The Plant Cell, Vol. 3, 1363–1378, December 1991 © 1991 American Society of Plant Physiologists

A *trans*-Splicing Model for the Expression of the Tripartite *nad5* Gene in Wheat and Maize Mitochondria

Anete Pereira de Souza,[a] Marie-France Jubier,[a] Evelyne Delcher,[a] Dominique Lancelin,[b] and Bernard Lejeune[a,1]

[a] Laboratoire de Biologie Moléculaire Végétale associé au Centre National de la Recherche Scientifique, UA 1128, Bâtiment 430, Université Paris XI, 91405 Orsay Cedex, France

[b] Biologie Cellulaire, Centre National de la Recherche Agronomique, route de Saint-Cyr, 78026 Versailles Cedex, France

The mitochondrial single-copy gene *nad5* of wheat and maize consists of 5 exons located on three widely separated regions of the genome that are independently transcribed. The first region contains exons I and II separated by an atypical group II intron; in the second region is exon III (only 22 bp long), which is flanked upstream by a maturase-related open reading frame (ORF) and exon e of the *nad1* gene, and downstream by a previously unidentified ORF (ORF143); in the third region are exons IV and V separated by a group II intron. In maize, this last domain is flanked upstream by the genes *rps12*, *nad3*, and tRNA^{Ser} and downstream by a chloroplast tRNA^{Cys}. RNA editing occurs in wheat exons IV and V as C-to-U changes. A detailed analysis of the transcription of the *nad5* gene in wheat and maize reveals that the exons are assembled into a 2.4-kb mRNA after two *cis*-splicing (between exons I and II and exons IV and V) and two *trans*-splicing events. The *trans*-splicing process involves the sequences flanking exons II, III, and IV that feature group II introns. A model is proposed for the assembly and maturation of the *nad5* transcripts.

INTRODUCTION

Higher plant mitochondrial genomes are characterized by their great length, ranging from about 200 to more than 2000 kb (for a review, see Newton, 1988), and also by a rather complicated molecular organization generated by sequence rearrangements through recombination processes (Gray, 1989; Palmer, 1990).

Several genes have been identified in higher plant mitochondrial genomes. rRNA genes for 5S, 18S, and 26S RNAs and tRNA genes have been directly located on the mitochondrial genome by hybridization with their own products. Other genes have been detected either by heterologous hybridization with mitochondrial sequences originating from other organisms such as fungi, yeast, or mammals or by a more direct approach using cDNAs (Hiesel et al., 1987; Schuster and Brennicke, 1987). These genes include several subunits of the inner membrane complexes I (NADH-ubiquinone oxidoreductase), III (ubiquinol-cytochrome-*c* oxidoreductase), IV (cytochrome-*c* oxidase), and ATP synthase and also other proteins such as ribosomal proteins Rps12, Rps13, and Rps14 (for a review, see Newton, 1988; Levings and Brown, 1989). The detection of the gene encoding the α-subunit of ATP synthase (Hack and Leaver, 1983), present in the nuclear genome in other organisms, demonstrates that the mitochondrial genome of higher plants can include more genes than its fungal or mammalian equivalents.

Among the genes cited above, cytochrome oxidase subunit II (*cox2*; Fox and Leaver, 1981) and NADH-ubiquinone reductase subunits 1 (*nad1*; Chapdelaine and Bonen, 1991; Wissinger et al., 1991), 4 (*nad4*; Lamattina et al., 1989), and 5 (*nad5*; Wissinger et al., 1988) contain intervening sequences. These intervening sequences belong to group II introns (Michel et al., 1983, 1989). When the exons and intron(s) constitute a unique transcription unit (*cox2*, for instance), introns are eliminated by a *cis*-splicing process. But when the exons are dispersed throughout the mitochondrial genome and correspond to separate transcription units, the formation of the mature messenger results from a *trans*-splicing process between the different transcripts. *trans*-Splicing was initially described in the chloroplast genome of Chlamydomonas for a gene encoding a polypeptide of the chlorophyll protein complex of photosystem I (*psaA*) (Kück et al., 1987) and in the chloroplast genome of higher plants for the gene encoding the ribosomal protein S12 (*rps12*) (Fukuzawa et al., 1986; Zaita et al., 1987; Hildebrand et al., 1988; Kohchi

[1] To whom correspondence should be addressed.

et al., 1988). *trans*-Splicing has recently been reported to occur also for the *nad1* gene in the mitochondrial genome of higher plants (Chapdelaine and Bonen, 1991; Wissinger et al., 1991).

The size of the mitochondrial genome in angiosperms suggests that more genes are still to be identified. Because comparison of coding sequences revealed higher levels of nucleotide conservation than other sequences in the mitochondrial genome (see Moon et al., 1985; Grabau, 1986; Bland et al., 1986, 1987; Falconet et al., 1988; Ecke et al., 1990), one can also expect that conserved sequences present in mitochondria of distantly related angiosperms represent genes or part of genes.

We report here the results of such a study involving a sequence present and transcribed in the mitochondrial genome of common wheat and selected initially for its conservation and transcription in several dicot species as well. We identified this sequence as part of the *nad5* gene. Comparison with published data showed that this *nad5* gene sequence has not yet been identified in plants. The complete structure of the gene was established in wheat and in maize; it consists of 5 exons dispersed in three different regions of the mitochondrial genome. The production of the mature RNA requires two *cis*- and two *trans*-splicing steps, for which a model is proposed.

RESULTS

Detection of a Sequence Homologous to the 3′ End of the *nad5* Gene

The high degree of sequence conservation observed for all the mitochondrial genes thus far identified in higher plants (see Moon et al., 1985; Grabau, 1986; Bland et al., 1986, 1987; Falconet et al., 1988; Ecke et al., 1990) suggests that conserved sequences in these genomes might correspond to actual genes. To isolate some of these sequences, restriction fragments ranging from 1.5 to 4 kb were prepared from cloned mitochondrial DNA (mtDNA) of wheat and used as probes in a series of heterologous hybridization experiments with mtDNAs from several dicots.

Fragments of wheat mtDNA giving positive signals and not related to previously identified genes were selected. Their possible expression was checked by RNA gel blot hybridization experiments using mtRNA from wheat. This analysis allowed us to select sequences that were both present in all mtDNAs tested and expressed in wheat. One of them, located on the SalI fragment G1 of the wheat mitochondrial genome (Quétier et al., 1985), was selected for further studies.

Preliminary mapping results indicated that sequence conservation extended over 2 kb when compared with the maize mtDNA. The conserved regions in wheat and maize were sequenced. Their detailed transcription pattern suggested the presence of 2 exons separated by 1 intron. Sequencing and transcription analysis allowed us to identify two potential open reading frames (ORFs) that were 396 and 147 bp long, respectively (the latter terminated by a TAA codon), and separated by an intron of 933 bp. Examination of the sequence of the putative intron revealed a typical group II intron structure. Although none of these exons had any homology with the published sequence of *nad5* of Oenothera (Wissinger et al., 1988) and sugar beet (Ecke et al., 1990), when compared with protein sequences in the National Biomedical Research Foundation Protein Identification Resource (NBRF-PIR) and the European Molecular Biology Organization Swiss-Prot data bases, both ORFs displayed significant identity with the C-terminal parts of *nad5* genes of both filamentous fungi Neurospora (Nelson and Macino, 1987) and Aspergillus (Brown et al., 1989). Hence, this novel sequence probably corresponds to the 3′ end of a putative *nad5* gene of wheat.

Localization of the 5 Exons of the *nad5* Gene in Wheat and Maize Mitochondrial Genomes

To characterize and to study the complete *nad5* gene of wheat, the corresponding cDNA sequence was established by a combination of several independent steps of

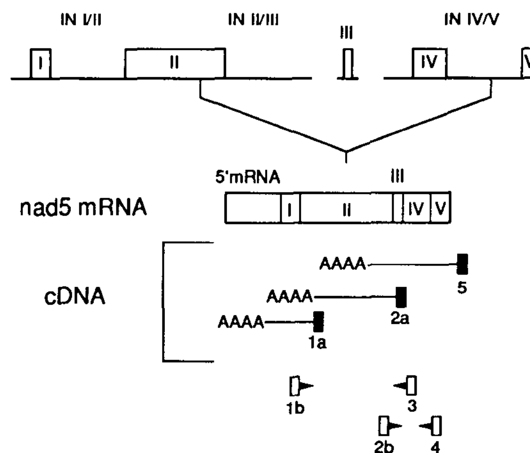

Figure 1. cDNA Cloning Strategy.

Different cDNA clones were required to obtain the complete cDNA sequence that allowed the location of exons I, II, and III. Three different oligonucleotides were used for primary cDNA synthesis and are indicated by black boxes. The other oligonucleotides used in this work are represented by open boxes. Details are given in Methods. Open boxes correspond to the exons designated by roman numerals; IN, intron.

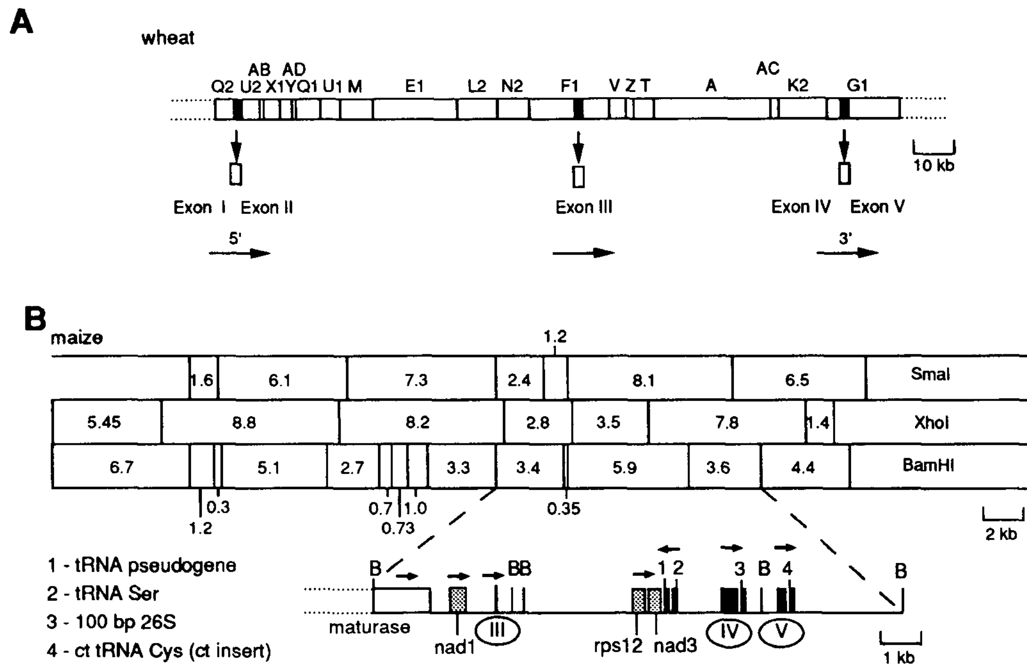

Figure 2. Localization of the *nad5* Gene in Mitochondrial Genomes of Maize and Wheat.

(A) Localization of the 5 exons of the *nad5* gene on the SalI restriction map of the mitochondrial genome of wheat (Quétier et al., 1985). Only the relevant part of the map is shown. Arrows indicate exon orientations.
(B) Upper portion: restriction map of cosmid N5G6 from a maize mitochondrial cosmid library (Lonsdale et al., 1984). Lower portion: localization and orientation of *nad5* exons III, IV, and V (black boxes and circled roman numerals). The genes or sequences are as follows: 1, tRNA pseudogene; 2, $tRNA^{Ser}$; 3, 100 bp homologous to the 26S rRNA gene (26S); 4, chloroplast (ct) $tRNA^{Cys}$. Open box, maturase; dotted boxes, *nad1* exon e, *rps12*, and *nad3* genes. Arrows indicate gene orientations; B, BamHI sites.

reverse transcription and polymerase chain reaction (PCR) amplification as shown in Figure 1.

By hybridizing the different cDNA clones obtained with wheat mtDNA, the position of all parts of the gene could be determined. The results can be summarized as follows: the gene consists of 5 exons, designated I to V, located on three different parts of the genome and separated from each other by about 50 kb. Figure 2A shows the location of these 5 exons on the wheat mtDNA map of Quétier et al. (1985): exons I and II are on SalI fragment Q2, the very short (22 bp) exon III is on fragment F1, and exons IV and V are on fragment G1. All these exons are single-copy sequences. The restriction map of the three regions involved is presented in Figure 3. Their assembly from these three separate locations is shown in Figure 4. Hybridization of maize mtDNA with the same cDNA or genomic probes revealed the same general organization in maize, as shown in Figure 2B, although only 6 kb separated the locations of exons III and IV according to the physical map of Lonsdale et al. (1984).

The three separate regions containing exons I and II, exon III, and exons IV and V, respectively, have been sequenced in wheat. The results are presented in Figure 5. Exons I (229 bp) and II (1217 bp) are separated by a group II intron (862 bp). Heterologous hybridization experiments or PCR amplification using suitable primers (not shown here) revealed an identical organization of this region for maize mtDNA. The sequence displays 97% identity with the corresponding sequences published for Oenothera (Wissinger et al., 1988) and sugar beet (Ecke et al., 1990). Comparison of the amino acid sequence with the products of the *nad5* genes of Neurospora (Nelson and Macino, 1987) and Aspergillus (Brown et al., 1989) is presented in Figure 6 and reveals 50% identity between these sequences.

The position of exon III, which is 22 bp long, is noteworthy. It lies on SalI fragment F1 and is located 372 bp downstream of the stop codon of exon e of the *nad1* gene; this exon is preceded by a "maturase-like" ORF (Wahleithner et al., 1990; Chapdelaine and Bonen, 1991). The 3′ end of exon III is 67 bp upstream of the ATG codon of a putative ORF encoding a 143-amino acid peptide (ORF143). This ORF could not be identified by comparison with protein sequences from data banks (NBRF and

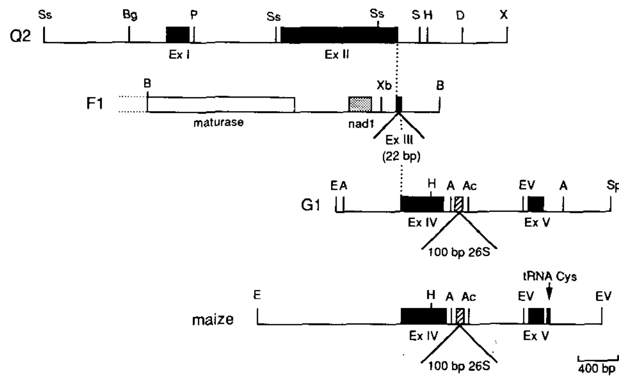

Figure 3. Restriction Map of the Three Regions of Wheat Mitochondrial Genome Containing the *nad5* Five Exons.

The 5 *nad5* exons (Ex) are shown as dark boxes. Open box, maturase; dotted box, *nad1* exon e; cross-hatched boxes, 100 bp homologous to 26S rRNA gene (26S rDNA); Q2, F1, and G1, Sall fragments of wheat mitochondrial genome. Ac, Accl; A, Aval; Bg, Bglll, E, EcoRl; EV, EcoRV; D, Dral; H, Hindlll; P, Pstl; S, Sall; Ss, Sstl; Xb, Xbal; X, Xhol. Comparison with the restriction map of the region including exons (Ex) IV and V in maize (lower line) shows restriction site conservation in the exons and in the intron.

Swiss-Prot). Probes specific to exon III, the maturase, exon e of the *nad1* gene, and ORF143 allowed us to show an identical organization of these sequences in maize mtDNA, with the corresponding region being mapped on maize cosmid N5G6 (Figure 2B). Exon III is not present in the published sequences of the *nad5* genes of Oenothera (Wissinger et al., 1988) and sugar beet (Ecke et al., 1990), but the 7 amino acids encoded by this exon display 70% identity with the two corresponding *nad5* fungal sequences, and 100% if conservative substitutions are allowed.

Exons IV (396 bp) and V (147 bp) are located on Sall fragment G1 and separated by a 933-bp intervening sequence. The same sequences (97% identity) with the same organization were found in maize, as shown in Figure 7, but in a different genomic environment: exon V in maize is immediately followed by a plastid $tRNA^{Cys}$ (Wintz et al., 1988), whereas 1.3 kb upstream of exon IV is a region encoding the *nad3*, *rps12*, and $tRNA^{Ser}$ genes (Gualberto et al., 1988). In wheat mtDNA, this same region is located on fragment L2, which is 90 kb away from exon IV (Figures 2A and 2B). Both exons were not present in the previously published *nad5* sequences of Oenothera (Wissinger et al., 1988) and sugar beet (Ecke et al., 1990). This raises the question of whether exons III, IV, and V are present in these plants as well. Noteworthy is the presence in the intron separating exons IV and V of two short sequences displaying homology with the wheat 26S rRNA gene (Falconet et al., 1988) (Figure 7). In the mtDNA of wheat, the 5 exons lie in the same order as they appear in the final transcript. This holds true also for the exons III, IV, and V in maize; in this species, the orientation of exons I and II has not been determined.

The presence of repeated sequences in the mtDNA of wheat (Quetier et al., 1985) and maize (Lonsdale et al., 1984) has been reported and is thought to be involved in the structural rearrangements observed in these genomes. In wheat, according to the physical map of the mitochondrial genome, two of these sequences are located within the 100 kb that separate exon I from exon V: one is located on fragment T (Figure 2A) and contains the 18S and 5S rRNA genes (Falconet et al., 1984; Lejeune et al., 1987) and the other, located on fragment M (Figure 2A), contains the 26S rRNA gene (Falconet et al., 1988).

The structural rearrangements that can be predicted from recombinations between the copies of these repeated sequences have been confirmed by the analysis of the cosmid library of mtDNA of wheat (Falconet et al., 1985; Quétier et al., 1985; Lejeune et al., 1987). None of these rearrangements leads to the formation of a shorter sequence allowing the synthesis of a unique precursor of the *nad5* gene. To check that such a rearrangement was not present in the mtDNA of wheat, hybridization experiments were performed (results not shown). The mtDNA was digested by several restriction enzymes, membrane blotted, and hybridized with probes taken from the exons and introns of the *nad5* gene. Only the hybridization signals expected from the locations of the probes on the different arrangements of the physical map mentioned above could be detected.

RNA Editing

RNA editing in mitochondrial transcripts of higher plants is well documented (for a review, see Mulligan, 1991; Schuster et al., 1991). Comparison of cDNA and the corresponding genomic sequences of the *nad5* gene in wheat revealed 11 C-to-U editing events (Figure 5) occurring only in exons IV (five changes) and V (six changes). Partially edited transcripts could not be found by comparison of several cDNA clones obtained independently. Ten amino acid changes resulted from these RNA editing events that rendered the putative protein more hydrophobic. Comparison with both fungal *nad5*-encoded amino acid sequences (Figure 6) showed that four of these changes reestablished amino acid identity between sequences, two abolished a preexisting amino acid identity, and the last four led to an unrelated amino acid.

Analysis of the Structure of the Intervening Sequences

According to the structural model proposed by Michel and Dujon (1983), the introns separating exons I and II or

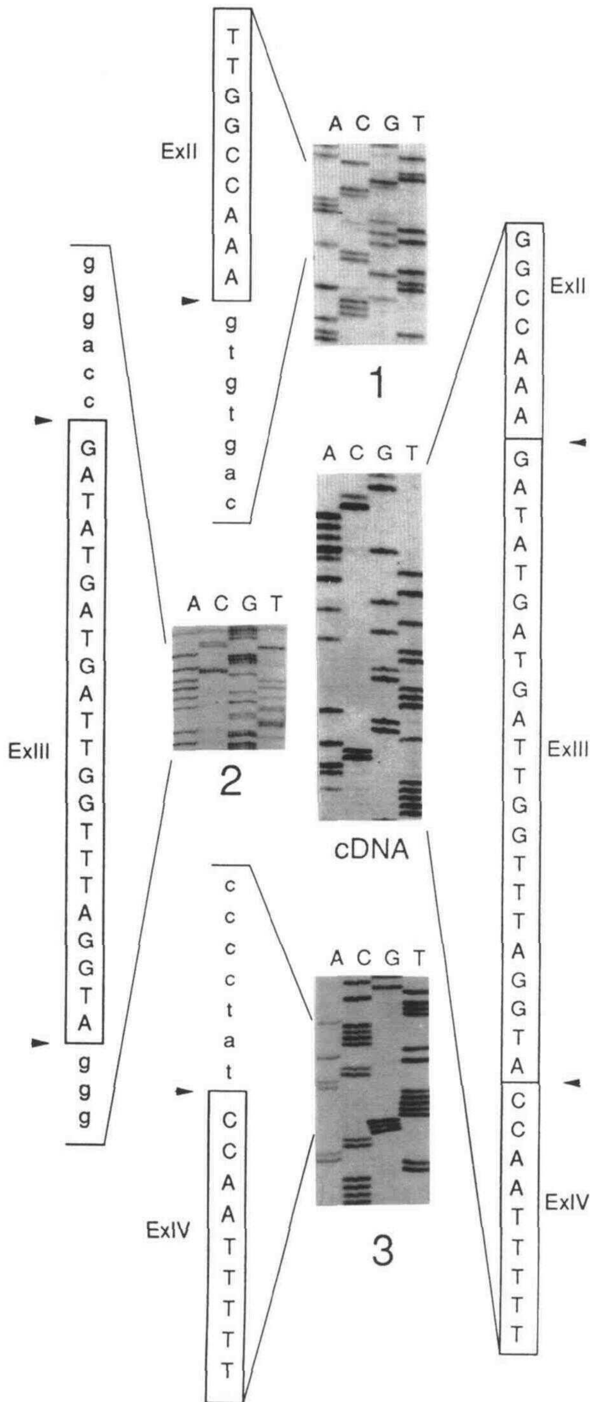

Figure 4. Comparison of Genomic and cDNA Sequences Showing the Exact Assembly of Exons II, III, and IV of the *nad5* Gene.

Section 1 shows the 3′ end of exon II and 5′ of the first interrupted intron; section 2, exon III (22 bp) and its flanking sequences; section 3, the 3′ end of the second interrupted intron and 5′ end of exon IV. The limits of exons in each case are indicated by arrows. cDNA, partial sequence of the cloned PCR product obtained with primers 4 and 2b. The arrows show the junction of exons (Ex) after the two *trans*-splicing events. Exon (Ex) regions are shown in uppercase letters and boxed.

exons IV and V belong to group II. Figures 8A, 8D, and 8F show the conserved structures of domains V of these introns. The intron separating exons IV and V displays all structural features predicted by the model of group II introns, but the first one (between exons I and II) does not follow the same rules. For instance, it is difficult to find a 5′ consensus sequence for splicing; similarly, the classical 3′ helical structure of domain VI cannot be detected, even if one takes into account the base redundancy at splicing sites of the intron. These differences seem to play a role in the splicing process of the introns separating exons I and II and exons IV and V according to the processing intermediate detected in both cases (see transcription experiments below and Discussion).

The physical separation of the three regions containing the different parts of the *nad5* gene implies that mature messenger processing requires two *trans*-splicing steps between three different primary transcripts. Such a processing has been shown to occur for the *nad1* gene in wheat (Chapdelaine and Bonen, 1991) and Oenothera (Wissinger et al., 1991).

For the *nad5* gene, the sequences involved in this *trans*-splicing process might be located 3′ of exon II and 5′ of exon III and, similarly, 3′ of exon III and 5′ of exon IV. The examination of these regions presented in Figures 8B, 8C, and 8E show helical domain V structures that are characteristic of class II introns: these helical structures have been detected 34 and 48 bp, respectively, upstream of exons III and IV (Figures 8B, 8C, and 8E). Exons II and III are prolonged on their 3′ end by conserved 5′ intron splicing sites (GTGTG and GGGCG, respectively). Our data suggest that these structures might represent the remains of two former group II introns that once linked exons II, III, and IV into a single functional unit. This unit was disrupted by genome rearrangement, resulting in the dispersion of the gene sections at three separate locations.

If this hypothesis is correct, the question arises regarding in which intron domain the discontinuity of both introns took place. Careful sequence examination in both cases revealed (Figure 5) that domain IV might be the region involved for the initial transcript junction leading to the *trans*-splicing. Two putative complementary short sequences, belonging to different regions, generate the helical stem of intron domains IV (Figures 8B and 8C). These short sequences seem to be the only remaining elements of domain IV. Jarrel et al. (1988) have shown that in vitro the presence of domain V is crucial for *trans*-splicing, whereas domain IV can be removed without abolishing

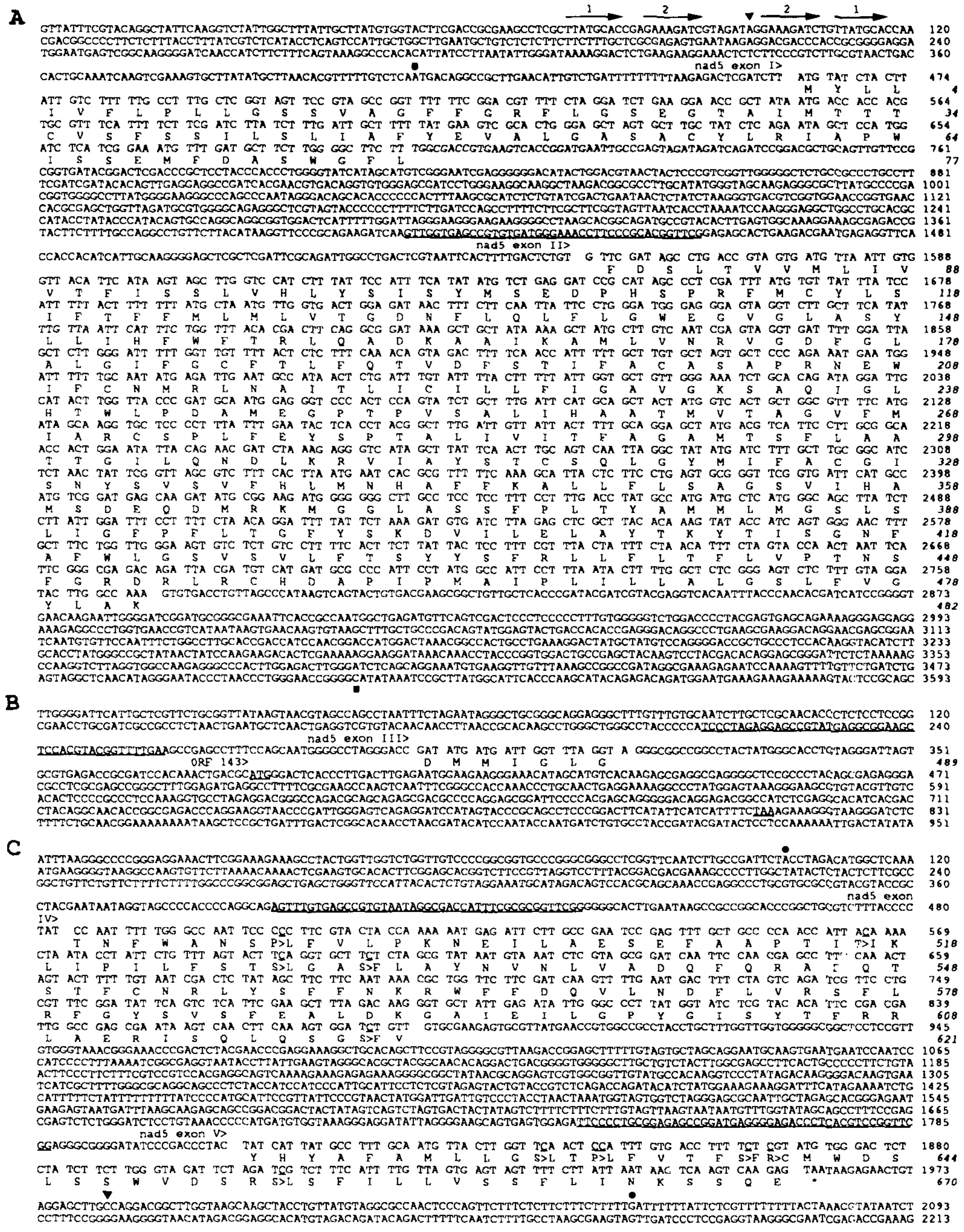

Figure 5. Sequences of the Genomic DNA and of the Corresponding cDNA of the *nad5* Gene and Its Flanking Regions in Wheat Mitochondrial Genome.

(A) Exons I and II separated by an intron and the flanking sequences.
(B) Exon III and the flanking sequences.
(C) Exons IV and V separated by an intron and the flanking sequences.
The amino acid sequence corresponding to each exon is given. The star indicates the stop codon. C-to-U changes after RNA editing are *underlined*; the predicted amino acids before and after editing are indicated. Positions of mRNA 5′ and 3′ termini are indicated by solid triangles. Short, direct repeats 1 and 2 on each side of the 5′ terminus are overlined with arrows. Homology break points are indicated by solid squares for *nad5* sequences of Oenothera (Wissinger et al., 1988) and sugar beet (Ecke et al., 1990) and by solid circles for the maize *nad5* region (exons IV and V) sequenced in this work. The domains V of the group II introns in the 5′ flanking regions of exons II, III, IV, and V are underlined. ORF143 is indicated downstream of exon III. Its start and stop codons have been underlined.

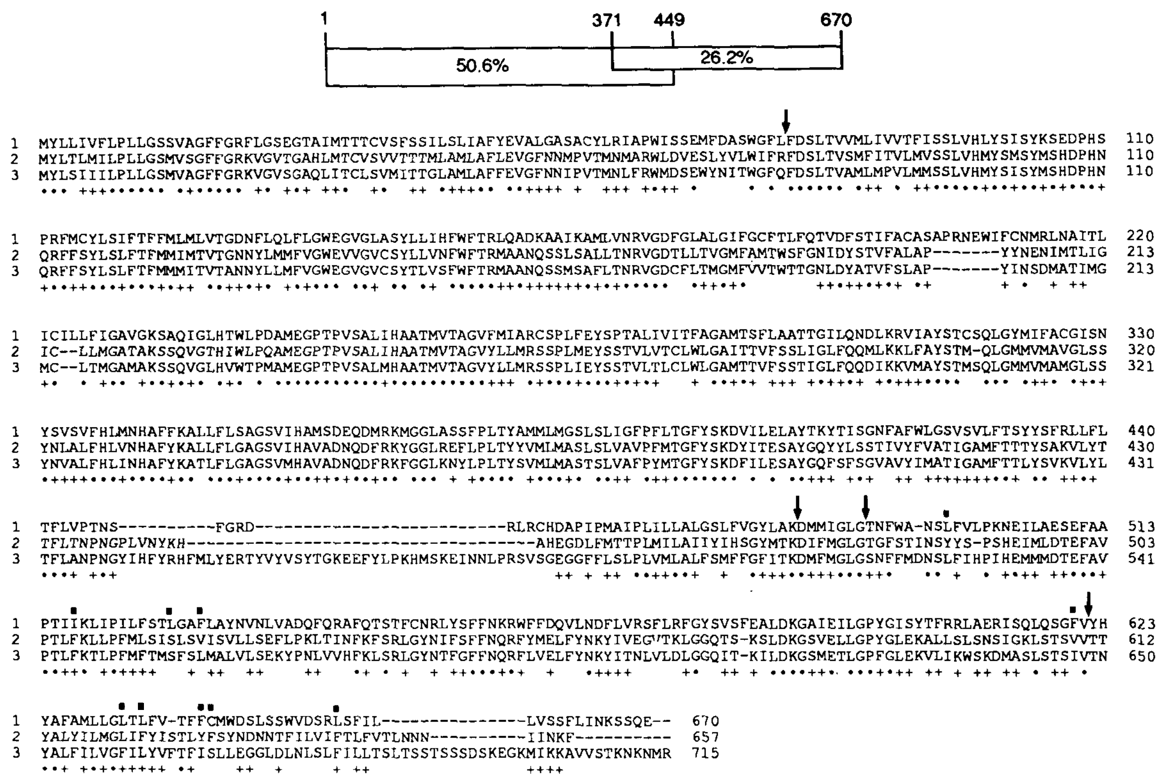

Figure 6. Amino Acid Homology among *nad5*-Encoded Polypeptides.

Nad5 protein sequences from (1), wheat (this work); (2), Aspergillus (Brown et al., 1989); (3), Neurospora (Nelson and Macino, 1987) have been aligned with the help of the Clustal program. In the lower line, the conserved amino acids are indicated by dots and the conservative amino acid substitutions by crosses. The limits of exons in wheat are signaled by vertical arrows. Black squares indicate the amino acid residue changes due to RNA editing.

splicing. Whereas the 3′ end of exon II still includes 717 bp, the 3′ end of exon III is only 58 bp away from the short domain IV sequence; the absence of most of the intron sequence indicates that the discontinuity resulted probably from more than one interruption of the intron sequence.

Size of the Transcript

S1 mapping and primer extension experiments performed in wheat (results not shown) demonstrate that the transcript starts 365 bp upstream the ATG codon (indicated in Figure 5). Despite the existence of a preferential initiation site, we observed a variation similar to previous reports indicating that multiple transcript termini, differing by a single nucleotide, can exist (see for instance 5′ termini of *atp1* in Mulligan et al., 1991). Interestingly, all 5′ termini are situated between the copies of direct repeated sequences (Figure 5). Moreover, the motif (G[A/T]$_{3,4}$) proposed to be involved in transcription initiation in maize (Mulligan et al., 1991) is present here in two copies upstream and two copies downstream of the 5′ terminus. The 3′ end of the transcript was determined by S1 mapping and shown to occur 22 bp downstream of the stop codon. According to the sequencing data, these results then predict a mature mRNA of 2.4 kb.

Transcription of the *nad5* Gene

To elucidate the complex maturation process of transcript joining and processing, a detailed study of the transcription of the *nad5* gene was done in wheat and maize and is presented in Figure 9. RNA gel blot hybridization experiments were done using short probes from all exons and introns (Figure 9B). Because no essential differences in the *nad5* gene organization or in the sequences of exons and introns have been detected in wheat and maize, only probes from wheat were used throughout this study.

Hybridization results in Figure 9A show patterns as complicated as those obtained by Hildebrand et al. (1988)

```
MAIZE : CATCTGCAACCTTCTGTCGAAAGAGTCCGCTTACGGGCGGGTGAGGATCAACACTCTGCCACCTAGACATGGCTCAAAACGAAGGGGTAAGGCCAAGTGTTCTTAAAACAAA  112
WHEAT : GGTTGGTCTGGTTGTCCCCGGCGGTGCCCGGGCGGGCCTCGGTTCAATCTTGCCGATTCT..................T...............................

ACTCTCAGTGCAC--TTCGGAGCACGGTCTTCCGTTAGGTCCTTTACGGACGACGAAAGCCAGCCCCTTGGCTATACTCTACTCTTCGCCTCCGG--CTGTTCTTTCCTTTTGGCTCGGC  228
....AG.......AC.........................................--..--.....................GG.T.TT.........T........C....

GGAGCTGAGCTGGGTTCCATTACACTCTGTAGGAAATGCATAGACAGTCCACGCAGCAAACCGAGGCCCTGCGTGCGCCGTACGTACCGCCTACGAATAATAGGTAGCCCCACCCCAGGC  348
.......................................................................................................................
                                                                                              nad5 exon IV>
AGAGTTTGTGAGCCGTGTAATAGGCGACCATTTCGCGCGGTTCGGGGGGCACTTGAATAAGCCGCCGGCACCCGGCTGCGTCTTTACCCCTAT CC AAT TTT TGG GCC AAT TCC  461
.............................................................................................. .. ... ... ... ... ... ...

CCC TTC GTA CTA CCA AAA AAT GAG ATT CTT GCC GAA TCC GAG TTT GCT GCT CCA ACC ATT ACA AAA CTA ATA CCT ATT CTG TTT AGT ACT  551
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...

TCA GGT GCT TCT CTA GCG TAT AAT GTA AAT CTC GTA GCG GAT CAA TTC CAA CGA GCC TTT CAA ACT AGT ACT TTT TGT AAT CGA CTC TAT  641
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

AGC TTC TTC AAT AAA CGC TGG TTC TTC GAT CAA GTT TTG AAT GAC TTT CTA GTC AGA TCG TTC CTG CGT TTC GGA TAT TCA GTC TCA TTC  731
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

GAA GCT TTA GAC AAG GGT GCT ATT GAG ATA TTG GGC CCT TAT GGT ATC TCG TAC ACA TTC CGA CGA TTG GCC GAG CGA ATA AGT CAA CTT  821
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

CAA AGT GGA TCT GTT  GTGCGAAGAGTGCGTTATGAACCGTGGCCGCCTACCTGCTTTGGTTGGTGGGGGCGGCTCCTCCGTTGTGGGTAAACGGGAAACCCGACTCTACGAACC  935
... ... ... ... ...  ...................................................................................................
 55bp 26S rDNA>                                                                                       48bp 26S
CGAGGAAAGGCTGCACAGCTTCCGTAGGGTCGTTAAGACCGGAGCTTTTTGTAGTGCTAGCAGGAATGCAAGTGAATGAATCCAATCCC--C---TTTCAAATCGGCGAGGTAATACCTT 1050
...........................G.....................................................AT.CCC...A....................
 rDNA>
ATTGAAGTAGGGCACGCTACGGCAACACAGGACTGACGGGGTGGGGGCTTGCTGTCTACTTGGCGAGCCTTCACTGCCCCCTTCTGTAACTTCCCTTCTTTCGTCCGTCCACGAGGCAGT 1170
.......................................................................................................................

CAAAAGAAAGAGAGAAAGGGGCGGCTATAACGCAGGAGTCGTGGCGGTTGTATGCCACAAGGTCCCTATAGACAAGGGGACAAGTGAATCATCGCTTTTGGGCGCAGGCAGCCCTCTACC 1290
.......................................................................................................................

ATCCATCCCATTGCATTCCTCTCATAGAGTACTGTACCGTCTCAGACCAGATACATCTATGGAAAGAAAGGATTTCATAGAAAATATTCATTTTTCTCTTTTTTTGGATCCCCATGCATT 1410
.......................G........................................................C.G........A.......TT............

CCGTTATTCCCGTAACTATGGATTGGATTGGATTGGATTGATTGATTGTCCCTACCTAACTAAATGGTAGTGGTCTAGGGAGCGCAATTGCTAGAGCACGGGAGAATGAAGAGTAATGAT 1530
........................-....--.---------------.......................................................................

TTAAGCAAGAGCAGCCGGACGGACTACTATAGTCAGTCTAGTGACTACTATAGTCTTTTCTTTCTTTGTAGTTAAGTAATAATGTTTGGTATAGCAGCCTTTCCGAGCGAGTCTCTGGGA 1650
.......................................................................................................................

TCTCCTGTAAACCCCCATGATGTGGTAAAGGGAGGGAGGATATTAGGGGAAGCAGTGAGTGGAGATTCCCCTGCGGAGAGCCGGATGAGGGGAGACCCTCACGTCCGGTTCGGAGGGCGG 1770
..................................-.---.................................................................................
    nad5 exon V>
GGATATCCCGACCCTAC    TAT CAT TAT GCC TTT GCA ATG TTA CTT GGT TCA ACT CCA TTT GTG ACC TTT TCT CGT ATG TGG GAC TCT CTA TCT 1862
.................    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

TCT TGG GTA GAT TCT AGA TCG TCT TTC ATT TTG TTA GTG AGT AGT TTT ATT ATT AAT AAG TCA AGT CAA GAA TAATAAGAGAACTGTAGGAGCTTG 1958
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... C.. ... ... ... ... ... ... ..G ........................
                                                                        ct tRNA Cys>
CCAGGACGGCTTGGTAAGCAAGCTACCTGTTATGTAGGCGCCAACTCCCAGTTCTTTCTCTTCTTTCTTTGGCGGCATGGCCAAGCGGTAAGGCAGGGGACTGCAAATCCTTTATCCCC 2078
.......................................................................ATTTTTTTATTCTCGTTTTTTTTACTAAACGTATAATCTCCTTTCCGG

AGTTCAAATCTGGGTGTCGCCTGACCCACAAAATACTTTTTATTTCTTATTGTACTGATT 2138
GGAAGGGGTAACATAGACGGAGGCACATGTAGACAGATACAGACTTTTTCAATCTTTTGC
```

Figure 7. Sequence Comparison of the Regions of Maize and Wheat mtDNAs Containing *nad5* Exons IV and V.

The upper line shows the maize sequence. Exons IV (nucleotides 442 to 836) and V (nucleotides 1788 to 1934) are separated by a group II intron. The lower line shows the corresponding sequence in wheat. Homologous nucleotides are designed by dots and deletions by dashes. In the intron sequence, two short sequences homologous to the 26S rRNA gene (26S rDNA), 55 bp and 48 bp, respectively, have been underlined. The chloroplast (ct) tRNA^{Cys} sequence downstream of exon V has been underlined.

for the maturation of transcripts of *rps12* gene in chloroplasts of tobacco. More transcription intermediates can be detected in maize than in wheat, suggesting faster mRNA processing in wheat. In both cases, probes taken from the 5 exons showed the predicted 2.4-kb mature transcript (Figure 9A, lanes 1, 3, 5, 7, and 10). Because the regions containing the three parts of the *nad5* gene are probably independently expressed, we attempted to determine the manner in which the different *nad5* RNAs are transcribed, assembled, and processed. A model taking into account the hybridization results is shown in Figure 9C.

Three independent RNA precursors are predicted. One of them corresponds to the 3.3-kb transcript including exons I and II, plus the intervening sequences (Figure 9A, lanes 1 to 4). The other putative precursors corresponding to exon III or exons IV and V could not be observed. The absence of such precursors could be the result of rapid processing. Alternative explanations exist for the nondetection of the precursor molecule for exons IV and V. First, its length, and hence its electrophoretic migration, could be similar to that of other transcription products of the same gene. In wheat, for example, probes 8 and 9

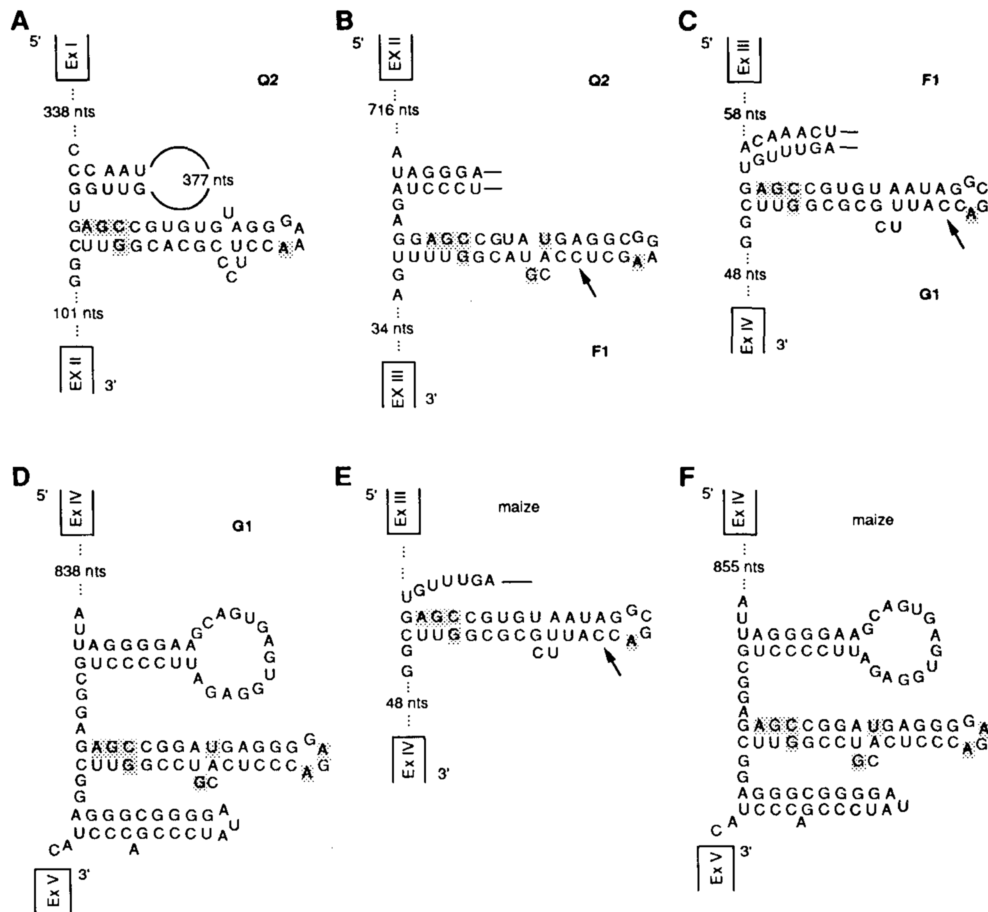

Figure 8. Intron Splicing Models for the *nad5* Gene.

(A) Domains IV and V of the atypical group II intron localized between exons I and II in wheat.
(B) Reconstitution of conserved domain V and discontinuous domain IV from the sequences flanking exons II and III, which are involved in the *trans*-splicing process in wheat.
(C) Reconstitution of conserved domain V and discontinuous domain IV from the sequences flanking exons III and IV, which are involved in the *trans*-splicing process in wheat.
(D) Domains IV, V, and VI of the group II intron localized between exons IV and V in wheat. This intron can be excised by *cis*-splicing and lariat formation.
(E) Reconstitution of conserved domain V and discontinuous domain IV from the sequence flanking exon IV, which is involved in the *trans*-splicing process in maize.
(F) Domains IV, V, and VI of the group II intron localized between exons IV and V in maize. This intron can be excised by *cis*-splicing and lariat formation.
Secondary structure models of the sequences flanking the exons (Ex) of the *nad5* gene in wheat **(A)** to **(D)** and in maize **(E)** and **(F)** have been drawn according to the model proposed by Michel and Dujon (1983) for the group II introns. Conserved nucleotides (nts) are shadowed. Arrows in **(B)**, **(C)**, and **(E)** point to the nucleotides that might be edited to stabilize the structure; Q2, F1, and G1, SalI fragments of wheat mitochondrial genome.

detected a 2.4-kb band (Figure 9A, lanes 8 and 9). This band could not possibly be the 2.4-kb mature transcript because probes 8 and 9 are entirely included in the second intron. This suggests that the precursor might be also a 2.4-kb molecule. Second, cotranscription with other genes might take place. This is possible in maize where exons IV and V are flanked upstream by the *rps12* and *nad3* genes, and downstream by a chloroplast $tRNA^{Cys}$ gene (Figure 2B). These genes could constitute one transcription unit. This unit would perhaps correspond to one of the

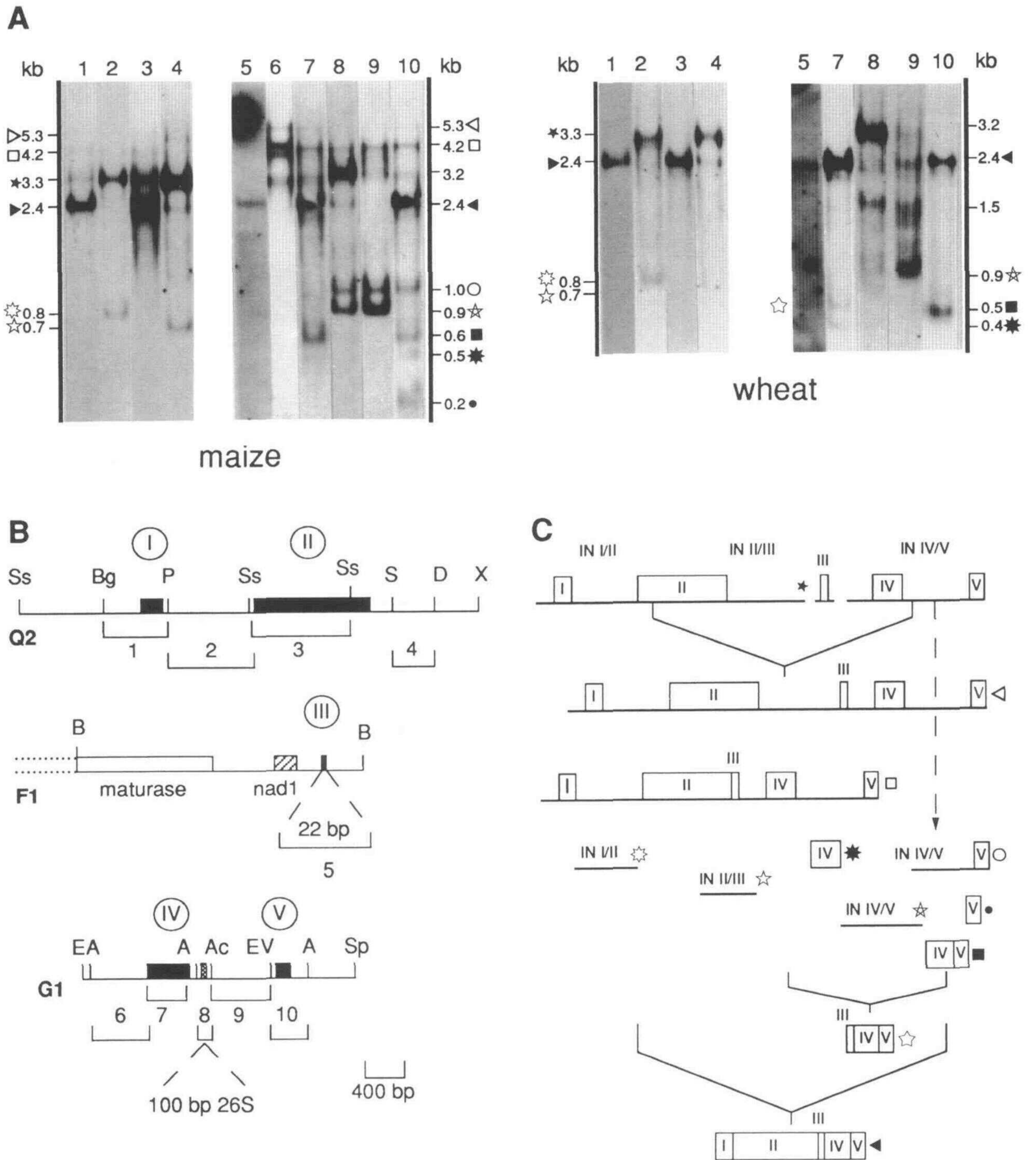

Figure 9. RNA Gel Blot Hybridizations of the *nad5* Transcripts in Wheat and Maize.

(A) Autoradiograms corresponding to the RNA gel blot hybridizations of total mtRNA of maize and wheat. Ten different probes, numbered from 1 to 10 according to their position on the *nad5* gene **(B)** have been used. Transcripts have been identified by their length and designed by symbols used in the model presented in **(C)**. As mentioned in the text, most processing intermediates can only be visualized in maize.

(B) Location of the 10 different probes (1 to 10) from the *nad5* gene of wheat used in the experiment. The 5 exons are presented as dark boxes and correspond respectively to probes 1 (I), 3 (II), 5 (III), 7 (IV), and 10 (V). Probes 2, 4, 6, 8, and 9 are strictly located on introns. Open box, maturase; cross-hatched box, *nad1* exon e; dotted box, 100 bp homologous to 26S rRNA gene (26S rDNA). Ac, AccI; A, AvaI;

four large transcripts detected in maize by probe 6 (Figure 9A, lane 6).

The assembly of the three precursors can generate a 5.3-kb RNA as detected by probes 1 to 4 and 6 to 10 (Figure 9A). A *trans*-splicing reaction between exons II and III eliminates the intervening sequence. The resulting 4.2-kb intermediate was detected by probes 1 to 3 and 6 to 10 (Figure 9A). Probe 4, corresponding to the intron, allowed the specific detection of excised molecules of 2.4, 1.2, and 0.7 kb, which might be *trans*-splicing intermediates (Figure 9A, lane 4). The intron located between exons I and II is removed by *cis*-splicing as a specific 0.8-kb band detected with probe 2 (together with the 5.3-, 4.2-, and 3.3-kb precursor molecules in Figure 9A, lane 2). There is no transcript revealing a *cis*-splicing reaction with lariat formation because no band of suitable size common to probes 1 and 3 or to probes 2 and 3 could be seen (Figure 9A, lanes 1 to 3). The weak signals obtained with probe 5, corresponding to exon III, allowed only the detection of the mature mRNA (2.4 kb) but not that of the 5.3- and 4.2-kb precursors (Figure 9A, lane 5). Nevertheless, the simultaneous presence of all other exons in these precursors also pleads for the inclusion of exon III in these precursors' molecules (Figure 9A, lanes 1, 3, 7, and 10).

The sequences downstream of exon III (preceding ORF143) and upstream of exon IV display the 5′ consensus and the domain V characteristics of a group II intron, but lack the elements necessary to reconstitute a Y-shaped molecule. The preliminary analysis of transcription of ORF143 (not shown here) suggests that this ORF could be involved in the constitution of a Y-shaped molecule. The typical group II intervening sequence between exons IV and V is eliminated by *cis*-splicing: the 0.9-kb intron sequence (probes 8 and 9) is present alone or linked with exon V as a 1.0 kb transcript, which implies the probable formation of a lariat (Figure 9A, lanes 8 and 9). The presence of two short sequences homologous to 26S ribosomal DNA explains the 3.2-kb band observed with probe 8. Figure 9A shows that exons IV and V (lanes 7 and 10) are present in the precursors and also either alone or linked in a 0.5-kb molecule in wheat and in a 0.6-kb molecule in maize. According to these results, *nad5* transcripts can be assembled and matured in either one of two ways, as shown in Figure 9C, that may not be exclusive of each other.

DISCUSSION

The gene encoding subunit 5 of NADH-ubiquinone reductase (*nad5*) has been detected in mitochondrial genomes of higher plants owing to its sequence conservation. This split gene consists of 5 exons and is dispersed at three different locations encompassing more than 100 kb in wheat. Although DNA rearrangements leading to the continuous transcription of the 5 exons cannot be excluded with certainty, this possibility is very unlikely. In the mtDNA of wheat, we have no evidence for the presence of a sequence allowing a unique precursor of the *nad5* gene to be synthesized.

The three different regions containing the 5 exons of *nad5* are separated by other genes and also by one copy of the direct repeated sequence containing the 18S and 5S rRNA genes as well as one copy of the invert repeated sequence harboring the gene encoding the 26S rRNA (Falconet et al., 1984, 1985; Lejeune et al., 1987). These genes have their own initiation and termination sites. Therefore, the synthesis of a continuous and very long primary transcript from exon I to exon V, encompassing genes encoding mitochondrial rRNAs, tRNAs, and polypeptides, seems unlikely. This structure makes improbable the assembly of the final messenger by way of the synthesis of a single transcript.

Our results suggest that the final messenger is obtained by the interaction of independent transcripts involving two *cis*- and two *trans*-splicing reactions. The same type of gene organization and the same maturation process involving *trans*-splicing have already been reported for the plastid genes *psaA* in Chlamydomonas (Kück et al. 1987) and *rps12* in higher plants (Fukuzawa et al., 1986; Zaita et al., 1987; Hildebrand et al., 1988) and for the mitochondrial *nad1* gene of wheat (Chapdelaine and Bonen, 1991) and of Oenothera (Wissinger et al., 1991).

Figure 9. (continued).

Bg, BglII; E, EcoRI; EV, EcoRV; D, DraI; P, PstI; S, SalI; Ss, SstI; X, XhoI.

(C) Model for the maturation of the different *nad5* transcripts. Open boxes correspond to the exons designed by roman numerals; IN, intron. For an easy identification of the different transcripts, the same symbols as in **(A)** have been used.

The model results from the analysis of the hybridization experiments above and presents two alternative ways for the assembly and processing of the transcripts: the three precursors are assembled as a 5.3-kb molecule (◁) that is eventually processed to the mature mRNA (◀), or through a two-step processing whereby exon III and exons IV (✱) and V (•) are assembled and processed, the product (✩) is linked to the first precursor (★), and then the intron (☼) between exons I and II is eliminated.

Gene Organization

The organization of the gene was studied in wheat and maize. Comparison of both results confirms the main structural features of the mitochondrial genome in higher plants: namely, the nucleotide sequence conservation of genes (allowing easy detection of the same gene between different angiosperm species) and the frequent sequence rearrangements. This last point is illustrated here by the different environment in which the same sequence is found. The region containing exons IV and V is found close to *rps12*, *nad3*, and tRNA genes in maize but not in wheat. Nevertheless some other gene arrangements seem to be conserved, for example the cluster (maturase, *nad1* exon e, *nad5* exon III, and ORF143) reported here for maize and wheat. It is not known whether this cluster is also conserved in dicots.

Other *nad5* genes show a great deal of structural variability, both in length and in internal structure. In Aspergillus, the gene encodes 657 amino acids (Brown et al., 1989) and in Chlamydomonas, it encodes 546 amino acids (Boer and Gray, 1986); in both species, the *nad5* gene has no intron, whereas in Neurospora (715 amino acids), there are 2 introns and 3 exons organized as one transcription unit (Nelson and Macino, 1987). When the predicted amino acid sequences encoded by the *nad5* gene of wheat and of other species are compared, the highest similarity is found with the filamentous fungi Aspergillus and Neurospora (Figure 6).

RNA Editing

As in most other instances of RNA editing reported for mitochondrial transcripts of higher plants (for review, see Mulligan, 1991; Schuster et al., 1991), only C-to-U changes have been observed. There is no explanation of why only the last 2 exons IV and V are affected by these changes. Editing events amounted to 11; no partially edited cDNA could be detected in wheat.

The amino acid changes caused by RNA editing in mitochondria of higher plants generally increase the conformity of an amino acid sequence of angiosperms to the same sequence in other groups of organisms (Mulligan, 1991; Schuster et al., 1991). By comparison with the fungal sequences, this is verified here only for 4-amino acid among the 10-amino acid changes observed (Figure 6). The six other changes either do not reestablish an amino acid similarity or even cancel a preexisting amino acid identity. One can then wonder whether the reason for RNA editing in the mitochondrial genome of higher plants is only to adapt the sequence to some hypothetical consensus model or whether it has structural and functional implications: many C-to-U changes, according to the universal code, lead to a more hydrophobic amino acid and, hence, increase the hydrophobicity of the protein. This is the case here for all changes observed in the amino acid sequence encoded by *nad5*.

RNA editing in the intervening sequences could not be observed because no cDNA clones containing intron sequences were obtained. Nevertheless the helical domain involved in the *trans*-splicing process (Figures 8B, 8C, and 8E) contains an A:C mismatch that could be eliminated by RNA editing, thus correcting the base pairing and stabilizing the structure. Such editing events correcting the intron structure have been reported in the *nad1* gene of Oenothera (Wissinger et al., 1991): they create A:U pairings instead of A:C mismatches in secondary structures involved in *trans*-splicing.

cis-Splicing between Exons I and II and Exons IV and V

Although both introns belong to group II, the comparative study of their splicing reveals differences that might be attributed to their structure. In the first intron, sequences at splicing sites are different from those found for typical group II introns. Previous studies have indicated that the structure of this region is important for splicing activity in vivo. As shown by van der Veen et al. (1987), mutations in the adenosine branch point of the fifth intron of the gene coding for the subunit I of cytochrome-*c* oxidase in yeast leads to excision without lariat formation. Schmelzer and Schweyen (1986) have shown also that some mutations can abolish in vitro and in vivo splicing. These results might thus constitute a model to explain why the *nad5* first intron, owing to its altered sequences, is spliced without lariat formation. As for the second intron, whose structure is typical of group II, the presence of a 1.0-kb intermediate transcript (Figure 9A, lanes 9 and 10) associating this second intron and exon V pleads for a conventional splicing process of the mitochondrial introns of group II.

trans-Splicing between Exons II and III and Exons III and IV

Evidence for *trans*-splicing processes can be found both in the structure of the involved sequences and in the transcription patterns obtained with suitable probes. As in the plastid *psaA* gene in Chlamydomonas (Kück et al., 1987; Choquet et al., 1988; Goldschmidt-Clermont et al., 1990) or the *rps12* gene in higher plants (Fukuzawa et al., 1986; Zaita et al., 1987; Hildebrand et al., 1988; Kohchi et al., 1988), intron features are characteristic of the sequences present at the 3′ and the 5′ ends, respectively, of 2 exons physically separated in the genome but linked in the mature mRNA. The *trans*-splicing process requires the formation of a Y-shaped molecule involving these sequences, either by direct interaction or by a mechanism involving a ligation step followed by *cis*-splicing with lariat

formation and specific processing of the lariat to yield a Y-shaped intron. These two possibilities have been previously described in Chlamydomonas by Goldschmidt-Clermont et al. (1991).

The results reported here for the *nad5* gene show undoubtedly a similar situation for sequences adjacent to the 3′ end of exon II and 5′ end of exon III. These sequences allow the reconstitution of a Y-shaped molecule, and the transcription results are in agreement with the hypothesis above.

The 3′ end of exon III and the 5′ end of exon IV, which contain intron-specific sequences, do not permit the reconstitution of a Y-shaped molecule because most of the intron sequence between domains I and IV is apparently missing. But the preliminary analysis of ORF143 transcription mentioned in Results suggests that this ORF could be involved in the *trans*-splicing process. Two hypotheses can explain its participation:

(1) Complementation between sequences 5′ of exon IV and 3′ of exon III would involve only the short sequences presented in Figure 8C; the Y-shaped molecule would then be formed by interaction of small mtRNA molecules with the elements flanking exon III and exon IV as in a model established for the first *trans*-splicing reaction of the *psaA* gene in Chlamydomonas (Goldschmidt-Clermont et al., 1991).

(2) This complementation results from a sequence located more than 700 bp downstream of exon III, according to our sequencing data. In this case, ORF143 would be included in the 5′ part of the interrupted intron.

More specific experiments are necessary to unveil the *trans*-splicing process necessary for the assembly of the 5 exons of the *nad5* gene.

Our results on transcription suggest that the efficiencies of the several splicing events are different. The 3.3-kb precursor of exons I and II is easily detected and, hence, seems to be processed more slowly (this leads to its accumulation) than the precursors for exon III and for exons IV and V. Moreover, the presence of some of the predictable intermediates of the *cis*- and *trans*-splicing reactions could not be demonstrated. This must be taken into account when analyzing the model presented in Figure 9C.

The transcription results of the *nad5* gene have led us to formulate the following hypothesis: (1) *cis*-splicing of exons I and II would require the presence of the 3 other exons, (2) exon III would interact with exon II if it had been first assembled to exons IV and V, (3) *cis*-splicing of exons IV and V might take place before or after their assembly with the other exons.

Taking this hypothesis into account has led us to propose the model for the assembly of the 5 exons presented in Figure 9C; in this model, the two proposed processing pathways may not be mutually exclusive. This model was based on the presence of the different intermediates and precursors shown by the use of multiple probes; the precursors that were very rapidly processed could not be detected. This preliminary model should be completed when RNA transcription has been analyzed in greater detail.

From an evolutionary point of view, the results described here for the *nad5* gene and those described by Chapdelaine and Bonen (1991) and Wissinger et al. (1991) for the *nad1* gene raise the question of how and when this gene organization originated. *trans*-Splicing is a necessary step in the reassembly of introns that have been partitioned through genome rearrangements. If the same repartition and association of exons are found in all higher plant species analyzed, it must then be concluded that they result from a unique and early event.

METHODS

Nucleic Acid Preparation and Analysis

Mitochondrial nucleic acids from wheat (*Triticum aestivum*) and maize (*Zea mays*) were prepared from purified mitochondria isolated from 6-day-old etiolated seedlings; mtDNA was obtained after mitochondrial lysis and CsCl-ethidium bromide centrifugation (Vedel and Quetier, 1978). Standard procedures were used for cloning and DNA gel blot analysis (Sambrook et al., 1989). The cosmid library of wheat mtDNA has been constructed in the laboratory (Quétier et al., 1985). The cosmid clone N5G6 of maize mtDNA has been kindly provided by C.M.R. Fauron (Howard Hughes Medical Institute, University of Utah, Salt Lake City).

Total mtRNA was isolated as described by Stern and Newton (1984), treated with RNase-free DNase I (Bethesda Research Laboratories) in standard conditions in the presence of RNase inhibitor (Bethesda Research Laboratories), phenol extracted, and precipitated. For RNA gel blot analysis, total mtRNA (5 μg) was fractionated on a 1.5% agarose vertical gel containing 2.2 M formaldehyde and transferred onto nitrocellulose filters as described by Sambrook et al. (1989).

DNA sequencing was performed by the dideoxy nucleotide chain termination method of Sanger et al. (1977) using wheat mtDNA or cDNA cloned in M13mp18 and M13mp19 vectors.

For S1 mapping experiments, probes were uniformly labeled by second-strand synthesis and precipitated with total mtRNA (15 μg). Hybridization was carried out overnight in standard buffer (Gualberto et al., 1988) at 48°C for the 3′ probe and at 45°C for the 5′ probe. Hybrids were digested with 1000 units per milliliter of S1 nuclease (Bethesda Research Laboratories) for 30 or 60 min at 37°C. The protected fragments were analyzed on a 6% acrylamide sequencing gel along with M13 sequencing reactions.

PCR-Based Specific cDNA Construction

cDNA first strand was synthesized from total mtRNA (500 to 800 ng) in a 20 μL of reaction mixture containing a specific primer (10 pmol) with 10 units of avian myeloblastosis virus reverse transcriptase (Pharmacia) in the recommended buffer in the presence of 10 units of RNase inhibitor (Bethesda Research Laboratories) at 42°C for 1 hr. The reaction was stopped by heating at 95°C and

loaded on a cDNA spun column (Pharmacia Sephagel 300) to remove excess linkers. The single-stranded cDNA was precipitated.

The anchored PCR procedure in the 5′ direction (Loh et al., 1989) was used to amplify the different batches of single-stranded cDNA. The single-stranded cDNA was first poly(dA) tailed in a 20-μL reaction volume with 10 units of terminal deoxyribonucleotide transferase (Bethesda Research Laboratories) in the appropriate buffer with 50 μM dATP for 5 min at 37°C. The reaction was stopped by heating to 65°C for 10 min and the volume adjusted to 250 μL. Ten-microliter aliquots were used as template in the PCR reaction.

Amplification was performed with 2.5 units of Taq DNA polymerase (Boehringer) in 100 μL of the recommended buffer and 1.5 mM concentration of each of the four deoxynucleotide triphosphates (dATP, dCTP, dGTP, and dTTP). The primers included one gene-specific primer (10 pmol) and the universal primer for the poly(dA) end that was a mixture of two different poly(dT) anchors, oligo(dT1) and oligo(dT2), at a ratio of 1:9 (Loh et al., 1989). The anchor of these two primers included restriction sites PstI, SstI, and XhoI. Unless otherwise indicated, after a preliminary denaturation step at 95°C for 5 min, Taq DNA polymerase was added and a 25-cycle amplification was performed (94°C, 1 min; 45°C or 50°C depending on the primer, 2 min; 72°C, 3 min) with a 10-min extension at 72°C of the final cycle. A preliminary assay was first performed on one-tenth of the sample that was separated on a 1% agarose gel and hybridized with an adequate probe. The rest of the amplification mixture was then separated on 1% agarose gel. According to the results of the test, DNA fractions of the appropriate size were electroeluted and ligated into M13 vectors (mp18 or mp19). After transformation of a recipient strain (JM109 or DH5α), single-stranded DNA was prepared from the transformed clones and sequenced.

Oligonucleotides

DNA oligonucleotides used in this work are as follows:
1a: 5′-TTCCGATGAGATCCATGGAGCTATT-3′, complementary to bases 641 to 666 (Figure 5A).
1b: 5′-TTGCTCGGTAGTTCCGTAGCCGGTT-3′, corresponding to bases 489 to 514 (Figure 5A).
2a: 5′-CCATTGCATCGGGTAACCAAGTATG-3′, complementary to bases 2039 to 2064 (Figure 5A).
2b: 5′-TGTGATCTTAGAGCTCGCTTACACA-3′, corresponding to bases 2530 to 2555 (Figure 5A).
3: 5′-TACCTAAACCAATCATCATATC-3′, complementary to bases 290 to 312 (Figure 5B).
4: 5′-CTTGTCTAAAGCTTCGAATGAGACT-3′, complementary to bases 764 to 739 (Figure 5C).
5: 5′-CTCCTACAGTTCTCTTATTACTCTT-3′, complementary to bases 1953 to 1979 (Figure 5C).
Oligo(dT1): 5′-CTCGAGCTCCTGCAG$(T)_{16}$-3′, linker.
Oligo(dT2): 5′-CTCGAGCTCCTGCAGTT-3′, linker.

GenBank Accession Numbers

The accession numbers for the sequences reported in this paper are M74157 (Figure 5A), M74158 (Figure 5B), M74159 (Figure 5C), and M74160 (maize sequence in Figure 7).

ACKNOWLEDGMENTS

The authors wish to thank Andrée Lepingle for her help during sequencing, Dr. Christiane M.R. Fauron for providing a maize cosmid, Dr. Francis Quétier, Dr. Martin Kreis, and Dr. Jean-Michel Grienenberger for helpful suggestions and discussion, and two anonymous reviewers for helpful comments. Anete Pereira de Souza is the recipient of a grant from the Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, Brasilia, Brazil. This work has been partially supported by funds from the French Ministère de la Recherche et de la Technologie (Paris, France, Contract 89.G.0453).

Received July 18, 1991; accepted October 14, 1991.

REFERENCES

Bland, M.M., Levings, C.S., and Matzinger, D.F. (1986). The tobacco mitochondrial ATPase subunit 9 gene is closely linked to an open reading frame for a ribosomal protein. Mol. Gen. Genet. **204,** 8–16.

Bland, M.M., Levings, C.S., and Matzinger, D.F. (1987). The ATPase subunit 6 gene of tobacco mitochondria contains an unusual sequence. Curr. Genet. **12,** 475–481.

Boer, P.H., and Gray, M.W. (1986). The URF 5 gene of *Chlamydomonas reinhardtii* mitochondria: DNA sequence and mode of transcription. EMBO J. **5,** 21–28.

Brown, T.A., Constable, A., Waring, R.B., Scazzocchio, C., and Davies, R.W. (1989). Nucleotide sequence of the *Aspergillus nidulans* mitochondrial gene for subunit 5 of NADH dehydrogenase. Nucl. Acids Res. **17,** 4371.

Chapdelaine, Y., and 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.

Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kück, U., Bennoun, P., and Rochaix, J.D. (1988). Mutant phenotypes support a *trans*-splicing mechanism for the expression of the tripartite *psaA* gene in the *C. reinhardtii* chloroplast. Cell **52,** 903–913.

Ecke, W., Schmitz, U., and Michaelis, G. (1990). The mitochondrial *nad5* gene of sugar beet (*Beta vulgaris*) encoding a subunit of the respiratory NADH dehydrogenase. Curr. Genet. **18,** 133–139.

Falconet, D., Lejeune, B., Quétier, F., and Gray, M.W. (1984). Evidence for homologous recombination between repeated sequences containing 18S and 5S ribosomal RNA genes in wheat mitochondrial DNA. EMBO J. **3,** 297–302.

Falconet, D., Delorme, S., Lejeune, B., Sévignac, M., Delcher, E., Bazetoux, S., and Quétier, F. (1985). Wheat mitochondrial 26S ribosomal RNA gene has no intron and is present in multiple copies arising by recombination. Curr. Genet. **9,** 169–174.

Falconet, D., Sévignac, M., and Quétier, F. (1988). Nucleotide sequence and determination of the extremities of the 26S ribosomal RNA gene in wheat mitochondria: Evidence for

sequence rearrangements in the ribosomal genes of higher plants. Curr. Genet. **13**, 75-82.

Fox, T.D., and Leaver, C.J. (1981). The *Zea mays* mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. Cell **26**, 315-323.

Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K., Umesono, K., Inokuchi, H., and Ozeki, H. (1986). Coding sequences for chloroplast ribosomal protein S12 from liverwort, *Marchantia polymorpha*, are separated far apart on different DNA strands. FEBS Lett. **198**, 11-15.

Goldschmidt-Clermont, M., Girard-Bascou, J., Choquet, Y., and Rochaix, J.D. (1990). *trans*-Splicing mutants of *Chlamydomonas reinhardtii*. Mol. Gen. Genet. **223**, 417-425.

Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Shirmer-Rahire, M., and Rochaix, J.D. (1991). A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell **65**, 135-143.

Grabau, E.A. (1986). Nucleotide sequence of the cytochrome oxidase subunit I gene from soybean mitochondria. Plant Mol. Biol. **7**, 377-384.

Gray, M.W. (1989). Origin and evolution of mitochondrial DNA. Annu. Rev. Cell. Biol. **5**, 25-50.

Gualberto, J.M., Wintz, H., Weil, J.H., and Grienenberger, J.M. (1988). The genes coding for subunit 3 of NADH dehydrogenase and for ribosomal protein S12 are present in the wheat and maize mitochondrial genomes and co-transcribed. Mol. Gen. Genet. **215**, 118-127.

Hack, E., and Leaver, C.J. (1983). The α-subunit of the maize F1-ATPase is synthesized in the mitochondrion. EMBO J. **2**, 1783-1789.

Hiesel, R., Schobel, W., Schuster, W., and Brennicke, A. (1987). The cytochrome oxidase subunit I and subunit III genes in Oenothera mitochondria are transcribed from identical promoter sequences. EMBO J. **6**, 29-34.

Hildebrand, M., Hallick, R.B., Passavant, C.W., and Bourque, D.P. (1988). *trans*-Splicing in chloroplast: The *rps12* loci of *Nicotiana tabacum*. Proc. Natl. Acad. Sci. USA **85**, 372-376.

Jarrel, K.A., Dietrich, R.C., and Perlman, P.S. (1988). Group II intron domain 5 facilitates a *trans*-splicing reaction. Mol. Cell. Biol. **8**, 2361-2366.

Kohchi, T., Umesono, K., Ogura, Y., Komine, Y., Nakahigashi, K., Komano, T., Yamada, Y., Ozeki, H., and Ohyama, K. (1988). A nicked group II intron and *trans*-splicing in liverwort, *Marchantia polymorpha*, chloroplast. Nucl. Acids Res. **16**, 10025-10036.

Kück, U., Choquet, Y., Schneider, M., Dron, M., and Bennoun, P. (1987). Structural and transcription analysis of two homologous genes for the P700 chlorophyll a-apoproteins in *Chlamydomonas reinhardtii*: Evidences for in vivo *trans*-splicing. EMBO J. **6**, 2185-2195.

Lamattina, L., Weil, J.H., and Grienenberger, J.M. (1989). RNA editing at a splicing site of NADH dehydrogenase subunit IV gene transcript in wheat mitochondria. FEBS Lett. **258**, 79-83.

Lejeune, B., Delorme, S., Delcher, E., and Quétier, F. (1987). Recombination in wheat mitochondrial DNA: Occurrence of nine different genomic contexts for the 18S-5S genes. Plant Physiol. Biochem. **25**, 227-233.

Levings, C.S., III, and Brown, G.G. (1989). Molecular biology of plant mitochondria. Cell **56**, 171-179.

Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier, L.L., and Davis, M.M. (1989). Polymerase chain reaction with single-sided specificity: Analysis of T cell receptor δ chain. Science **243**, 217-220.

Lonsdale, D.M., Hodge, T.P., and Fauron, C.M.R. (1984). The physical map and organization of the mitochondrial genome from fertile cytoplasm of maize. Nucl. Acids Res. **12**, 9249-9261.

Michel, F., and Dujon, B. (1983). Conservation of RNA secondary structure in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. EMBO J. **2**, 33-38.

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

Moon, E., Kao, T., and Wu, R. (1985). Pea cytochrome oxidase subunit II gene has no intron and generates two mRNA transcripts with different 5′-termini. Nucl. Acids Res. **13**, 3195-3212.

Mulligan, R.M. (1991). RNA editing: When transcript sequences change. Plant Cell **3**, 327-330.

Mulligan, R.M., Leon, P., and Walbot, V. (1991). Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression. Mol. Cell. Biol. **11**, 533-543.

Nelson, M.A., and Macino, G. (1987). Structure and expression of the overlapping ND4L and ND5 genes of *Neurospora crassa* mitochondria. Mol. Gen. Genet. **206**, 307-317.

Newton, K.J. (1988). Plant mitochondrial genomes: Organization, expression and variation. Annu. Rev. Plant Physiol. Plant Mol. Biol. **39**, 503-532.

Palmer, J.D. (1990). Contrasting modes and tempos of genome evolution in plant organelles. Trends Genet. **6**, 115-120.

Quétier, F., Lejeune, B., Delorme, S., Falconet, D., and Jubier, M.F. (1985). Molecular form and function of the wheat mitochondrial genome. In Molecular Form and Function of Plant Genome, Vol. 83, ASI Series A: Life Sciences, L. van Vloten-Doting, G.S.P. Groot, and T.C. Hall, eds (New York: Plenum Press), pp. 413-420.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**, 5463-5467.

Schmelzer, C., and Schweyen, R.J. (1986). Self-splicing of group II introns in vitro: Mapping of the branch point and mutational inhibition of lariat formation. Cell **46**, 557-565.

Schuster, W., and Brennicke, A. (1987). Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera*: Is genetic information transferred between organelles via RNA? EMBO J. **6**, 2857-2863.

Schuster, W., Wissinger, B., Hiesel, R., Unseld, M., Gerold, E., Knoop, V., Marchfelder, A., Binder, S., Schobel, W., Scheike, R., Gronger, P., Ternes, R., and Brennicke, A. (1991). Between DNA and protein—RNA editing in plant mitochondria. Physiol. Plant. **81**, 437-445.

Stern, D.B., and Newton, K.J. (1984). Isolation of intact plant mitochondrial RNA using aurintricarboxylic acid. Plant Mol. Biol. Rep. **2**, 8-15.

van der Veen, R., Kwakman, J.H.J.M., and Grivell, L.A. (1987). Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. EMBO J. **6**, 3827-3831.

Vedel, F., and Quétier, F. (1978). Hydrolyse spécifique de l'ADN chloroplastique et de l'ADN mitochondrial des végétaux supérieurs par les enzymes de restriction. Physiol. Vég. **16,** 411–425.

Wahleithner, J.A., MacFarlane, J.L., and Wolstenholme, D.R. (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.

Wintz, H., Grienenberger, J.M., Weil, J.H., and Lonsdale, D.M. (1988). Location of two tRNA genes and a tRNA pseudogene in the maize mitochondria genome: Evidence for the transcription of a chloroplast gene in mitochondria. Curr. Genet. **13,** 247–254.

Wissinger, B., Hiesel, R., Schuster, W., and Brennicke, A. (1988). The NADH-dehydrogenase subunit 5 gene in *Oenothera* mitochondria contains two introns and is co-transcribed with 18S-5S rRNA genes. Mol. Gen. Genet. **212,** 56–65.

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

Zaita, N., Torozawa, K., Shinozaki, K., and 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.