Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort

O. Malek and V. Knoop

RNA 1998 4: 1599-1609

RNA (1998), *4*:1599–1609. Cambridge University Press. Printed in the USA.

Trans-splicing group II introns in plant mitochondria: The complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort

OLAF MALEK and VOLKER KNOOP
Allgemeine Botanik, Universität Ulm, D-89069 Ulm, Germany

ABSTRACT

The fragmentation of group II introns without concomitant loss of splicing competence is illustrated by extraordinary gene arrangements in plant mitochondrial genomes. The mitochondrial genes nad1, nad2, and nad5, all encoding subunits of the NADH dehydrogenase, require *trans*-splicing for functional assembly of their mRNAs in flowering plants. Tracing the origins of *trans*-splicing group II introns shows that they have evolved from formerly *cis*-arranged homologs whose descendants can still be identified in lineages of early branching land plants. In this contribution we present the full set of ancestor introns for all five conserved mitochondrial *trans*-splicing positions. These introns are strikingly small in the quillwort *Isoetes lacustris*, the continuous nad2 gene intron in this species representing the smallest (389 nt) land plant group II intron yet identified. cDNA analysis shows correct splicing of the introns in vivo and also identifies frequent RNA editing events in the flanking nad gene exons. Other representatives of the ancestral *cis*-arranged introns are identified in the fern *Osmunda regalis*, the horsetail *Equisetum telmateia*, and the hornwort *Anthoceros crispulus*. Only the now identified intron in *Osmunda* carries significant traces of a former maturase reading frame. The identification of a continuous homolog in *Anthoceros* demonstrates that intron invasion into the affected genes in some cases predated the split of vascular and nonvascular plants more than 400 million years ago. As an alternative to disruption after size increase, the respective introns can get secondarily lost in certain lineages.

Keywords: intron evolution; recombination; RNA editing

INTRODUCTION

Group II introns are typical components of contemporary organelle genomes in plants, algae, fungi, and protists. Their modes of evolution are of particular relevance to hypotheses elaborating a phylogenetic relationship to the widespread nuclear spliceosomal introns of eukaryotes. Fragmentation of the characteristic group II secondary structure (Michel et al., 1989) may have led to the separate snRNAs in the nuclear spliceosome, now interacting with a largely unstructured intron sequence (Jacquier, 1990; Sharp, 1991). Whether the observed similarities between the two intron types actually indicate common origin or just chemical determinism (Weiner, 1993) is an open question (Michel & Ferat, 1995).

Possibly originating in eubacteria, where group II introns are now identified in increasing numbers (see Yeo et al., 1997, for a recent example), group II introns may have reached the eukaryotic cell in the genomes of the bacterial endosymbionts giving rise to mitochondria and chloroplasts. The absence of group II introns from animal mitochondria would, in the above scenario, be explained by secondary losses in line with the strict size reduction of these genomes. However, intron mobility (Grivell, 1996) and horizontal sequence transfer may likewise have played a major role in establishing the group II intron distribution observed today. The hypothetical evolutionary route from the conventional group II intron to the interacting individual snRNAs of the nuclear spliceosome is thus an attractive, but unproven, working hypothesis.

However, both in vivo and in vitro observations show that group II introns can retain splicing competence after disruption with the resulting separate RNAs interacting *in trans*. Most notably, a relevant group II intron structure, subdomain ID3, has recently been functionally replaced in in vitro reconstitution experiments by its potential spliceosomal counterpart, the U5 snRNA (Hetzer et al., 1997).

Reprint requests to: PD Dr. Volker Knoop, Allgemeine Botanik, Universität Ulm, 89069 Ulm, Germany; e-mail: volker.knoop@biologie.uni-ulm.de.

In the living world, examples of *trans*-splicing group II introns have initially been identified in chloroplast genomes of algae (Kück et al., 1987) and plants (Fukuzawa et al., 1986; Zaita et al., 1987). *Trans*-splicing group II introns are particularly frequent in mitochondria of flowering plants (angiosperms) with five to six documented examples of broken group II introns in three genes: nad1, nad2, and nad5 (Fig. 1). The peculiar gene arrangements of these three nad genes, encoding subunits of complex I, are conserved between dicot (Knoop et al., 1991; Wissinger et al., 1991; Binder et al., 1992; Handa et al., 1992) and monocot (Chapdelaine & Bonen, 1991; Pereira de Souza et al., 1991) species. Their origin thus likely predates this splitting among flowering plants at least 140 million years ago. The only as yet available complete seed plant chondriome sequence (Unseld et al., 1997) has not revealed further examples of *trans*-arrangements among its 23 group II introns. Group II intron positions are well conserved among flowering plants but strikingly different in the only other plant mitochondrial genome for which the sequence is fully available, the one of the liverwort *Marchantia polymorpha* (Oda et al., 1992). *Trans*-splicing is absent from the *Marchantia* chondriome.

The evolution of group II introns may alternatively follow different pathways in other genomic environments. In the chloroplast genome of the unicellular protist alga *Euglena gracilis*, group II introns are strikingly small and degenerate and apparently have given rise to a unique group of even smaller introns designated group III (Thompson et al., 1997; Doetsch et al., 1998). Both evolutionary routes, the general reduction of size and secondary structure features in *Euglena* chloroplasts and the frequent fragmentations observed in plant mitochondria, underline the evolutionary plasticity of group II introns.

We have recently started to elucidate the evolutionary history of the peculiar group II intron arrangements in the mitochondrial nad genes of angiosperms (Malek et al., 1997). These experiments had identified the first

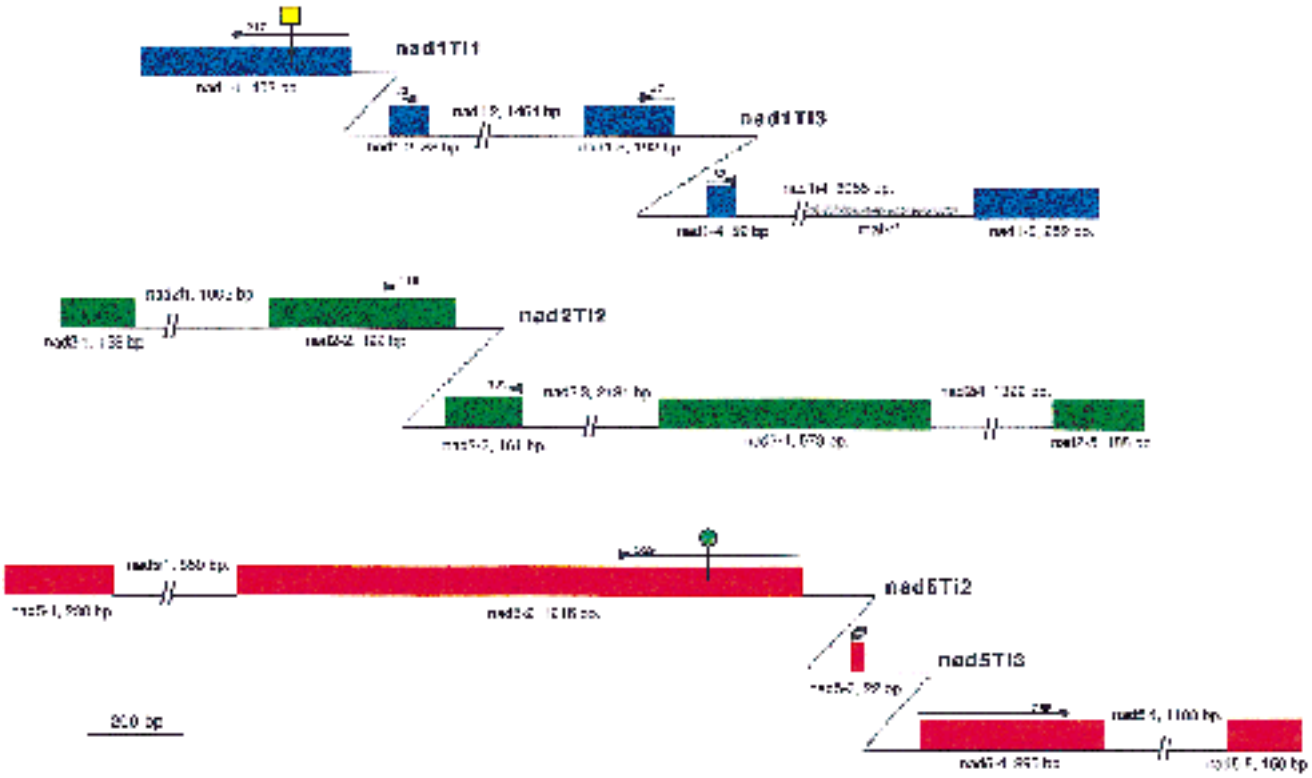


FIGURE 1. The *trans*-splicing genes nad1, nad2, and nad5 of flowering plants with exons drawn approximately to scale (sequences of the flowering plant *Oenothera berteriana* (evening primrose) are taken as reference). All intron sequences are group II. Only intron nad2i3 is positionally conserved in the liverwort *Marchantia polymorpha*. Intron nad1i4 carries the only known maturase reading frame (mat-r*) in flowering plant mitochondria. It is also the only known example of an intron that can be variably *cis*- or *trans*-arranged among flowering plants (Chapdelaine & Bonen, 1991; Conklin et al., 1991; Wissinger et al., 1991). Approximate locations of oligonucleotide primer sequences (see Materials and Methods) for the amplification approaches are indicated (arrowheads) and distances of primer ends to the splice sites are given except for exon nad5-3, the sequence of which had to be used entirely for primer design. *Cis*-arranged homologs for all *trans*-splicing introns nad1Ti1 through nad5Ti3 are now identified in nonseed plants as described in the text (Table 1). Serendipitously, group II introns at novel positions were identified in two cases: An intron detected in the nad1Ti1 identification approach is so far unique to the moss *Ceratodon purpureus* (square), another group II intron (circle) is common to most fern species so far investigated (S. Vangerow, T. Teerkorn and V. Knoop, unpubl.).

cis-arranged homologs to the *trans*-splicing introns in flowering plants. We now report the discovery of ancestral *cis*-arranged intron homologs for all known mitochondrial *trans*-splicing introns.

RESULTS

The experimental strategy to identify *cis*-arranged homologs to the *trans*-arranged introns in mitochondria of flowering plants (Malek et al., 1997) includes the selective amplification of PCR products from primers binding in the flanking nad gene exons, which can only be expected for continuous exon–intron–exon arrangements (Fig. 1). An appropriate size increase relative to the intronless loci (e.g., in the liverwort *Marchantia*) or to products derived from cDNA then suggests the presence of a *cis*-arranged intervening sequence in the amplified region. A designation is introduced for the purpose of this article to clearly identify the respective *trans*-splicing events in flowering plants and their *cis*-arranged counterparts reported here: nad1Ti1, nad1Ti3, nad2Ti2, nad5Ti2, and nad5Ti3.

The nad reading frames are noninterrupted at the respective positions in the mosses *Frullania tamarisci* and *Ceratodon purpureus* as judged from cDNA-sized products, the nature of which was confirmed in Southern blot hybridizations (not shown). The mosses thus reflect the noninterrupted liverwort gene arrangements in these regions. The single exception of a size-increased product among the mosses was identified for the nad1Ti1 amplification in *Ceratodon*. The corresponding PCR product was cloned and sequenced. In this case, the reading frame is noninterrupted at the *trans*-splicing position of angiosperms, but is interrupted by a different group II intron of 766 bp at a novel upstream position (Fig. 1 and database entry Y17810). Functional splicing of this intron was shown by RT-PCR and sequencing of the cDNA product.

The sequence conservation of nad genes was sufficient to allow amplification of cDNA-sized products in some cases from algae of the Charales group, which are believed to be closely related to land plants. PCR products were cloned and sequenced for nad1Ti1, nad5Ti2, and nad5Ti3 from *Chara corallina* and *Lamprothamnium papulosum* to confirm identity of the continuous and well conserved reading frames.

Continuous group II introns homologous to all five *trans*-splicing mitochondrial introns of flowering plants are now identified in lower, that is cryptogamous, plants (Table 1).

nad1Ti1

The first (and *trans*-splicing) intron nad1Ti1 in the nad1 gene of angiosperms is highly divergent from the group II intron consensus (Michel et al., 1989) in sev-

TABLE 1. Overview of identification of *cis*-arranged homologs to the *trans*-splicing introns of angiosperms.[a]

Intron	Species	Presence	Intron size	Accession number
nad1Ti1	*Isoetes lacustris*	present	520	Y17812
	Equisetum telmateia	absent	—	—
	Ceratodon purpureus	absent	—	Y17810
	Lamprothamnium papulosum	absent	—	AJ010253
	Chara corallina	absent	—	AJ010252
nad1Ti3	*Asplenium nidus*	absent	—	Y17972
	Osmunda regalis	present	2,809	Y17815
	E. telmateia	present	1,553	Y17811
	algae/mosses	absent	—	—
nad2Ti2	*A. nidus*	present	2,343	Y07910
	Marsilea drummondii	present	1,342	Y07911
	E. telmateia	absent	—	—
	I. lacustris	present	389	Y17813
	algae/mosses	absent	—	—
nad5Ti2	*A. nidus*	present	1,825	Y07912
	mosses	absent	—	—
	L. papulosum	absent	—	AJ010254
nad5Ti3	*A. nidus*	absent	—	AJ010257
	I. lacustris	present	434	Y17814
	Anthoceros crispulus	present	2,391	Y17809
	mosses	absent	—	—
	L. papulosum	absent	—	AJ010255
	C. corallina	absent	—	AJ010256

[a] All of these introns are known to be absent from the liverwort *Marchantia polymorpha* chondriome (accession M68929). Where no sequence information is available, evidence for intron absence comes from cDNA-sized PCR products obtained from DNA, the nature of which was confirmed by Southern blot analyses. Algae stands for *Lamprothamnium* and *Chara* and mosses stands for *Ceratodon* and *Frullania*, respectively, where hybridization data are available.

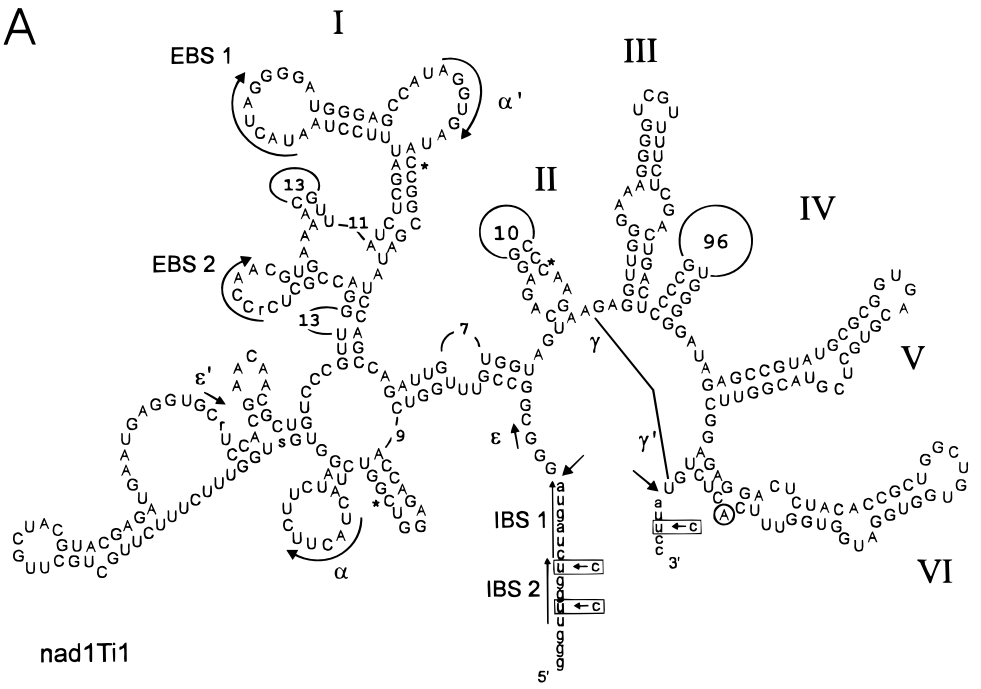


FIGURE 2. A: The secondary-structure model of the 520-nt group II intron nad1Ti1 in the quillwort *Isoetes lacustris* conforms much better to the consensus model than the *trans*-arranged counterparts in flowering plants. The intron sequence is given in capital letters and exon sequences in lower case letters. Three sequence ambiguities (r = G/A, s = G/C) could not be resolved in direct sequencing of the PCR product which was recalcitrant to cloning. The six typical domains radiating from a central wheel are designated by Roman numerals. Tertiary structure interactions α–α', γ–γ', ϵ–ϵ', and EBS–IBS (exon and intron binding sites 1 and 2) are indicated, a δ–δ' interaction is not discernible. The potential lariat branch point (adenosine) is encircled. Asterisks indicate potential sites of C-to-U RNA editing that could improve helical base-pairings. **B**: Functional splicing of the intron is shown by RT-PCR cDNA analysis, which also identifies 28 RNA editing events in the flanking exon sequences. Two of these affect the IBS2 binding site in the model above. C-to-U exchanges are identified by black boxes, U-to-C exchanges by grey boxes, respectively. The *Marchantia polymorpha* chondriome sequence (M68929) is shown for comparison to indicate the increase in sequence similarity resulting from RNA editing in the quillwort. Database accession number of the *Isoetes lacustris* sequence is Y17812.

eral aspects. Like the intron itself, these highly unusual features are present both in dicotyledonous (Conklin et al., 1991; Wissinger et al., 1991) and monocotyledonous (Chapdelaine & Bonen, 1991) flowering plants. In the evening primrose *Oenothera berteriana*, the conserved 5′ intron end (usually GtGcG) is replaced by a different sequence (ATTAGG), the presumed branching A nucleotide in domain VI is located at position −10 instead of −7 or −8, and an insertion of 14 nt extends the highly conserved domain V loop from the consensus of 4 to 18 nt.

In our selection of plants, only in the quillwort *Isoetes lacustris* was a *cis*-arranged group II intron identified as counterpart to this *trans*-splicing intron. The nad1Ti1 PCR product of *Isoetes* was recalcitrant to different approaches of cloning for reasons not yet understood and had accordingly to be sequenced directly. The sequence shows a 520-bp group II intron sequence with

conserved secondary- and primary-structure features much closer to the consensus than its *trans*-arranged counterparts in flowering plants (Fig. 2). The sequence abnormalities in the latter are now easily explained by small sequence insertions into the consensus structure after branching of *Isoetes* in the land plant phylogeny. The hexamer sequence ATTAGG is inserted before the conserved 5′ intron end (GGGCG) and the triplett CTC behind the branching A motif is duplicated in the stem of domain VI in the *Oenothera* sequence. The extra 14 nt in domain V are inserted in the middle of the conserved GUGA tetranucleotide loop motif. The secondary structure model supports the hypothesis that breakage of the *trans*-splicing introns in the angiosperm lineage has occurred in the variable loop of domain IV. The downstream parts of the angiosperm introns extending from the broken domain IV stem and including the signature domains V and VI show more than 80% nucleotide identity to the novel *Isoetes* intron sequence. Aligning the upstream part is more ambiguous because of large indels and a similarly high sequence identity is restricted to the first 30 nt of the intron sequences.

Splicing of the intron at the positions predicted from the secondary structure was confirmed by RT-PCR in *Isoetes*. The quillwort had earlier been identified as a species with a particularly high frequency of RNA editing in its mitochondria (Malek et al., 1996). The cDNA analysis in nad1 confirms this assessment by identifying the pyrimidine exchanges typical for plant mitochondrial sequences in the nad1 exon sequences (Fig. 2B). Twenty-five of 54 cytidine residues (46%) in an nad1 exon region of 240 bp are converted to uridines. Three further editing events occur in the reverse direction (U to C). Two editing events may influence splicing by affecting the IBS2 sequence in the secondary structure model.

No PCR products were obtained in the nad1Ti1 amplification approach with DNA of fern and gymnosperm species. An alternative upstream primer binding in a conserved region of the downstream half of the nad1Ti1 intron in front of exon nad1-2 (Fig. 1) allowed amplification of a product from DNA of the fern *Asplenium nidus* even with the downstream primer used for analysis of nad1Ti3 (see below). Cloning and sequencing showed the presence of the conserved downstream nad1Ti1 intron sequence without the characteristic insertions in domain V and VI of the angiosperm nad1Ti1 introns. The *Asplenium* sequence confirms absence of nad1Ti3 (see below) and simultaneously also of intron nad1i2 (Fig. 1) conserved among seed plants. Southern hybridization experiments suggest a *trans*-arrangement of intron nad1Ti1 also in this fern (not shown). The disruption of intron nad1Ti1 leading to the *trans*-arrangement may thus have taken place after separation of lycopods from the main vascular plant lineage but before separation of ferns and seed plants in this case (Fig. 3).

nad1Ti3

Continuous pendants to the *trans*-splicing intron nad1Ti3 of flowering plants were isolated from the fern *Osmunda regalis* and the horsetail species *Equisetum hyemale* and *E. telmateia*. PCR products were cloned and fully sequenced for *O. regalis* and *E. telmateia* and

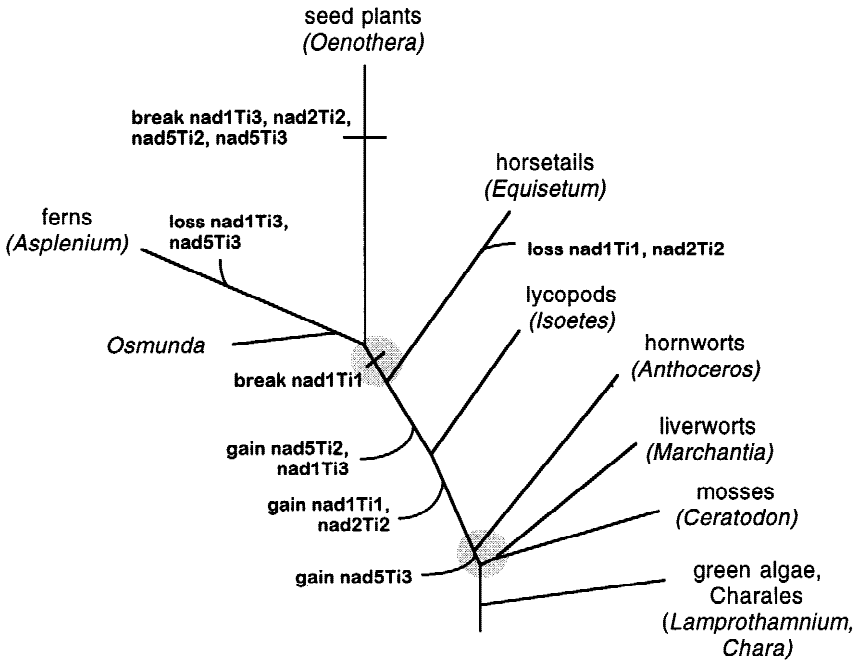


FIGURE 3. A possible scenario of *trans*-splicing intron evolution. The presented data suggest intron appearance, disappearance, and breakage as mapped on a current view of land plant phylogeny in a parsimonious interpretation. Intron loss or breakage is indicated where sequence data or Southern hybridization data are available. Additional intron breakage may have occurred for intron nad5Ti2 in the *Osmunda* lineage and for introns nad5Ti2 and nad5Ti3 in the *Equisetum* lineage where no PCR products have been obtained. Grey shadings indicate areas where information on branching pattern is currently lacking. A junction of liverworts and mosses is supported by mitochondrial sequence data (Beckert et al., in press), the branching order in relation to hornworts remains unclear. Names of model genera investigated here are given below designations of the main lineages. Recent additional group II intron disruptions have occurred independently in the nad1i4 intron in wheat (Chapdelaine & Bonen, 1991) and *Petunia* (Conklin et al., 1991). Intron nad5Ti3 got disrupted secondarily in domain I leading to a tripartite intron structure in the flowering plant *Oenothera berteriana* (Knoop et al., 1997).

clearly identify conserved group II introns with sizes of 2.8 and 1.5 kbp, respectively. The large size variation is restricted to domain IV, both introns being highly similar to each other and to the angiosperm introns in the structurally relevant parts. Comparison of the fern ally and fern sequences to those of angiosperms reveal sequence identities exceeding 80% both in the downstream part and first 100 bp of the upstream part and reflect the closer phylogenetic relatedness in comparison to the *Isotes* nad1Ti1 intron. The deduced secondary structures feature most consensus elements (Fig. 4) with the exception of a CG replacing the conserved AY dinucleotide motif at the downstream intron end. This deviation is conserved with the *trans*-splicing homologs of the angiosperms (Chapdelaine & Bonen, 1991; Wissinger et al., 1991). Maturase homologies extending over essentially the entire loop of domain IV in the *Osmunda* sequence are clearly identified in BLASTX (Altschul et al., 1997) searches. The highest similarity (BLAST estimate of chance similarity = 3×10^{-49}) is observed with the maturase of the third intron in the mitochondrial cob gene of *Marchantia* (accession S25952). Six frameshifts and two in-frame stop codons in a total region of 2,030 bp including the maturase homology preclude functional expression in *Osmunda*, however.

Functional splicing of the *Osmunda* nad1Ti3 intron at the sites predicted from the secondary structure model was shown by cDNA analysis. The cDNA sequence additionally identifies three typical RNA editing events of the C to U type (Fig. 4B).

In this case, it is interesting to note that cDNA-sized products have been obtained from other ferns (*Sphaeropteris, Marsilea,* and *Asplenium*) and exemplary sequencing of the corresponding product in *Asplenium* confirmed identity of the product and absence of the corresponding intron. With the eusporangiate fern *Osmunda* and the fern ally *Equisetum* definitely occu-

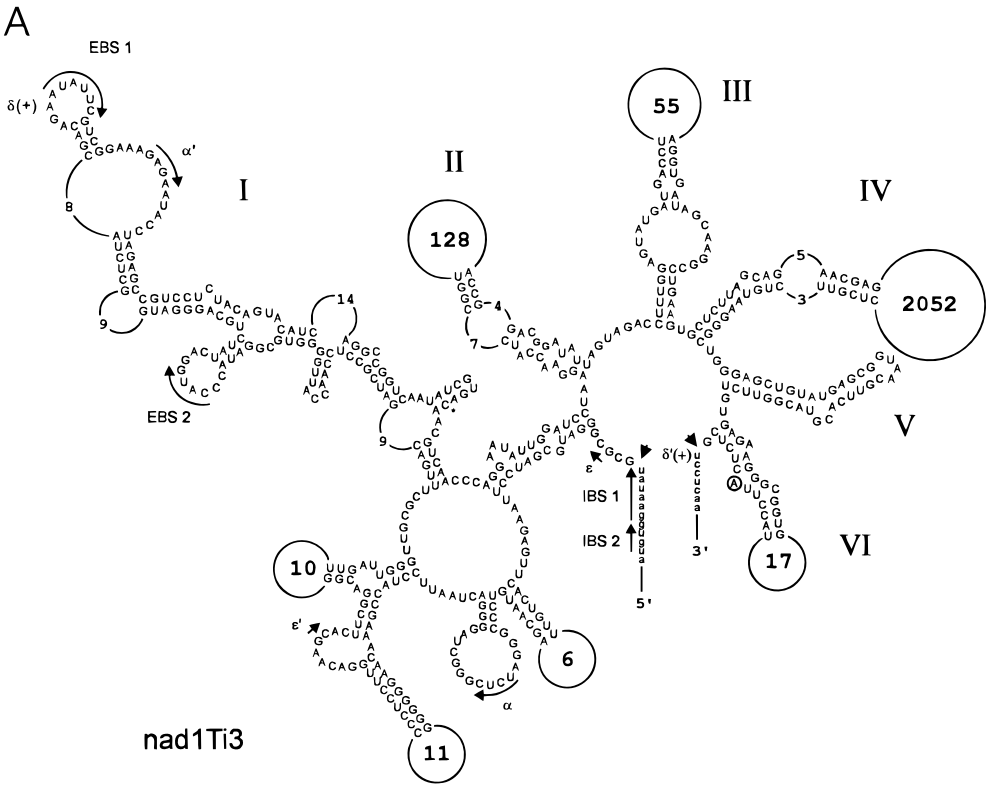


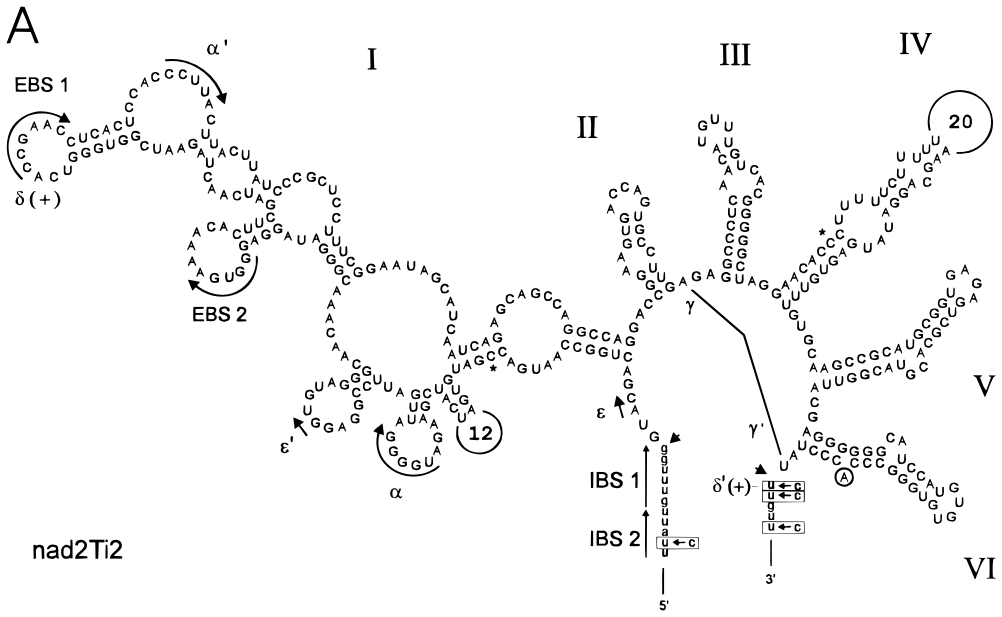
FIGURE 4. **A**: The intron nad1Ti3 secondary structure is shown for the sequence of the fern *Osmunda regalis.* Differences in the nad1 gene sequence of the horsetail *Equisetum telmateia* are restricted to regions with minor influence on the secondary structure shown. Database accession numbers are Y17815 and Y17811, respectively. Designation of intron features is as in Figure 2. **B**: Functional splicing of the intron is shown by RT-PCR and cloning of the respective product obtained from *Osmunda* cDNA. Editing positions are indicated as in Figure 2.

pying more basal positions in the plant phylogeny, it is obvious that intron nad1Ti3 fares differently after occupying this position early in land plant evolution (Fig. 3). It got secondarily lost in the lineage to the advanced (leptosporangiate) ferns like *Asplenium*, but was disrupted by recombination events in the lineage leading to seed plants.

nad2Ti2

Cis-arranged counterparts to the *trans*-splicing intron nad2Ti2 of seed plants were first identified in the two leptosporangiate ferns *A. nidus* and *Marsilea drummondii* (Malek et al., 1997). The homologous introns are also present in the ferns *Osmunda, Sphaeropteris*, and *Angiopteris,* identified by cloning and partial sequence analysis (not shown). The *cis*-arranged mitochondrial nad2Ti2 intron thus appears to be common among ferns.

A particularly small PCR product of 700 bp was obtained in the nad2Ti2 amplification approach from the quillwort *Isoetes*. Cloning and sequencing identified the smallest yet known group II intron in plants (389 bp) which still conforms to all signature elements of group II introns (Fig. 5). The compact *Isoetes* homolog even allows definition of an α–α' interaction not easily deducible from the fern sequences. The extraordinary small domain IV loop (20 nt) is extended more than 80-fold in *Asplenium*. Sequence comparisons reveal significant identity with the angiosperm introns in the downstream intron halves of the latter and exceed 80%, but unambiguous aligning is again restricted in



B

```
Marchantia  CTT CAA AGT TTA TGT TTT TAT GTG ATC GCA GCA TCA AAA AGA GAC TCT GAA TTT TCC ACA  60
  Isoetes   C■■ CAG AGT T■A ■GT T■T ■AT GTG ATC GCA GCA ■CA AGA AGA GAC ■CT GAA TTT TCC ACG
            P>L         S>L R>C S>F H>Y               P>S           P>S

            GAA GCC GGC TTA AAA TAT TTT ATT TTA GGT GCT TTC TCC TCT GGA ATA TTA TTG TTT GGT 120
            GAA GCC GGC T■A AAA TAT T■T ATT TTA GGC GCA TTC TCC ■CC GGA ATA T■A TTG TTT GG■
                        S>L         S>F                         P>S         S>L
                                                                          IBS2  IBS1

            TGT TCT ATG ATT TAT GGG TTT ACT GGA GTT ACC AAC TTT GAG GAA TTA GCT AAG ATT TTC 180
            ■GT ■CC A■G ATC TAT GGA TTG ACT GGA GTG ACT GAC T■T GAG GAA T■G GCT AAG ATT TTC
            R>C P>S T>M                                   S>F         S>L

            ACT GGA TAT GAA ATA ACT TTG TTC GGT GCT CAA TCT AGT GGT ATC TTT ATG GGA ATT TTA 240
            GCT GGA TAT GAA ATC ACT TTG TTC AGT GCT ▒AA TCT AGT GGT CTC TTT C■G GGG ATC TTA
                                              *>Q                       P>L

            TTT ATC GCT GTA GGT TTT TTA TTT 264
            TTT ATA GCC GTA GGT ■TT ■C■ CTT
                              L>F P>S
```

FIGURE 5. **A**: The 389-nt nad2Ti2 intron in *Isoetes lacustris*. Designation of secondary- and primary-structure features is as in Figure 2; accession number is Y17813. **B**: Splicing of the intron is confirmed by cDNA analysis, which also identifies RNA editing events (indicated as in Figure 2) in the flanking nad2 exon sequences.

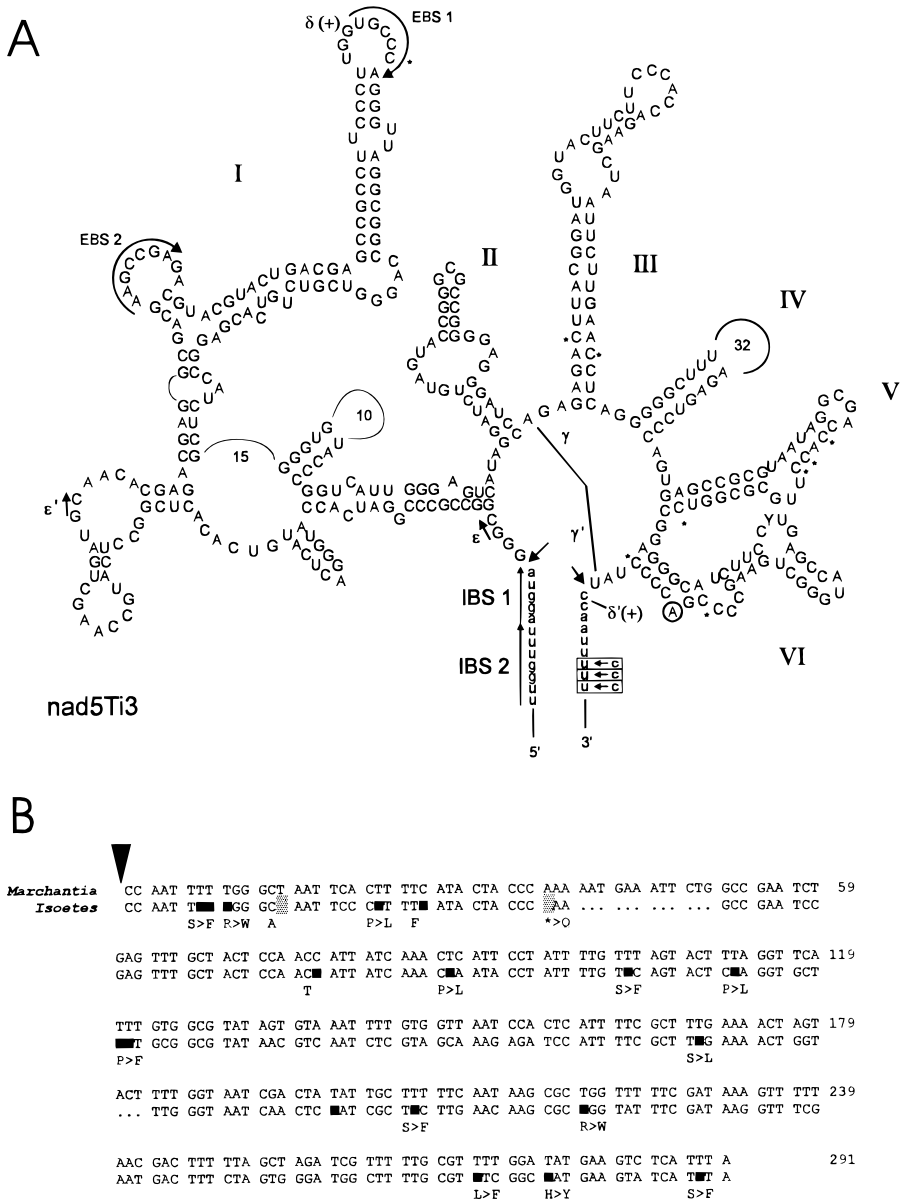


FIGURE 6. A: The 434-nt nad5Ti3 intron in *Isoetes lacustris* (database accession Y17814). Designation of primary- and secondary-structure features is as in Figure 2. Asterisks indicate positions of potential C-to-U editing. IBS sequences are covered by the 5′ amplification primer extending over the entire small upstream exon. The upstream sequence shown is taken from *Oenothera* (Knoop et al., 1991). Homologous introns are present in the hornwort *Anthoceros crispulus* and the fern *Osmunda regalis*. The *Anthoceros* intron homolog (not shown; accession Y17809) is significantly larger (2,391 nt) and in combination with observations of a particular high frequency of C-U RNA editing in both directions (Steinhauser et al., in press) the secondary-structure folding is more ambiguous. **B**: Splicing of the intron is confirmed by cDNA analysis, which also identifies RNA editing events (indicated as in Figure 2) in the flanking nad5 exon sequences.

the upstream region. The noncanonical upstream intron end GTACG in the fern and angiosperm sequences is, however, conserved in *Isoetes*. The abnormality of a looped-out guanosine nucleotide in domain VI of the fern nad2Ti2 introns is not shared by *Isoetes,* which features the consensus A at the branch site for lariat formation. A mismatch replacing the canonical GC pair at the base of domain V is, however, conserved between *Isoetes*, the ferns, and the angiosperms. Further minor improvements of the secondary structure may be achieved by RNA editing events reconstituting base pairings in the stems of domain I, IV, and V. However, editing in intron sequences has only been observed rarely and cannot be reliably predicted from secondary structure modeling (Carrillo & Bonen, 1997).

Splicing of the nad2Ti2 intron in *Isoetes* was shown by RT-PCR at the sites predicted from the secondary structure (Fig. 5B). The cDNA analysis concomitantly identified 23 RNA editing events in the flanking nad2 exon sequences of *Isoetes*. The reconstitution of con-

served codon identities includes removal of a genomically encoded stop codon from the nad2 reading frame in this region.

nad5Ti2

A *cis*-arranged homolog of the nad5Ti2 *trans*-splicing intron was initially identified in *Asplenium* (Malek et al., 1997). This approach had serendipitously revealed a novel group II intron insertion 212 bp upstream of nad5Ti2 (Fig. 1). A phylogenetic survey of the nad5 gene has shown that this intron is generally present in fern species but absent from mosses and liverworts (unpubl. data). Contrary to the nad2Ti2 example, no PCR products were obtained from other ferns, possibly indicating that the *cis*-arranged nad5Ti2 intron in *Asplenium* is an exception and that disruption events like the one in angiosperms have occurred independently in the other lineages.

nad5Ti3

The amplification approach for nad5Ti3 revealed intron sequences for the widest phylogenetic spectrum, including for the first time a nonvascular plant, the hornwort *Anthoceros crispulus.* A particularly small intron was again identified in *I. lacustris*. With a length of 434 nt, it is only slightly larger than the nad2Ti2 intron described above. All relevant consensus elements are present in the deduced secondary structure (Fig. 6). Modeling of the intron had to take into account the upstream nad5 exon sequences of flowering plants, since the upstream PCR primer covers this small 22-bp exon entirely (Fig. 1). Functional splicing of the *Isoetes* intron was again confirmed by cDNA analysis after cloning of RT-PCR products, identifying 20 RNA editing events in the downstream nad5 exon (Fig. 6B). The homologous introns in the hornwort *A. crispulus* and the fern *O. regalis* are significantly larger with sizes of 2.4 and 2.8 kbp, respectively. The *Anthoceros* intron was fully sequenced and shows sequence similarity exceeding 80% with the *Isoetes* and angiosperm sequences in the downstream region. The highly condensed upstream region of the *Isoetes* intron has lost most similarities to its homologs while the *Anthoceros* nad5Ti3 sequence with a total length of 2,391 bp is significantly larger, and alternative secondary structure foldings of domains I through IV can be envisioned, most notably when the high potential for C-U RNA editing in both directions characteristic for *Anthoceros* (Malek et al., 1996, Steinhauser et al., in press) is taken into account. An additional split in domain I of the already *trans*-splicing intron nad5Ti3 has resulted in a tripartite RNA structure in *O. berteriana* (Knoop et al., 1997).

DISCUSSION

The complete set of *cis*-arranged counterparts to all five *trans*-splicing mitochondrial introns conserved among seed plants has been identified. The continuously arranged, and most likely evolutionary ancestral, homologs have survived in at least one or the other lineage of cryptogamous plants: ferns, horsetails, lycopods, and hornworts (Fig. 3). All *trans*-splicing introns in plant mitochondria are thus obviously products of intron disruptions by genomic recombinations in the advanced vascular plant lineage(s). The particularly small introns in the quillwort *Isoetes* suggest that the chondriome in this species has taken a different route of evolution in comparison to the more complex vascular plants. A secondary reduction of intron size in this species after or in line with a loss of intron-encoded maturase reading frames is likely. A size increase is, on the other hand, apparent in the (still) *cis*-arranged fern (and hornwort) introns and was also the likely prerequisite to offer an easy target for the recombinatorial activity in the seed plant lineage.

The possibility that the here-described *cis*-arranged intron homologs to the *trans*-splicing introns of angiosperms are not orthologous but represent separate, independent insertion events can never be ruled out with ultimate confidence. Even significant intron similarities could possibly be explained by the conservation of a yet unidentified intron source locus leading to independent copies at distant time points in evolution. It clearly remains to be determined where the introns that occupied the nad gene loci very early in land plant diversification originated. Integration into the new loci by reverse splicing followed by reverse transcription employing intron-encoded maturase activities is an obvious possible explanation. Only intron nadTi3 in *Osmunda,* however, clearly documents remnants of a maturase reading frame rendered nonfunctional by frameshift mutations.

Based on two efficiently self-splicing yeast mitochondrial prototypes, group II introns are often generally considered autocatalytic. However, no self-splicing activity has as yet been reported for any land plant group II intron. This lack of activity has sometimes been ascribed to degenerate secondary structures or an absolute dependence on protein cofactors. The particularly small nad2Ti2 intron in the quillwort *Isoetes* with its clearly defined secondary-structure features appeared as a good novel candidate for detection of self-splicing activity. However, no autocatalytic activity is detected after incubation of the precursor intron in diverse buffer conditions known to promote group II intron self-splicing (S. Steinhauser & V. Knoop, unpubl. observations).

The models presented here suggest that RNA editing events may affect the RNA secondary structures in some cases. While prediction of RNA editing positions in introns has turned out to be unreliable in seed plants

(Carrillo & Bonen, 1997), a different picture may emerge for *Isoetes* with its particularly frequent RNA editing in coding regions.

The intron secondary-structure models confirm earlier notions that intron disruption is targeted to domain IV. Most likely, the current evolutionary status quo in seed plants reflects that only disruptions in this variable intron domain can be compensated by tertiary interactions (e.g., between domain V and domain I). However, although DNA recombination in seed plants is frequent enough to account even for infraspecific genome variations, only the same set of five introns has so far been identified as *trans*-splicing in all species investigated. These five introns, the evolutionary history of which is traced here, may carry particularly strong or additional, yet to be identified, tertiary structure interactions that can more efficiently compensate for the loss of a continuous arrangement. At least one of them, nad5Ti3, tolerates a secondary disruption in domain I (Knoop et al., 1997). The resulting tripartite group II intron structure strongly reminds us of the only other known example for a twice-fragmented intron in the *Chlamydomonas reinhardtii* chloroplast (Goldschmidt-Clermont et al., 1991). Alternatively, and in line with the observed lack of self-splicing activity, protein cofactors play the essential role both in *cis*-splicing and in the bi- or trimolecular reactions of *trans*-splicing group II introns in plant organelles.

MATERIALS AND METHODS

Nucleic acid preparation

Approximately 0.5–3 g of plant leaf material was used for total cellular DNA preparations by the CTAB (Cetyltrimethylammoniumbromide) method (Doyle & Doyle, 1990). Plant species under investigation are given in Table 1 and detailed in the text. Total nucleic acids were fractionated into RNA and DNA by differential precipitation in the presence of 2 M lithium acetate. Subsequently, the crude DNA preparations were treated with RNaseA and the RNA fraction with RNase-free DNase.

Molecular biology techniques

Synthesis of cDNA was done with the respective kit from Boehringer Mannheim in the presence of random hexamer primers as recommended by the manufacturer. Sequences of the oligonucleotides flanking the *trans*-splicing sites of angiosperm nad genes as outlined in Figure 1 were as follows (5′ to 3′):

nad1Ti1 upstream: gttacaacctgcagcagatggtttg, downstream: ccatttgagctgcagatcgtaatgc,
nad1Ti3 upstream: gaaactaatcgagctccgtttgatc, downstream: ctcattaagatcttattggcatactc,
nad2Ti2 upstream: attgccatggatttagctattgag, downstream: gaaaaggaactgcagtgatctt,
nad5Ti2 upstream: gtgattcatgccatggcggatgagc, downstream: tacctaaaccaatcatcatatc,
nad5Ti3 upstream: gatatgatgattggtttaggta, downstream: caat agcacctttgtctaaagctt.

Oligonucleotide pairs were used for PCR amplification in a Biomed waterbath thermocycler with annealing temperatures of 45–50 °C. PCR products were cloned into the Bluescript SKII+ (Stratagene) vector except in one case where the product had to be sequenced directly (see text). Sequencing of the cloned products was done by the dideoxy method in the presence of α-[^{35}S]-dATP or with Cy5-fluorescence-labeled primers with subsequent electrophoresis in the ALF express sequencer (Pharmacia). Sequence handling was done with the UWGCG (Genetics Computer Group, University of Wisconsin) software package 8 for VAX and UNIX (Devereux et al., 1984). For Southern Blot analyses, samples of the PCR reactions were run on 0.8% agarose gels prior to blotting onto nylon membranes. Cloned PCR products were used as probes after radioactive labeling with α-[^{32}P]-dCTP. Blot membranes (PALL Biodyne B, 0.45 μm) were used according to recommendations of the manufacturer and washed at 60 °C in 0.1× SSC prior to autoradiography.

ACKNOWLEDGMENTS

The work presented was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) to V.K. (Kn411-1 and Kn411-2), a Landesforschungsschwerpunkt (LFSP) in Baden-Württemberg and from the Universität Ulm. We gratefully acknowledge continuous support by Axel Brennicke and the skillful technical assistance of Dagmar Pruchner and Kathrin Lättig.

Received July 21, 1998; returned for revision August 12, 1998; revised manuscript received September 1, 1998

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Beckert S, Steinhauser S, Muhle H, Knoop V. A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial nad5 gene. *Plant Syst Evol.* In press.

Binder S, Marchfelder A, Brennicke A, Wissinger B. 1992. RNA editing in *trans*-splicing intron sequences of nad2 mRNAs in *Oenothera* mitochondria. *J Biol Chem 267*:7615–7623.

Carrillo C, Bonen L. 1997. RNA editing status of nad7 intron domains in wheat mitochondria. *Nucleic Acids Res 25*:403–409.

Chapdelaine Y, Bonen L. 1991. The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: A *trans*-splicing model for this gene-in-pieces. *Cell 65*:465–472.

Conklin PL, Wilson RK, Hanson MR. 1991. Multiple *trans*-splicing events are required to produce a mature nad1 transcript in a plant mitochondrion. *Genes & Dev 15*:1407–1415.

Devereux R, Haeberli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res 12*:387–395.

Doetsch NA, Thompson MD, Hallick RB. 1998. A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: Are group III introns the chicken or the egg? *Mol Biol Evol 15*:76–86.

Doyle JJ, Doyle, JL. 1990. Isolation of plant DNA from fresh tissue. *Focus 12*:13–15.

Fukuzawa H, Kohchi T, Shirai H, Ohyama K, Umesono K, Inokuchi H, Ozeki H. 1986. Coding sequences for chloroplast ribosomal protein S-12 from the liverwort Marchantia polymorpha are separated far apart on the different DNA strands. *FEBS Lett 198*: 11–15.

Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD. 1991. A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. *Cell 65*:135–143.

Grivell LA. 1996. Transposition: Mobile introns get into line. *Curr Biol 6*:48–51.

Handa H, Mizobuchi-Fukuoka R, Pinyarat W. 1992. The rapeseed mitochondrial gene for subunit 2 of the NADH dehydrogenase complex: A *trans*-spliced structure is conserved in one of the smallest plant mitochondrial genomes. *Curr Genet 31*:336–342.

Hetzer M, Wurzer G, Schweyen RJ, Mueller MW. 1997. *Trans*-activation of group II intron splicing by nuclear U5 snRNA. *Nature 386*:417–420.

Jacquier A. 1990. Self-splicing group II and nuclear pre-mRNA introns: How similar are they? *Trends Biochem Sci 15*:351–354.

Knoop V, Altwasser M, Brennicke A. 1997. A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet 255*:269–276.

Knoop V, Schuster W, Wissinger B, Brennicke A. 1991. *Trans* splicing integrates an exon of 22 nucleotides into the nad5 mRNA in higher plant mitochondria. *EMBO J 10*:3483–3493.

Kück U, Choquet Y, Schneider M, Dron M, Bennoun P. 1987. Structural and transcription analysis of two homologous genes for the P700 chlorophyll alpha apoproteins in *Chlamydomonas reinhardtii*. Evidence for in vivo *trans*-splicing. *EMBO J 6*:2185–2196.

Malek O, Brennicke A, Knoop V. 1997. Evolution of *trans*-splicing plant mitochondrial introns in pre-Permian times. *Proc Natl Acad Sci USA 94*:553–558.

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, Ferat JL. 1995. Structure and activities of group II introns. *Annu Rev Biochem 64*:435–461.

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

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. A primitive form of plant mitochondrial genome. *J Mol Biol 223*:1–7.

Pereira de Souza A, Jubier MF, Delcher E, Lancelin D, Lejeune B. 1991. A *trans*-splicing model for the expression of the tripartite nad5 gene in wheat and maize mitochondria. *Plant Cell 3*:1363–1378.

Sharp PA. 1991. "Five easy pieces." *Science 254*:663.

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. Plant mitochondrial RNA editing—extreme in hornworts and dividing the liverworts? *J Mol Evol.* In press.

Thompson MD, Zhang L, Hong L, Hallick RB. 1997. Extensive structural conservation exists among several homologs of two Euglena chloroplast group II introns. *Mol Gen Genet 257*:45–54.

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.

Weiner AM. 1993. mRNA splicing and autocatalytic introns: Distant cousins or the products of chemical determinism? *Cell 72*:161–164.

Wissinger B, Schuster W, Brennicke A. 1991. *Trans* splicing in *Oenothera* mitochondria: nad1 mRNAs are edited in exon and *trans*-splicing group II intron sequences. *Cell 65*:473–482.

Yeo CC, Tham JM, Yap MW, Poh CL. 1997. Group II intron from *Pseudomonas alcaligenes* NCIB 9867 (P25X): Entrapment in n plasmid RP4 and sequence analysis. *Microbiology 143*:2833–2840.

Zaita N, Torazawa K, Shinozaki K, Sugiura M. 1987. *Trans* splicing in vivo joining of transcripts from the divided gene for ribosomal protein S12 in the chloroplasts of tobacco. *FEBS Lett 210*:153–156.