© 1993 Oxford University Press

Nucleic Acids Research, 1993, Vol. 21, No. 5 1297–1305

Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts

Eiji Ohta, Kenji Oda, Katsuyuki Yamato, Yasukazu Nakamura, Miho Takemura, Naoko Nozato, Kinya Akashi, Kanji Ohyama* and François Michel[1]

Laboratory of Plant Molecular Biology, Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, 606-01, Kyoto, Japan and [1]Centre de Genetique Moleculaire du CNRS, Laboratoire associe a l'Universite Pierre et Marie Curie, 91190 Gif-sur-Yvette, France

Received August 20, 1992; Revised and Accepted January 21, 1993

GenBank accession no. M68929

ABSTRACT

The complete nucleotide sequence of the mitochondrial DNA (mtDNA) from a liverwort, *Marchantia polymorpha*, contains thirty-two introns. Twenty-five of these introns possess the characteristic secondary structures and consensus sequences of group II introns. The remaining seven are group I introns, six of which happen to interrupt the gene coding for subunit 1 of cytochrome oxidase (*cox1*). Interestingly, the insertion sites of one group II and four group I introns in the *cox1* gene coincide with those of the respective fungal mitochondrial introns. Moreover, comparison of the four group I introns with their fungal counterparts shows that group I introns inserted at identical genomic sites in different organisms are indeed related to one another, in terms of the peptide sequences generated from the complete or fragmental ORFs encoded by these introns. At the same time, the liverwort introns turned out to be more divergent from their fungal cognates than the latter are from one another. We therefore conclude that vertical transmission from a common ancestor organism is the simplest explanation for the presence of cognate introns in liverwort and fungal mitochondrial genomes.

INTRODUCTION

Group I and group II introns were originally described as two families of introns which, in addition to encoding proteins, possess unique secondary structures (1). Since then, the structured RNA components of group I introns have been shown to be responsible for the self-splicing phenomenon discovered by Cech and his collaborators (2), while various functions are ascribed to the proteins encoded by some of these introns. In the cases of the intron of the large ribosomal RNA gene of *Saccharomyces cerevisiae* and the 4th intron (ai4) in the gene (*cox1*) for subunit 1 of cytochrome oxidase from the same organism, the intron-encoded proteins have been shown to act as site-specific DNA endonucleases which recognize and cleave intron-less copies of the genes (3–5). Thus, they promote the insertion of introns into genes that lack them, and this ensures that the introns are retained during evolution. In other cases, e.g. the 2nd (bi2) and 4th (bi4) introns of the cytochrome *b* gene of *S.cerevisiae*, genetic data imply that the intron encoded proteins are required for splicing (hence their designation as maturases) (6–9). Since the proteins encoded by introns ai4 and bi4 of *S.cerevisiae* are homologous, the endonuclease and maturase functions encoded by the respective introns must also be somehow related (10). While all known group II introns are confined to organelle genomes (11), the known distribution of group I introns is much broader in living systems: group I introns have been found not only in mitochondria and chloroplasts, but also in the nuclear-encoded ribosomal RNA genes of protists (12–14), bacteriophages (15), cyanobacteria (16, 17), and more recently, other eubacteria (18). Rather paradoxically however, only group II introns have been described so far from plant mitochondria (19–25). We now report the presence of six group I introns in the *cox1* gene of the liverwort, *Marchantia polymorpha*. This paper is devoted to the description of these introns and a discussion of the problem of their origins.

MATERIALS AND METHODS

The complete nucleotide sequence of the liverwort mitochondrial DNA was determined in our laboratory (accession number M68929 in GenBank Data Library) (26, 27). Computer aided analysis of nucleotide sequences was carried out using the Hitachi DNASIS program on an NEC-9801VM computer, and the IDEAS program on a FACOM M-780 computer (Data Processing Center, Kyoto University).

* To whom correspondence should be addressed

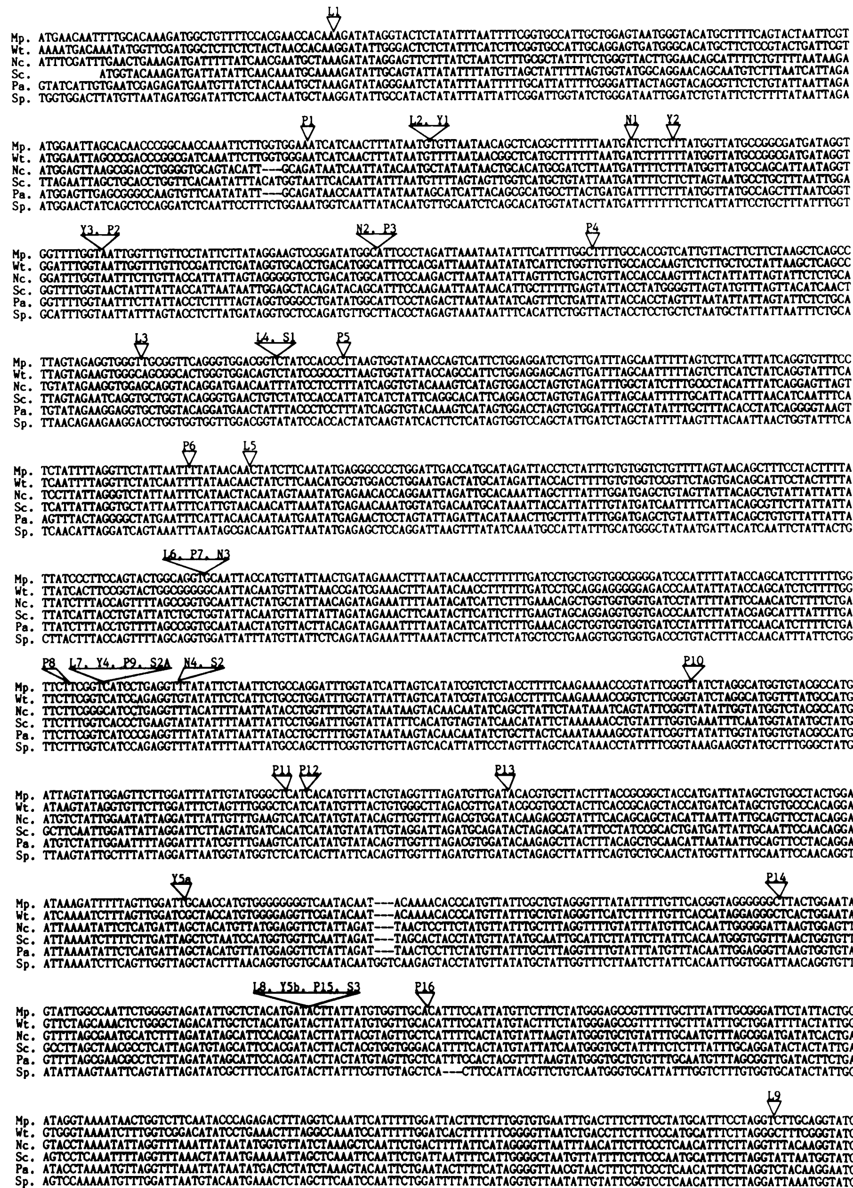

Figure 1. Alignments of *cox1* genes with sites of insertion of introns from liverwort (Mp)(this work), wheat (Wt)(28), *Neurospora crassa* (Nc)(29), *Saccharomyces cerevisiae* (Sc)(30, 31), *Podospora anserina* (Pa)(32), and *Schizosaccharomyces pombe* (Sp)(33, 34). Sites of intron insertion are marked by arrowheads with letters L1 to L9 for liverwort, N1 to N4 for *N.crassa*, Y1 to Y5b for *S.cerevisiae*, P1 to P16 for *P.anserina*, and S1 to S3 for *S.pombe*.

RESULTS

The split *cox1* gene of the liverwort mitochondrial genome

The 186608 base pairs (bp) of the liverwort mitochondrial genome potentially code for 94 genes, 17 of which are interrupted by a total of 32 introns (26, 27). Twenty-five of these introns were assigned to group II and the remaining seven belong to group I, based on their sequence and structure analysis (this paper and unpublished data). The seven group I introns are the 3rd, 4th, and 6th to 9th introns (abbreviated as ai3 to ai9) of the *cox1* gene coding for subunit 1 of cytochrome oxidase, and the sole intron in the *nad5* gene for subunit 5 of the respiratory-chain NADH dehydrogenase.

The exon–intron boundaries of the nine *cox1* introns were unambiguously determined by first comparing the liverwort DNA sequences with those of fungal and plant *cox1* genes (Figure 1), and then taking advantage of invariant structural and sequence features of group I and group II introns. Thus all known group I introns end with a G residue, and nearly all of them begin after an U residue (35). In addition, that U must be paired with a G within the P1 stem. Finally, most group I introns include a pairing (P10), whose function is to align the 3′ exon with the 5′exon at the ligation step (Figure 1). Note also that five out of six group I introns in the liverwort *cox1* gene begin after GGU and the remaining one begins after GAU.

Interestingly, while the *cox1* genes of higher plants contain no introns at all, more than half of the liverwort *cox1* introns happen

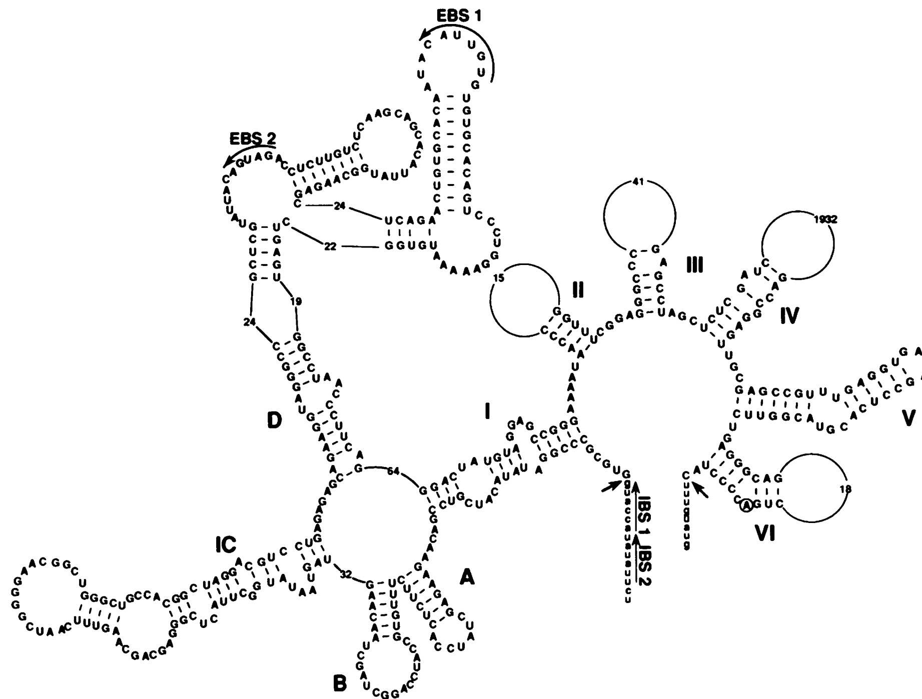

Figure 2. A secondary structure model of liverwort intron ai2 in the *cox1* gene. Thick arrows point to intron–exon boundaries. The pairing between the exon-binding (EBS1) and intron binding (IBS1) sites is indicated by thin arrows.

Table 1. Introns inserted at the same sites as liverwort introns in the mitochondrial *cox1* genes from other organisms.

Organisms	Number of introns group I	group II	Introns inserted in the same sites					ref.
Mp	6	3	ai2	ai4	ai6	ai7	ai8	
Wt	0	0	–	–	–	–	–	28
Cr	0	0	–	–	–	–	–	37
Nc	3	1	–	–	ai3	–	–	29
Pa	14	2	–	–	ai7	ai9	ai15	32
An	3	0	–	–	ai2	–	–	36
Sc	4	3	ai1	–	–	ai4	ai5b	30, 31
Sp	4	0	–	ai1	–	ai2a	ai3	33, 34
Kl	3	1	–	–	–	–	ai4	38
Sd	4	0	–	ai1	–	–	ai4	39
Tr	0	0	–	–	–	–	–	40
Hm	0	0	–	–	–	–	–	41
Pu	0	0	–	–	–	–	–	42
Tb	0	0	–	–	–	–	–	43

Wt wheat, Cr *Chramydomonas reinhardtii*, Nc *Neurospora crassa*, Pa *Podospora anserina*, An *Aspergillus nidulans*, Sc *Saccharomyces cerevisiae*, Sp *Schizosaccharomyces pombe*, Pu *Paramecium aurelia*, Kl *Kluveromyces lactis*, Sd *Saccharomyces douglasii*, Tr *Trichophyton rubrum*, Tb *Trypanosoma brucei* and Hm human. Some other plant *cox1* genes have been sequenced and all have no intron; these are rice (44), maize (45), pea (46), *Beta* (47), *Oenothera* (48) and radish (49). All but liverwort ai2 and S.c ai1 belong to group I introns.

to be inserted at sites where introns have been reported to exist in the genes of fungal mitochondrial genomes (Table 1).

Analysis of the liverwort group II intron ai2

Of the three group II introns (ai1, ai2 and ai5) in the liverwort *cox1* gene, only intron ai2 was analyzed here in order to allow comparison with yeast intron ai1. The secondary structure of intron ai2 (Figure 2) was established by comparative analysis (11), taking advantage of the presence of two rather close relatives of intron ai2 in the liverwort mitochondrial genome. One of them is intron 1 in the *cox3* gene, and the other (somewhat less similar) one is intron ai5 located further downstream in the *cox1* gene (in preparation). Comparison of the structures and sequences in Figure 2 with consensus structures and sequences of subgroup IIA and IIB introns, described by Michel *et al.* (11), showed that like yeast intron ai1, the three liverwort introns probably belong to the former, rather than the latter subgroup, in spite of the somewhat unusual structure of domain IC and the seven rather than six nucleotides separating the bulging A residue in helix VI from the 3′ intron–exon boundary. However, comparison of liverwort intron ai2 with yeast intron ai1 (1) and other fungal mitochondrial group II introns (11) did not reveal a particularly

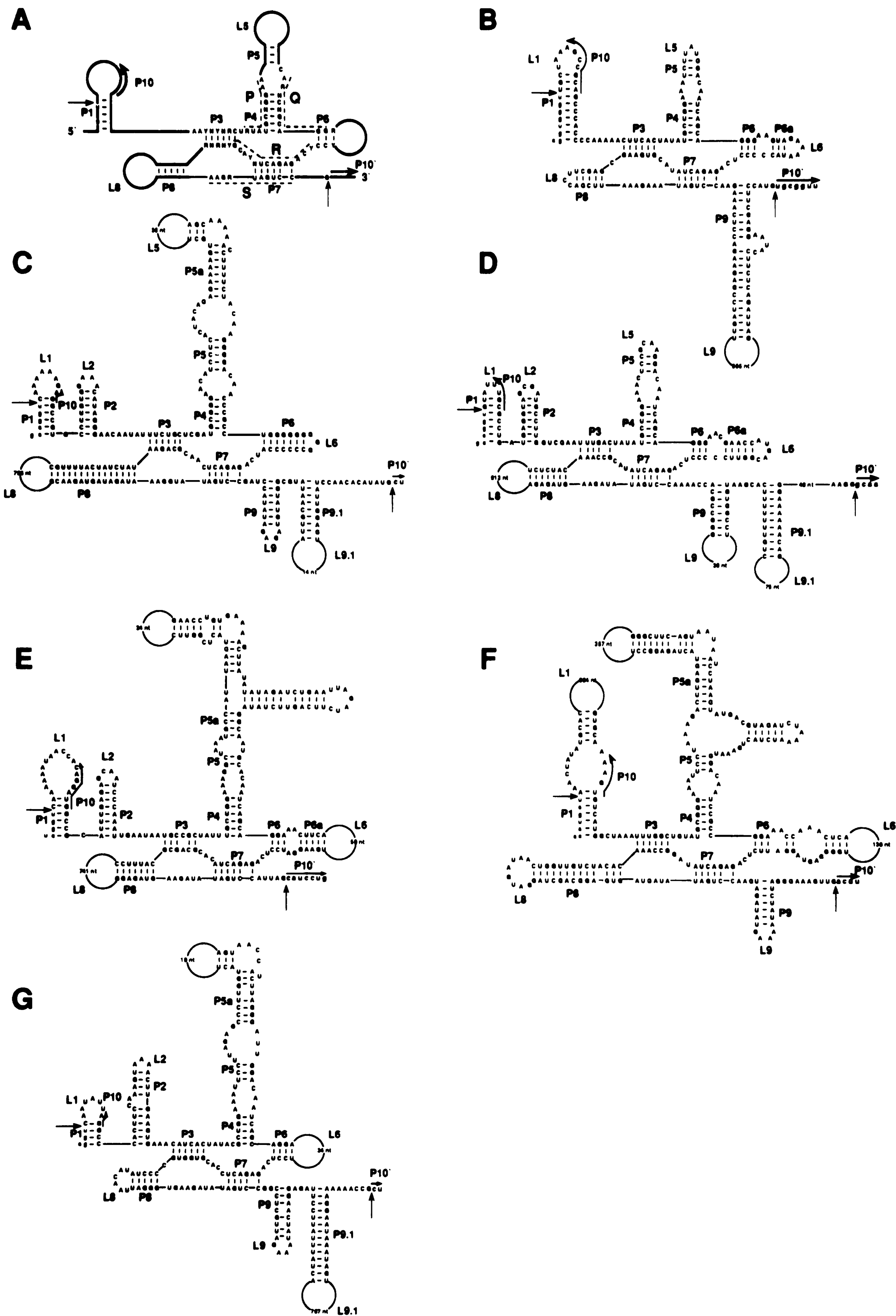

Figure 3. Secondary structure models of liverwort group I introns. Letters P and L refer to base paired stems and loops, respectively (51). Arrows point to intron–exon boundaries. **A)** Consensus secondary structure of the liverwort group I introns. Thin lines represent connections between adjacent nucleotides, whereas thicker lines indicate various lengths of nonconserved sequences. **B** to **H**). Secondary structure models of individual introns ai3, ai4, ai6, ai7, ai8 and ai9 in the *coxI* gene, respectively.

close relationship, in terms of either primary sequences or detailed secondary structure. In fact, mitochondrial subgroup IIA introns from such diverse organisms as yeast, fission yeast and filamentous fungi were significantly more closely related to one another than to those three liverwort introns.

Like all fungal subgroup IIA introns (11), liverwort intron ai2 potentially encodes a protein related to reverse transcriptases (50). The deduced ai2 protein sequence can be aligned over nearly its entire length with the ones translated from fungal group II introns (not shown), but the liverwort intron ai2 sequence did not show a particularly close relationship with any of the fungal ones.

Secondary structure analysis of liverwort group I introns

The 87 group I introns known in 1990 were arranged into four major subgroups as described by Michel and Westhof (35). Secondary structure models of liverwort group I introns (Figure 3) are all typical of subgroup IB, and the same is true of all the fungal introns inserted at the same sites as the liverwort introns ai4, ai6, ai7 and ai8. However, only in the case of intron ai4, was there evidence of a very close connection between the structured sections of a liverwort intron and its fungal counterpart, intron ai1 of *S.pombe*. Although these two introns were somewhat divergent in primary sequence, even within the conserved group I core, they shared many features which set them apart from most other group I introns: the sequence of the P3-P4 junction (UCGA) is unique; both have an U:C odd pair at the base of the P4 stem (as is also the case for liverwort intron ai8); the sequence of the P5 stem is the same (UCCUC:GGGGA); the P6 stem lacks an internal loop (a feature otherwise confined to subgroup ID); and the P7 stem ends with a C:U odd pair (the only other case is provided by liverwort intron ai9). Another counterpart of liverwort intron ai4 is intron ail of the *coxI* gene of *S.douglasii*, which also has an U:C odd pair at the base of the P4 stem and lacks an internal loop in its P6 stem.

The secondary structures and primary sequences of introns ai6 to ai8 (Figures 3 and 4) could be compared with those of their fungal counterparts as well as a broader selection of subgroup IB introns (35), but such comparisons remained inconclusive. The liverwort intron ai6 shares a short P5 stem with its three counterparts in filamentous fungi, but lacks a P9 stem with a GNRA loop, whereas all three fungal introns have one. Also, intron ai6 has a P2 stem, whereas its three counterparts lack one. The fungal counterparts of liverwort intron ai7 have no clear distinctive features and poorly conserved P2 and P9 stems. By contrast liverwort intron ai7 (like liverwort introns ai4 and ai6, and over 28 other group I introns) exhibits the most canonical form of the P2 stem, with a GNRA terminal loop and a distance of 12 bp between the L2 loop and the U:G pair of stem P1 that marks the 5′ splice site (35) but has no P9 stem at all. Note that while the absence of a P9/P9.0 pairing is unique among group I introns, intron ai7 does have an extra stable P10 pairing (7 Watson–Crick base pairs, including four G:C pairs) and this should suffice to ensure efficient ligation (35). Finally, the liverwort intron ai8 lacks the two extra stems between the P3 and P4 stems as well as the GUA sequence of the P6-P7 junction that are so characteristic of its five fungal counterparts (Figure 4), but shares with two of them (*P.anserina* ai15 and *K.lactis* ai4) a large insertion at the tip of the P1 stem, most of which consists of an ORF (see below).

```
             P            Q            R              S

Mp.ai4    CGAUGCGGGAA  AACCCGCCUG  UCAGAGACUCAA  AAGGUAUAGUCC
Sp.ai1    :::::::AA::  ::UU::::::  ::::::::::GU  ::::::::::::
Sd.ai1    UA::::AUA::  ::UAU:::AA  :::::::::A::  :::AC:::::::

Mp.ai6    AUAUGCUGGAA  AAUCAGCAGG  UCAGAGACUAUA  AAGAUAUAGUCC
An.ai2    ::::A::::::  ::::::U:::  ::::::::::::  ::A:::::::::
Pa.ai7    :AU::::::G:  ::::::::::  :::::G::::::  :::::::G:C::
Nc.ai3    :AU::::::::  ::::::::::  :::::G::::::  :::::::G::::

Mp.ai7    AUUUGCUGGAA  AAUCAGCAGG  UCAGAGACUACA  AAGAUAUAGUCC
Sc.ai4    :AA::::::::  ::::::::::  ::::::::::::  ::::::::::::
Pa.ai9    ::AA::GA:G:  :::UC:::A:  ::::::::C:U:  :::::::G::::
Sp.ai2a   ::AA:::::::  :::::::U::  ::::::::::U:  ::::G:::::::

Mp.ai8    GUAUGCGGGAA  UAUCAGCAGG  UCAGAGACUAUA  AUGAUAUAGUCC
Sc.ai5b   C::ACG:::UU  A:::CCGU:A  GU::::::::A:  :A::G:::::::
Pa.ai15   C::ACG:U:::  A::ACCGU::  GU::::::::A:  :A:U::::::::
Sp.ai3    C::ACG:U:::  A::ACCGU::  GU::::::::A:  :A:U::::::::
Kl ai4    C::ACGAA:::  A::UUCGU::  GU::::::::::  :A::C:::::::
Sd ai4    C::ACG:::::  A:::CCGU:A  GU::::::::A:  ::AU:U::::::
```

Figure 4. Alignment of conserved group I core sequences P, Q, R, S (see Figure 3A) in liverwort introns ai4, ai6, ai7 and ai8 and their counterparts in fungal *coxI* genes (29, 30, 32–34, 36, 38, 39). Abbreviation of organisms is the same as in Table 1.

Analysis of intronic open reading frames

Many group I introns potentially encode proteins and these have been arranged into several independent families based on the conservation of amino-acid sequences. The largest of these families is characterized by two related dodecapeptide motifs (1, 31, 52) and has been subdivided into many smaller groups (32). Two of the liverwort introns, ai4 and ai8, have long open reading frames continuous with the upstream exons (exonic reading frame) and both proteins translated from these reading frames exhibited characteristic dodecapeptide motifs. We have searched our database of intron-encoded proteins for possible close relatives of the ai4 and ai8 putative proteins. In both cases, the only sequence with which these ORFs could be aligned over nearly their entire length turned out to come from one of the fungal introns inserted at the same site, that is, *S.pombe* intron ai1 for liverwort intron ai4 (101 identical amino acids out of 233 aligned residues, see Figure 5A), and *P.anserina* intron ail 5 for liverwort intron ai8 (84 identical amino acids out of 248 aligned residues, see Figure 5D). This also coincided with the fact that, of the five fungal mitochondrial introns present at the same site as liverwort intron ai8, only *P.anserina* ai15 has an ORF that is inserted at the tip of stem P1 and at the same time has dodecapeptide motifs. Intron ai4 of *K.lactis* also has an ORF at the tip of P1 stem, but it doesn't have dodecapeptide motifs.

Neither of the two remaining liverwort introns (ai6 and ai7) with counterparts in fungal *coxI* genes contained a complete uninterrupted ORF continuous with the 5′ exonic reading frame or with an AUG codon close to its 5′ end. Rather, they each comprised several ORF fragments separated by stop codons or frameshifts (at least one in intron ai7, at least five in intron ai6; see Figures 5B and 5C). In view of that, one might be tempted to suggest that, as is the case in higher plants, the mitochondrial compartment of the liverwort includes an RNA editing machinery which would somehow remove all blocks to translation in these two introns. However, we believe this is an unlikely possibility, first, due to the variety of defects involved, and second, although the RNA/cDNA sequences of the interrupted reading frames encoded by the introns were not directly tested for RNA editing, due to our inability to detect RNA editing whenever we checked the questionable points of the liverwort sequences in the other genes by comparing the DNA and RNA/cDNA sequences (26, 54, Oda *et al.* unpublished results). Therefore, we regard these

Figure 5. Alignment of the complete or fragmentary ORFs in the liverwort introns. **A.**) ai4; **B.**) ai6; **C.**) ai7; and **D.**) ai8, with similar ORFs in fungal group I introns (29, 30, 32–34, 36, 53). Abbreviation of organisms is the same as in Figure 1. Amino acid residues are written in one letter symbols, and dashes are used to maximize the matching of amino acid sequences. Asterisks stand for stop codons or frameshifts (frameshifts were assigned to positions that maximize similarity between the fragmented liverwort ORFs and their fungal cognates: none of the frameshifts can be moved by more than a few codons without losing homology and/or introducing additional stop codons). The dodecapeptide motifs are boxed.

Table 2. Matrix of amino acid identities for intronic ORFs related to the one in liverwort intron ai6.

	Liverwort ai6	*A. nidulans* ai2	*N. crassa* ai3	*P. anserina* ai7a
A. nidulans ai2	78/252	–		
N. crassa ai3	75/252	166/274	–	
P. anserina ai7a	77/252	167/274	251/279	–
P. anserina ai7b	63/252	91/272	89/273	90/272

Numbers indicate amino acid identities in pairwise comparisons.

ORFs as remnants of a functional protein gene and note that a similar situation has been described in a number of fungal group I introns (55). In any case, the point here is, that just like for introns ai4 and ai8, the (pseudo-) proteins generated by gluing these ORFs together on paper are closely related to the ones translated from fungal introns inserted at the same place in the exonic sequence. As seen in Figure 5B and Table 2, the pseudo-protein translated from liverwort intron ai6 shares 78 amino acids with the one encoded in *A. nidulans* intron ai2, 75 amino acids with that from *N. crassa* intron ai3, and 77 and 63 amino acids with the two proteins potentially encoded in *P. anserina* intron ai7. As for the liverwort ai7 pseudo protein (Figure 5C and Table

Table 3. Matrix of amino acid identities for intronic ORFs related to the one in liverwort intron ai7.

	Liverwort ai7	*P.anserina* ai9	*P.anserina* ai10	*S.cerevisiae* ai4
P.anserina ai9	57/222	–		
P.anserina ai10	54/234	116/231	–	
S.cerevisiae ai4	60/221	109/231	92/231	–
S.cerevisiae bi4	62/219	106/229	94/229	151/232

Numbers indicate amino acid identities in pairwise comparisons.

3), it shares 57 and 60 amino acids, respectively, with *P.anserina* intron ai9 and yeast intron ai4 (and 54 and 62 amino acids, respectively, with closely related introns ai10 of *Podospora* and intron bi4 of yeast).

DISCUSSION

No group I intron has yet been reported from a higher plant mitochondria. Assuming group I introns are indeed absent from the mitochondrial compartment of these organisms, two contrasting explanations may account for their presence in the mitochondrial genes of the liverwort, *Marchantia polymorpha*. Either group I introns were present in the last common ancestor of bryophytes and angiosperms, and got lost in the branch leading to the latter; or they were missing from that common ancestor, and have subsequently been acquired somewhere along the branch leading to bryophytes. The latter proposal might seem to be far-fetched. However, the known homing properties of endonuclease-encoding group I introns, (i.e. their ability to invade intron-less copies of the gene (56)), may make them reasonable candidates for a successful landing in foreign lands. Moreover, there is precedence for what is likely to represent horizontal transmission of group I introns during the evolution of fungi. Probably the most compelling case comes from comparison of the mitochondrial genomes of *Aspergillus nidulans* and *S.pombe* (33). These two widely divergent organisms within the fungal kingdom have group I introns inserted at the same site in their *coxI* genes, but the point here is the fact that these introns are so similar in primary sequence that they are actually significantly less divergent than surrounding exons.

The first question that had to be asked in this context is whether the liverwort introns which happen to be inserted at the same locations in the *coxI* gene as fungal introns, are true cognates of the latter. Analysis of intron-contained ORFs provided a clear answer to this problem. We searched a database of over 50 group I intronic ORFs with so-called dodecapeptide motifs for relatives of the complete and fragmentary ORFs contained in liverwort introns ai4, ai6, ai7 and ai8. It was found (Figure 5) that the sequences most closely related to the liverwort ORFs were in all four cases associated with either those fungal introns that share the same insertion site, or an intron that clearly results from a relatively recent duplication of one of the latter (intron ai10 from intron ai9 of *Podospora anserina* and intron bi4 from intron ai4 of *S.cerevisiae* in Figure 5C). In the case of liverwort introns ai6 and ai7, there exist several ORFs in the database that can be aligned over most of their length with the liverwort sequences, so that rooted evolutionary trees can readily be generated by using the number of amino-acid identities in pairwise comparisons (Tables 2 and 3). The resulting trees have the liverwort sequences in an outgroup position, just as should be the case if they had been vertically inherited from a common ancestor organism. It also should be noted that the *Podospora* and *Neurospora* sequences are more similar to each other (Table 2) than to the one from *Aspergillus*, and those from *S.cerevisiae* are still very distant when compared to *Podospora* (Table 3), again as expected from the known phylogeny of these organisms in the absence of horizontal transmission events. Assuming the evolution of all ORFs was reasonably chronological, the relative timing of the transposition events can be estimated from the data in Tables 2 and 3. Thus, transposition-duplication of intron ai4 to a cytochrome *b* site in the yeast lineage would be after the date of divergence of *Aspergillus* from *Neurospora-Podospora*, whereas the duplication event that generated the ancestor of *Podospora* intron ai10 from that of intron ai9 would have occurred shortly after the divergence of yeast from the common ancestor of *Aspergillus*, *Neurospora* and *Podospora*. Finally, the event that created the ancestor of *Podospora* ai7b ORF and consisted in transposition-duplication of the ancestral *Podospora* ai7a ORF to a new location downstream from the P7 stem of the same intron must have barely post-dated the separation of plants from fungi. This is also supported by the fact that the compound ORF of liverwort intron ai6 is located at the tip of stem P8, as are the ORFs in *Podospora* ai7a, *Aspergillus* ai2, and *Neurospora* ai3 introns. One possible objection to this analysis is that the ORFs in the liverwort introns ai6 and ai7 are actually vestigial. This could lead to serious overestimates of phylogenetic distances. However, it is a fact that these ORFs have from 54 to 78 amino acids in common with their cognates in filamentous fungi (Tables 2 and 3), so that they are actually no more divergent than the presumably functional ORFs in liverwort intron ai4 and *S.pombe* intron ai1 (101 amino acid identities, Figure 5A), or than the ORFs in liverwort intron ai8 and *Podospora* intron ai15 (84 amino acid identities, Figure 5D).

Contrary to intron-encoded ORFs, the structural components of liverwort group I introns provide little ground for phylogenetic reasoning. Only in the case of liverwort intron ai4 and *S.pombe* intron ai1 do a liverwort intron and its fungal counterpart share a significant number of distinctive features, the combination of which sets them apart from all other group I introns. Other liverwort group I introns have no obvious affinities beyond their subgroup IB membership. This may simply mean that the rate of divergence of structural components and especially those nucleotides that are not constrained by multiple interactions (35) is high in comparison to the time elapsed since the divergence

of the exonic sequences in host organisms. One can not exclude the possibility that the ORFs and structured cores have had independent evolutionary histories (57). As pointed out by Belfort (58), a reasonable model for the homing phenomenon by which endonuclease-encoding group I introns invade the intron-less copies of their host genes allows occasional cleavage of the DNA copy of an intron and its subsequent invasion by the endonuclease-coding sequence. This is followed by a second transposition event of the whole intron with its immediate exonic surroundings to the (exonic) cleavage site preferred by the endonuclease. Thus, while the ORF would tend to home back to the same exonic site, it could bring to that site a variety of intron cores during evolution on the sole condition that the old and new surroundings were sufficiently similar for conversion to have been initiated, and splicing and gene function to remain active.

All that seems to be ruled out in this context is a plant chloroplast origin for the liverwort mitochondrial group I introns, as the only group I intron in plant chloroplasts, whose origins can be traced back to cyanobacteria (16, 17), belongs to a different subgroup (35). Algal chloroplast genomes also have group I introns and two of them belong to subgroup IB; these are intron 2 of the *psb*A gene of *Chlamydomonas moewusii* (35) and intron 5 of the LSU rRNA gene of *Chlamydomonas eugametos* (35). Both contain ORFs, but in L9.2 and L6, respectively, and these positions are different from those of liverwort ai4, ai6, ai7 and ai8. The ORF in the *C.moewusii* intron is unrelated to any of the group I intron-encoded proteins (59). Also, the ORF in the *C.eugametos* intron has only one dodecapeptide-like motif, while these motifs are usually repeated twice, and that ORF is related to those encoded in *Podospora* ND3, ND5, ND4L1, ND4L2 and COI5 introns (60), which all belong to subgroup IC. In conclusion, the relationship of algal introns with liverwort introns is not clear, since the only feature these introns have in common is subgroup IB membership. While we can not formally exclude that some ancestral chloroplast introns were inherited by mitochondria only to be lost on the way to higher plants, the available evidence makes it much more likely that the introns shared by liverwort and fungal mitochondria were vertically transmitted during evolution.

Finally, one may wonder why six out of the seven mitochondrial group I introns of liverwort happen to interrupt its *cox1* gene. Actually, most fungal mitochondrial genomes also have a disproportionate number of their known group I introns in the *cox1* gene (3 out of 5 in *Aspergillus*, 4 out of 8 in *S.cerevisiae*, 14 out of 33 in *Podospora*). This phenomenon is just as striking for group II introns, since 6 out of 9 known insertion sites of fungal group II introns are in the *cox1* gene. These introns might have been acquired by the ancestral *cox1* gene and maintained in the various extant descendant lineages, but why then were there so many introns in that ancestor gene? Since no answer is known to that riddle, two speculations are just worthwhile. The first one, invokes the fact that most of the intron-encoded proteins are translated from the 5′ exon, probably for regulatory purposes; perhaps it is statistically easier for them to retain their individuality and perform their functions when at least initially fused to *cox1* protein fragments. The second one postulates that the *cox1* gene was assigned a master role in mitochondria by the nucleus of the host cell. Thus, the expression of the *cox1* gene was regulated by the nucleus, and regulated in turn the biogenesis of mitochondria. Splicing, managed by some nuclear factors, would have provided a ready means of post transcriptional regulation.

ACKNOWLEDGEMENTS

This research was in part supported by grant-in-aids for science research in Priority Areas and for General Scientific Research from the Ministry of Education, Science and Culture, Japan, and by a Research grant from the Yamada Science Foundation. The sequence has been submitted to GenBank data library (accession number M68929). We also thank L.Jaeger, Centre de Genetique Moleculaire du CNRS, for his help in analyzing the secondary structure of introns.

REFERENCES

1. Michel,F., Jacquier,A. and Dujon,B. (1982) *Biochimie* **64**, 867–881.
2. Cech,T.R. (1988) *Gene* **73**, 259–271.
3. Colleaux,L., d'Auriol,L., Betermier,M., Cottarel,G., Jacquier,A., Galibert,F. and Dujon,B. (1986) *Cell* **44**, 521–523.
4. Delahodde,A., Goguel,V., Becam,A.M., Creusot,F., Perea,J., Banroques,J. and Jacq,C. (1989) *Cell* **56**, 431–441.
5. Wenzlau,J.M., Saldanha,R.J., Butow,R.A. and Perlman,P.S. (1989) *Cell* **56**, 421–430.
6. Lazowska,J., Jacq,C. and Slonimski,P.P. (1980) *Cell* **22**, 333–348.
7. de la Salle,H., Jacq,C. and Slonimski,P.P. (1982) *Cell* **28**, 721–732.
8. Anziano,P.Q., Hanson,D.K., Mahler,H.R. and Perlman,P.S. (1982) *Cell* **30**, 925–932.
9. Weiss-Brummer,B., Rodel,G., Schweyen,R.J. and Kaudewitz,F. (1982) *Cell* **29**, 527–536.
10. Delahodde,A., Goguelm,V., Becam,M., Creusot,F., Perea,J., Banroques,J. and Jacq,C. (1989) *Cell* **56**, 431–441.
11. Michel,F., Umesono,K. and Ozeki,H. (1989) *Gene* **82**, 5–30.
12. Wild,M.A. and Gall,J.G. (1979) *Cell* **16**, 565–573.
13. Nomiyama,H., Kuhara,S., Kukita,T., Otuka,T. and Sasaki,Y. (1981a) *Nucl. Acids Res.* **9**, 5507–5520.
14. Nomiyama,H., Sakaki,Y. and Takagi,Y. (1981b) *Proc. Natl. Acad. Sci. USA* **78**, 1376–1380.
15. Shub,D.A., Gott,J.M., Xu,M-Q., Lang,B.F., Michel,F., Tomaschewski,J., Pedersen Lane,J. and Belfort,M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1151–1155.
16. Kuhsel,M.G., Stickland,R. and Palmer,J.D. (1990) *Science* **250**, 1570–1573.
17. Xu,M-Q., Kathe,S.D., Goodrich-Blair,H., Nierzwicki-Bauer,S.A. and Shub,D.A. (1990) *Science* **250**, 1566–1570.
18. Reinhold-Hurek,B. and Shub,D.A. (1992) *Nature* **357**, 173–176.
19. Fox,T.D. and Leaver,C.J. (1981) *Cell* **26**, 315–323.
20. Wahleithner,J.A., Macfarlane,J.L. and Wolstenholme,D.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 548–552.
21. Chapdelaine,Y. and Bonen,L. (1991) *Cell* **65**, 465–472.
22. Knoop,V., Schuster,W., Wissinger,B. and Brennicke,A. (1991) *EMBO J.* **10**, 3483–3493.
23. Lamattina,L. and Grienenberger,J.M. (1991) *Nucl. Acids Res.* **19**, 3275–3282.
24. Wissinger,B., Schuster,W. and Brennicke,A. (1991) *Cell* **65**, 473–482.
25. Hunt,M.D. and Newton,K.J. (1991) *EMBO J.* **10**, 1045–1052.
26. Oda,K., Yamato,K., Ohta,E., Nakamura,Y., Takemura,M., Nozato,N., Akashi,K., Kanegae,T., Ogura,Y., Kohchi,T. and Ohyama,K. (1992) *J. Mol. Biol.* **223**, 1–7.
27. Oda,K., Yamato,K., Ohta,E., Nakamura,Y., Takemura,M., Nozato,N., Akashi,K., Kanegae,T., Ogura,Y., Kohchi,T. and Ohyama,K. (1992) *Plant Mol. Biol. Rep.* **10**, 105–163.
28. Bonen,L., Boer,P.H., McIntosh,J.E. and Gray,M.W. (1987) *Nucl. Acids Res.* **15**, 6734–6734.
29. Field,D., Sommerfield,A., Saville,B. and Collins,R.A. (1989) *Nucl. Acids Res.* **22**, 9087–9099.
30. Bonitz,S.G., Coruzzi,G., Thalenfeld,B.E., Tzagoloff,A. and Macino,G. (1980) *J. Biol. Chem.* **255**, 11927–11941.
31. Hensgens,L.A., Bonen,L., deHaan,M., van der Horst,G. and Grvell,L.A. (1983) *Cell* **32**, 379–389.
32. Cummings,D.J., Michel,F. and McNally,K.L. (1989) *Curr. Genet* **16**, 381–406.
33. Lang,B.F. (1984) *EMBO J.* **3**, 2129–2136.
34. Trinkl,H. and Wolf,K. (1986) *Gene* **45**, 289–297.
35. Michel,F. and Westhof,E. (1990) *J. Mol. Biol.* **216**, 585–610.
36. Waring,R.B., Brown,T.A., Ray,J.A., Scazzocchio,C. and Davies,R.W. (1984) *EMBO J.* **3**, 2121–2128.

37. Boer,P.H. and Gray,M.W. (1986) *Nucl. Acids Res.* **14**, 7506–7507.
38. Hardy,C.M. and Clark-Walker,G.D. (1991) *Curr. Genet.* **20**, 99–114.
39. Tian,G.-L., Michel,F., Macadre,C. and Lazowska,J. (1992) *Gene*, in press.
40. de Bievre,C. and Dujon,B. (1992) *Curr. Genet.* **22**, 229–234.
41. Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J. and Roe,B.A. (1980) *J. Mol. Biol.* **143**, 161–178.
42. Pritchard,A.E., Seilhamer,J.J., Mahalingam,R., Sable,C.L., Venuti,S.E. and Cummings,D.I. (1990) *Nucl. Acids Res.* **18**, 173–180.
43. Hensgen,L.A.M., Brakenhoff,J., De Vries,B.F., Sloof,P., Tromp,M.C., Van Boom,J.H. and Benne,R. (1984) *Nucl. Acids Res.* **19**, 7327–7344.
44. Kadowaki,K., Suzuki,T., Kazama,S., Oh-fuchi,T. and Sakamoto,W. (1989) *Nucl. Acids Res.* **17**, 7519–7519.
45. Isaac,P.G., Jones,V.P. and Leaver,C.J. (1985) *EMBO J.* **4**, 1617–1623.
46. Kemmerer,E.C., Kao,T, Deng,G. and Wu,R. (1989) *Plant Mol. Biol.* **13**, 121–124.
47. Senda,M., Harada,T., Mikami,T., Sugiura,M. and Kinoshita,T. (1991) *Curr. Genet.* **19**, 175–181.
48. Hiesel,R., Schobel,W., Schuster,W. and Brennicke,A. (1987) *EMBO J.* **6**, 29–34.
49. Makaroff,C.A., Apel,I.J. and Palmer,J.D. (1991) *Curr. Genet.* **19**, 183–190.
50. Michel,F. and Lang,B.F. (1985) *Nature* **316**, 641–643.
51. Burke,J.M., Belfort,M., Cech,T.R., Davies,R.W., Schweyen,R.J., Shub,D.A., Szostak,J.W. and Tabak,H.F. (1987) *Nucl. Acids Res.* **15**, 7217–7221.
52. Waring,R.B., Davies,R.W., Scazzocchio,C. and Brown,T.A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6332–6336.
53. Nobrega,F.G. and Tzagoloff,A. (1980) *J. Biol. Chem.* **255**, 9828–9837.
54. Ohyama,K., Ogura,Y., Oda,K., Yamato,K., Ohta,E., Nakamura,Y., Takemura,M., Nozato,N., Akashi,K., Kanegae,T. and Yamada,Y. (1991) Evolution of Life Eds. Osawa,S. and Honjo,T. 187–198.
55. Michel,F. and Cummings,D.J. (1985) *Curr. Genet.* **10**, 69–79.
56. Dujon,B. (1989) *Gene* **82**, 91–114.
57. Mota,E.M. and Collins,R.A. (1988) *Nature* **332**, 654–656.
58. Belfort,M. (1990) *Annu. Rev. Genet.* **24**, 363–385.
59. Turmel,M., Boulanger,J. and Lemieux,C. (1989) *Nucl. Acids Res.* **17**, 3875–3887.
60. Turmel,M., Boulanger,J., Schnare,M.N., Gray,M.W. and Lemieux,C. (1991) *J. Mol. Biol.* **218**, 293–311.