1i624*, the homologue of *aI6* from *Marchantia*, present high identity from P1 to P8 regions. The L8 loop is totally different in primary sequence and size compared to *aI6*, leading to a divergent secondary structure of the P9 region between the two species (Fig. 2b and accompanying Supplementary material). As *Marchantia aI6*, *cox1i624* is a non-encoding intron.

Intron 3 (*cox1i747*) presents all the characteristics of a group II intron. It can be folded into a canonical secondary structure with six conserved helical domains I–VI (D1–D6) linked to a central core (Michel et al. 1989). Moreover, some motifs such as the EBS1-IBS1 and EBS2-IBS2 base pair interactions between D1 and exon1 region, the A residue in a bulge on D6 involved in the first transesterification step and the tertiary structure interactions $\alpha$–$\alpha'$, $\varepsilon$–$\varepsilon'$ and $\gamma$–$\gamma'$ characteristic of functional group II introns are clearly distinguished (Fig. 3). Domain V differs from the consensus structure with a short distal helix and an unusual 8nt loop instead of the GAAA-tetraloop conserved motif. This variation has been observed in some other group II introns (Lang et al. 2007).

The location and the coding properties of *cox 1* introns from different plant species is drawn in Fig. 4.

As indicated above, *E. arvense cox2* is interrupted by a 1,383 bp intron with no homologue in *M. polymorpha* mitochondrial genome. Gene bank blast analysis revealed that *cox2i373* has high nucleotidic sequence identity with other group II introns of *cox2* mitochondrial genes from many vascular plants located at identical insertion site. In fact, *cox2i373* can be folded in a canonical secondary structure with tertiary interaction motifs characteristics of group II introns (for details see Supplementary material S6).

## Discussion

### The horsetail *E. arvense cox1* and *cox2* are interrupted by introns

In bryophytes, some mitochondrial genes present a complex organization. In the liverwort *M. polymorpha, cox1*, *cox2*, *cob* and *atp9* are interrupted by nine, two, three and one introns, respectively (Oda et al. 1992). Nine of them correspond to group II and six to group I introns. Moreover, five out of these nine group II introns carry maturase-reverse transcriptase type ORFs and two group I introns encode for homing endonucleases of the LAGLIDADG type. This situation is radically modified in higher plants where all these introns, particularly the group I introns, are not present. This study focuses on the horsetail *cox1*, *cox2, cob,* and *atp9* mitochondrial genes. *E. arvense* is one of the ancient ferns, placed at the interface between non-vascular and vascular plants in the conquest of land during evolution (Pryer et al. 2001). The structure of the four *Equisetum* genes is less complex than the homologous *M. polymorpha* counterparts. Three introns interrupt *cox1* and one split *cox2*, while no introns were found in *cob* and *atp9.* Interestingly, *cob* and *atp9* present the intronless configuration found in higher plants. *E. arvense atp1* (Wikström and Pryer 2005) and *rrn18* (Duff and Nickrent 1997; accession no. AF058663) have no intron. It should be noted that the data presented in this report concern the functional introns defined by the actual splicing sites and exon-joining boundaries resulting from comparison between the gene with the respective mature transcript cDNA sequences.

### RNA editing involves C-to-U but no U-to-C conversions in *Equisetum arvense* mitochondria

The RNA editing process occurs by C-to-U modifications in all land plant organelles with the exception of the marchantiid subclass of liverwort (Malek et al. 1996). In addition, U-to-C RNA editing has been reported in hornworts (Steinhauser et al. 1999; Kugita et al. 2003), isoetes and ferns (Vangerow et al. 1999) where it modifies codons and suppresses genomic stop codons.

Sequence analysis of cDNA clones from *E. arvense* transcripts showed that *cox1, cob* and *atp9* mRNAs are edited by C-to-U conversions at 47, 10, and 7 residues, respectively. In all cases, the codon changes produced by RNA editing lead to a higher identity of the encoded protein with plant and non-plant homologues (Gualberto et al. 1989; Bégu et al. 1990; Araya et al. 1994). The *E. arvense cox1* mRNA is extensively edited as observed for the same transcript in some gymnosperms (Chaw et al. 2008), and higher than reported for angiosperms *cox1* transcripts (Giegé and Brennicke 1999; Notsu et al. 2002; Handa 2003). For *cob* and *atp9* mRNAs, editing level is similar to those disclosed for higher plants transcripts. No editing events were observed in mature *cox2* transcripts, an uncommon situation already observed for a few transcripts in seed plants (Lu et al. 1998; Giegé and Brennicke 1999). The fact that U-to-C changes were not detected in none of the transcripts analyzed is a strong indication that this kind of editing event does not occur in this clade. Additionally, no stop codon interrupts the mitochondrial ORFs, a fact that could foresee such possibility. The results presented here validate the prediction, based on gene sequence analyses, that only C-to-U RNA editing occurs in this species (Malek et al. 1996; Dombrovska and Qiu 2004; Qiu et al. 2006).

### Two group I introns are present in *cox1*

*Equisetum cox1* introns *cox1i*395 and *cox1i*624, the homologues of the fourth and sixth *M. polymorpha* introns,

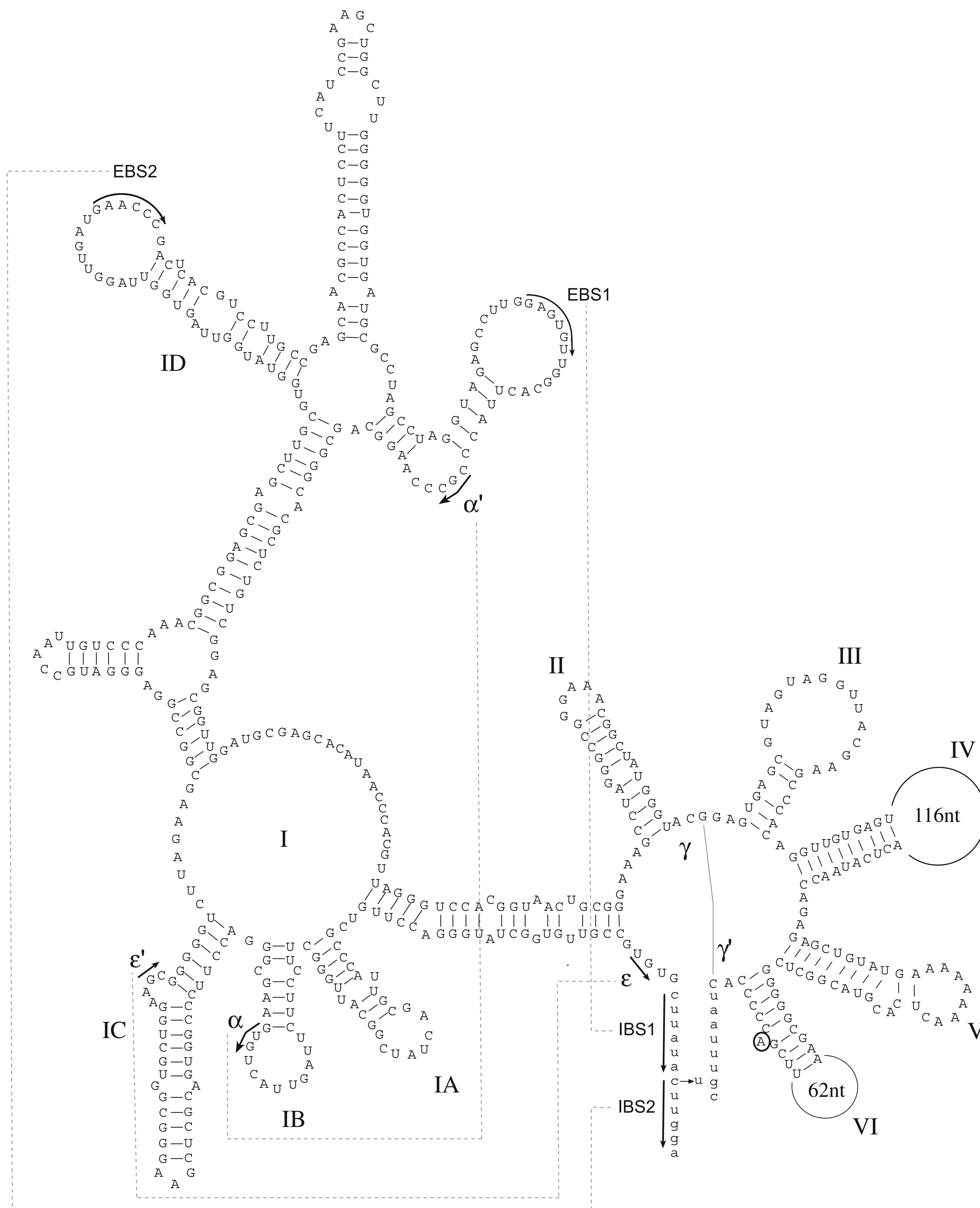



**Fig. 3** Secondary structure of the 710 nt *cox1747* intron of the horsetail *E. arvense*. Roman numerals (*I–VI*) indicate the conserved domains characteristic of group II introns. Structural elements α–α′, ε–ε′, γ–γ′, EBS and IBS implicated in putative tertiary interactions are connected by *dotted lines*. The nucleotides in lowercase letters correspond to the 3′-end of exon 3 and the 5′-end of exon 4, respectively. The C→U editing event in IBS2 is indicated. Some characteristics defining *cox1712* as group IIA intron are a bulging "A" (*circled*) in domain VI that is located 7 nt sequence YAY (γ′ region) according to Michel et al. (1989)

respectively, are located at similar position in both species. It should be noted that *cox1i*395 is absent, but *cox1i*624 is conserved in the moss *P. patens* (Terasawa et al. 2007) (Figs. 2b, 4). Group I introns are present in bryophytes, but are absent in seed plants. In fact all mitochondrial introns described so far in vascular plants are group II introns, with one exception, a group I intron acquired by an angiosperm through horizontal transfer from a fungal donor (Cho et al.

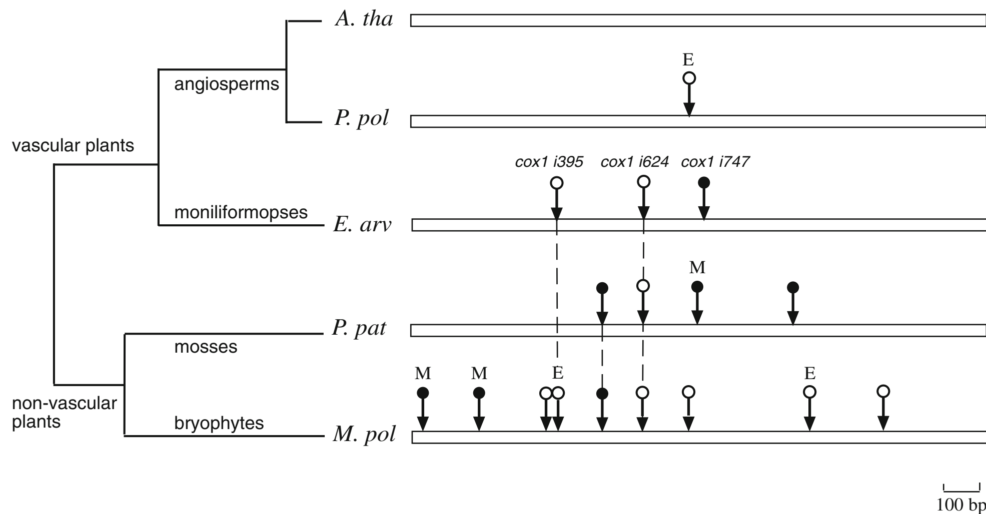



**Fig. 4** Diagram of introns inserted within *cox1* ORF found in different land plants: *Arabidopsis thaliana* (*A.tha*) (Unseld et al. 1997) *Peperomia polybotrya* (*P.pol*) (Vaughn et al. 1995), *Equisetum arvense* (*E.arv*) (this study), *Physcomitrella patens* (*P. pat*) (Terasawa et al. 2007) and *Marchantia polymorpha* (*M.pol*) (Oda et al. 1992). *Vertical arrows* indicate the intron insertion sites. *Arrows with open circles* signal group I introns and *arrows with filled circles* indicate group II introns. *M* group II intron-encoded RT-maturase; *E* group I intron-encoded LAGLIDADG endonuclease. The name of introns found in *E. arvense* is indicated

1998). Both, *cox1i*395 and *cox1i*624 introns can be drawn in a secondary structure characteristic of group I introns (Ohta et al. 1993; Vicens and Cech 2006). The elements involved in intron splicing are clearly identified: (a) the conserved G-U pair in P1 in the 5′ splice site involved in the first transesterification step, (b) the P10 sequence, complementary to the first residues of the downstream exon, required for the second transesterification step, (c) the conserved sequence GACU in P7 which forms the GTP binding site, (d) the domains P1, P3, P4, P5, P6, P7, and P8 forming the catalytic core.

While the first intron has a high identity with the *M. polymorpha* intron 4, *cox1i*624 sequence is conserved only at regions spanning from P2 to P8 base pair segments, but diverge in their L8 loop and P9 region (Fig. 2b). Interestingly, the *P. patens*, homologous *cox1* intron (Terasawa et al. 2007) also presents a divergent sequence after P8 region, exactly at the same position where the *Marchantia* and *Equisetum* sequences diverge (Fig. 2b). This situation results in a different stem-loop structure of the L9–P9 domain (see accompanying Supplementary material). The situation of *cox1i*624 suggests that a rearrangement might have occurred during transfer of the introns in these species. These regions may be a useful species marker for phylogenetic studies. This is the first evidence of group I introns probably transferred from non-vascular plant ancestors to a vascular plant mitochondrial genome. Considering the basal position of *Equisetum* in moniliformopses, it should be interesting to know whether these mitochondrial group I introns are conserved in other ferns, and transmitted to lycophytes.

The group II intron, *cox1i747,* is found only in *Equisetum* mitochondria

The third intron of *cox1*, *cox1i747,* is a group II intron of the subclass A as revealed by the canonical secondary structure that can be drawn from the nucleotide sequence (Fig. 3). The different domains D1–D6 radiating from a central wheel, and the tertiary structure interactions are clearly identified following the rules proposed by Michel et al. (1989). With the exception of some restricted conserved motifs common to group II introns, no sequences related to *cox1i747* are found in *M. polymorpha.* Moreover, it is inserted 18 downstream compared to the insertion position of *ai7*, a group I intron in *M. polymorpha cox1* (Fig. 4). Interestingly, a group II intron, referred as *i10*, in the moss *P. patens* also differs of *E. arvense* and *Marchantia* intron in sequence and insertion site (Terasawa et al. 2007). Contrary to the description given by the authors, *i10* has no homologue with any *cox1* intron from *M. polymorpha,* but corresponds to a group II intron encoding for a putative maturase-like protein. No intron from either fungi or protist was found at this insertion site in *cox1* sequences present in databases. However, a group I intron is found 7 nt upstream the i747 in *cox1* of several fungi (Gonzalez et al. 1998). It is striking that a restricted region of *cox1* has been targeted by unrelated intervening sequences in different species. In contrast, no introns were found in 25 different plants, including one fern, one psilophyte, and two lycophytes, by analyzing a region of *cox1* encompassing the insertion site (Sper-Whitis et al. 1996). Moreover, analysis of gene data bank indicates that no such intron is present in land plant

mitochondrial genomes described so far. This suggests that *cox1i747* is a new intron acquired by horizontal transfer during evolution of Equisetaceae, but was not transmitted to related pteridophytes.

*cox2i373* is highly conserved among land plants

*Marchantia* mitochondria share one group II intron with angiosperms, and nine are common with the moss *P. patens* introns (Chaw et al. 2008; Terasawa et al. 2007). *Equisetum cox2i*373 is conserved in *P. patens* and several vascular plants (Terasawa et al. 2007). Remarkable is the strong conservation of this intron throughout evolution. Indeed, the primary and secondary structure, even in the complex domain I, are very close to wheat mitochondria *cox2* intron (Farré and Araya 2002; Supplementary material S6). The main difference was found in domain IV which may vary from 212 in *Zea mays* (Covello and Gray 1990) up to 1,647 nt in *Huperzia lucidula* (Qiu et al. 2006 and accession no. DQ677486). Another group II intron (*cox2i696)* has been found in some angiosperms (reviewed by Bonen 2008), the lycophyte *H. lucidula* (accession no. DQ677486) and the moss *P. patens* (Terasawa et al. 2007). Our results clearly show that this intron is absent in *cox2* from horsetail *E. arvense* mitochondria.

It should be noted that ambiguous responses can be obtained when comparing either the *cox1* group I introns or the *cox2* group II intron of horsetails with other plant species. In the first case, *E. arvense* will be placed close to liverworts, while in the second, *Equisetum* will be closer to angiosperms. This particular situation point out that it is important to use a multifaceted approach in phylogenetic studies. Our results suggest that the *cox2* intron was acquired very early during vascular plant evolution.

Intron-encoded ORFs have been lost in the horsetail *Equisetum arvense*

Intron-encoded proteins are required for splicing (Saldanha et al. 1993). In *Marchantia*, ten ORFs were found in introns from six genes: two code for endonucleases and two for maturases in *cox1*, two encode for maturase-like protein in *atp1*, and one maturase-like ORF is present in *cox2*, *cob*, *atp9* and *rrn18* introns (Oda et al. 1992). In all higher plant mitochondria investigated so far, a single *Mat-r* like ORF is encoded in domain IV of one *nad1* intron (Bégu et al. 1998; Zhu et al. 2007). The endonucleases encoded for by group I introns have two LAGLIDADG motif characteristics of active fungal endonucleases (Colleaux et al. 1986). *Equisetum cox1i395* has only one such motif. Translation of all three reading frames is prevented by the deletion of four nucleotides at the very beginning of the intron, creating a premature stop codon (Fig. 2a). In some plant species, the presence of stop codons inside ORFs may be solved by U-to-C RNA editing (Malek et al. 1996). This is not the case since, as earlier discussed, *E. arvense* does not undergo U-to-C editing at least for the four transcripts studied here.

In many instances, maturase-related ORFs are found associated to group II introns and seem to be required for splicing (Lazowska et al. 1980). Group II introns from *E. arvense cox1* and *cox2* do not bear *mat-r* ORFs, a situation also found in three group II introns that interrupt *nad1* (Dombrovska and Qiu 2004). Moreover, no introns were found in *cob* and *atp9* (this work), *atp1* (Wikström and Pryer 2005) and *rrn18* (Duff and Nickrent 1997, accession no. AF058663). Taken together, these results indicate that compared to liverwort, *E. arvense* mitochondrial genome lacks of a large part of the introns, in particular the coding introns, and that the remaining ones are devoid of coding capacity. This situation raises the question of how mitochondrial introns are excised in the absence of intron ORFs. In higher plants, the transfer of mitochondrial genetic information into the nucleus is an ongoing process (Nugent and Palmer 1991). Interestingly, four ORFs presenting a relevant homology with the *mat-r* like proteins encoded in group II introns, *cox2i250* and *cobi824* of *Marchantia* are nuclear encoded in the seed plant *A. thaliana*, and potentially targeted to mitochondria (Mohr and Lambowitz 2003). One of them has been demonstrated to be involved in the splicing of a mitochondrial transcript (Nakagawa and Sakurai 2006). Moreover, a nuclear-encoded gene is responsible for the splicing of the unique group I intron present in land plant chloroplast (Asakura and Barkan 2007). It is tempting to propose that transfer from mitochondria to the nucleus of intron might have already occurred in the sphenophyte clade. If the evidence presented here support the hypothesis of a loss of coding introns in pteridophyte, it remains an open question until the complete sequence of the mitochondrial genome will be available.

**Acknowledgments** The authors acknowledge Christophe Hubert for helpful assistance in sequence analysis, and are greatly indebted to Simon Litvak for constant encouragements and support. This research was supported by the Centre National de la Recherche Scientifique, the Université Victor Segalen Bordeaux 2, France and the Ministère de l'Enseignement Supérieur et de la Recherche. DNA Sequencing was performed at the Genotyping and Sequencing Facility of Bordeaux (grants from the Conseil Régional d'Aquitaine no. 20030304002FA and 20040305003FA and from the European Union, FEDER no. 2003227).

## References

Araya A, Bégu D, Litvak S (1994) RNA editing in plants. Physiol Plant 91:543–550

Asakura Y, Barkan A (2007) A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts. Plant Cell 19:3864–3875

Bégu D, Graves PV, Domec C, Arselin G, Litvak S, Araya A (1990) RNA editing of wheat mitochondrial atp synthase subunit 9: direct protein and cDNA sequencing. Plant Cell 2:1283–1290

Bégu D, Mercado A, Farré JC, Moenne A, Holuigue L, Araya A, Jordana X (1998) Editing status of *mat-r* transcripts in mitochondria from two plant species: C-to-U changes occur in putative functional RT and maturase domains. Curr Genet 33:420–428

Bonen L (2008) *Cis*- and *trans*-splicing of group II introns in plant mitochondria. Mitochondrion 8:26–34

Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW, Tabak HF (1987) Structural conventions for group I introns. Nucl Acids Res 15:7217–7221

Chaw S-M, Shih AC-C, Wang D, Wu Y-W, Liu S-M, Chou T-Y (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, BpU sequences, and abundant RNA editing sites. Mol Biol Evol 25:603–615

Cho Y, Qiu YL, Kuhlman P, Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci USA 95:14244–14249

Colleaux L, d'Auriol L, Betermier M, Cottarel G, Jacquier Agalibert F, Dujon B (1986) Universal code equivalent of a yeast mitochondrial reading frame is expressed into *E. coli* as a specific double stranded endonuclease. Cell 44:521–533

Covello PS, Gray MW (1990) Differences in editing at homologous sites in messenger RNAs from angiosperm mitochondria. Nucleic Acids Res 18:5189–5196

Dombrovska O, Qiu Y-L (2004) Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol 32:246–263

Duff RJ (2006) Divergent RNA editing frequencies in hornwort mitochondrial *nad5* sequences. Gene 366:285–291

Duff RJ, Nickrent DL (1997) Characterization of mitochondrial small-subunit ribosomal RNAs from holoparasitic plants. J Mol Evol 45:631–639

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

Farré J-C, Araya A (2002) RNA splicing in higher plant mitochondria: determination of functional elements in group II intron from a chimeric *cox II* gene in electroporated wheat mitochondria. Plant J 29:203–214

Giegé P, Brennicke A (1999) RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFS. Proc Natl Acad Sci 96:15324–15329

Gonzalez P, Barroso G, Labarère J (1998) Molecular analysis of the split *cox1* gene from the Basidiomycota *Agrocybe aegerita*: relationship of its introns with homologous Ascomycota introns and divergence levels from common ancestral copies. Gene 220:45–53

Gualberto J-M, Lamattina L, Bonnard G, Weil J-H, Grienenberger J-M (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature 341:660–662

Groth-Malonek M, Pruchner D, Grewe F, Knoop V (2005) Ancestors of *trans*-splicing mitochondrial intron support serial sister group relationships of hornworts and mosses with vascular plants. Mol Biol Evol 22:160–174

Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Res 31:5907–5916

Henrick P, Crane PR (1997) The origin and early evolution of plants on land. Nature 389:33–39

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46:123–139

Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K (2003) RNA editing in hornwort chloroplasts makes more than half of the genes functional. Nucleic Acids Res 31:2417–2423

Lambowitz AM, Zimmerly S (2004) Mobile group II introns. Annu Rev Genet 38:1–35

Lang BF, Laforest M-J, Burger G (2007) Mitochondrial introns: a critical view. Trends Genet 23:119–125

Lazowska J, Jacq C, Slonimski PP (1980) Sequence of introns and flanking exons in wild-type and *box3* mutants of cytochrome *b* reveals an interlaced splicing protein coded by an intron. Cell 22:333–348

Lu MZ, Szmidt AE, Wang XR (1998) RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *cox1* gene. Plant Mol Biol 37:225–234

Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15:1403–1411

Michel F, Jacquier A, Dujon B (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie 64:867–881

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

Mohr G, Lambowitz AM (2003) Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. Nucleic Acids Res 32:647–652

Nakagawa N, Sakurai N (2006) A mutation in At-nMat 1a, which encodes a nuclear gene having similarity to group II intron maturase, causes impaired splicing of mitochondrial NAD4 transcript and altered carbon metabolism in *Arabidopsis thaliana*. Plant Cell 47:772–783

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki N, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA acquisition and loss during the evolution of flowering plants. Mol Genet Genomics 268:434–445

Nugent JM, Palmer JD (1991) RNA-mediated transfer of the gene cox-II from the mitochondrion to the nucleus during flowering plant evolution. Cell 66:473–481

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. J Mol Biol 223:1–7

Ohta E, Oda K, Yamato K, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K, Michel F (1993) Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares 5 intron positions with its fungal counterparts. Nucleic Acids Res 21:1297–1305

Pryer KM, Schneider H, Smith AR, Cranfilll R, Wolf PG, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409:618–622

Qiu YL, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA 103:15511–15516

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

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. Mol Biol Evol 25:1762–1777

Sper-Whitis GL, Russell AL, Vaughn JC (1994) Mitochondrial RNA editing of cytochrome-*c*-oxidase subunit II (coxII) in the primitive vascular plant *Psilotum nudum*. Biochim Biophys Acta 1218:218–220

Sper-Whitis GL, Moody JL, Vaughn JC (1996) Universality of mitochondrial RNA editing in cytochrome-*c* oxidase subunit I (*coxI*) among land plants. Biochim Biophys Acta 1307:301–308

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V (1999) Plant mitochondrial RNA editing. J Mol Evol 48:303–312

Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwata M, Sato N (2007) The mitochondrial genome of the moss *Physcomitrella patens* sheds a new light on mitochondrial evolution in land plants. Mol Biol Evol 19:24–38

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 15:57–61

Vangerow S, Teerkorn T, Knoop V (1999) Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. Plant Biol 1:235–243

Vaughn JC, Mason MT, Sper-Whitis GL, Kulman P, Palmer JD (1995) Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *cox1* gene of *Peperomia*. J Mol Evol 41:563–572

Vicens Q, Cech TR (2006) Atomic level architecture of group I introns revealed. Trends Biol Sci 31:41–51

Wahleithner JA, Macfarlane JL, Wolstenholme DR (1990) A sequence encoding a maturase-related protein in a group-II intron of a plant mitochondrial *nad1* gene. Proc Natl Acad Sci USA 87:548–552

Waring RB, Davies RW, Scazzochio C, Brown TA (1982) Internal structure of a mitochondrial intron of *Aspergillus nidulans*. Proc Natl Acad Sci USA 79:6332–6336

Wikström N, Pryer KM (2005) Incongruence between primary sequence data and the distribution of a mitochondrial *atp1* group II intron among fern and horsetails. Mol Phylogenet Evol 36:484–493

Zhu X-Y, Chase MW, Qiu Y-L, Kong H-Z, Dilcher DL, Li J-H, Chen Z-D (2007) Mitochondrial matR sequences help to resolve deep phylogenetic relationships in rosids. BMC Evol Biol 7:217–231

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415