STRUCTURE AND EXPRESSION OF THE TRANS-SPLIT GENE FOR NADH DEHYDROGENASE SUBUNIT I IN WHEAT MITOCHONDRIA.

Yvan Chapdelaine

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfilment of the requirements for the
Ph. D. degree in the

Ottawa-Carleton Institute of Biology

ACKNOWLEDGMENTS.

I am grateful to Dr. Linda Bonen for her technical and theoretical guidance, and for her great interest in the project. It has been a pleasure to work with her and to learn from her the basis of scientific research. I have also appreciated the time and the interest of the members of my advisory committee, Dr. Carmody, Dr. Hickey and Dr. Johnson.

I thank Pascale Garber, Karen Williams, Caroline Wood, Degen Zhuo and other undergraduate students that have worked in the laboratory for their help. A special thanks to Sharon Bird for her discipline (and tips and tricks!) that is so indispensable for successful research in molecular biology.

Je tiens à remercier également le personnel de soutiens du Département de Biologie pour le support logistique et leur dévouement dans le bon déroulement du programme de Doctorat, spécialement Paul Brunon et Jacques Hélie. De même, pour le partage des joies et des peines communes, je souligne la solidarité de mes collègues, particulièrement Michèle Rouleau, Madeleine Lévesque, Dave Knox et Michel Gilbert.

ABSTRACT.

Plant mitochondrial genomes are known to undergo frequent DNA rearrangements during evolution. Such DNA rearrangements have been observed within the extreme 5' or 3' termini of plant mitochondrial genes. The present studies of the gene for NADH dehydrogenase subunit I (*nad1*) in wheat demonstrate that DNA rearrangements can even occur at internal sites.

This work shows that the *nad1* gene is comprised of five single-copy exons in wheat mitochondria. The predicted NAD1 amino acid sequence is closely related to that from non-plant and chloroplast counterparts. Somewhat surprisingly, these *nad1* sequences are scattered at four distantly-separated sites in the wheat mitochondrial genome. The analysis of regions flanking *nad1* coding segments revealed the presence of sequence motifs and helical structures that are hallmarks of group II introns. A model is proposed for *nad1* gene expression in which one *cis*- and three *trans*-splicing events are necessary for the production of *nad1* mRNAs.

To investigate *nad1* gene expression at the RNA level, transcripts arising from the four *nad1* coding regions were analyzed. Northern blot hybridizations and cDNA sequence analysis show that stable transcripts contain all five correctly-linked *nad1* exons. S1 nuclease analysis in the regions flanking the *nad1* coding segments also revealed the presence of stable transcripts which would be large enough to contain group II intron structures of normal sizes (<3 kb). As has been observed for virtually all plant mitochondrial mRNAs, the maturation of *nad1* transcripts also involves RNA editing events. Interestingly, one of these RNA editing sites converts an ACG codon to AUG to create an initiation codon, and this

suggests that RNA editing at this site is obligatory for translation of *nad1* mRNAs to proceed. RNA editing is also observed within the discontinuous *nad1*a/b intron in wheat (which is of particular interest because it lacks several key features of group II introns) but any contribution to intron structure or splicing is as yet uncertain.

The wheat *nad1* gene organization seems to have resulted from DNA rearrangements that occurred within previously continuous introns. This organization of the *nad1* gene in wheat, a monocot, differs from that of dicot *nad1* genes and this suggests that some DNA rearrangements have occurred relatively recently during plant evolution. The unusual *nad1* gene structure shows that DNA rearrangements can alter not only gene order but also gene structure, provided that scattered gene pieces are properly transcribed and spliced.

RÉSUMÉ.

Les génômes mitochondriaux des plantes sont reconnus pour subir de fréquents réarrangements d'ADN durant l'évolution. De tels réarrangements ont déjà été observés aux extrémités 5' et 3' de certains gènes mitochondriaux de plantes. La présente étude du gène *nad1*, qui code pour la sous-unité I du complexe de la NADH déshydrogénase du blé, démontre que des réarrangements d'ADN peuvent même se produire à des sites internes.

Ce travail montre que le gène *nad1* comprend cinq exons ayant chacun une seule copie dans la mitochondrie du blé. La séquence anticipée de la protéine NAD1 est très rapprochée des séquences d'autres types d'organismes et de celles des chloroplastes. Il est intéressant de noter que, contrairement à ceux de la plupart des gènes, les exons du gène *nad1* sont dispersés en quatre sites éloignés dans le génôme mitochondrial du blé. Des structures hélicoïdales, qui sont caractéristiques des introns du groupe II, sont néanmoins présentes dans les régions adjacentes à ces exons. L'intron *nad1*a/b comporte un intérêt particulier parce qu'il ne possède pas plusieurs éléments-clés des introns du groupe II. Un modèle est proposé dans lequel trois réactions de *trans*-épissage et une réaction de *cis*-épissage sont nécessaires pour la liaison des parties codantes du gène *nad1*.

Afin d'examiner l'expression du gène *nad1* au niveau de l'ARN, les produits de transcription provenant des quatre régions codantes ont été analysés. Par des hybridations ADN-ARN et par le séquençage d'ADN-c, des transcrits stables, qui contiennent les cinq exons du gène *nad1*, ont été observés. Une analyse des les régions adjacentes aux exons du gène *nad1* à l'aide de la nucléase S1 a aussi révélé que des produits de

transcription qui seraient assez longs pour contenir des structures d'introns du groupe II de taille normale (<3 kb). Comme il a été observé pour virtuellement tous les ARN messagers mitochondriaux de plantes, la maturation des transcrits du gène *nad1* implique aussi plusieurs sites d'édition d'ARN. Il est intéressant de constater que l'un de ces sites correspond à la conversion d'un codon ACG en AUG pour créer un codon d'initiation, ce qui suggère que l'édition d'ARN à ce site est indispensable pour la traduction des ARN messagers du gène *nad1*. Un site d'édition d'ARN est aussi observé dans l'intron discontinu *nad1*a/b du blé, mais la contribution de ce changement à la structure secondaire de l'intron ou à l'épissage demeure incertaine.

L'organisation du gène *nad1* du blé semble être le résultat de réarrangements d'ADN qui se sont produits dans des introns préalablement continus. Cette organisation du gène *nad1* du blé, une monocotylédone, diffère de celle du gène *nad1* de dicotylédones, ce qui indique que certains réarrangements d'ADN se sont produits dans les introns du gène *nad1* durant l'évolution des angiospermes. La structure inhabituelle du gène *nad1* montre que des réarrangements d'ADN peuvent modifier non seulement l'ordre des gènes, mais aussi leur structure, à la condition que les morceaux dispersés d'un gène soient correctement transcrits et épissés.

TABLE OF CONTENTS.

LIST OF FIGURES.

LIST OF TABLES..

ABBREVIATIONS.

A	adenosine
AMV	avian myeloblastosis virus
atp	genes for ATP synthase subunits
bp	base pairs
C	cytidine
cDNA	complementary DNA
CMS	cytoplasmic male sterility
cob	gene for cytochrome$_{b-c1}$ subunit
cox	genes for cytochrome oxidase subunits
ddNTP	dideoxynucleotide triphosphate
dNTP	deoxynucleotide triphosphate
EDTA	ethylendiaminetetraacetic acid
G	guanosine
I	inosine
kb	kilobases
mRNA	messenger RNA
mtDNA	mitochondrial DNA
mtRNA	mitochondrial RNA
Myr	million years
nad	genes for NADH dehydrogenase subunits (ND is used for non-plant organisms, and *ndh* for chloroplast homologues)
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
rpl	genes for protein of the large ribosomal subunits
rps	genes for protein of the small ribosomal subunits
rrn	genes for ribosomal RNAs
rRNA	ribosomal RNA
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SSC	150 mM NaCl, 32 mM sodium acetate, pH 7.0
T	thymidine
TE	10 mM Tris-HCl pH 7.5, 1 mM EDTA
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
U	uridine

CHAPTER 1. LITERATURE REVIEW.

1.1. ORGANIZATION OF THE GENETIC INFORMATION IN THE PLANT CELL.

The genetic information in the plant cell is divided among three compartments (nucleus, mitochondria and chloroplast) that can replicate their own DNA and synthesize proteins. Organellar genomes encode only a modest, but crucial, portion of the proteins involved in their biogenesis and function. Many of these proteins are components of the respiratory chain in the mitochondria, and of the photosynthetic apparatus in the chloroplast. Based on genetic studies, it has been estimated that approximately 90% of the mitochondrial proteins are nuclear-encoded and imported from the cytoplasm (Neupert and Schatz, 1981). The nucleus plays a central role in the regulation of organellar activities.

It is generally accepted that the high degree of dependence of mitochondria and chloroplasts on nuclear gene products has increased during evolution due to massive gene transfer from organelles to the nuclear genome. This idea is based on the endosymbiont hypothesis, which proposes that the contemporary organellar DNA is the remainder of the genome of prokaryotic-like symbionts that invaded primitive eukaryotic cells. Phylogenetic studies suggest that the most closely-related free living relatives of mitochondria and chloroplasts are the purple bacteria (alpha subgroup) and blue-green algae (cyanobacteria) (Gray et al. 1989; Van de Peer et al. 1990).

1.2. MITOCHONDRIAL GENES.

Studies on mitochondrial genomes from different organisms have shown that they contain a similar, yet not identical, set

of genes. However, independent gene transfer events that have occurred during evolution have led to certain differences in mitochondrial gene contents in different lineages (reviewed by Gray, 1989a). This variation is evident by examining the complete sequence of mitochondrial genomes from man (Anderson et al. 1981), several other species of the animal kingdom (Attardi and Schatz, 1988), *Podospora anserina* (Cummings et al. 1990), *Paramecium* (Pritchard et al. 1990), *Marchantia polymorpha* (Oda et al. 1992) and virtually the complete mitochondrial genomes from *Saccharomyces cerevisiae* (Wolf and Del Giudice, 1988) and *Chlamydomonas reinhardtii* (Michaelis et al. 1990). In Table 1, the mitochondrial protein-coding genes that have been identified are indicated for man, *S. cerevisiae* and *Podospora*. Protein-coding genes that have been identified so far in certain angiosperm mitochondrial genomes are also shown.

Mitochondrial genes can be classified according to the respiratory or structural function of their products. The genes involved in respiration encode a number of components of the electron transport chain complexes; these include the NADH dehydrogenase, cytochrome$_{b-c1}$, cytochrome oxidase and ATP synthase complexes. Genes for structural components comprise ribosomal RNA (rRNA) and transfer RNA (tRNA) genes, and genes for ribosomal proteins. The rRNA and tRNA genes are almost universally present in mitochondrial genomes although some tRNAs are known to be imported from the cytoplasm in several organisms, such as plant and protozoans (Nagley, 1989, Small et al. 1992).

The genes encoding subunits of the cytochrome oxidase complex (*coxI, coxII, coxIII*) and the cytochrome$_{b-c1}$ complex (*cob*) are present in each mitochondrial genome shown in Table

Table 1. Compilation of the mitochondrial protein-coding genes in different organisms.

	Homo sapiens[a]	Saccharomyces cerevisiae[b]	Podospora anserina[c]	Marchantia polymorpha[d]	Angiosperms[e]
Cytochrome oxidase					
coxI	+	+	+	+	+
coxII	+	+	+	+	+[g]
coxIII	+	+	+	+	+
Cytochrome b					
cob	+	+	+	+	+
ATP synthase					
atp6	+	+	+	+	+
atp8	+	+	+	–	?
atp9	–	+	–	+	+
atpA	–	–	–	+	+
NADH dehydrogenase					
nad1, nad2 nad3, nad4	+	–	+	+	+
nad4L	+	–	+	+	?
nad5, nad6	+	–	+	+	+
nad7	–	–	–	+[f]	+
Ribosomal proteins					
rps3, rps7, rps12, rps14,rps19, rpl16	–	–	–	+	+[g]
rps13	–	–	–	+	+[g]
rps1, rps10, rpl5, rpl6	–	–	–	+	?
rps2, rps4, rps8, rps11, rpl2	–	–	–	+	?
var1	–	+	–	–	?
RNA maturases and/ or endonucleases	–	+	37	11	+
Open Reading Frames	–	2-3	32	31	+
Genome Size	16 kb	75 kb	94 kb	187 kb	200-2400 kb

[a] Anderson et al. (1981); [b] Wolf and Del Giudice (1988); [c] Cummings et al. (1990); [d] Oda et al. (1992); [e] Bonen (1991); [f] Pseudogene; [g] Genes not present in all plants examined.

1; however the *coxII* and *coxIII* genes are absent from *C. reinhardtii* mitochondrial DNA (mtDNA) (Michaelis et al. 1990), as well as *coxII* in *Paramecium* (Pritchard et al. 1990). Similarly, four subunits of the ATP synthase complex are encoded in certain mitochondrial genomes but their presence varies greatly in different organisms. The *atp6* gene, which is the only one that is mitochondrially-encoded in all organisms shown in Table 1, is absent from *C. reinhardtii* mtDNA (Michaelis et al. 1990). Nine genes for subunits of the NADH dehydrogenase complex have been found so far in the mitochondrial genomes of different organisms. Eight of these genes are listed in Table 1, and the gene for another subunit (ND8) has been identified in the protist *Paramecium* (Pritchard et al. 1990). Some of these NADH dehydrogenase genes are absent from the mitochondrial genome of human and *Podospora* (ND7, ND8), *C. reinhardtii* (ND3, ND4L, ND7, ND8) and all of them in yeast. In trypanosomes, only ND1,ND2, ND4, ND5 and ND7 have been identified to date (reviewed in Bonen, 1991).

Ribosomal protein genes form a group that is absent from animal mitochondrial genomes, rare in fungi and more abundant in plants (see Table 1 and Bonen, 1991). The liverwort mitochondrial genome contains at least sixteen ribosomal protein genes based on their homology to bacterial ones (Oda et al. 1992). Seven of the ribosomal protein genes found in liverwort are known to be present in certain angiosperm mitochondrial genomes (*rps3, rps7, rps12, rps13, rps14, rps19, rpl16*). The genes for ribosomal proteins S12, S14, L2 and L14 have also been found in *Paramecium* mtDNA (Pritchard et al. 1990).

Based on the complete sequence of the chloroplast genomes in liverwort (Ohyama et al. 1986) and tobacco (Shinozaki et

al. 1986), chloroplasts also contain genes that are evolutionarily related to mitochondrial ones. This group includes the genes for the four ATP synthase subunits and the NADH dehydrogenase subunits shown in Table 1, as well as subunit 8 which was shown to be encoded by the chloroplast *psbG* gene (Nixon et al. 1989). There are also seven ribosomal proteins (S3, S7, S12, S14, S19, L2, L16) that are encoded by both chloroplast and mitochondrial genomes.

When this work began in 1987, only a few genes had been found in plant mtDNA: essentially genes for cytochrome oxidase, cytochrome$_{b-c1}$ and ATP synthase subunits. Estimations based on *in organello* protein synthesis in isolated mitochondria in maize (Forde and Leaver, 1980) or transcriptional studies in *Brassica* (Makaroff and Palmer, 1987), predicted that 20-30 proteins would be encoded in plant mitochondrial genomes. As listed in Table 1, twenty-one protein-coding genes have been identified in angiosperm mitochondria to date. The complete sequence of the *Marchantia polymorpha* mitochondrial genome indicates the presence of approximately 30 protein-coding genes of known functions, three ORFs with homology to previously identified mitochondrial ORFs in other organisms and 28 additional ORFs longer than 60 codons (Oda et al. 1992). Thus, the liverwort mitochondrial genome could potentially encode as many as 60 proteins, which is somewhat more than initially expected for angiosperms. It is likely that several liverwort mitochondrial ORFs will also be present in angiosperm mitochondrial genomes.

1.2.1. Gene transfer from mitochondria to nucleus.

There are only two protein-coding genes, namely *coxI* and *cob* genes, that are present in all mitochondrial genomes that have been examined. This suggests that most genes that were originally present in the endosymbiont have been either lost or transferred to the nucleus. For successful gene transfer, nuclear copies must be properly transcribed and specific signals must be acquired for their expression. The nuclear gene product must also be targeted to mitochondria and this involves an intricate protein import machinery (reviewed by Pfanner and Neupert, 1990).

One may wonder why all the mitochondrial genetic information has not been completely transferred to the nucleus. It appears costly to maintain protein synthesis and DNA replication machineries in organelles to express only a small number of genes. Interestingly, in *Epifagus virginiana*, a nonphotosynthetic root parasite, the plastid genome has been reduced considerably over a short period of time (5-50 Myr) and has retained only rRNA genes, ribosomal protein genes and some ORFs of unknown function (dePamphilis and Palmer, 1990). These ORFs might be involved in essential nonphotosynthetic functions that would justify the maintenance of this plastid genome. This raises the possibility that constraints may exist for maintaining certain genes within organelles, but it remains puzzling as to which genes those are (if any) and what their functions are.

Well-documented examples of gene transfer events suggest a model to describe how such events occur. One example is the *atp9* gene, which is mitochondrially-encoded in yeast but not in *Podospora* (Cummings et al. 1990) or animals (Attardi and Schatz, 1988). In *Neurospora*, both nuclear and mitochondrial

genomes contain one copy of the *atp9* gene, but only the nuclear copy has been demonstrated to be expressed (Mishra, 1991). Similarly, the *coxII* gene is found in almost all plant mitochondrial genomes, but is nuclear-encoded in mung bean and cowpea and is present in both genomes of certain legumes (Nugent and Palmer, 1991). These two examples (*atp9* and *coxII* genes) suggest that the nuclear gene copy becomes functional before the loss of the mitochondrial copy. Interestingly, based on Southern blot hybridizations, each ribosomal protein-coding gene that has been identified in the mitochondrial genome of a specific angiosperm appears to be absent from some other ones (Table 1), raising the possibility that many of these genes have been transferred to the nuclear genome during plant evolution.

1.3. RATES OF NUCLEOTIDE SUBSTITUTION IN MITOCHONDRIAL GENOMES.

Another source of diversity among mitochondrial genomes lies in the extremely variable rates of nucleotide substitution as inferred from DNA sequence data. Based on restriction site mutations, plant mtDNA appears to have a very slow rate of nucleotide substitution in both coding and non-coding regions (Palmer and Herbon, 1988). Similar differences in the rates of nucleotide substitution are observed in rRNA gene sequences (Gray et al. 1989). Moreover, in protein-coding genes, synonymous and nonsynonymous positions are estimated to change approximately three times slower in mtDNA than in plastid DNA, which itself is believed to evolve two times more slowly than nuclear sequences (Wolfe et al. 1987).

There are also major differences in the rates of nucleotide substitution among mitochondrial genomes from

different organisms. In contrast to the slow rate observed in plant mtDNA, the converse situation is seen in mammals where mtDNA evolves faster than nuclear DNA. Consequently, plant and mammalian mtDNAs differ in their respective rates of silent substitutions by 10 to 100 times (Wolfe et al. 1987). The reasons behind the slow evolution of plant mtDNA sequences are not known, but they may be due in part to differences in the DNA repair systems and in the fidelity of replication. However, Gray and colleagues (1989) suggested that a high level of conservation of DNA sequences in plant mitochondria alone cannot explain these differences, and that mitochondrial genomes may have a polyphyletic origin. Whether mitochondria and chloroplasts have monophyletic or polyphyletic origins has been debated in the literature. Arguments for one view or the other view are based on the diversity of membrane ultrastructure and pigment composition, and the question is whether these features have emerged before or after endosymbiosis (reviewed by Gray, 1989b; 1991).

1.4. PLANT MITOCHONDRIAL GENOMES.

It is clear from the above sections that mitochondrial genome features display considerable variation in different organisms. In this section, an overview of the plant mitochondrial genome organization is presented in terms of size, structure and recombinogenic nature. This section also illustrates that mitochondrial genomes are in many respects more complex in plants than in other organisms.

1.4.1. Plant mitochondrial genome sizes.

One distinctive feature of plant mitochondrial genomes is their large size. The smallest plant mitochondrial genome is found in *Marchantia polymorpha* (187 kb, Oda et al. 1992), and the smallest one among angiosperms has been reported in *Brassica hirta* (208 kb, Palmer and Herbon, 1987). Thus all plant mitochondrial genomes are more than 12 times larger than human mtDNA which is organized in a compact circular molecule of approximately 16 kb (reviewed in Attardi and Schatz, 1988). There are also certain animal mitochondrial genomes that are as small as 14.3 kb or as large as 32 kb (reviewed in Gray, 1989a). Plant mitochondrial genomes are also larger than most fungal ones which vary from 17.6 kb in *Schizosaccharomyces pombe* EF1 (Wolf and Del Giudice, 1988) to 176.3 kb in *Agaricus bitorquis* (Hintz et al. 1985).

Another hallmark of plant mtDNA is that sizes vary greatly between closely-related species. The most striking variation in size has been observed in the family *Cucurbitaceae*. The watermelon mitochondrial genome is approximately 330 kb which is about seven times smaller than the 2400-kb mitochondrial genome of muskmelon (Ward et al. 1981). Such differences were initially unexpected since these plants presumably have a very similar gene content.

The large size of plant mtDNA compared to other organisms is mostly due to non-coding DNA, to the presence of introns and to the presence of plant-specific genes. Non-coding DNA is abundant in plant and fungal mtDNAs (Gray, 1989a). In contrast, non-coding regions represent only a very small portion of animal mitochondrial genomes. Introns are also important contributors to fungal and plant genome sizes compared to animals in which introns have not yet been

reported. Although the number of plant mitochondrial genes might be substantially larger than that of non-plant organisms (section 1.2), coding regions are expected to cover only a small part of most plant mitochondrial genomes.

One may also wonder what are the reasons for these differences in size among plant mitochondrial genomes. Based on solution hybridization data, only 5-10% of extra DNA consists of repetitive sequences (Ward et al. 1981). The integration of DNA from nuclear and chloroplast sources also appears to contribute to the large size of plant mitochondrial genomes. Chloroplast DNA sequences are frequent and widespread in mitochondrial genomes (Stern and Palmer, 1984). The presence of sequences closely-related to various plastid ones at different sites of plant mitochondrial genomes suggests that multiple DNA transfer events from the chloroplast to the mitochondria have occurred recently during evolution (review by Newton, 1988). Except for some chloroplast tRNA genes that are functional in plant mitochondria (Joyce and Gray, 1989), most chloroplast DNA sequences that have been inserted in mitochondrial genomes appear to be non-functional.

1.4.2. Plant mitochondrial genome structure.

When plant mtDNA is digested with restriction enzymes, DNA fragments of various abundances are observed. Certain low level molecules can be detected only by hybridization and have been termed "sublimons" (Small et al. 1989). These different stoichiometries and the large sizes of plant mitochondrial genomes have hindered a rapid characterization of their structure.

Restriction mapping analysis of plant mtDNA predicts the presence of large circles in wheat (430 kb; Lejeune and

Quétier, 1988), maize (570 kb; Lonsdale et al. 1984), *B. campestris* (218 kb; Palmer and Shields, 1984), *B. hirta* (208 kb; Palmer and Herbon, 1987), sugar beet (386 kb; Brears and Lonsdale, 1988), *Petunia hybrida* (443 kb; Folkerts and Hanson, 1989) and liverwort (187 kb; Oda et al. 1992). Plant mitochondrial genomes also contain a certain number of repeated elements which vary from none in *Brassica hirta* (Palmer and Herbon, 1988) to as many as ten in wheat (Lejeune and Quétier, 1988). In *B. campestris* for example, the 218-kb mitochondrial genome contains two copies of a directly repeated sequence of approximately 2 kb. A model has been proposed in which these repeated elements recombine to generate two subgenomic circles of 135 kb and 83 kb (Palmer and Shields, 1984). Restriction mapping analysis is also consistent with the presence of such subgenomic circles.

However, except in liverwort, none of the master circles that are inferred from restriction mapping in different plants have been observed by electron microscopy, and only in rare cases have subgenomic circles been physically characterized (reviewed by André et al. 1992). Using pulsed-field gel electrophoresis and Eckhardt gels, Levy and colleagues (1990) have shown that a large circular chromosome of 120 kb in maize (Black Mexican Sweet) is abundant. They suggested that this chromosome has been isolated from the rest of the genome because it lacks recombinational repeats (see also André et al. 1992). In watermelon, electron microscopy and gel electrophoresis data indicate that linear molecules are more abundant than circular ones, and that some linear molecules are four times longer than the 330-kb mitochondrial genome (Bendich and Smith, 1990). It is therefore unclear whether the circular master chromosome exists in certain plants, and it is

certainly not abundant in most plants. While animal and fungal mtDNAs are usually circular, linear mtDNAs have been reported in some, but not all, protists such as *Tetrahymena* (55 kb), *Paramecium* (41 kb) and *C. reinhardtii* (15.8 kb; reviewed by Gray, 1989a).

The organization of plant mitochondrial genomes in multiple circular and/or linear molecules raises the problem as to how it is ensured that all genetic information is replicated and transmitted to progeny. Palmer and Shields (1984) have suggested that the mtDNA replication system may operate on multiple circles or by suppression of recombination and only the master chromosome would be replicated (see also Lonsdale, 1984). However, little is known about how and when recombinational events occur, and how the level of each molecular form is regulated. In yeast, mitochondrial growth and division is independent of mtDNA replication since strains without mtDNA can still undergo mitochondrial division. This indicates that mitochondrial division is controlled by the nucleus (Attardi and Schatz, 1988). In plants, it has been shown that isolated mitochondria are capable of both replicative and repair DNA synthesis (reviewed in Levings and Brown, 1989), but many questions remain to be addressed regarding the replication origin(s) and the factors involved in the replication process.

1.4.3. Plant mitochondrial DNA rearrangements.

Although the primary sequence of mtDNA is remarkably well conserved among plants, the gene order varies considerably in different species because of frequent DNA rearrangements that have occurred during evolution. This is in contrast with animal mitochondrial genomes in which the gene order is nearly

identical between widely divergent species (Attardi and Schatz, 1988), but is similar to fungal ones among which the gene order differs considerably (Scazzocchio, 1987; Lonsdale, 1989). In plants, the frequency of DNA rearrangements is such that, even between closely related species, gene flanking sequences are often completely different, and breakpoints in homology are found very close to or even within coding regions. Interestingly, the liverwort mitochondrial genome contains certain ribosomal protein operons with a similar gene order to those of the chloroplast and *Escherichia coli* (Oda et al. 1992). This observation suggests that plant mitochondrial genome organization may have been stable prior to the divergence of bryophytes and angiosperms.

The reasons behind the recombinogenic nature of mtDNAs in angiosperms have been extensively discussed in the literature (Lonsdale, 1989; André et al. 1992). Short repeated sequences are often found at sites of DNA rearrangements, and sometimes correspond to gene pieces that have been duplicated in plant mitochondrial genomes. These short repeats are ubiquitous in plant mtDNA and are clearly active in recombination (André et al. 1992). Interestingly, fungal mitochondrial genomes also contain short repeats that can recombine occasionally (Wolf and Del Giudice, 1988). In both yeast and maize, a three-stage recombination model, which involves subsequent recombination between long and short repeated elements, has been proposed to explain duplication and deletion of mtDNA sequences during evolution (Small et al. 1989).

André and colleagues (1992) have proposed another model which may explain the origin of some of the short repeated sequences. This model suggests that reverse transcription of "nonfunctional" RNA molecules may occur and these sequences

could be subsequently integrated in plant mitochondrial genomes. Although reverse transcriptase activity has not yet been detected in plant mitochondria, several reverse transcriptase-homologous sequences are present (Brennicke and Schuster, 1987; Newton, 1988; Wahleithner et al. 1990). Thus, frequent DNA rearrangements and short repeats are possibly linked, and this raises questions whether homologous recombination is general or site-specific (discussed in Joyce et al. 1988). As yet, there is no evidence suggesting that sequence motifs are necessary for recombination to occur.

1.4.4. Cytoplasmic male sterility.

Mitochondrial DNA rearrangements are intimately associated with deficiencies leading to certain types of cytoplasmic male sterility (CMS) in plants. The net result of CMS is that pollen is defective or simply absent in an otherwise normal plant. Analysis of specimens shows that novel chimeric genes are often generated by mtDNA rearrangements. The *urf13-T* and *pcf* genes are well-documented examples of chimeric genes involved in CMS (reviewed by Newton, 1988; Hanson, 1991). Although the mechanism underlying CMS is not well understood, it appears that chimeric gene products affect the respiratory capacity of mitochondria. The use of male-sterile lines has contributed greatly to plant breeding and to the production of hybrid seeds. Our understanding of how CMS is occurring will benefit from a better knowledge of the fundamental aspects of plant mtDNA structure and gene expression.

1.5. PLANT MITOCHONDRIAL TRANSCRIPT ANALYSIS.

Rearrangements in plant mtDNA often occur in the neighbourhood of translated sequences or within transcribed regions. Consequently, regulatory elements that are usually found upstream and downstream of coding sequences are often located within unrelated regions in different plants. Despite the large genome size and the abundance of spacer DNA, most mtDNA regions appear to be transcribed although at different levels among plants (reviewed in Gray et al. 1992). In *B. campestris*, Northern blot hybridizations indicate that approximately 30% of the 218-kb mitochondrial genome is transcribed into 24 abundant stable RNAs and most of the rest of the genome is transcribed at a lower level (Makaroff and Palmer, 1987). Similarly, *in organello* transcription studies suggest that the maize mitochondrial genome is virtually entirely transcribed, but some regions have no stable transcripts *in vivo* (Finnegan and Brown, 1990).

1.5.1. Plant mitochondrial transcription units.

Transcript analysis of mitochondrial genes from different plants reveals that some genes are cotranscribed: e.g. *nad3-rps12* in wheat (Gualberto et al. 1988), *ORF25-coxIII* in rice (Liu et al. 1992), *rps14-cob* in broad bean (Wahleithner and Wolstenholme, 1988), *rrn18-rrn5-nad5* in *Oenothera* (Wissinger et al. 1988). However, genes that are transcriptionally linked in one plant are often unlinked in others because of DNA rearrangements. For example, the *rps13* gene is cotranscribed with the *coxI* gene in *Oenothera* (Wissinger et al. 1990), and the *atp9* gene in tobacco (Bland et al. 1986). In wheat, the *rps13* gene is preceded by the *atp6* gene but appears to have no stable transcript (Bonen, 1987).

Transcription in plant mitochondria involves multiple promoters (Gray et al. 1992). This is in contrast with animal mitochondria which have only one promoter for each DNA strand (reviewed in Clayton, 1991), but is similar to yeast where at least 19 promoters are seen (reviewed in Wolf and Del Giudice, 1988). Plant mitochondrial transcription units, so far, do not exceed two or three genes, but the liverwort mtDNA organization suggests that longer operons are present (Oda et al. 1992).

1.5.2. Potential transcription initiation sites in plant mitochondria.

The identification of transcription initiation sites constitutes an important step in our understanding of plant mitochondrial gene expression. A number of 5' termini have been mapped by primer extension but the lack of consensus in the corresponding regions has raised the possibility that some of these transcript ends could result from RNA processing (Gray et al. 1992). To get around this problem, the guanylyltransferase enzyme has been used to label specifically primary transcripts which have di- or triphosphate at their 5' termini. The results led to the identification of a consensus sequence in wheat (Covello and Gray, 1991), maize (Mulligan et al. 1991) and soybean (Brown et al. 1991).

Wheat:	rAaannG**CrTA**tArtragt	Covello and Gray, 1991
Maize:	$^{t}_{a}$**CRTA**$^{G}_{t}$A$^{a}_{t}$AAa	Mulligan et al. 1991
Soybean:	yrAAATnn**CRTA**AGAGAAGAAAG	Brown et al. 1991

These consensus promoters are similar but share only the short CRTA motif (in bold). In addition, the position of the transcription initiation site (underlined) varies relative to the CRTA motifs, and this variation is also seen for different maize promoters (Mulligan et al. 1991). In similar experiments in soybean, the identification of an unrelated transcription initiation site for the primary "RNA *e*" (which does not fit the soybean consensus) has led Brown and colleagues (1991) to propose that different types of promoters could be used in plant mitochondria.

In maize, site directed mutagenesis indicates that the $G(AT)_{3-4}$ motif proposed earlier (Mulligan et al. 1988) is dispensable *in vitro*, and a stretch of 11 nucleotides (ACGTATTAAAA) constitutes an essential promoter element for the *atpA* gene (Rapp and Stern, 1992). It appears, however, that this 11-bp sequence alone is insufficient for promoter activity. These observations contrast with the *S. cerevisiae* mitochondrial consensus promoter (ATATAAGTA) which is also well conserved in other yeast species (Wolf and Del Giudice, 1988; Gray et al. 1992). Thus, initiation of transcription appears to occur at specific sequences in plant mitochondria, but these may vary between species, and one species may have multiple types.

1.5.3. RNA processing at 5' and 3' termini.

RNA processing also contributes to the complexity of transcript profiles in plant mitochondria (reviewed by Levings and Brown, 1989; Gray et al. 1992). This is similar to observations made for the chloroplast in which polycistronic transcripts are extensively processed at 5' and 3' termini as well as between genes (reviewed by Gruissem et al. 1988). In

animal mitochondria, protein-coding genes are interspersed with tRNA genes which appear to serve as RNA processing signals (reviewed by Clayton, 1991), and a site-specific endoribonuclease plays a major role in RNA processing. In *Neurospora*, long stem-loop structures flank tRNA genes at RNA processing sites (Burger et al. 1985), and tRNA structures have been shown to act as processing signals in *S. pombe* (Wolf and Del Giudice, 1988). In plant mitochondria, both stem-loop and tRNA structures appear to be used in processing (Gray et al. 1992). Single and double stem-loop structures have been observed downstream of certain genes. Some of these structures are located immediately upstream of the position of transcript termini established by S1 nuclease mapping. Although these structures may resemble bacterial transcription termination sites (Schuster et al. 1986), it is more likely that they are involved in RNA processing which appears to be predominant over transcription termination in plant mitochondria (reviewed in Levings and Brown, 1989; Gray et al. 1992).

1.6. **RNA EDITING.**

RNA editing involves posttranscriptional modifications that can specifically alter the coding information that is present in the genomic DNA. Different types of RNA editing have been found in a wide variety of genetic systems, including the mammalian nucleus, in the mitochondria of trypanosomes, *Physarum* and plants, and in plant chloroplasts (reviewed in Cattano, 1991; Gray et al. 1992).

In trypanosomal mitochondria, RNA editing is a developmentally-regulated process that results in extensive insertions and deletions of uridines (U). These changes are necessary at the RNA level to generate the correct reading

frame. Some mRNAs contain as many as 40% of their bases either inserted or deleted (Feagin et al. 1988). In slime mold mitochondria (*Physarum polycephalum*), cytidines (C) are inserted in both mRNAs and rRNAs, and RNA editing sites are rather evenly spaced (Mahendran et al. 1991).

Another type of RNA editing involves tissue-specific cytidine to uridine (C-to-U) changes that create a stop codon in mammalian apolipoprotein B transcripts. *In vitro* studies indicate that the process occurs by deamination (Hodges and Scott, 1992). Similarly, RNA editing is crucial in the conversion of a CAG codon (glutamine) to CGG codon (arginine) in a glutamate gated channel (Sommer et al. 1991). The authors have suggested that deamination may in fact convert an adenosine (A) to inosine (I). In these two cases of editing in mammalian nuclear genes, like in trypanosomes, the process is also developmentally regulated.

Mammalian editing in apolipoprotein B transcripts is conceptually similar to the C-to-U RNA editing process that was recently discovered in plant mitochondria (Gualberto et al. 1989; Covello and Gray, 1989; Hiesel et al. 1989), although plant mitochondrial RNA editing is much more extensive (see below). Interestingly, several C-to-U changes have also been observed in chloroplast RNAs but appear to be less frequent than in plant mitochondrial transcripts (Hoch et al. 1991; Maier et al. 1992).

Although these modifications are all grouped under the label "RNA editing" and although many of them are found in the mitochondria of different organisms, they are different with respect to the type of changes involved, their frequency and presumably the mechanisms underlying the process. These observations and the sporadic distribution of RNA editing in

different lineages led Gray et al. (1992) to conclude that the different types of mtRNA editing are likely to have evolved relatively recently.

1.6.1. RNA editing in plant mitochondria.

Plant mitochondrial C-to-U RNA editing sites are seen in the coding regions of most mRNAs, and some editing sites have also been observed in non-translated regions. Approximately 2-15% of the codons are affected, and usually the predicted amino acid sequence similarity with homologues from other organisms is increased (reviewed in Walbot, 1991; Gray et al. 1992; Bonnard et al. 1992). A large portion of the RNA editing sites are plant-specific, so that it is possible to predict the position of certain sites by a comparison of plant genomic DNA sequences (Covello and Gray, 1990). Uncommon U-to-C changes called reverse editing have been described (reviewed by Schuster et al. 1991a; Bonnard et al. 1992). These U-to-C changes are only present in rare cDNA clones, except one silent U-to-C editing site that has been observed in all wheat *coxIII* transcripts examined (Gualberto et al. 1990).

The analysis of RNA editing in transcripts at different stages during RNA maturation revealed that higher levels of editing are observed after splicing (Yang and Mulligan, 1991; Sutton et al. 1991), and in polysomal RNA (Gualberto et al. 1991). These observations indicate that a temporal relationship exists between RNA editing and transcript maturation, and supports the view that mRNAs are for the most part edited during early stages of the RNA maturation process.

1.6.2. Mechanisms underlying RNA editing.

Although the phenomenon of RNA editing in plant mitochondria has been extensively described since its discovery three years ago, the mechanisms by which RNA editing occurs and how editing sites are specifically recognized remain obscure. There are no consensus and/or secondary structures that have been identified in the sequences flanking the edited sites, except that deviations from expected nucleotides frequencies have been observed at surrounding positions (Covello and Gray, 1990), and that a guanosine (G) rarely precedes an edited C in *Oenothera* (Schuster et al. 1991a). These data suggest that constraints exist on sites to be edited, but they are not sufficient alone to distinguish those nucleotide positions to be edited.

The C-to-U changes seen in plant mitochondria could occur by base modifications, base substitutions or nucleotide substitutions. Editing in mammalian apolipoprotein B mRNAs occurs by deamination, which would correspond to a base modification mechanism. In trypanosomal mitochondria, guide RNAs are responsible for the recognition of the sites to be edited and for providing Us to be inserted, and this appears to be consistent with a model involving a double transesterification reaction (Blum et al. 1990; 1991; Harris and Hajuk, 1992). *In vitro* experiments have shown that RNA editing activity is detected in wheat mitochondrial extracts, and that the machinery involved is nuclease and protease sensitive (Araya et al. 1992), but more work will be necessary to elucidate the components involved in plant mtRNA editing.

1.7. RNA SPLICING.

Organellar introns are usually divided into different categories according to their structural characteristics. Group I and group II introns can be folded into distinct secondary structures (Michel et al. 1982; see Burke, 1988 and Michel et al. 1989 for reviews). Other short AT-rich introns found in *Euglena* chloroplast clearly share some structural similarities with group II introns (Michel et al. 1989). These three types of introns are different from the most studied pre-mRNA introns found in nuclear genes (Ruby and Abelson, 1991). It should be noted, however, that group II and nuclear pre-mRNA introns have biochemical characteristics in common between their splicing mechanisms (Jacquier, 1990). A fourth category consists of short introns in tRNA genes that are excised by a different RNA processing mechanism (reviewed by Rogers, 1990).

Some group I and group II introns have been demonstrated to undergo self-splicing (Cech, 1986) and reverse-splicing *in vitro* (Woodson and Cech, 1989; Augustin et al. 1990). However, efficient splicing *in vivo* is dependent on specific protein factors that are encoded by intronic ORFs (maturases) or by nuclear genes (reviewed in Lambowitz and Perlman, 1990). Maturase ORFs are present in some, but not all, group I and group II introns and may be necessary for the splicing of more than one intron. In group I introns, another class of ORFs encodes endonucleases that are directly involved in intron mobility (Delahodde et al. 1989, see below). In group II introns, maturase ORFs have blocks of homology with retroviral reverse transcriptases (Michel and Lang, 1985).

The majority of the nuclear-encoded proteins involved in splicing of group I and group II introns appear to be

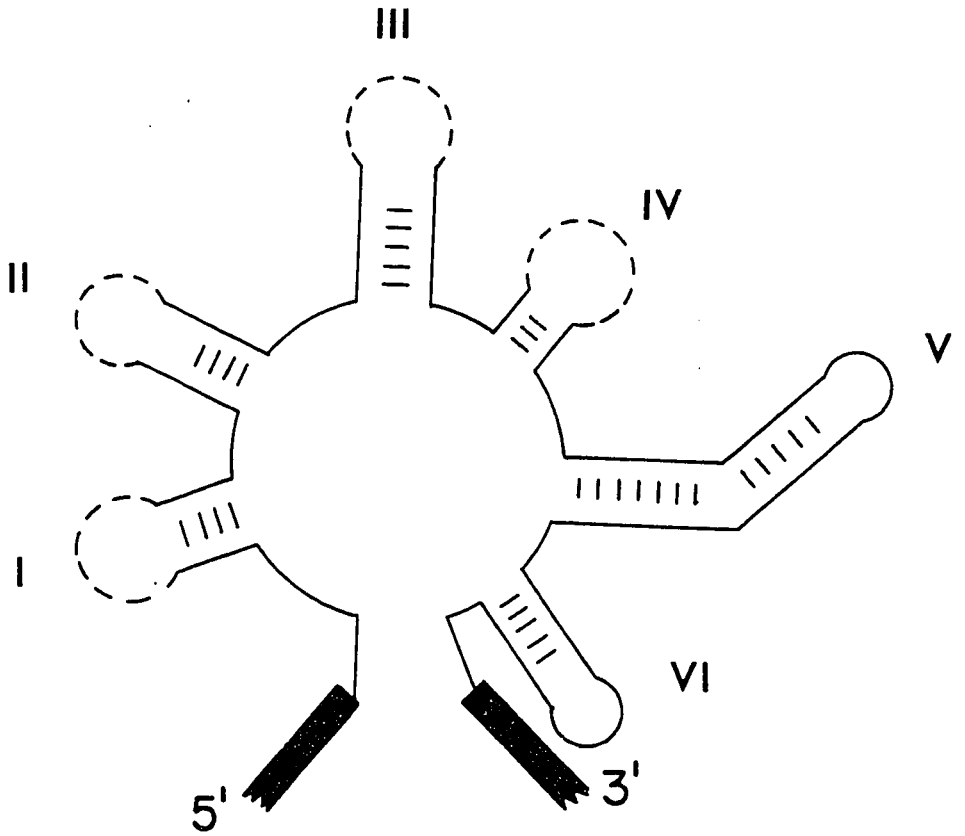
multifunctional (Lambowitz and Perlman, 1990). The most well-studied ones are the leucyl-tRNA synthetase in yeast and the tyrosyl-tRNA synthetase in *Neurospora*. The identification of a number of additional genetic loci in yeast and *Neurospora* that are important in intron splicing (reviewed by Lambowitz and Perlman, 1990) illustrates the complexity of the machinery involved in this process.

The number of group I introns in organelles is exceptionally high compared to the rare examples found in nuclear and eubacterial genomes (Cavalier-Smith, 1991; Reinhold-Hurek and Shub, 1992). Group II introns have as yet been found only in organellar genomes. The idea that most group I and group II introns were inserted in organellar genomes recently has gained much support during the last few years (reviewed in Palmer and Logsdon, 1991). This idea is based on their biased phylogenetic distribution and on the fact that introns can behave as mobile genetic elements. There are many group I and some group II introns that have been shown to be mobile (reviewed by Dujon, 1989; Perlman and Butow, 1989; Schmidt et al. 1990).

1.7.1. Structural characteristics of group II introns.

Group II introns can be folded into a secondary structure which includes a central core surrounded by six helical domains that are represented schematically in Figure 1 (Michel et al. 1982). Mutagenesis studies have confirmed predictions that were initially based on phylogenetic conservation. Consensus sequences are found at the 5' splice site (GTGCG or GCGCG) and at the 3' splice site (YAY in subgroup IIA or YAR in subgroup IIB). The members of these two subgroups are classified according to structural and/or sequence motifs

Figure 1. Schematic representation of the group II intron secondary structure. The intron core structure is surrounded by six helical domains (I-VI) and is flanked by the 5' and 3' exons (solid blocks). The position of the bulging A in domain VI is not shown.


III
IV
II
V
I
VI
5'
3'

found in domains I, III and VI (Michel et al. 1989). The domain V usually forms a hairpin structure of 14 bp containing two bulging nucleotides on the 3' side of the helix, and has a purine-rich tetraloop. These are among the most conservative and easily identifiable elements of group II introns.

In addition, several long-range interactions are crucial to establish appropriate intron structures. Among these, exon-binding sequences (EBS1, EBS2), that are located in domain I, are complementary to intron-binding sequences (IBS1, IBS2) located at the 3' end of the upstream exon. IBS1-EBS1 interaction is believed to be important for the precision of splicing at the 5' splice site and for maintaining the splicing intermediates in close proximity (Michel and Jacquier, 1987).

Based on *in vitro* self-splicing studies, another long-range interaction has been found between domain VI and the 5' splice site (van der Veen et al. 1986). Group II introns, like spliceosomal introns (Jacquier, 1990), form a lariat structure as a by-product of splicing. Lariat formation results from a 2'-5' phosphodiester bond between an internal 2' hydroxyl group and the 5' splice site. This hydroxyl group is provided by a bulging A located 7 or 8 nucleotides upstream of the 3' splice site in domain VI. The bulging A is among the most conservative positions of group II introns.

In vitro self-splicing properties of group II introns and site-directed mutagenesis have also been used to test which structures are essential for catalytic functions. In these experiments, structural elements have been removed individually (domain II, III, IV or VI) or in blocks (domains II, IV and VI) without abolishing self-splicing *in vitro* (Bachl and Schmelzer, 1990; Koch et al. 1992). In spite of all

the long-range interactions that are known, the specific role of domain V remains obscure. It is clear, however, that domain V forms a key element which is probably central to the tertiary folding of group II introns. Recent studies of the intron 5-gamma of the yeast *coxI* gene (Koch et al. 1992) indicate that domain V interacts directly with domain I, and that these two domains together with the upstream exon constitute the only essential structures that are necessary for catalytic activity of group II introns.

1.7.2. *Trans*-splicing *in vivo*.

The term "*trans*-splicing" is used to designate the covalent linkage of RNAs arising from different transcription units. In the more commonly observed *cis*-splicing process, the spliced RNAs arise from a single transcript. *Trans*-splicing is involved *in vivo* in the linkage of a 39 nt leader sequence to all pre-mRNAs in trypanosomes and related organisms. The process involves pre-mRNA-like intron structures and components of the spliceosome (Agabian, 1990). *Trans*-splicing is also involved in the maturation of a minority of mRNAs in nematodes (reviewed by Sharp, 1987; Laird, 1989), as well as in *Euglena* (Tessier et al. 1991).

Another *trans*-splicing model has been proposed for two chloroplast genes. In these cases, secondary structural features characteristic of group II introns appear to interact specifically for the junction of coding sequences. The *rps12* gene in liverwort and tobacco chloroplasts contains exons that are located at two widely separated sites and on different strands (Fukuzawa et al. 1986, Zaita et al. 1987; Hildebrand et al. 1988). This unusual gene organization results from a discontinuity within domain III of the first intron, and base

pairing in this domain may constitute an important *trans*-interaction for the proper recognition of introns halves (Kohchi et al. 1988). Similarly, the *C. reinhardtii psaA* gene has three exons and requires two *trans*-splicing events. Exons 1 and 2 are located on opposite strands and are also far apart from exon 3; each exon appears to be transcribed independently (Kück et al. 1987).

The isolation of *C. reinhardtii* mutants suggests that at least fourteen nuclear loci are required for the maturation of *psaA* mRNAs (Goldschmidt-Clermont, 1988). Mutations in these loci affect the splicing of either intron 1 or intron 2, or both, showing that specific and common factors are necessary. A chloroplast genetic locus (designated *tscA*) also appears to be essential for *trans*-splicing. It has been proposed that the *tscA* gene product is a short RNA molecule that would be directly involved in the formation of intron 1 core structure (Goldschmidt-Clermont et al. 1991). These findings illustrate the complexity of the components (RNA and/or protein) involved in the splicing of distant coding sequences in the chloroplast genome.

1.8. OBJECTIVES.

This overview has illustrated that plant mitochondrial genomes have both unique and common features compared to those of non-plant organisms. One of the most prominent features of plant mitochondrial genomes is their highly recombinogenic nature. It is therefore particularly interesting to study the influence of DNA rearrangements on genome organization, gene structure and gene expression during evolution.

This work aims to elucidate aspects of the structure and expression of the wheat mitochondrial gene for subunit I of

the NADH dehydrogenase complex (*nad1*). The NADH dehydrogenase complex is comprised of about 25 subunits in *Neurospora* mitochondria (Weiss et al. 1987), and probably at least as many in plants (Douce and Neuberger, 1989); and it constitutes the largest component of the electron transport chain. Respiration in plant mitochondria is more complex than in animals because of alternative pathways such as the rotenone-resistant NADH dehydrogenase and the cyanide-insensitive oxidase (Douce and Neuberger, 1989).

Somewhat unexpectedly, genes homologous to the mitochondrial NADH dehydrogenase genes are present in the chloroplast genomes (Matsubayashi et al. 1987). The presence of genes for NADH dehydrogenase components in chloroplast genomes has raised the hypothesis that chloroplasts may contain an NADH dehydrogenase activity and perhaps a respiratory chain similar to the mitochondrial one, which could be involved in chlororespiration. Chloroplast respiratory activity has been reported in *Chlamydomonas* (Bennoun, 1982) and is expected to be present in land plants (Meng et al. 1987; Wu et al. 1989; dePamphilis and Palmer, 1991).

Prior to the study presented here, sequences homologous to short segments of the *nad1* gene had been found in the mitochondrial genomes of several plants (Bland et al. 1986; Stern et al. 1986; Bonen 1987). However, because these sequences corresponded only to small internal portions, it was not clear whether they were parts of pseudogenes. The functionality of these *nad1* sequences was also doubtful because, interestingly, mitochondrial genes for NADH dehydrogenase subunits are absent in the yeast mitochondrial genome (Table 1), and have not been found in the nuclear

genome (Marres et al. 1991). Assuming the hypothesis that these plant *nad1* sequences are not functional, the gene could have been transferred to the nucleus during evolution and the protein imported into the mitochondrion. Another theoretical possibility is that the mitochondrial NAD1 protein is perhaps imported from the chloroplast where it could be encoded by the evolutionarily-related *ndhA* gene (Makaroff and Palmer, 1987).

Alternatively, these *nad1* sequences could be functional but unusual *cis*- or *trans*-splicing would be required for their expression. In the present study, the hypothesis that the rest of the *nad1* gene sequences are present and functional in wheat mitochondria was tested at the DNA and RNA levels. The objectives of this research have been (1) to determine whether the whole *nad1* gene is present and, if so, (2) to examine whether it is functional in wheat mitochondria, (3) to study its expression at the RNA level both with respect to RNA splicing and RNA editing, and (4) to analyze *nad1* transcripts to gain information about potential intronic structures.

CHAPTER 2. MATERIALS AND METHODS.

2.1. NUCLEIC ACID ISOLATION.

The methods used for the isolation of DNA and RNA from wheat mitochondria (*Triticum aestivum* var. Frederick) were derived from Bonen and Gray, (1980) and Wilson and Chourey (1984). Prior to germination, seeds were sterilized for 5 min. in 1% NaOCl and 10 min. in 0.01 N HCl, and rinsed with sterile water 3-5 times after each treatment. For DNA isolation, dissected embryos were germinated for 24 hours in the dark on two layers of Whatman 3MM paper soaked in sterile 1% glucose in Petri dishes. For RNA isolation, 50 g of seeds/tray (25 cm x 25 cm) were germinated in approximately 1 inch of autoclaved vermiculite soaked with water over 3 days in the dark. Pea (*Pisum sativum* var. Homesteader) seeds were treated as for wheat, and germinated for 6 days in the dark prior to isolating mtDNA and mtRNA.

Tissues (dissected embryos for wheat mtDNA, and shoot tissue for wheat and pea mtRNA and pea mtDNA) were harvested and ground in buffer I (0.44 M sucrose, 50 mM Tris-HCl pH 8.0, 3 mM EDTA, 1 mM ß-mercaptoethanol, 0.1% bovine serum albumin) with a mortar and a pestle on ice, and filtered through four layers of cheesecloth and one layer of Miracloth (Calbiochem). To isolate mitochondria, centrifugation was performed at 500 x g for 5 min., and the supernatant was centrifuged at 12 000 x g for 30 min. The crude mitochondrial pellet was resuspended in buffer I and the centrifugation cycles were repeated once.

For DNA isolation, mitochondria were resuspended in 0.6 ml buffer II (50 mM Tris-HCl pH 8.0, 20 mM EDTA), followed by lysis in 1.2 ml buffer III (0.2 M Tris-HCl pH 8.0, 0.1 mM EDTA, 0.2 M NaCl, 2% SDS, 0.2 M ß-mercaptoethanol) and

incubated at 65°C for 20 min. Then, 0.6 ml 5 M potassium acetate were added and samples were left on ice for 30 min. Samples were spun for 3 min. and the DNA (in the supernatant) was precipitated with 120 ul 5 M ammonium acetate and 1.2 ml cold isopropanol at -20°C for 30 min. The samples were then centrifuged for 5 min. The DNA pellet was washed with 70% ethanol, dried and resuspended in 1.4 ml buffer II. DNA was reprecipitated with 150 ul 3 M sodium acetate and 1.0 ml isopropanol (room temperature), spun for 30 seconds, washed with 70% ethanol, dried and stored in 50 ul TE buffer (10 mM Tris-HCl pH 7.5, 1.0 mM EDTA).

For RNA isolation, mitochondria were resuspended in 0.2 ml buffer IV (10 mM Tris-HCl pH 8.5, 50 mM KCl, 10 mM $MgCl_2$). Then 0.2 ml buffer IV containing 8% Triton-X 100 were added. Mitochondria were lysed by adding 0.6 ml detergent mix (2% tri-isopropylnaphthalene sulfanate, 12% sodium p-aminosalicylate, 0.1 M NaCl, 20 mM Tris-HCl pH 7.4), followed by phenol extraction (0.7 ml phenol containing 0.1% 8-hydroxyquinoline and saturated with TE buffer; Maniatis et al. 1982). 120 ul 5 M NaCl and 0.7 ml phenol were added to the aqueous phase for a second phenol extraction step. RNA was precipitated by the addition of two volumes of 95% ethanol and stored at -20°C.

2.2. OLIGONUCLEOTIDE SYNTHESIS.

The following oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and have been designated according to the *nad1* exon (or region) that they specify. The SalI, XmnI, XhoI and BglII restriction sites are underlined within oligomers A3, A4, B3 and E, respectively, and were used for cloning the PCR amplification products into M13 vectors.

A1:	CCTGCTATAATTATTCCATAAACGC	(Figure 3A, positions 507-483)
A2:	AAGGCTACTCCTAGTAGAAG	(Figure 3A, 198-179)
A3:	TCGGGTCGACCAGGTCAGGC	(Figure 3A, 86-105)
A4:	ATTTTAGAACCCTTTCTC	(Figure 3A, 899-882)
B1:	ACTTCATAAGGGACCATTTG	(Figure 3B, 676-657)
B2:	CGTAATGCTCCTAGAAAGGC	(Figure 3B, 646-627)
B3:	TTGGGTTGGGGTCTCGAGCCG	(Figure 3B, 508-528)
B4:	GAAGCTGTCGCTTGACGGAC	(Bonen, 1987; 615-596)
C1:	ATCTGCTTTTGCGCCATGAC	(Figure 3B, 2182-2163)
D1:	TAAGATCATATTGGCATACT	(Figure 3C, 255-236)
D2:	GGCGAACATCTAATCTAG	(Figure 3C, 909-926)
E1:	AATGGCAAGATCTAGGATAG	(Figure 3D, 2829-2810)
E2:	GAGCTAATGATAGAGGCAAGAACAC	(Figure 3D, 2980-2956)
E3:	CCTTCAGAAGAAACTTCCTG	(Figure 3D, 3059-3040)

The oligomer E2 sequence was derived from a conserved region within the watermelon mitochondrial ORF36 (Stern et al. 1986) and oligomer D1 from the broad bean *nad1* sub-terminal exon (Wahleithner et al. 1990). The position of oligomer D2 is based on the sequence of only one strand. An oligomer that specifies *nad5c* was derived from the *Oenothera* counterpart sequence (see Figure 3B positions 151-131 in Knoop et al. 1991). Two primers flanking the multiple cloning site are based on the sequence of primers #1212 (universal sequencing primer) and #1201 (reverse sequencing primer, New England Biolab).

2.3. CLONING OF PLANT MITOCHONDRIAL DNA.

Wheat mtDNA (approximately 5 ug) was digested with various restriction enzymes (from different suppliers), separated by electrophoresis on agarose gels (1%) in Tris-acetate buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 20 mM sodium acetate), recovered using GeneClean (BIO/CAN Scientific) and ligated with T4 DNA ligase (Boehringer Mannheim) into dephosphorylated pUC plasmid vectors (15 ng) for

transformation of *E. coli* strain TB1. Clones were retrieved from banks using colony hybridization methods (Maniatis et al. 1982). Plasmid DNA was prepared by the boiling method (Maniatis et al. 1982).

For DNA subcloning, plasmid restriction fragments of interest (approximately 100 ng) were isolated from low melting point agarose gels (1%, BRL), ligated into 2.5 ng M13 DNA previously digested with the appropriate restriction enzymes for transformation of *E. coli* strain JM101 (Maniatis et al. 1982). M13 single stranded DNA was prepared using the polyethylene glycol method (Maniatis et al. 1982).

2.4. RADIOACTIVE LABELLING OF NUCLEIC ACIDS.

Synthetic oligomers (100 ng) were labelled at their 5' termini with T4 polynucleotide kinase (Pharmacia, 7 units), gamma-^{32}P-ATP (3000 Ci/mmole, NEN/Dupont, 50 uCi), in kinase buffer (50 mM Tris-HCl pH 9.5, 10 mM $MgCl_2$, 5 mM dithiothreitol), and spun through 1 ml Sephadex G-50 columns (Pharmacia) previously equilibrated in TE buffer (Maniatis et al. 1982). The second 50-ul eluate was used as probe or as primer.

Probes specific to certain M13 DNA inserts were obtained by the synthesis of the complementary strand using the universal sequencing primer, dNTPs (G, C, T), α-^{32}P-dATP (3000 Ci/mmole, NEN/Dupont) and Klenow fragment polymerase (Pharmacia). Probes from denatured double-stranded DNA fragments (approximately 50 ng) were synthesized by the random priming method using hexamers (Pharmacia), dNTPs (G, C, T), α-^{32}P-dATP and Klenow fragment polymerase (Sambrook et al. 1989).

For S1 nuclease mapping, DNA size markers were obtained by filling recessed 3' ends, generated by cleavage of pBR322 DNA with the restriction enzyme HinfI, using the Klenow fragment polymerase (Pharmacia) and α-^{32}P-dATP (Maniatis et al. 1982). Plasmid digests as well as isolated DNA fragments were also radioactively labelled by this method for S1 mapping. To obtain uniformly labelled probes that were specific to short regions, M13 DNA inserts were radioactively labelled by the polymerase chain reaction (PCR). Each PCR-amplification mix contained 125 ng of each primer, 5 uM dATP, 15 uM dGTP, 15 uM dCTP, 15 uM dTTP, 1.5 ul α-^{32}P-dATP, 5 ul 10 x amplification buffer (see section 2.7), 5 ul M13 DNA (approximately 0.1 ng), and 2 units of Taq DNA polymerase (BIO/CAN). Typically, 25 cycles (30 sec. at 94°C, 1 min. at 47°C, and 1 min. at 72°C each) were performed on a Perkin Elmer Cetus thermal cycler.

2.5. NORTHERN AND SOUTHERN BLOT HYBRIDIZATIONS.

For Northern blots, wheat mtRNA (approximately 5 ug RNA/sample, stored in ethanol) was spun 15 min. (Eppendorf 5415, 14000 revolutions per minute, washed with 70% ethanol, dried and dissolved in 4.5 ul TE buffer, 2 ul 0.1 M $NaPO_4$ pH 7.0, 3.5 ul 37% formaldehyde (BDH) and 10 ul deionized formamide. The samples were incubated at 60°C for 5 min. Then 2 ul of sterile RNA loading buffer (50% glycerol, 1 mM EDTA, bromophenol blue) were added. The gels contained 0.01 M $NaPO_4$ pH 7.0, 1.25% agarose (BRL) and 7% formaldehyde (BDH). Electrophoresis was performed at 200 volts for 4-5 hours. One lane was stained by three washes for 30 min. in 0.1 M ammonium acetate (only the second wash includes 10 ug/ml ethidium bromide) and photographed using Polaroid film 57. The transfer

of RNA to nylon membrane (ICN) was carried out by capillary action in 20 x SSC (3 M NaCl, 0.65 mM sodium acetate, pH 7.0) overnight (Maniatis et al. 1982). Nucleic acids were crosslinked to the nylon membrane by ultraviolet illumination (254 nm) for 5 min.

For Southern blots, DNA was digested with restriction enzymes prior to electrophoresis on 1% agarose gel in Tris-acetate buffer. Nucleic acids were stained with 10 ug/ml ethidium bromide (Maniatis et al. 1982) for photography, prior to denaturation (1.5 M NaCl, 0.5 M NaOH for 25 min.) and neutralization (3 M sodium acetate pH 5.5 for 25 min.). For genomic DNA, the denaturation step was preceded by an acid hydrolysis step (0.25 N HCl, 15 min.). DNA was transferred to nylon membranes and crosslinked as described for Northern blots.

Oligonucleotide hybridizations were carried out in 5 x SSC, and 5% deionized formamide (BDH), 0.1% SDS and 50 ug/ml tRNA at 39°C overnight (derived from Choquet et al. 1988). Membranes were washed twice in 2 x SSC, 0,1% SDS for 15 min. at room temperature. Hybridizations using other types of probes were carried out overnight at 42°C in 50% deionized formamide (BDH), 5 x SSC, 0.5% SDS, 50 mM $NaPO_4$ pH 7.0, 250 ug/ml sheared denatured calf thymus DNA and 5 x Denhardt's (Maniatis et al. 1982). Membranes were rinsed in 2 x SSC, 0.1% SDS, washed in 0.2 x SSC, 0.1% SDS for 30 min. at room temperature and for 30 min. at 50°C. Washing solutions were changed every 15 min. For autoradiography, membranes were exposed to XAR films (Kodak) at -80°C with an intensifying screen for the appropriate length of time.

2.6. NUCLEIC ACID SEQUENCE ANALYSIS.

DNA sequence analysis was carried out by the dideoxynucleotide chain termination method (Sanger et al. 1977) using α-^{35}S-dATP or α-^{32}P-dATP (Amersham and NEN/Dupont) with either Sequenase (USB) or Klenow polymerase (Amersham). The sequencing gel (0.4 mm thickness) contained 7% acrylamide (Maniatis et al. 1982), 7 M urea and Tris-borate buffer (45 mM Tris-HCl pH 8.3, 45 mM boric acid, 1.3 mM EDTA). After pre-electrophoresis (1000 volts for one hour) on a sequencing apparatus (BRL, model S2), electrophoretic separation of nucleic acids was carried out at 1800 volts for 2 hours and at 2000 volts for 2 hours. Then a second aliquot of the sample was loaded and electrophoresis was continued for approximately 1.5 hours at 2200 volts. For all sequences presented, both strands were sequenced, restriction sites were crossed and dITP was used to resolve ambiguities. The sequence data were analyzed using Microgenie (Beckman) programs and GenBank/NBRF data banks.

The method used for sequencing PCR products was derived from Hsiao (1991). The PCR products were precipitated with one volume of 5 M ammonium acetate and 3 volumes of 95% ethanol for 15 min at room temperature, and spun for 10 min. The pellet was dried and resuspended in 5.0 ul TE buffer. Samples were separated by electrophoresis on a low melting point agarose gel (BRL), and DNA was isolated following standard protocols (Maniatis et al. 1982). Fifty percent of the purified PCR product (approximately 1 ug) was incubated with the appropriate oligomer (10 ng in 1 ul) and 1.5 ul 1.0 N NaOH at 37°C for 30 min. The denaturation was stopped by adding 1.5 ul 1.0 N HCl and sequencing was carried out with Sequenase

following the recommended protocol of the supplier, except that the annealing step was omitted.

The method used for DNA sequencing with the avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences) was derived from Graham et al. (1986). Approximately 200 ng of M13 single stranded DNA (1.5 ul), 0.8 ul RT buffer (340 mM Tris-HCl pH 8.3, 500 mM NaCl, 60 mM $MgCl_2$, 50 mM dithiothreitol), 0.5 ul 17 uM dATP, 1.0 ul primer (10 ng/ul), 2.5 ul H_2O and 2.5 ul α-^{32}P-dATP were incubated at 65°C and cooled to 40°C. Then 0.8 ul RT buffer and 5.0 ul of the deoxynucleotide mix (0.5 mM dGTP, 0.5 mM dTTP and 0.5 mM dCTP) were added, and divided in four tubes (3.3 ul each) containing 1 ul of one dideoxynucleotide solution (0.01 mM ddATP, 0.125 mM ddGTP, 0.25 mM ddCTP or 0.5 mM ddTTP). 1 ul of the enzyme mix (0.4 ul RT buffer, 2.8 ul H_2O and 0.8 ul 17 units/ul reverse transcriptase) was added to each tube for incubation at 42°C for 15 min., and then 3.0 ul chase mix (1.25 mM of each dNTP in 0.1 x RT buffer) for an additional 15 min. The reaction was stopped with 7 ul formamide-loading buffer (80% v/v deionized formamide, 50 mM Tris-HCl pH 8.3, 1 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue).

RNA sequences were obtained by the reverse transcriptase dideoxy chain termination method described by Geliebter and colleagues (1986, procedure 2), except that actinomycin D was omitted. End-labelled oligonucleotides (approximately 10 ng in 10 ul, section 2.4) were incubated with approximately 25 ug of wheat mtRNA (washed with 70% ethanol and dried) and 2.4 ul 5 x annealing buffer (1.25 M KCl, 50 mM Tris-HCl pH 8.3). The annealing temperature was calculated according to the formula 4(G+C)+2(A+T)-5 (Geliebter et al. 1986). After incubation for 45 min., the annealing mixes were divided into four tubes

containing the AMV reverse transcriptase enzyme (3 units each, Life Sciences), 3.3 ul of the RT-nucleotide buffer (24 mM Tris-HCl pH 8.3, 16 mM $MgCl_2$, 8 mM dithiothreitol, 0.4 mM dATP, 0.4 mM dCTP, 0.8 mM dGTP, 0.4 mM dTTP), and 1 ul of one of the four dideoxynucleotide mixes (1 mM dATP, 1 mM dCTP, 1 mM dGTP and 2 mM dTTP). The samples were incubated at 50°C for 45 min. Products were resolved on 7% acrylamide sequencing gels. The gels were dried and exposed for autoradiography on TM-trilite film (3M) at room temperature for the appropriate length of time.

2.7. **PCR AMPLIFICATION.**

For PCR amplification, cDNAs were prepared using the method described above for reverse transcriptase dideoxy sequencing, except that the dideoxynucleotide mixes were omitted. The products were also incubated for one additional hour with 2 ul 2 mM dNTPs. The cDNAs were purified by phenol extraction and sodium-acetate-ethanol precipitation (Maniatis et al. 1982), and dissolved in 100 ul TE buffer. Each amplification tube contained 600 ng of each primer (approximately 100 pmole), 10 ul of amplification buffer (100 mM Tris-HCl pH 9.0, 500 mM KCl, 15 mM $MgCl_2$, 0.1% gelatin and 1% Triton; BIO/CAN), 10 ul dNTPs (2 mM each), 2 units Taq DNA polymerase (BIO/CAN) and 10 ul DNA template (mtDNA or cDNA) in a final volume of 100 ul. Thirty cycles were performed on a Perkin Elmer Cetus DNA thermal cycler: 1 min. at 94°C, 2 min. at the melting temperature, which was calculated according to the formula 4(G+C)+2(A+T), and 3 min. at 72°C. A final extension was then performed at 72°C for 5 min.

2.8. S1 NUCLEASE MAPPING.

The method used for S1 nuclease mapping was derived from Maniatis et al. (1982). Approximately 25 ug of wheat mtRNA were dissolved in 5 ul H_2O and 3 ul 5 x PIPES buffer (0.2 M piperazine-N,N'-bis(2-ethanesulfonic acid) disodium salt pH 6.5, 5 mM EDTA, 2 M NaCl) and dried. The pellet was resuspended in 3 ul denatured radio-labelled DNA fragment (approximately 10 ng, section 2.4) and 12 ul deionized formamide (BDH), incubated at 85°C for 15 min., and then at 49°C overnight. Then the mixture was allowed to cool slowly to 43°C, and put on ice. 300 ul of S1 buffer (0.25 M NaCl, 30 mM sodium acetate pH 5.5, 1 mM $ZnSO_4$, 20 ug/ml carrier DNA) were added and the mix was divided in four tubes (80 ul each) containing 0, 50, 200 and 500 units of S1 nuclease (Pharmacia or Boehringer Mannheim). After incubation at 30°C for 30 min. The reactions were stopped by adding 22 ul stop solution (4 M ammonium acetate, 50 mM EDTA and 50 ug/ml carrier tRNA) and 215 ul 95% ethanol. Protected nucleic acids were precipitated at -20°C for 30 min. The samples were spun for 15 min., washed with 95% ethanol, dried and the pellet was resuspended in 6 ul TE buffer of which 2 ul were typically resolved on 7% polyacrylamide sequencing gels (section 2.6).

CHAPTER 3. RESULTS.

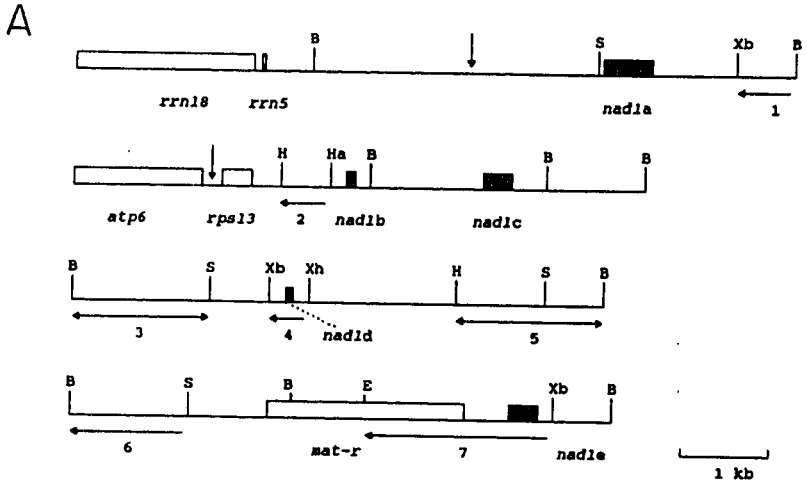
3.1. CHARACTERIZATION OF THE *nad1* CODING SEQUENCES IN WHEAT MITOCHONDRIA.

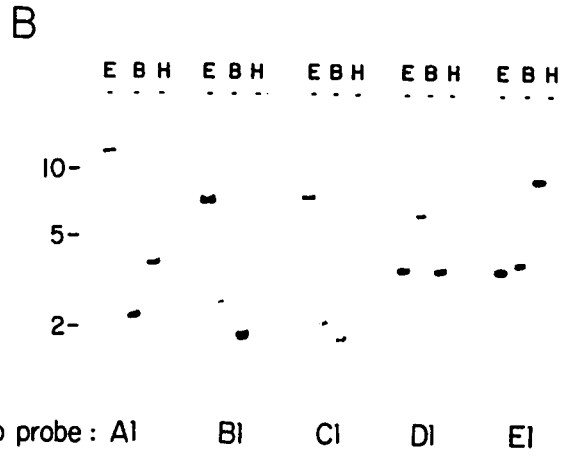
Prior to the present studies, short mitochondrial ORFs with homology at the protein level to the NADH dehydrogenase subunit I had been identified in several plants. These *nad1* sequences were located approximately 1 kb downstream of the *rps13* gene in wheat (Bonen, 1987), maize and tobacco (Bland et al. 1986). These sequences, however, could potentially encode only an internal segment of the anticipated NAD1 polypeptide. In the absence of ORFs with homology to the N-terminal portion of the NAD1 polypeptide in the upstream region, it had been suggested that these *nad1* sequences might be pseudogenes and that the functional NAD1 protein could be encoded in the nucleus or the chloroplast (Lonsdale, 1989; Makaroff and Palmer, 1987). In this section, another possibility is examined, that is, that the remainder of the gene is encoded at other site(s) in the wheat mitochondrial genome.

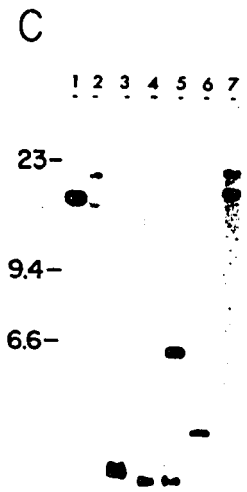
3.1.1. Identification of *nad1* gene segments.

DNA sequence analysis of the region extending approximately 5 kb downstream of the *rps13* gene has led to the identification of two *nad1* exons (retrospectively named *nad1*b and *nad1*c; Figure 2). *Nad1*b is located 946 bp downstream of the *rps13* gene (Bonen, 1987). These two exons encode 28 and 64 amino acids, respectively, which correspond to the central one-third of the NAD1 polypeptide. They are 96% identical at the DNA level to watermelon counterparts which have been found on an EcoRI DNA fragment of 4.5 kb (Stern et al. 1986) and share 51% derived amino acid sequence identity with the human

Figure 2. Restriction map and Southern hybridization analysis of the wheat mitochondrial *nad1* gene. **A.** The restriction map shows the four wheat mitochondrial genomic regions containing *nad1* coding segments (solid boxes). The *rrn18, rrn5, atp6, rps13* and *mat-r* genes are indicated by open bars. All BamHI (B) and SalI (S) restriction sites are shown. Only EcoRI (E), HaeIII (Ha), HindIII (H), XbaI (Xb) and XhoII (Xh) restriction sites that delimit the probes used in panel **C** are shown. Probes were labelled by second strand synthesis of M13 cloned DNA (arrows), or random priming on isolated restriction fragments (double arrows). Vertical arrows denote the 3' ends of recombinational repeat elements located upstream of *nad1*a and *nad1*b, respectively. **B.** Hybridization experiments using exon-specific oligonucleotides as probes were performed on EcoRI, BamHI and HindIII restriction digests of wheat mtDNA. **C.** Hybridization experiments using probes from *nad1* coding and flanking sequences were performed on SalI restriction digests of wheat mtDNA. The lane numbers correspond to the probes shown in panel **A.** The size of markers is indicated in kilobases.


A
B
S
Xb
B
rrn18
rrn5
nad1a
1
H
Ha
B
B
B
atp6
rps13
2
nad1b
nad1c
B
S
Xb
Xh
H
S
B
3
4
nad1d
5
B
S
B
E
Xb
B
6
mat-r
7
nad1e
1 kb


B
E B H E B H E B H E B H E B H
10-
5-
2-
Oligo probe : A1 B1 C1 D1 E1

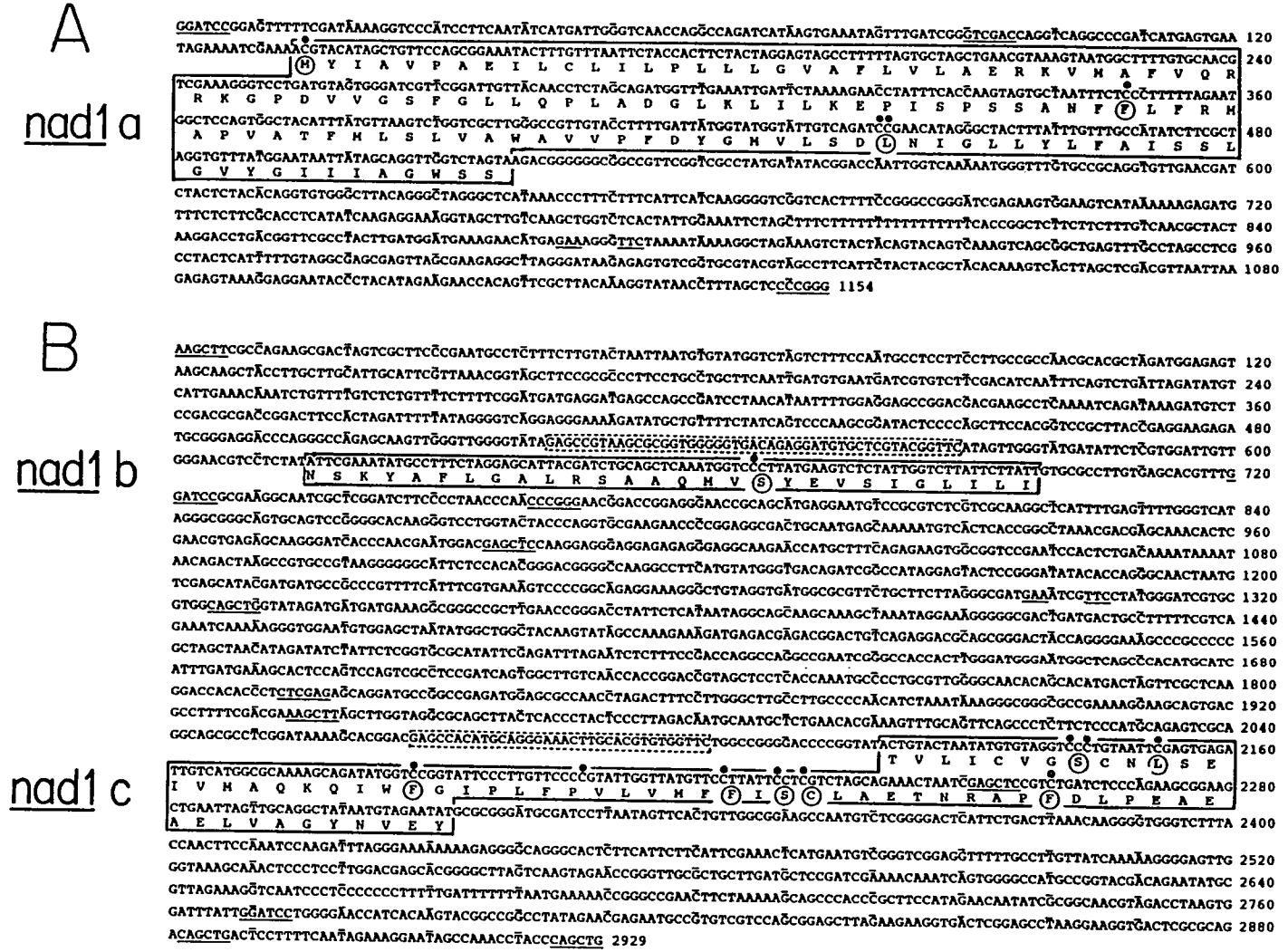

C
1 2 3 4 5 6 7
23-
9.4-
6.6-

ND1 counterpart (Anderson et al. 1981). The wheat *nad1*b and *nad1*c exons are separated by a group II intron of 1422 bp (Figure 2 and 3B, see also section 3.6) that is approximately 80% identical to the corresponding watermelon sequences. The sequence homology with watermelon, however, abruptly ends 1.2 kb downstream of *nad1*c, and no sequence homology is observed for at least 1 kb further downstream (Appendix IB). This suggests that the 3' end of the *nad1* gene is unlikely to be encoded in this region (section 3.1.3.).

Because the *atp6* and *rps13* genes are located less than 3 kb upstream of *nad1*b sequences in wheat, and because translation in three frames of this region did not reveal any ORF with significant homology to the N-terminal portion of the ND1 polypeptide from other organisms, the following approaches were undertaken to examine whether the 5' *nad1* coding sequences might be located at another site in the wheat mitochondrial genome. Heterologous hybridizations using a probe from the *C. reinhardtii* ND1 gene (Boer and Gray, 1988a) were unsuccessful. In a second approach, a synthetic oligomer that is complementary to *nad1*b (oligomer B2; note that, through this work, oligomers are designated according to the exon or region of the *nad1* gene that they specify), was used as primer on wheat mtRNA template in reverse transcriptase dideoxy sequencing experiments. Approximately equimolar amounts of two sequences were observed (Figure 4A), one corresponding to an unspliced RNA and the second, to a processed RNA species.

This information was used for the synthesis of another oligomer (oligo A1) which served as a hybridization probe on Southern blots of restricted wheat mtDNA (Figure 2B). The putative 5' *nad1* coding region was obtained from a partial

Figure 3. Nucleotide sequence of the wheat *nad1* gene segments and flanking regions: (**A**) *nad1*a, (**B**) *nad1*b-*nad1*c, (**C**) *nad1*d and (**D**) *nad1*e (boxes). Amino acid sequences were derived using the universal genetic code, but altered amino acids (circled) are shown at positions of C-to-U RNA editing sites (dots; see section 3.3). Group II intronic domain V sequences are indicated by dashed boxes. Solid arrowheads show the breakpoints in homology with broad bean (Wahleithner et al. 1990) and *Oenothera* (Wissinger et al. 1991) sequences. Restriction sites: HindIII, BamHI, SstI, SalI, EcoRI, PvuII, XhoI, SmaI, XbaI and XmnI are underlined in the sequence. The HindIII site at the 5' end of the sequence shown in panel **B** corresponds to the 3' HindIII site (position 680) in Figure 2 of Bonen (1987). The EMBL data bank accession numbers for these sequences are X57965-X57968.


A
nad1 a
B
nad1 b
nad1 c

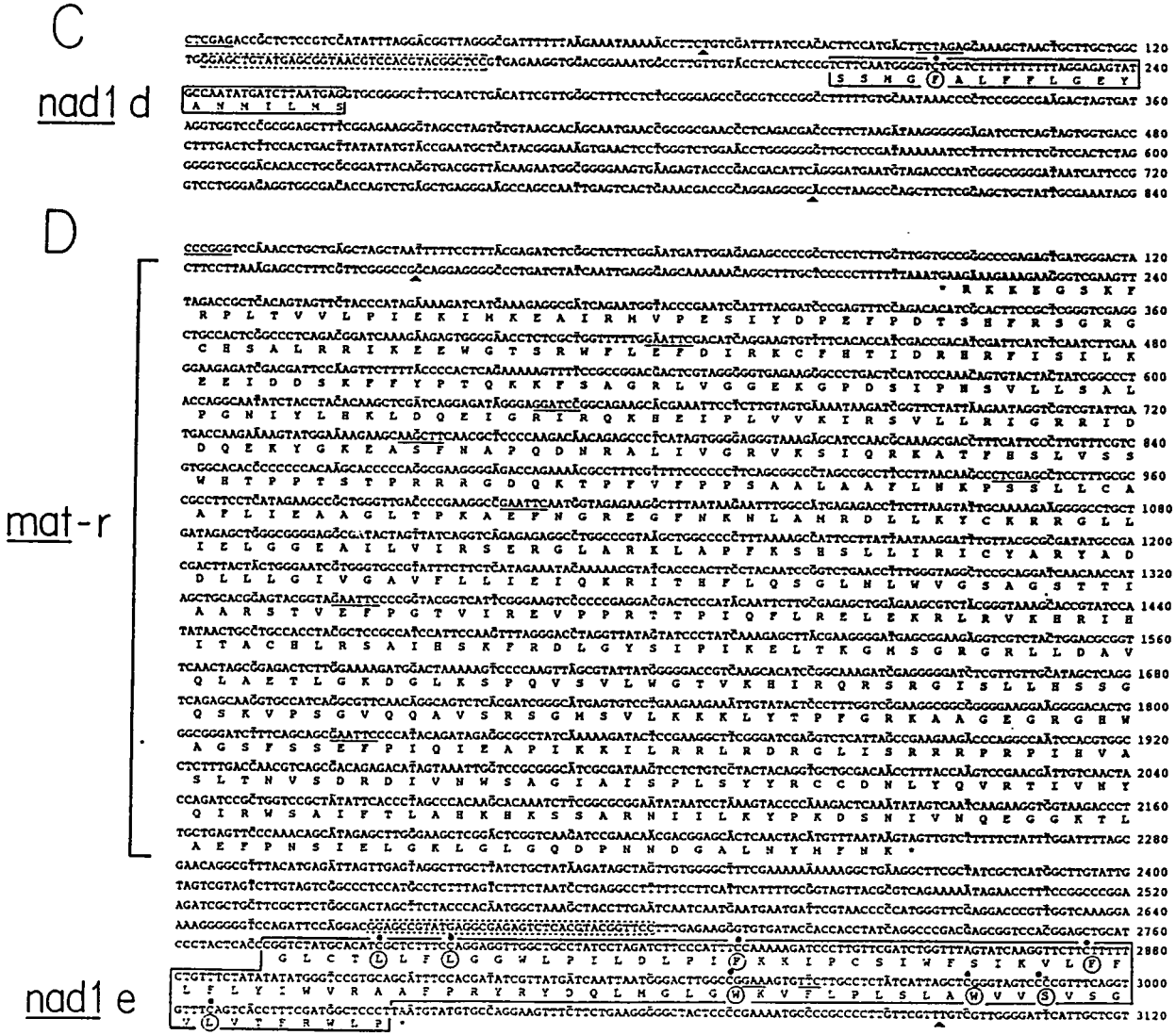
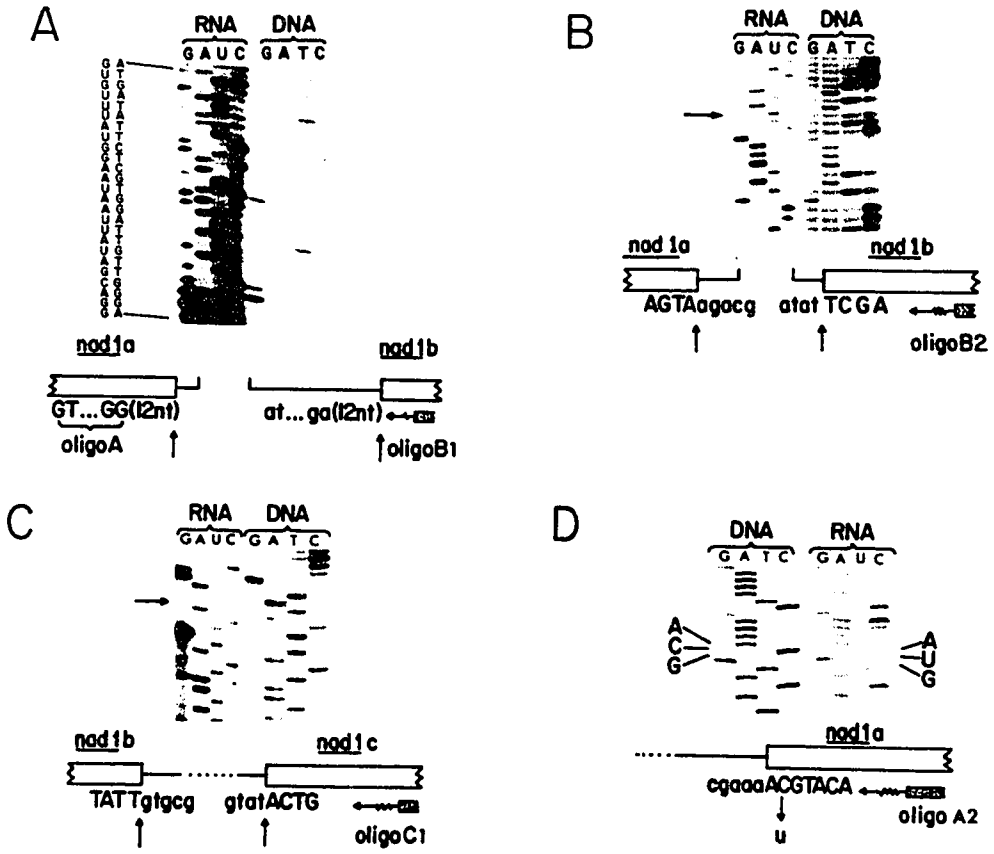

C
nad1 d
D
mat-r
nad1 e

Figure 4. Autoradiographs of direct cDNA-sequencing gels across *nad1* exon junctions and around the initiation codon. Primer extension experiments using reverse transcriptase dideoxynucleotide sequencing were performed directly on *nad1* transcripts (**A**) upstream of *nad1*b for the identification of *nad1*a (oligomer B2, not B1 as shown), (**B**) upstream of *nad1*b for the identification of the *nad1*a/*nad1*b junction (oligomer B1, not B2 as shown), (**C**) upstream of *nad1*c for the identification of the *nad1*b/*nad1*c junction (oligomer C1) and (**D**) upstream of *nad1*a for the identification of the initiation codon (oligomer A2). In panels **A**, **B** and **C**, arrows indicate exon junctions. Sequences using the same primers on cloned genomic DNA templates are shown within each panel.


A
RNA
DNA
G A U C
G A T C
nod1a
nod1b
GT... GG(12nt)
at... ga(12nt)
oligoA
oligoB1
B
RNA
DNA
G A U C
G A T C
nod1a
nod1b
AGTAagacg
atat TCGA
oligoB2
C
RNA
DNA
G A U C
G A T C
nod1b
nod1c
TATTgtgcg
gtatACTG
oligoC1
D
DNA
RNA
G A T C
G A U C
A
C
G
A
U
G
nod1a
cgaaaACGTACA
oligo A2
u

plasmid clone bank of wheat mitochondrial EcoRI and BamHI restriction fragments of approximately 15 kb and 2.2 kb, respectively. DNA sequence analysis of the hybridizing region confirmed the presence of a reading frame for the amino-terminal portion of an NAD1 polypeptide (Figure 3A). This segment (named *nad1*a) contains 128 codons and shows 50% amino acid identity with its human mitochondrial ND1 counterpart (Anderson et al. 1981). This level of conservation is in the range expected for comparisons of mitochondrial proteins between plants and animals (Dawson et al. 1984), which supports the view that these sequences are functional.

The 3' *nad1* coding segments (designated *nad1*d and *nad1*e) were identified by heterologous hybridization experiments using synthetic oligomers D1 and E2 which were designed from published plant *nad1*-homologous segments, namely, the broad bean *nad1* sub-terminal exon (Wahleithner et al. 1990) and the watermelon ORF36 (Stern et al. 1986), respectively. These experiments led to the identification and cloning of BamHI and HindIII fragments of approximately 5.8 and 8.0 kb containing *nad1*d and *nad1*e, respectively. They encode 20 and 85 codons and share 98 and 92% nucleotide sequence identity with their broad bean counterparts, respectively. Together, they correspond approximately to the 3' one-third of the *nad1* gene.

In summary, the wheat mitochondrial genome contains five *nad1* coding segments, designated *nad1*a-*nad1*e, that are 385, 83, 192, 59 and 256 bp in length, respectively. Together, they are capable of encoding a polypeptide of 325 amino acids. A comparison of the predicted wheat NAD1 polypeptide sequence with other organisms, taking into account RNA editing, will be presented in section 3.4.

3.1.2. Southern blot hybridizations.

Southern blot analysis has been carried out for the four regions encoding *nad1* gene segments to determine whether they are present as single copies in the wheat mitochondrial genome. In these experiments, end-labelled oligonucleotides specifying the five *nad1* exons were used as probes on EcoRI, BamHI, HindIII restriction digests of wheat mtDNA (Figure 2B), as well as labelled DNA fragments on SalI and XhoI restriction digests (Figure 2C and data not shown). For each oligomer, only one hybridizing fragment was seen for at least two different restriction enzymes. Within the limitations of this approach, this indicates that each exon is present as a single genomic copy in wheat mitochondria, and precludes the possibility of additional copies that could be located on short DNA restriction fragments or the presence of smaller exons. Certain lanes contain two hybridizing fragments that are due to the closeness of *nad1* segments to recombinational repeats containing the *rrn18-rrn5* genes, the *atp6* gene and the repeat "9" (see Appendix II, Lejeune and Quétier, 1988) in the vicinity of the *nad1*a, *nad1*b-*nad1*c and *nad1*e regions, respectively (Figure 2B, oligos A1 and B1; Figure 2C, lanes 2 and 7).

These data were used to establish the minimum distances that separate the *nad1* coding sequences in wheat mtDNA. Except for *nad1*b and *nad1*c, the restriction maps surrounding the different regions could not be linked. *Nad1*a is located a minimum of 20 kb away from *nad1*b and, moreover, at least two genes, namely *atp6* and *rps13*, are located between them (Figure 2A; Bonen, 1987; Bonen and Bird, 1988). These experiments also show that *nad1*c and *nad1*d are separated by at least 7 kb. Similarly, *nad1*d and *nad1*e are at least 12 kb apart (Figure

2). Consequently, the five wheat *nad1* coding segments are scattered over four distant sites and the distance between *nad1*a and *nad1*e is a minimum of 40 kb. If one also adds the size of the SalI fragments that contain *nad1* coding segments and the flanking ones that have been found by hybridization, the minimum distance between *nad1*a and *nad1*e is about 60 kb.

To examine further the *nad1* gene organization in the wheat mitochondrial genome, Southern blot hybridizations were performed on SalI restriction digests of wheat mtDNA (Figure 2C). These results were correlated with the 430 kb SalI restriction map (Lejeune and Quétier, 1988). Based on the sizes of hybridizing SalI fragments, the *nad1*a, *nad1*b-*nad1*c and *nad1*e exons are located on fragments H1, I and F1, which are approximately 16.8, 15.5 and 19.1 kb in size, respectively (lanes 1, 2 and 7). The *nad1*e upstream region contains a SalI fragment of 4.0 kb that can be assigned to fragment V (lane 6). The probes from the *nad1*d region hybridize to three adjacent SalI fragments of approximately 3.6, 3.2, and 6.0 kb that are tentatively assigned to W, X or Y and S, respectively (lanes 3, 4 and 5). These fragments, however, are distant on the physical map as proposed by Lejeune and Quétier (1988); thus, it remains somewhat uncertain where exactly *nad1*d is located (Appendix II). According to this map, *nad1*a, *nad1*b-*nad1*c and *nad1*e regions are transcribed from the same strand. Distances of more than 200 kb are predicted between *nad1*a and *nad1*e and, furthermore, numerous mitochondrial genes are anticipated between *nad1* segments. The large distances between the *nad1* gene pieces that are predicted by the physical map of the wheat mitochondrial genome could be reduced by recombination events, but, as indicated by hybridization data, such distances could not be less than 60 kb.

3.2. ANALYSIS OF SPLICED *nad1* TRANSCRIPTS.

3.2.1. Direct cDNA sequencing analysis of *nad1* transcripts.

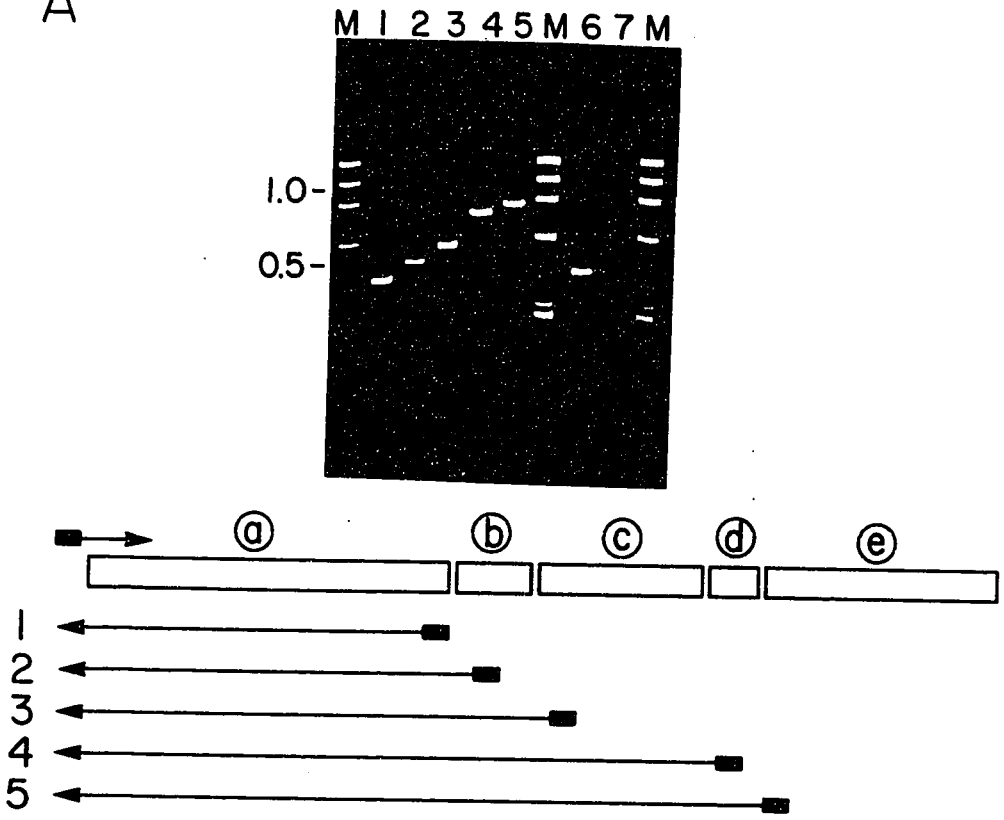
After the identification of *nad1* gene pieces that are distantly separated, it was important to address the issue of whether these segments are linked at the RNA level. Two experimental approaches have been used to identify the precise *nad1* exon junctions and to determine if they are correctly spliced. Direct cDNA sequence analysis of wheat *nad1* transcripts, as used in the identification of *nad1*a (Figure 4A), indicates that *nad1*b precursor transcripts and *nad1*a-*nad1*b spliced transcripts are approximately equally abundant (Figure 4A and B). Similarly, direct cDNA sequence analysis upstream of *nad1*c revealed a mixture of precursor and spliced transcripts (panel C). These experiments indicate that *nad1*a, *nad1*b and *nad1*c are correctly linked at the RNA level to generate the *nad1* reading frame (Figure 3). Direct cDNA sequencing across the *nad1*c/d junction was unclear (see section 3.3.5) and across *nad1*d/e junction it was unsuccessful (not shown); therefore, those proposed junctions are based on sequencing of cloned PCR-amplified cDNAs.

3.2.2. Analysis of PCR-amplified reverse transcripts.

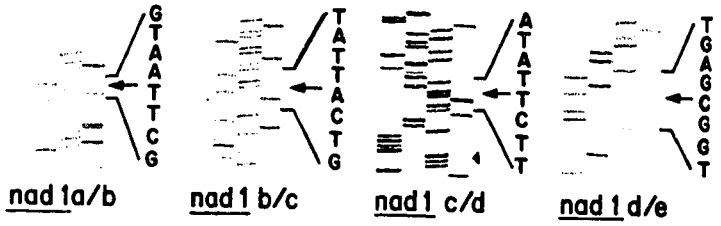
The results obtained by direct RNA sequencing with reverse transcriptase (Figure 4A, B and C) were corroborated by the analysis of PCR-amplified reverse transcripts. Oligomers A1, B1, C1, D1 and E1, which are specific to the respective *nad1* exons, were used as primers on wheat mtRNA and the resulting cDNAs were PCR-amplified using oligomer A3 as the second primer (Figure 5, schematic). The products, shown in Figure 5A (lanes 1 to 5), correspond approximately to the

Figure 5. Characterization of PCR-amplified *nad1* cDNAs. **A.** PCR-amplified products of five *nad1* cDNAs (lanes 1-5) synthesized using nested, exon-specific primers (oligos A1, B1, C1, D1 and E1) with oligomer A3 shown in the schematic. Wheat mtDNA was used as template with oligomer sets A1/A3 (lane 6) and B1/A3 (lane 7). **B.** Sequence analysis of cloned PCR products at the four *nad1* exon junctions (arrows). The arrowhead indicates an edited U at position 215 within *nad1*d (Figure 3C). Lane order: G,A,T,C.

A


M 1 2 3 4 5 M 6 7 M
1.0-
0.5-
a
b
c
d
e
1
2
3
4
5

B


G T A A T T C G
T A T T A C T G
A T A T T C T T
T G A G C G G T
nad 1a/b
nad 1 b/c
nad 1 c/d
nad 1 d/e

predicted sizes of 422, 494, 578, 762 and 825 bp, respectively. There is no evidence for the presence of transcripts containing incorrectly spliced *nad1* segments since, in each case, single PCR-generated products were seen (Figure 5A).

It could be argued that *nad1*a and *nad1*b may be close together in low level subgenomic circles (sublimons) which would not have been detected by hybridization analysis. When wheat mtDNA, rather than RNA (or cDNA), was used with oligomers A1 and A3 in PCR experiments, the predicted product (cf. 422 bp) was seen (Figure 5A, lane 6). In contrast, when oligomers B1 and A3 were used as primers, no product was seen (Figure 5A, lane 7). Under similar conditions, with oligomers C1 and B3, PCR products of approximately 1.7 kb were easily obtained. Therefore, with oligomers B1 and A3, considering that *nad1*a is approximately 0.4 kb long, any genomic arrangement in which *nad1*a and *nad1*b are separated by less than 1.3 kb should have been detected. These data, in conjunction with Southern hybridization data (Figure 2), are consistent with the view that *nad1*a and *nad1*b are not closely linked in the wheat mitochondrial genome.

The sequence analysis of cloned cDNAs, which were derived from PCR products seen in lane 5 (Figure 5A), allowed the assignment of precise exon boundaries (Figure 5B) and confirmed the results obtained by direct cDNA sequencing described for the *nad1*a/b and *nad1*b/c junctions. It should be noted that two-nucleotide redundancies occur at the *nad1*a/b and *nad1*b/c borders (Figure 3). The sites proposed here are based on group II intron consensus sequences (see section 3.6.). Taken together, these observations clearly indicate

that the correct reading frame for NAD1 protein is generated at the RNA level.

3.3. RNA EDITING IN *nad1* TRANSCRIPTS.

RNA editing which was recently discovered in plant mitochondria (Gualberto et al. 1989; Covello and Gray, 1989; Hiesel et al. 1989) alters the coding information that could be deduced from genomic DNA sequences, and amino acid sequences can only be accurately derived from RNA or cDNA sequence information. Using the same type of experiments described in section 3.2, *nad1* RNA sequences were inferred by direct cDNA sequencing and by the analysis of cloned PCR-amplified cDNAs. These sequences were then compared to mtDNA genomic sequences to identify the positions of RNA editing sites in the wheat *nad1* coding sequences.

3.3.1. Identification of the initiation codon of the *nad1* gene.

The deduced amino acid sequence from the wheat *nad1*a region shows 50% similarity to its human ND1 counterpart (section 3.1); however, it lacks an ATG triplet in the vicinity of the predicted start codon. Initiation of translation theoretically could involve (1) a non-conventional initiator, (2) additional splicing events to generate the extreme 5' coding sequences possessing an AUG codon, or (3) RNA editing to create an AUG codon. To distinguish among these possibilities, the synthetic oligomer A2 was used in direct cDNA sequencing experiments. The result, shown in Figure 4D, demonstrates the presence of RNA editing that changes the ACG codon at position 134-136 (Figure 3) to AUG in approximately 50% of the RNA molecules. This C-to-U conversion is typical of

RNA editing in plant mitochondria and, in this case, the RNA editing event is in fact obligatory for protein synthesis to proceed.

3.3.2. Distribution of RNA editing sites in the wheat mitochondrial *nad1* coding regions.

A comprehensive analysis of RNA editing sites within the wheat mitochondrial *nad1* coding region of 975 bp was carried out. PCR amplification products of cDNAs derived from fully-spliced transcripts were cloned, sequenced and compared to their genomic counterparts[1]. A total of 23 C-to-U RNA editing sites were observed, with each of the five *nad1* exons containing at least one editing site, namely 4, 1, 9, 1 and 8 in *nad1*a-e, respectively (Figure 3, Figure 6, Table 2). Two codons (103 and 164) are modified at two positions, and in all but one of the 21 different codons involved, RNA editing is predicted to result in amino acid changes (Table 2).

The RNA editing sites are asymmetrically distributed with approximately 40% being located within a 120 nt stretch in *nad1*c (Figure 6). In contrast, *nad1*a, which is 385 nt in length, contains only four sites, including the initiation codon. Consequently, only 2% of the codons are affected in *nad1*a compared to 13% in *nad1*c.

---

[1] The cDNAs derived from fully spliced transcripts were obtained by primer extension on total mtRNA with the oligomer E3 (located downstream of *nad1*e), and amplified with oligomers E3 and A3 (located upstream of 5' exon *nad1*a). The PCR products were restricted and cloned into appropriate M13 vectors. Nine clones covered the region from EcoRV to SalI site, 8 from BglII to SstI and 3 from BglII to EcoRII. The relative positions of these sites are shown in Figure 6 and Appendix VI.

Figure 6. Schematic representation of RNA editing sites in plant mitochondrial *nad1* coding regions. The schematic shows the five *nad1* exons (boxes) in wheat, *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991), with the vertical bars indicating the positions of RNA editing sites. Dashed lines represent sites in common. The solid triangle indicates the position of a stretch of 10 Ts in *nad1*d (see section 3.3.5.). Restriction sites: BglII (Bg), EcoRII (Ec), EcoRV (Ev), SalI (S), and SstI (Ss).

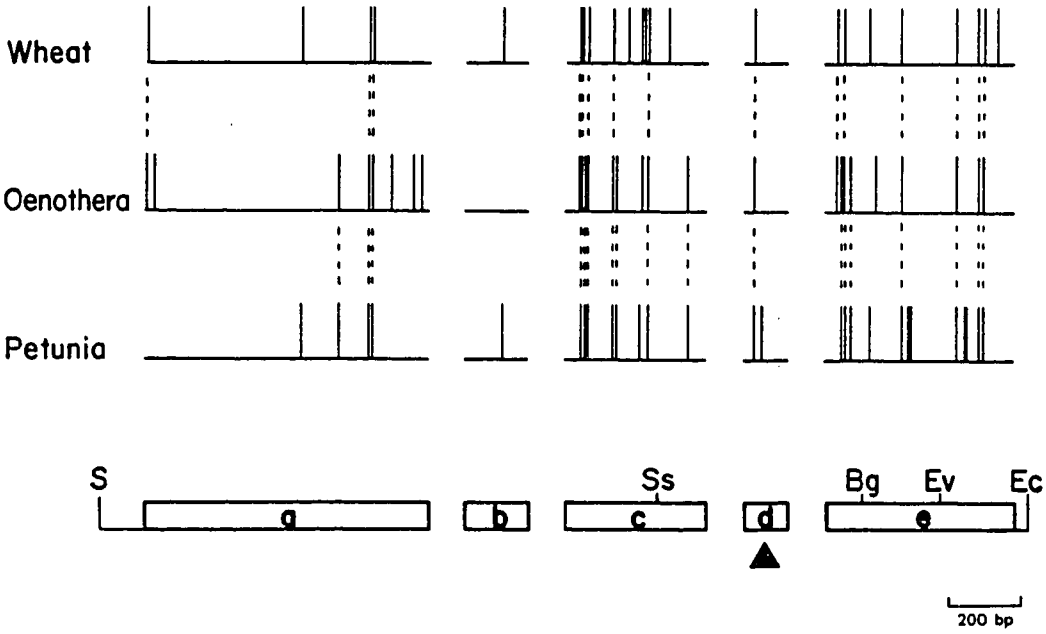

Wheat
Oenothera
Petunia
S
Ss
Bg
Ev
Ec
a
b
c
d
e
200 bp

Table 2. Compilation of *nad1* RNA editing sites in wheat, *Oenothera*[a] and petunia[b].

Codon	Nucleotide position	Wheat	Oenothera	Petunia	Amino acid changes	Shared sites[c]
nad1a						
1	2	ACG-AUG	ACG-AUG	ATG	T-M	-
14	40	CTA	CTA-UUA	CTA	L	-
72	215	TCC-UUC	TCC	TCC-UUC	S-F	-
89	265	TGG	CGG-UGG	CGG-UGG	R-W	-
103	307-8	CCG-UUG	CCG-UUG	CCG-UUG	P-L	+, +
112	336	GCC	GCC-GCU	GCC	A	-
123	368	ATA	ACA-AUA	ATA	T-I	-
126	376	TGG	CGG-UGG	TGG	R-W	-
nad1b						
146	436	CCT-UCU	TCT	CCT-UCU	P-S	-
nad1c						
164	490-2	CCC-UCU	CCC-UCU	TCC-UCU	P-S	-, +
165	493	TGT	CGT-UGU	CGT-UGU	R-C	-
167	500[d]	TCG-UUG	TCG-UUG	TCG-UUG	S-L	+
179	536-7	TCC-UUC	TCC-UUU	TCC-UUU	S-F	+, -
185	555	CCC-CCU	CCT	CCC	P	-
190	570	TTC	TTC	TTC-UUU	F	-
191	571-3	CTT-UUU	TTC-UUU	TTT	L-F	-, -
193	577	CCT-UCU	TCT	TCT	P-S	-
194	580	CGT-UGU	CGT-UGU	CGT-UGU	R-C	+
203	608	TCT-UUU	TTT	TTT	S-F	-
212	635	TTA	TCA-UUA	TCA-UUA	S-L	-
nad1d						
225	674	TCT-UUU	TCT-UUU	TCT-UUU	S-F	+
228	683	TTT	TTT	TCT-UUU	S-F	-
nad1e						
245	734	TCG-UUG	TCG-UUG	TCG	S-L	-
247	740	TTT	TCT-UUU	TCT-UUU	S-F	-
248	743	CCA-CUA	CCA-CUA	CCA-CUA	P-L	+
252	754	CTG	CCG-CUG	CCG-CUG	P-L	-
260	779	TCC-UUC	TTC	TCC-UUC	S-F	-
263	789	ATC	ATC-AUU	ATC	I	-
275	823	CTT-UUU	CTC-UUC	CTT-UUU	L-F	+
278	834	TTT	TTC	TTC-UUU	F	-
279	835	CTA	CTA	CTA-UUA	L	-
300	898	CGG-UGG	CGG-UGG	CGG-UGG	R-W	+
303	909	TTC	TTC	TTC-UUU	F	-
306	916	CTA	CTA	CTA-UUA	L	-
310	928	CGG-UGG	CGG-UGG	CGG-UGG	R-W	+
313	937	CCC-UCC	CCC-UCC	CCC-UCC	P-S	+
318	953	TCA-UUA	TTA	TTA	S-L	-

[a] Wissinger et al. (1991). [b] Conklin et al. (1991). [c] "+" denotes sites that are common to the three plants. [d] Also shared with maize chloroplast *ndhA* gene (Maier et al. 1992).

3.3.3. Comparison of *nad1* RNA editing sites among wheat, *Oenothera* and petunia.

It is helpful at this stage to state the chronological identification of *nad1* homologous sequences in plant mitochondria. Short *nad1* gene sequences (mentioned in section 3.1) had been found in tobacco and maize (*nad1b*, Bland et al. 1986), in wheat (*nad1b*, Bonen, 1987), in watermelon (*nad1b-nad1c*, Stern et al. 1986) and more recently in broad bean (*nad1d-nad1e*, Wahleithner et al. 1990). Independent work by Brennicke's group in Germany, conducted in parallel with the present study in wheat (published in part, Chapdelaine and Bonen, 1991, Appendix VII), has led to the identification of the *nad1* gene in *Oenothera* (Wissinger et al. 1991). The *nad1* gene has also been characterized subsequently in petunia (Conklin et al. 1991) and in liverwort (Oda et al. 1992). In this section, the positions of the wheat *nad1* RNA editing sites are compared with those of two dicots, *Oenothera* and petunia (Figure 6 and Table 2).

Oenothera and petunia *nad1* coding sequences each have a total of 27 RNA editing sites and the number of predicted amino acid changes, 20 and 19 respectively, is similar to that of wheat because they have more silent sites. The two dicots share only a slightly higher number of sites (18) than either has in common with the monocot, wheat (15). There are 12 sites that are shared by all three plants, and only one is silent. Three of those sites occur within the two doubly-modified codons, and several of the shared sites are found within clusters, particularly at the 5' ends of *nad1c* and *nad1e* (Figure 6). Interestingly, one of the shared sites (Table 2, codon 167) has recently been shown to undergo C-to-U editing in the chloroplast *ndhA* counterpart in maize (Maier et al.

1992). It is worth noting that both the asymmetric distribution of RNA editing sites and their clustering are features common to all three plants (Figure 6), even though only half of the sites are in common.

There are additional regions in which editing occurs at close but not identical sites in different plants. For example, of the five editing sites within codons 190-194 of *nad1*c, only one is held in common by all three plants. There are two codons (72 and 245) in which editing is seen in two of the three plants. Their absence in the third plant is unexpected since it results in radical amino acid substitutions (Table 2).

3.3.4. Analysis of partially edited *nad1* cDNAs.

Although most RNA editing events are predicted to be crucial for generating transcripts capable of encoding functional proteins, partially edited transcripts have been observed for several plant mitochondrial genes (Schuster et al. 1991a; Bonnard et al. 1992). Among the cDNA clones derived from fully-spliced wheat *nad1* transcripts (see footnote page 52), only two clones lacked complete editing, namely at position 734 (codon 245) in both clones and at position 823 (codon 275) in one clone. Both sites are predicted to result in non-silent amino acid changes, but they are located in a poorly-conserved region of the gene. In addition, four cDNA

clones have been derived from precursor transcripts[2], and each clone was fully edited.

The high level of editing in *nad1* spliced transcripts contrasts with the results obtained by direct cDNA sequencing of the total population of *nad1*a transcripts in which only approximately half the molecules are edited at the initiation codon (Figure 4D). These observations suggest that most, if not all, editing events occur prior to the completion of splicing events. The level of RNA editing in wheat *nad1* transcripts also contrasts with that described in *Oenothera* and petunia *nad1* transcripts in which sites were often partially edited (cf. Wissinger et al. 1991; Conklin et al. 1991). The source of plant or tissue culture material, or the gene examined may influence the degree of editing observed. Incidentally, *nad5* coding sequences in wheat showed no partial editing (De Souza et al. 1992).

3.3.5. Other rare changes in *nad1* cDNA sequences.

Differences between the DNA and cDNA sequences that do not fall in the category of the C-to-U form of RNA editing were occasionally observed in *nad1* coding regions. A total of 20 positions were found out of more than 10 000 bp of *nad1* cDNA sequence examined and each was represented in only one cDNA clone. These changes involve 11 T-to-C, 2 C-to-U, 3 A-to-G, 1 A-to-U, 2 A insertions, and one U deletion (Appendix VI).

---

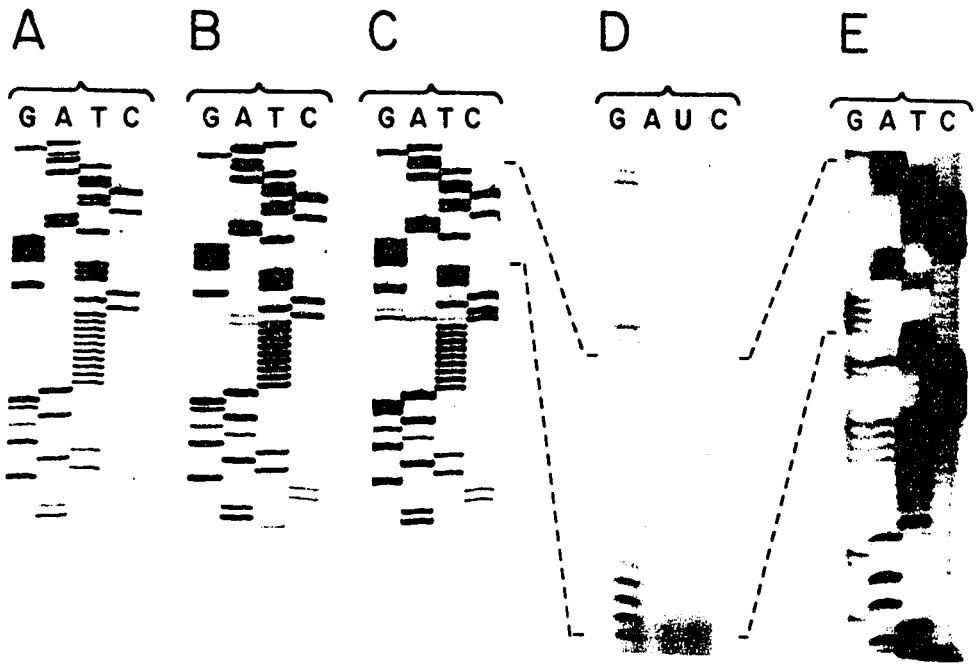
[2] Oligomer D1 was used in primer extension experiments on wheat mtRNA and cDNAs were amplified with oligomers D1 and B3. Therefore, initial transcripts contained at least *nad1*b-d and possibly *nad1*e, but not *nad1*a. The cDNAs were amplified and the PCR products were digested with the restriction enzymes SstI and XhoI which cut within *nad1*c and oligomer B3, respectively. Oligomer B3 is located immediately upstream of domain V of the *nad1*a/b intron. The position of the SstI site is shown in Figures 3 and 6.

Reverse transcriptase and Taq DNA polymerase have combined error frequencies estimated to be less than 4×10^{-4} (Kunkel and Eckert, 1989) which is considerably lower than the rate of changes observed here (approximately 2×10^{-3}). Such low-frequency differences have been reported for other plant mitochondrial mRNAs (reviewed by Schuster et al. 1991a; Bonnard et al. 1992), except for a T-to-C modification which was found in all twelve *coxIII* cDNAs examined in wheat (Gualberto et al. 1990). The biological meaning of these changes is uncertain since they generally do not seem to increase the predicted polypeptide sequence similarity with other organisms.

Another unexpected difference between the *nad1* RNA and DNA sequences occurred within a stretch of 10 Ts in *nad1d* (positions 680-689 in Table 2, triangle in Figure 6). Of the 17 cDNA clones examined in this region, 7 clones are identical to the genomic DNA (Figure 7A), 8 clones contain only 9 Ts (Figure 7B), one clone contains 11 Ts instead of 10 (not shown) and one clone contains a T-to-C conversion (Figure 7C). The latter T-to-C conversion results in a non-conservative change in codon 227 from leucine to proline.

These differences could be due to errors made by reverse transcriptase or Taq DNA polymerase *in vitro*, or alternatively, aberrant RNA editing intermediates which occasionally lead to C rather than U at the second position of codon 227. When wheat mtRNA was used as template for direct cDNA sequencing with reverse transcriptase, a double sequence was seen in the *nad1d* region after the T-stretch (Figure 7E) suggesting that a mixture of cDNA molecules differing in length by one nucleotide within the T-stretch is present. These observations are consistent with the results of the

Figure 7. Sequence analysis of a T-stretch region in *nad1*d. **A-C.** Three different cloned cDNAs were sequenced using Sequenase enzyme (USB): **A** no deletion (10 Ts), **B** one nucleotide deletion (9 Ts) and **C** a T is changed to a C (9 Ts + 1 C). **D.** Wheat mtRNA template was sequenced using AMV reverse transcriptase (Life Sciences) and oligo D1. **E.** Cloned DNA containing 10 Ts was sequenced using AMV reverse transcriptase.


A
B
C
D
E
G A T C
G A T C
G A T C
G A U C
G A T C

cloned cDNA sequence analysis. However, when reverse transcriptase was used on cloned DNA template containing the stretch of 10 Ts, again a double sequence was seen after but not before the T-stretch (Figure 7E). This result suggests that the reverse transcriptase has difficulties in faithfully copying this region *in vitro*. It is worth noting, however, that a stretch of 8 Ts located in *nad1e* (position 821-828 in Table 2) was never found to have deletions in 12 cDNA clones examined.

Thus, ten of the 17 cDNA clones examined differed from the genomic sequence in the stretch of 10 Ts in *nad1d*. Nucleotide deletions within strings of Ts in the wheat *nad3* and *rps12* genes have been observed five times in a string of 9 Ts as well as in other strings of Ts (Gualberto et al. 1991). These deletions coincide with the position of RNA editing sites and were suggested to represent RNA editing intermediates, which would support the view that the RNA editing mechanism involves nucleotide replacement rather than modification (see also Chasan, 1991; Bonnard et al. 1992). In the *Oenothera cob* gene (Schuster et al. 1991b), one clone out of 48 showed a deletion in a stretch of 7 Ts. Although the results using DNA template controls in reverse transcriptase experiments rather point to *in vitro* errors being made in homopolymeric regions, they do not rule out nucleotide replacement as a plausible mechanism for RNA editing in plant mitochondria.

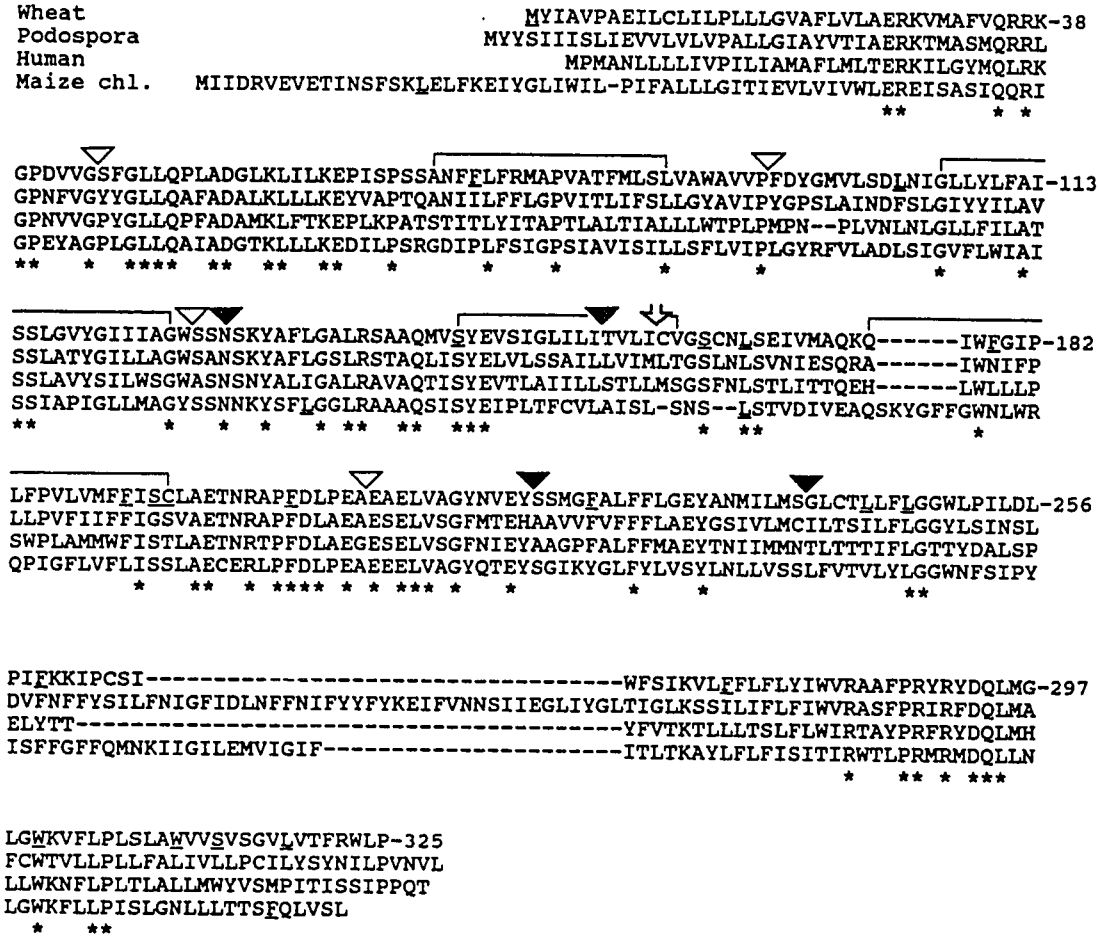
3.4. ANALYSIS OF AMINO ACID SEQUENCES DEDUCED FROM THE *nad1* AND *mat-r* GENES.

3.4.1. The *nad1* gene.

The five *nad1* exons in wheat are predicted to encode 128, 28, 64, 20 and 85 amino acids, respectively (Figure 3, Figure 8), taking into account exon junctions (section 3.2) and RNA editing to generate the initiation codon (sections 3.3.1). The wheat NAD1 amino acid sequence, as deduced from the edited cDNA sequence, shares 47, 45 and 38% identity with its counterparts in *Podospora* (Cummings et al. 1988), human (Anderson et al. 1981) and maize chloroplast (Maier et al. 1992), respectively (Figure 8). Even without considering RNA editing, the plant mitochondrial *nad1* genes are remarkably highly conserved at the DNA level. For example, the wheat and *Oenothera* coding sequences share 96% nt identity and have no deletion/insertions, compared to 93% identity for the *coxI* gene (Kemmerer et al. 1989) which has traditionally been regarded as the most conservative mitochondrial protein-coding gene (cf. Bibb et al. 1981). Identity increases to 97% or 98%, when the *nad1* sequences are compared at the RNA or amino acid levels, respectively. The wheat *nad1* cDNA sequence is 79% identical to the liverwort *nad1* gene (which does not appear to require RNA editing, Oda et al. 1992), and identity is 84% at the amino acid level between wheat and liverwort predicted NAD1 sequences.

The NAD1 protein is rich in hydrophobic residues which presumably reflects the interactions of the NAD1 polypeptide with the mitochondrial inner membrane. Three calculated transmembrane helices (positions 69-86, 106-125, 146-162) and one surface helix (176-194, Weiss et al. 1991) are flanked by

Figure 8. Alignment of the predicted NAD1 amino acid sequences. The sequences were derived from mitochondrial genes in wheat (cDNA, this work), *Podospora anserina* (Cummings et al. 1988) and human (Anderson et al. 1981); and from maize chloroplast (cDNA, Maier et al. 1992). The position of introns in wheat (solid triangle), *P. anserina* (open triangle) and maize chloroplast (open arrow) are shown. Amino acid positions that require RNA editing in plant mitochondria and chloroplast are underlined in wheat and maize. Asterisks denote residues that are conserved in all four sequences. Calculated transmembrane and surface helices are indicated by brackets (as shown by Weiss et al. 1991).


Wheat MYIAVPAEILCLILPLLLGVAFLVLAERKVMAFVQRRK-38
Podospora MYYSIIISLIEVVLVLVPALLGIAYVTIAERKTMASMQRRL
Human MPMANLLLLIVPILIAMAFLMLTERKILGYMQLRK
Maize chl. MIIDRVEVETINSFSKLELFKEIYGLIWIL-PIFALLLGITIEVLVIVWLEREISASIQQRI
GPDVVGSFGLLQPLADGLKLILKEPISPSSANFFLFRMAPVATFMLSLVAWAVVPFDYGMVLSDLNIGLLYLFAI-113
GPNFVGYYGLLQAFADALKLLLKEYVAPTQANIILFFLGPVITLIFSLLGYAVIPYGPSLAINDFSLGIYYILAV
GPNVVGPYGLLQPFADAMKLFTKEPLKPATSTITLYITAPTLALTIALLLWTPLPMPN--PLVNLNLGLLFILAT
GPEYAGPLGLLQAIADGTKLLLKEDILPSRGDIPLFSIGPSIAVISILLSFLVIPLGYRFVLADLSIGVFLWIAI
SSLGVYGIIIAGWSSNSKYAFLGALRSAAQMVSYEVSIGLILITVLICVGSCNLSEIVMAQKQ------IWFGIP-182
SSLATYGILLAGWSANSKYAFLGSLRSTAQLISYELVLSSAILLVIMLTGSLNLSVNIESQRA------IWNIFP
SSLAVYSILWSGWASNSNYALIGALRAVAQTISYEVTLAIILLSTLLMSGSFNLSTLITTQEH------LWLLLP
SSIAPIGLLMAGYSSNNKYSFLGGLRAAAQSISYEIPLTFCVLAISL-SNS--LSTVDIVEAQSKYGFFGWNLWR
LFPVLVMFFISCLAETNRAPFDLPEAEAELVAGYNVEYSSMGFALFFLGEYANMILMSGLCTLLFLGGWLPILDL-256
LLPVFIIFFIGSVAETNRAPFDLAEAESELVSGFMTEHAAVVFVFFFLAEYGSIVLMCILTSILFLGGYLSINSL
SWPLAMMWFISTLAETNRTPFDLAEGESELVSGFNIEYAAGPFALFFMAEYTNIIMMNTLTTTIFLGTTYDALSP
QPIGFLVFLISSLAECERLPFDLPEAEEELVAGYQTEYSGIKYGLFYLVSYLNLLVSSLFVTVLYLGGWNFSIPY
PIFKKIPCSI----------WFSIKVLFFLFLYIWVRAAFPRYRYDQLMG-297
DVFNFFYSILFNIGFIDLNFFNIFYYFYKEIFVNNSIIEGLIYGLTIGLKSSILIFLFIWVRASFPRIRFDQLMA
ELYTT----------YFVTKTLLLTSLFLWIRTAYPRFRYDQLMH
ISFFGFFQMNKIIGILEMVIGIF----------ITLTKAYLFLFISITIRWTLPRMRMDQLLN
LGWKVFLPLSLAWVVSVSGVLVTFRWLP-325
FCWTVLLPLLFALIVLLPCILYSYNILPVNVL
LLWKNFLPLTLALLMWYVSMPITISSIPPQT
LGWKFLLPISLGNLLLTTSFQLVSL

well-conserved hydrophilic domains (Figure 8). In addition, a block of conserved residues close to the C-terminus is preceded by a region that is known to be variable in length (Burger and Werner, 1985).

The positions of the intervening sequences in the wheat *nad1* gene do not correspond to the locations of introns identified in evolutionarily-related counterparts (Figure 8). Some of these introns also belong to different classes; the plant mitochondrial and chloroplast ones belong to group II (section 3.6, Matsubayashi et al. 1987), and those from *Podospora* to group I (Cummings et al. 1988). Interestingly, the third of the four group I introns in the *Podospora* mitochondrial ND1 gene is close to the position of the wheat *nad1*a/b intron, and the single chloroplast *ndhA* intron is near the site of the wheat *nad1*b/c intron (Figure 8). It is worth mentioning that these introns are located near the end of the second and the third transmembrane domains discussed above (Weiss et al. 1991). The close proximity of these presumably unrelated introns may reflect preferential sites of insertion. While mobility of group I introns is well documented, less is known about how group II introns are inserted during evolution (Dujon, 1989).

Most of the 20 amino acid changes due to RNA editing within the wheat *nad1* coding region are predicted to increase the similarity with counterparts from other organisms and from the chloroplast (Figure 8). Five RNA editing sites convert the predicted amino acids to identity in these four organisms, including a serine residue that corresponds to an RNA editing site in maize chloroplast *ndhA* transcripts (section 3.3, Maier et al. 1992). Five other sites generate identity with at least one of the three other sequences. These observations highlight

the crucial role of RNA editing in the production of functional polypeptides.

3.4.2. The *mat-r* gene.

The region preceding the wheat *nad1*e exon contains an ORF of 678 codons (Figure 3D) that is very closely related to the 687-codon maturase-related (*mat-r*) ORF that was previously identified in broad bean (Wahleithner et al. 1990). The wheat and broad bean *mat-r* genes are located at the same genomic site (i.e. upstream of *nad1*e), and are 86% identical at the amino acid level compared to 90% between wheat and broad bean *nad1*d and *nad1*e exons. The pattern of nucleotide substitutions observed between these two *mat-r* genes (approximately 45% occurring in the third position) supports the view that these sequences are under functional constraint. Interestingly, the third position is occupied by T in only 18% of the codons. In contrast, 40% of codons in the *nad1* reading frame end by T, which is typical among plant mitochondrial genes (Dawson et al. 1984). The analysis of the broad bean *nad1*d-*nad1*e region led to similar conclusions (Wahleithner et al. 1990). Such differences in the codon usage of group II intronic ORFs has also been observed in fungal mitochondria (Michel and Lang, 1985; Cummings et al. 1990) but the underlying reasons as to the cause of these differences remain obscure. It has been suggested that the low frequency of codons ending with T might reflect a recent invasion of fungal mitochondrial genomes by these introns (Lambowitz, 1989).

Both wheat and broad bean *mat-r* genes are related to fungal mitochondrial group II intronic ORFs (cf. Lang et al. 1985) and possess six of the nine "invariant" amino acid residues characteristic of reverse transcriptases (Michel and

Lang, 1985; Wahleithner et al. 1990). Interestingly, the plant *mat-r* gene which lacks an initiation codon and homology to the N-terminal portion of the putative reverse transcriptase in wheat (Figure 3D), broad bean (Wahleithner et al. 1990) and *Oenothera* (Wissinger et al. 1991), does not contain any codon near its 5' terminus which could be converted to AUG by RNA editing. As yet, it is not known how translation of the *mat-r* reading frame is initiated.

3.5. *Nad1* TRANSCRIPT ANALYSIS.

The unusual organization of the wheat *nad1* gene raises fundamental questions about how gene pieces that are physically very far apart in the mitochondrial genome can give rise to functional mRNAs. This section will focus on an analysis of the size and the relative abundance of stable *nad1* transcripts, including precursor RNAs, mature RNAs and splicing by-products. Because *nad1* coding segments are present as single copies, they can be analyzed directly by Northern blot hybridization and S1 nuclease mapping. These two approaches have been taken in a partial characterization of stable *nad1* transcripts.

3.5.1. Northern blot hybridizations.

Several types of radio-labelled probes have been used for Northern blot hybridizations with wheat mtRNA (oligomers, cloned M13 DNAs and isolated DNA fragments, section 2.4). Among these experiments, end-labelled oligonucleotides gave the best resolution of hybridizing signals. It should be noted, however, that efficiency of labelling and/or hybridization seems to vary between different oligomers (see

Figure 9, lanes 5 and 6 for example). Consequently, lanes were exposed for autoradiography for different lengths of time.

3.5.1.1. Transcripts common to the five *nad1* exons.

Northern blot hybridizations were performed with a series of eight ^{32}P-end-labelled oligonucleotides specifying *nad1* exons, as well as *nad1*-flanking regions downstream of *nad1*a, *nad1*d and *nad1*e, and upstream of *nad1*b (Figure 9). All exon-specific oligomers and oligomer E3 hybridize to three major transcripts of approximately 1.5, 1.7 and 1.9 kb (lanes 2, 5, 6, 7 and 9), although they are not as clearly visible in lane 9. These RNA molecules appear to contain the five *nad1* exons and can potentially encode the 325 amino acid NAD1 protein. These three putative *nad1* mRNAs in wheat are larger than the single 1.2 kb mRNA found in *Oenothera* (Wissinger et al. 1991).

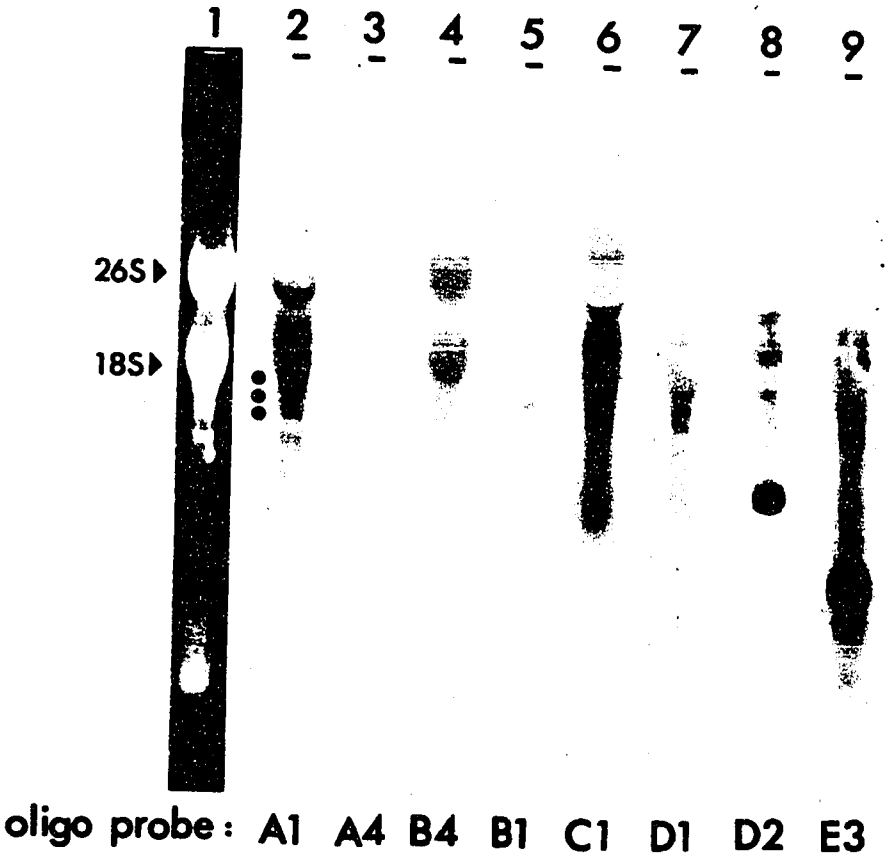
3.5.1.2. The *nad1*a region.

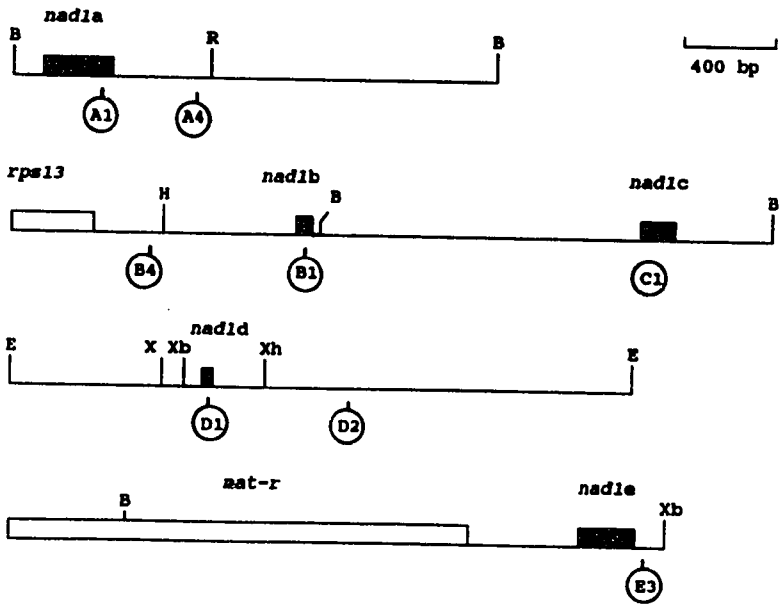
In addition to the three putative mRNAs, other transcripts of various abundances and sizes are observed for particular regions of the *nad1* gene and presumably correspond to precursors or splicing by-products. A transcript of approximately 3.2 kb is detected with both oligomers A1 and A4 (Figure 9, lanes 2 and 3), and likely corresponds to a precursor of the *nad1*a region. The 1.9 kb transcript seen in lane 3 may correspond to a splicing product of the *nad1*a/b intron, although it comigrates with one of the putative mRNAs.

3.5.1.3. The *nad1*b-*nad1*c region.

To examine transcripts arising from the *nad1*b-*nad1*c region, oligomers B4, B1 and C1 were used in Northern blot hybridizations. Oligomer B4, which is located 0.7 kb upstream

Figure 9. Northern blot hybridization analysis of *nad1* transcripts. Total wheat mtRNA was separated by electrophoresis on a 1.25% agarose-formaldehyde gel (lane 1), transferred to a nylon membrane and probed with ^{32}P-end-labelled oligomers specifying *nad1* coding (lanes 2, 5, 6 and 7) or flanking regions (lanes 3, 4, 8 and 9). The 18S (1.9 kb) and 26S (3.5 kb) rRNAs are shown, and dots indicate the three transcripts common to exon-specific probes. A restriction map of *nad1* coding regions (solid boxes) shows the position of oligomers (circled letters) that were used as probes. The *rps13* and *mat-r* genes are also shown (open boxes). Some restriction sites are shown as landmarks: BamHI (B), EcoRI (E), HindIII (H), RsaI (R), XbaI (Xb) and XhoI (X). The relative exposure times are in approximate ratios of 1 (lanes 2, 3, 5, 6 and 7): 5 (lane 4): 15 (lanes 8 and 9).


1
2
3
4
5
6
7
8
9
26S
18S
oligo probe : A1 A4 B4 B1 C1 D1 D2 E3


nad1a
B
R
B
400 bp
A1
A4
rps13
nad1b
nad1c
H
B
B
B4
B1
C1
nad1d
E
X
Xb
Xh
E
D1
D2
mat-r
nad1e
B
Xb
E3

of *nad1*b, hybridizes to two high molecular weight transcripts of approximately 2.1 and 3.7 kb (Figure 9, lane 4). It is possible that these signals are due to non-specific hybridization to rRNAs, which may explain their diffuse appearance. Alternatively, they could be precursor transcripts of the *nad1*b-*nad1*c region and the removal of the 1422 bp *nad1*b/c intron may explain their difference in size. However, this appears doubtful because the 3.7 kb transcript is detected neither with oligomer B1 nor oligomer C1 (lanes 5 and 6). It is more likely that the large transcripts of approximately 2.9 kb and >4 kb in the lanes 5 (longer exposure not shown) and 6 are precursors of this region. Transcripts of similar sizes had been previously detected with a probe covering the region upstream of the HindIII site which is located 0.6 kb upstream of *nad1*b (Bonen, 1987). There are other hybridizing transcripts of approximately 2.5 kb in lanes 5 and 6, and of 0.9 kb in lane 6. It remains to be determined whether they could represent splicing intermediates or by-products.

3.5.1.4. The *nad1*d and *nad1*e regions.

In contrast to the *nad1*b-*nad1*c region which has a complex transcript profile (Figure 9; lanes 4, 5 and 6), only the three mRNAs common to the five *nad1* exons are prominent when the *nad1*d-specific oligomer D1 is used (lane 7). In the *nad1*d downstream region, a putative precursor transcript of approximately 0.9 kb is detected (oligomer D2, lane 8), which is also seen in lane 7 (longer exposure not shown). These transcripts would be large enough to cover the *nad1*d-flanking regions that are homologous in wheat, broad bean and *Oenothera* (Wahleithner et al. 1990; Wissinger et al. 1991). These

hybridizing transcripts (lanes 7 and 8) could also represent comigrating RNAs. Other minor transcripts larger than 2 kb (lane 8) perhaps arise from longer precursors or may correspond to splicing intermediates. This possibility has to be kept in mind since group II intron splicing by-products, because of their branched structures, may have reduced mobility on agarose gels (van der Veen et al. 1986).

The three putative *nad1* mRNAs are not resolved as clearly in the *nad1*e region (lane 9), but the data are consistent with their presence. In addition, a short transcript of 200-400 nt is more abundant than other *nad1* transcripts and has been detected with other probes (oligomer E1 located at the 5' end of *nad1*e, and probe 7 from Figure 2, data not shown). Another short transcript was observed when a probe for *nad1*c was used (Figure 9, lane 6). In watermelon, a probe containing ORF36 sequences, which are homologous to *nad1*e, also hybridizes to short and abundant transcripts (Stern et al. 1986). It is also worth noting that several short transcripts (but not nearly as abundant) have been seen with several *nad5* exon- and intron-specific probes in wheat and maize (De Souza et al. 1991). De Souza and colleagues suggested that these RNA molecules could arise from free *nad5* exons or introns. Similarly, the short transcripts that are observed in the wheat *nad1*e region may correspond to the "free *nad1*e exon" (256 bp coding + 3' tail), or to the *nad1*e exon plus short upstream and/or downstream sequences. Further experiments will be required to distinguish among these possibilities.

In summary, Northern blot hybridizations have led to the identification of three putative *nad1* mRNAs of approximately 1.5, 1.7 and 1.9 kb. These experiments and reverse transcriptase sequencing across exon junctions (section 3.2,

Figure 4A, B and C) demonstrate that the five *nad1* gene segments are actively transcribed in wheat mitochondria, and suggest that correctly joined *nad1* transcripts are relatively abundant. The ratios of *nad1* mRNAs to precursor transcripts may vary from approximately 1:1 in the *nad1*a region to 10:1 in the *nad1*d region. The complex and distinct transcript profiles that were observed for the different *nad1* coding regions are consistent with the idea that these regions are encoded by different precursor RNA molecules. There is no evidence for long precursor transcripts that would contain all five *nad1* exons. In *Oenothera*, complex *nad1* transcript profiles are also seen (Wissinger et al. 1991), as well as for the watermelon *nad1*b-*nad1*c region (Stern et al. 1986). The complexity of *nad1* transcript profiles in plants at least in part reflects the unusual structure of the *nad1* gene.

3.5.2. S1 nuclease analysis of *nad1* transcripts.

To examine further the transcripts arising from the *nad1* coding regions, S1 nuclease mapping experiments were carried out. These experiments were limited mainly to regions downstream of *nad1*a, *nad1*c and *nad1*e, as well as upstream of *nad1*b. These regions were the most interesting ones for future correlation with intron structures (section 3.6).

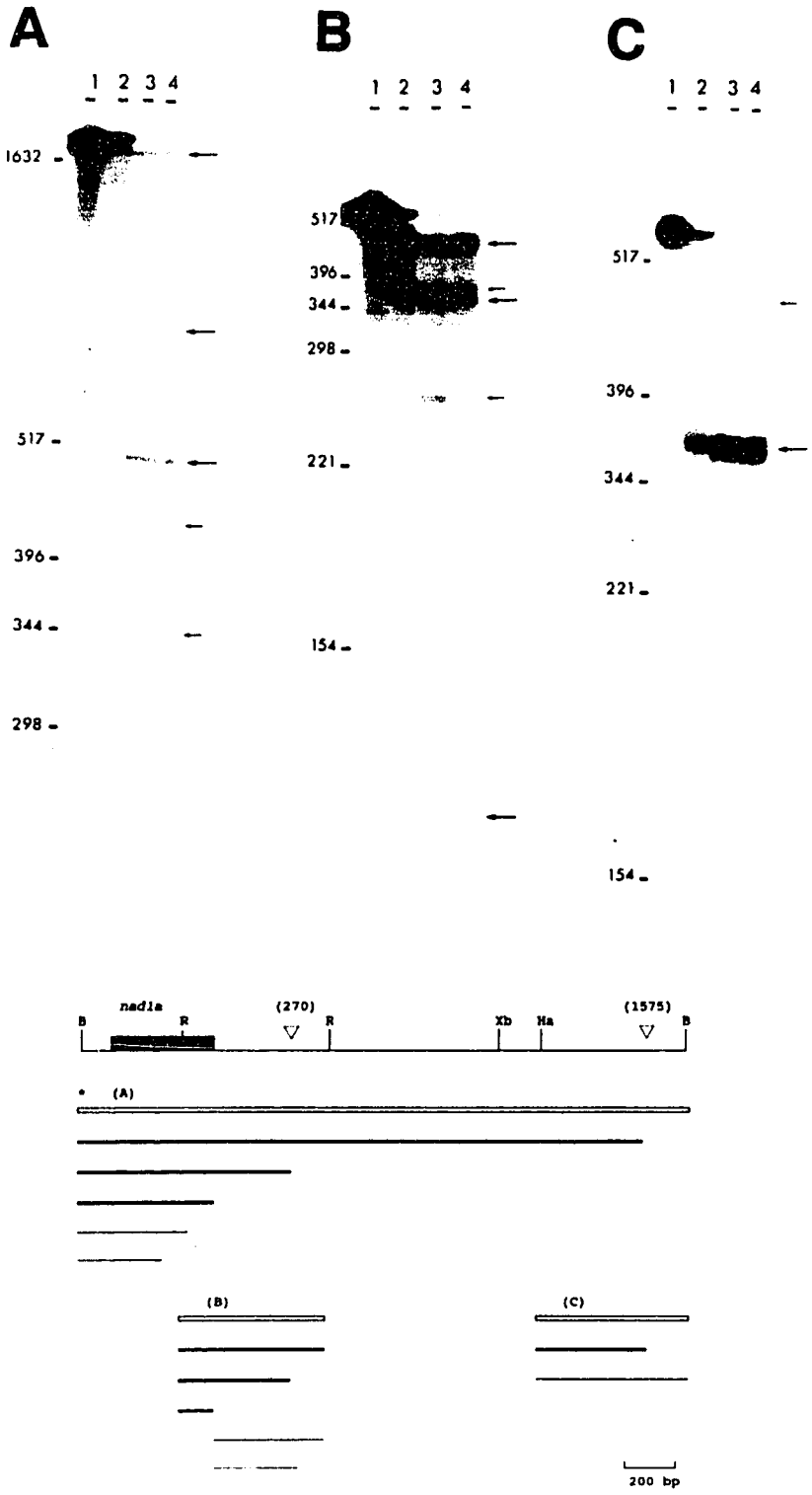
The radio-labelled probes used for these S1 mapping experiments can be divided in two types: uniformly-labelled and end-labelled. The uniformly-labelled probes were obtained from PCR-amplified M13 clones using sequencing and reverse sequencing primers (flanking the multiple cloning site) and α-^{32}P-dATP was included in the reaction. PCR products are therefore some 100 bp larger than the cloned insert, and these vector sequences can be used as an internal control to

distinguish between fully protected inserts and DNA-DNA hybrids. The end-labelled probes were obtained by extension of recessed 3' ends generated by restriction enzymes and labelling with α-^{32}P-dATP and the Klenow fragment polymerase; this type of probe is indicated by asterisks on schematics (Figure 10, 11 and 12). It should be noted that in contrast to end-labelled probes, the radioactive signals of S1 products derived from uniformly-labelled ones are proportional to the length of the protected fragments.

3.5.2.1. The *nad1*a region.

To analyze the transcripts arising from the *nad1*a region, a BamHI restriction fragment of 2.2 kb was used in S1 nuclease mapping experiments (Figure 10A). Three major S1 products of approximately 515, 800 and 2000 nt were observed and analyzed more precisely in two other experiments shown in panels B and C. Of the three main protected fragments seen in panel B, the largest is approximately 520 bp in size and corresponds to the fully protected DNA. The 120 and 380 nt products appear to correspond to the 515 and 800 nt products, respectively, seen in panel A. The former marks the 3' end of *nad1*a, which presumably arises from hybrids with transcripts in which *nad1*a and *nad1*b are spliced. The latter represents a S1 nuclease sensitive site located approximately 270 nt downstream of *nad1*a. In panel C, a major S1 product of 360 nt is observed and appears to correspond to the 2000 nt product seen in panel A. This suggests that most stable transcripts end approximately 1575 bp downstream of *nad1*a, although a minority of RNA molecules extend further downstream as suggested by the fully protected fragment of 480 nt. This result was also confirmed using a probe from the XbaI site to the BamHI site

Figure 10. S1 nuclease analysis of transcripts arising from the *nad1*a region. The schematic summarizes the S1 nuclease mapping experiments shown in panels **A**, **B** and **C**. DNA fragments that were used as probes (open bars), end-labelled (asterisk) or uniformly labelled, are shown with the letters corresponding to the different panels. Thick and thin lines refer to major and minor S1 nuclease products (indicated by large and small arrows in panels, respectively). Open triangles denote transcript termini with the distances to the upstream *nad1* exon in parentheses (bp). Restriction sites: BamHI (B), HaeIII (Ha), RsaI (R) and XbaI (Xb). **A.** A BamHI fragment of approximately 2.2 kb was end-labelled. Cloned DNA fragments RsaI-RsaI (**B**) and HaeIII-BamHI (**C**) were uniformly labelled by PCR amplification. In each lane (1-4), 0, 50, 200 and 500 units of S1 nuclease were used, respectively. The sizes of markers, which were obtained from HinfI restriction digests of pBR322 plasmids, are indicated (nt).


A
1 2 3 4
1632
517
396
344
298
B
1 2 3 4
517
396
344
298
221
154
C
1 2 3 4
517
396
344
221
154
nadia
B
R
(270)
R
Xb
Ha
(1575)
B
(A)
(B)
(C)
200 bp

(data not shown).

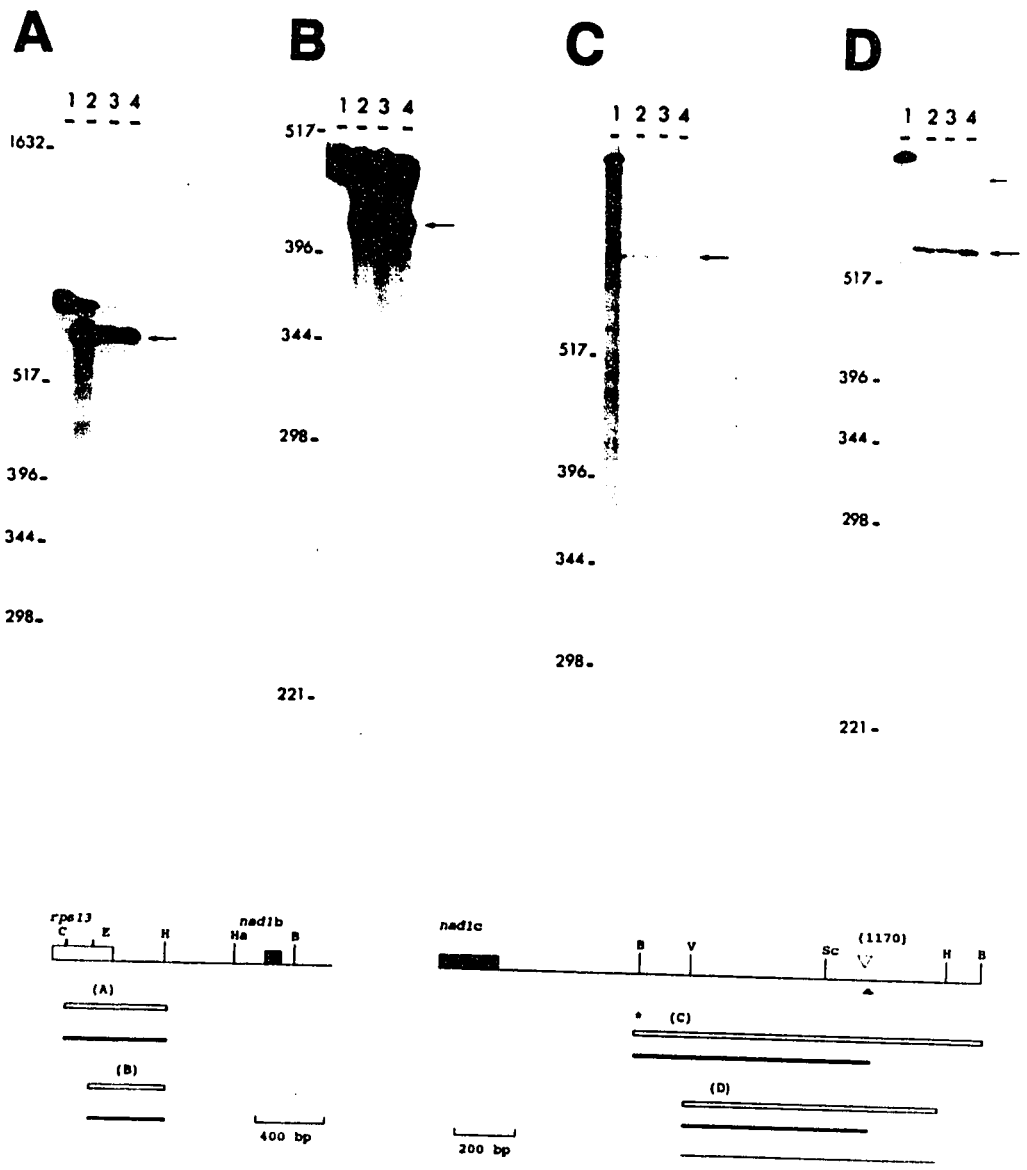
In addition to these major protected DNA fragments, other less abundant S1 products were seen (Figure 10A and B, small arrows). In panel A, two products of 340 and 430 nt may correspond to mis-matches due to RNA editing sites in codons 72 and 103 (Table 2), respectively. This suggests that RNA editing sites can be detected under the conditions that were used. However, only one other editing site is present in *nad1*a (in the initiation codon) but it was not detected. Consistently, more abundant products are observed where two adjacent mis-matches are present (codon 103) than only one (codon 72). It is worth mentioning that, since mature transcripts are more frequently edited than precursors (Bonnard et al. 1992), the ratio of spliced transcripts to precursors could be underestimated by S1 mapping.

Minor protected fragments of approximately 270 and 400 nt seen in panel B (small arrows) are most easily explained by the formation of DNA-RNA hybrids with excised introns (as suggested by the schematic in Figure 10, thin lines). Northern hybridizations revealed an intron-specific transcript of 1.9 kb, and this supports the idea that such stable splicing by-products of the *nad1*a/b intron may be present (section 3.5.1.2). The use of additional probes in Northern blots and S1 nuclease analysis would be necessary to verify this interpretation of the results.

3.5.2.2. <u>The *nad1*b-*nad1*c region</u>.

Previous Northern blot hybridizations upstream of *nad1*b had shown that *nad1* transcripts begin upstream of the HindIII site (Bonen, 1987). These results were confirmed by S1 nuclease mapping using a HindIII-HaeIII DNA fragment which was

Figure 11. S1 nuclease analysis of transcripts arising from the *nad1b-nad1c* region. Symbols are as defined in Figure 10. The restriction sites are BamHI (B), ClaI (C), EcoRI (E), HaeIII (Ha), HindIII (H), PvuII (V) and ScaI (Sc). The solid triangle in the schematic marks the breakpoint in homology with watermelon (Stern et al. 1986) and *Oenothera* sequences (Wissinger et al. 1991). Cloned DNA fragments ClaI-HindIII (**A**), EcoRI-HindIII (**B**) and PvuII-HindIII (**D**) were uniformly ^{32}P-labelled by PCR amplification. **C.** An isolated BamHI fragment was end-labelled. In each lane (1-4), 0, 50, 200 and 500 units of S1 nuclease were used, respectively. The sizes of markers are indicated (nt).


A
B
C
D
1 2 3 4
1632
517
396
344
298
221
rps13
C
E
H
nad1b
Ha
B
(A)
(B)
400 bp
nad1c
B
V
Sc
(1170)
H
B
(C)
(D)
200 bp

found to be fully protected (data not shown). In Figure 11A and B, ClaI-HindIII and EcoRI-HindIII DNA fragments of approximately 580 and 420 nt, respectively, also appear to be fully protected. This suggests that the 5' terminus of transcripts would be located upstream of the ClaI restriction site.

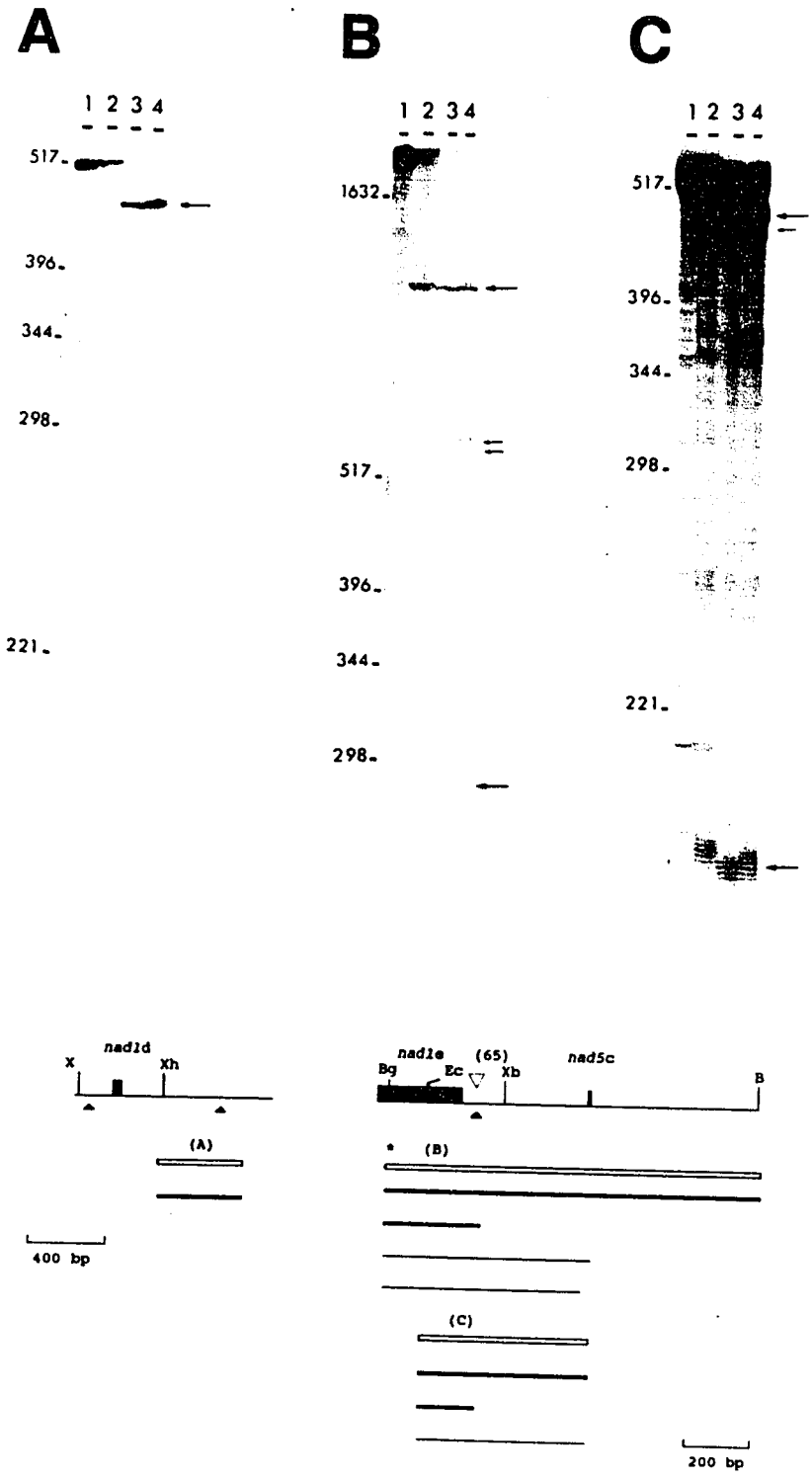
Downstream of *nad1c*, the S1 nuclease analysis, using either ^{32}P-end-labelled (Figure 11C) or uniformly-labelled probes (panel D), shows products of approximately 700 and 570 bp, respectively. These data indicate that most transcript termini are located approximately 1170 bp downstream of *nad1c* (Figure 11C and D). These results were also confirmed using a ScaI-HindIII DNA fragment as a probe (data not shown), which was used to determine the location of transcript termini more precisely. A minor S1 product of approximately 750 nt in panel D corresponds to fully protected DNA fragments, which suggests that some transcripts extend farther downstream. The presence of putative transcript termini between *nad1c* and the BamHI site appears unlikely since (1) most of important core sequences of the *nad1c/d* group II intron are expected to be located downstream of *nad1c* rather than upstream of *nad1d* (see section 3.6), and (2) there are sequences farther downstream of the BamHI site that are well-conserved among wheat (Appendix IIB), *Oenothera* (Wissinger et al. 1991) and watermelon (Stern et al. 1986). In fact, the breakpoint in sequence homology between wheat and the two dicots occurs approximately 1190 bp downstream of *nad1c* (solid triangle, Figures 3 and 11), which is only 20 bp downstream from the estimated position of the transcript termini.

3.5.2.4. The *nad1*d and *nad1*e regions.

Only one S1 nuclease experiment is presented for the *nad1*d region, and a fully-protected product of 470 nt is seen (Figure 12A). These results suggest that transcripts extend at least 650 bp downstream of *nad1*d coding sequences; this is at least 120 bp beyond the breakpoint in homology between the wheat, broad bean (Wahleithner et al. 1990) and *Oenothera* sequences (Wissinger et al. 1991; solid triangle, Figure 12).

DNA sequence analysis downstream of *nad1*e revealed the presence of a short exon of 22 bp encoding a central portion of the *nad5* gene (*nad5*c) in wheat (De Souza et al. 1991). The *nad1*e-*nad5*c region was analyzed by S1 nuclease mapping using end-labelled and uniformly-labelled DNA fragments (Figure 12B and C). In panel B, an S1 product of approximately 1050 nt covers the entire restriction fragment used, indicating that *nad1*e is cotranscribed with *nad5*c. This confirms that stable transcripts are present in the region of 0.4 kb downstream of *nad5*c in wheat (De Souza et al. 1991). This result is also consistent with a fully protected fragment of 495 nt seen in panel C. Another S1 product of approximately 290 nt is observed in panel B, which suggests that *nad1* transcript termini are located approximately 65 bp downstream of *nad1*e stop codon. This result is also supported by a product of 175 nt seen in panel C. Because the *nad5*c probe also hybridizes to the three putative *nad1* mRNAs (L. Bonen, unpublished), these transcript termini may correspond to the 3' ends of the 200-400-bp transcripts spanning *nad1*e (cf. Figure 9, lane 9). However, more experiments are necessary to clarify this hypothesis because the signal from the 290 nt product is weaker than expected from such abundant transcripts, and because the background is high in the panel C. Therefore, the

Figure 12. S1 nuclease analysis of transcripts arising from the *nad1*d and *nad1*e regions. Symbols are as defined in Figure 10. Restriction sites are BamHI (B), BglII (Bg), EcoRV (Ev), XbaI (Xb), XhoI (X) and XhoII (Xh). The solid triangles in the schematic mark the breakpoints in homology with broad bean (Wahleithner et al. 1990) and *Oenothera* sequences (Wissinger et al. 1991). **A.** A cloned DNA fragment XhoII-HindIII DNA fragment was uniformly labelled (using the sequencing primer and oligomer D2) from the XhoII restriction site to the position specified by oligomer D2, approximately 440 bp farther downstream. **B.** A pUC plasmid containing a 3.5 kb BamHI insert was linearized by cleavage with BglII restriction enzyme, and end-labelled. **C.** A cloned EcoRV-BamHI DNA fragment was uniformly labelled (using the sequencing primer and an oligomer specifying *nad5*c) from the EcoRV restriction site to *nad5*c. In each lane (1-4), 0, 50, 200 and 500 units of S1 nuclease were used, respectively. The sizes of markers are indicated (nt).


A
B
C
1 2 3 4
517
396
344
298
221
1632
nad1d
X
Xh
(A)
400 bp
nad1e
(65)
nad5c
Bg
Ec
Xb
B
(B)
(C)
200 bp

quality of these results does not exclude the presence of additional transcript termini.

Panel B also revealed two minor S1 products of approximately 570 and 600 nt, which could correspond to transcript termini that are located in the vicinity of *nad5*c borders. A minor product of approximately 480 nt, as seen in panel C, is consistent with the one of 570 nt seen in panel B. These results suggest that *nad5*b/c and *nad5*c/d introns can undergo splicing prior to processing of *nad1*e-*nad5*c precursor transcripts. Transcripts containing *nad5*c but not *nad1*e could not be detected in panel C. However, as mentioned above, the quality of the results does not permit the absolute exclusion of their presence. These observations may reflect that the processing of the *nad5* transcripts is faster than that of *nad1*. This hypothesis is also consistent with the suggestion that splicing of *nad5*c is an early event in the *nad5* transcript maturation (De Souza et al. 1991).

In summary, a 3' transcript terminus is located approximately 65 nt downstream of *nad1*e, and it coincides with the breakpoint in homology between wheat, broad bean (solid triangle, Figure 3 and 12; Wahleithner et al. 1990) and *Oenothera* (Wissinger et al. 1991). The breakpoint in homology between wheat and petunia sequences (Conklin et al. 1991) is located 15 bp farther downstream. In *Oenothera*, the 3' mRNA terminus is located within 100 bp downstream of *nad1*e (Wissinger et al. 1991). The variability in length of the three putative *nad1* mRNAs in wheat (described in section 3.5.1) may occur at either end of the gene. The region of 1 kb upstream of *nad1*a was sequenced, but no consensus transcription initiation site, as proposed for wheat mitochondrial genes (Covello and Gray, 1991), could be found

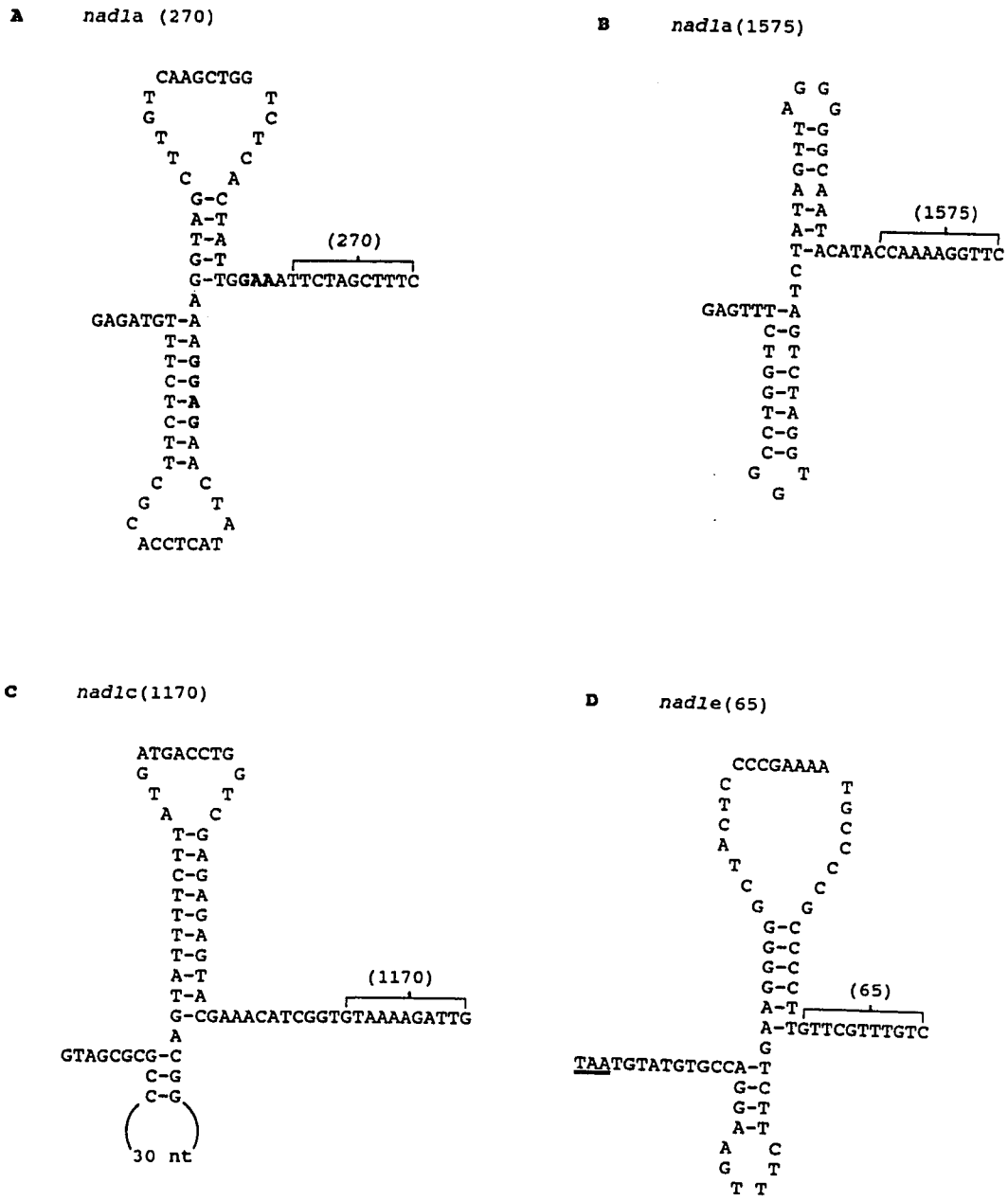
(Appendix IA). One possibility is that the *nad1*a stable transcripts are processed forms of longer precursors and that transcription is initiated in the region of the $tRNA^{pro}$ gene, which is located 1.2 kb upstream of *nad1*a (Runeberg-Ross et al. 1987). Run-on transcription experiments and guanylyltransferase capping experiments support the presence of transcription initiation sites in this region (P. Garber and L. Bonen, unpublished). Alternatively, the *nad1*a region could be cotranscribed with the *rrn18* gene which is located 3.3 kb upstream of *nad1*a (Figure 2).

3.5.3. Potential processing signals of *nad1* precursor transcripts.

The discrete transcript termini which were found 270 and 1575 bp downstream of *nad1*a, 1170 bp downstream of *nad1*c and 65 bp downstream of *nad1*e appear to be RNA processing sites since, in each case, transcription extends further downstream, although it cannot be excluded that they result from multiple transcription termination signals. Double stem-loop structures have been found a few nucleotides upstream of transcript termini for a number of plant mitochondrial genes (reviewed by Gray et al. 1992).

The regions immediately upstream of each of the four *nad1* transcript termini that were found (section 3.5.2) were analyzed to examine whether such double stem-loop structures could be present. In Figure 13, the structures shown are designated according to the distances that separate them from the upstream exons. Among these, only *nad1*e(65) is well conserved with *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991) sequences. The *nad1*a(270) and *nad1*c(1170) double stem-loop structures are poorly conserved

Figure 13. Potential double stem-loop structures located immediately upstream of 3' termini of *nad1* transcripts. The numbers represent distances to the upstream exons and the brackets indicate the flanking 5-nt upstream and downstream of the transcript termini determined by S1 nuclease mapping. The GAGG and GAA motifs are in bold (**A**) and the stop codon in *nad1*e is doubly underlined (**D**). In the regions shown in panels **B and C**, only one strand was sequenced. Although shown as DNA, Us would be substituted by Ts in the RNA structures.


A nad1a (270)
CAAGCTGG
(270)
GAGATGT-A
ACCTCAT
B nad1a(1575)
(1575)
GAGTTT-A
C nad1c(1170)
ATGACCTG
(1170)
GTAGCGCG-C
30 nt
D nad1e(65)
CCCGAAAA
(65)
TAATGTATGTGCCA-T

in *Oenothera* (Wissinger et al. 1991), and a loop of 30 nt in *nad1*c(1170) seems unlikely to play a role *in vivo* (Figure 13C). It is worth noting that GAGG and GAA motifs flank the *nad1*a(270) single stem-loop (nucleotides in bold), and resemble the ones proposed downstream of the *atp9* gene in maize and tobacco (Bland et al. 1986). The *nad1*a(1575) structure contains two mis-matches, and counterpart sequences are not available for other plants.

In summary, S1 nuclease mapping experiments have led to the identification of discrete transcript termini downstream of *nad1*a, *nad1*c and *nad1*e in wheat. Taken together with Northern hybridizations, these data strengthen the hypothesis that the different *nad1* coding segments are located on different precursor RNAs since transcripts arising from regions flanking the *nad1* exons cover only regions of 1-1.5 kb. The transcript termini downstream of *nad1*c coincide closely with the breakpoints in homology between plants which suggests that important intronic sequences are likely to be present between this site and the upstream coding sequences.

3.6. STRUCTURAL ANALYSIS OF *nad1* INTRONS.

In the previous sections, it was shown that (1) all *nad1* coding sequences are present and transcribed in wheat mitochondria, (2) they are correctly linked at the RNA level and (3) *nad1* transcripts cover regions of 1-1.5 kb downstream of *nad1*a, *nad1*c, *nad1*d as well as upstream of *nad1*b. The next questions that will be addressed are how these *nad1* coding transcripts are joined together and what structural elements could be involved. This problem is intimately related to the nature of the *nad1* intervening sequences. The flanking regions of *nad1* coding segments were analyzed in order to determine if

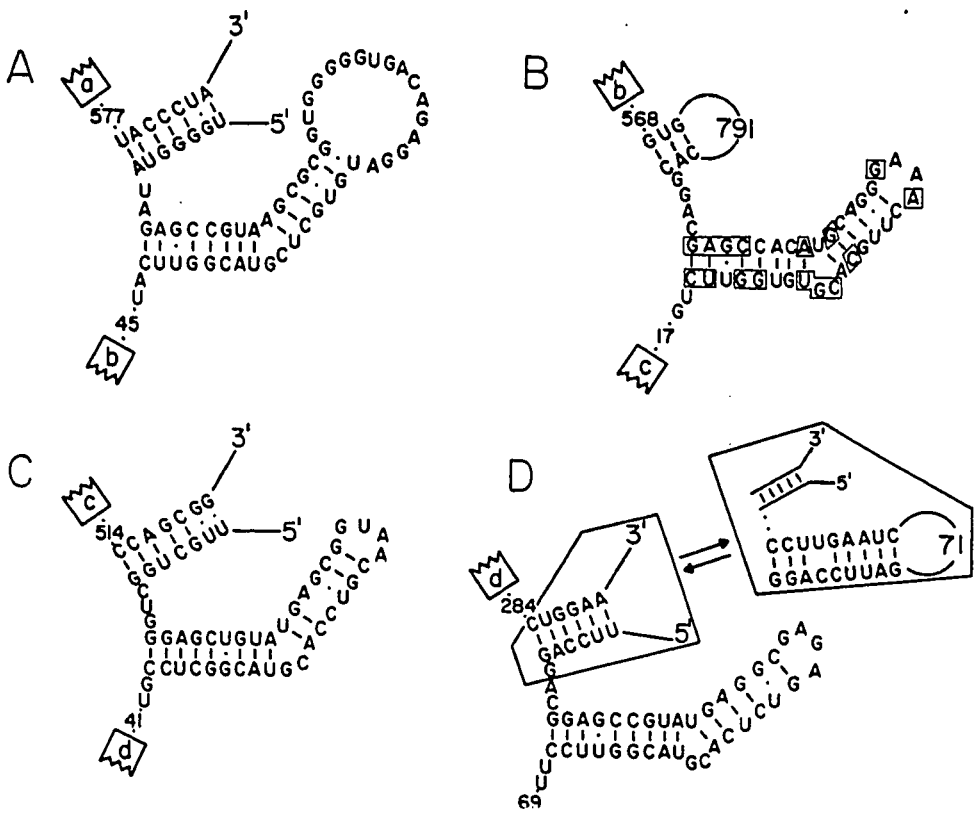
sequences or structural motifs that are hallmarks of organellar introns could be identified.

3.6.1. Group II intron structures flanking *nad1* coding sequences.

The *nad1* gene contains four introns that are designated *nad1*a/b, *nad1*b/c, *nad1*c/d and *nad1*d/e. When sequences flanking the wheat *nad1* coding segments were examined, features were found that are clearly diagnostic of group II introns (Figure 3, dashed blocks, Figure 14). Group II introns contain a core structure that is surrounded by six helical domains (Figure 1). In the 100-bp regions upstream of *nad1*b, *nad1*c, *nad1*d and *nad1*e, stem structures with two bulging nucleotides which resemble closely the typical domain V structure of group II introns were found (Michel et al. 1989). The blocked nucleotides (Figure 14B) depict 16 positions common to the domain V of these four *nad1* introns. These positions emphasize the strong conservation of this structure. The domain V loop structure typically contains four nucleotides, as is the case for three of the four *nad1* introns (Figure 14B, C and D); however the *nad1*a/b intron, instead, has a 17-nt loop structure (Figure 14A) and lacks several other features which will be dealt with separately (section 3.6.2).

Typical sequences at the 5' splice site (GTGCG or GCGCG; Michel et al. 1989) and the domain VI helical structure near the 3' splice site are present in the *nad1*b/c, *nad1*c/d and *nad1*d/e introns. However, in the case of *nad1*c/d, three editing events must be invoked at positions -4, -6 and -8. A bulging adenosine located 7 nt from the 3' splice site is also present in domain VI of the *nad1*c/d and *nad1*d/e introns (Figure 3); this would place these introns in subgroup IIA

Figure 14. Secondary structural models of *nad1* group II intron structures, illustrating putative *trans*-splicing (**A**, **C** and **D**) and *cis*-splicing (**B**) of wheat *nad1* exons (broken blocks). The only features of the group II core structure (Michel et al. 1989) shown are the well-defined domain V (hairpin) and the upstream domain IV, with a potential discontinuity within the looped region (**A, C, D**) or within domains I, II or III (inset in **D**). The structure in (**B**) is shown as proposed for watermelon (Michel et al. 1989) and positions conserved among the four wheat *nad1* domain V sequences are boxed. The numbers indicate distances to the exons.


A
B
C
D
a
b
c
d
3'
5'
577
45
568
791
17
514
41
284
71
69

(Michel et al. 1989). Although the *nad1*b/c intron lacks the bulging A, it is also a member of subgroup IIA based on other criteria. Exon binding sequences proposed in the domain I of the watermelon *nad1*b/c intron (Michel et al. 1989) are present in wheat (Figure 3B positions 839-845 and 914-919), as is the exon binding sequence proposed for the broad bean *nad1*d/e intron (Wahleithner et al. 1990; Figure 3C positions 518-522). In addition, the structure proposed for the watermelon *nad1*b/c intron (Michel et al. 1989) is conserved in wheat (Appendix III). These data indicate that *nad1* intervening sequences are related to group II ones, as defined by Michel and colleagues (1989).

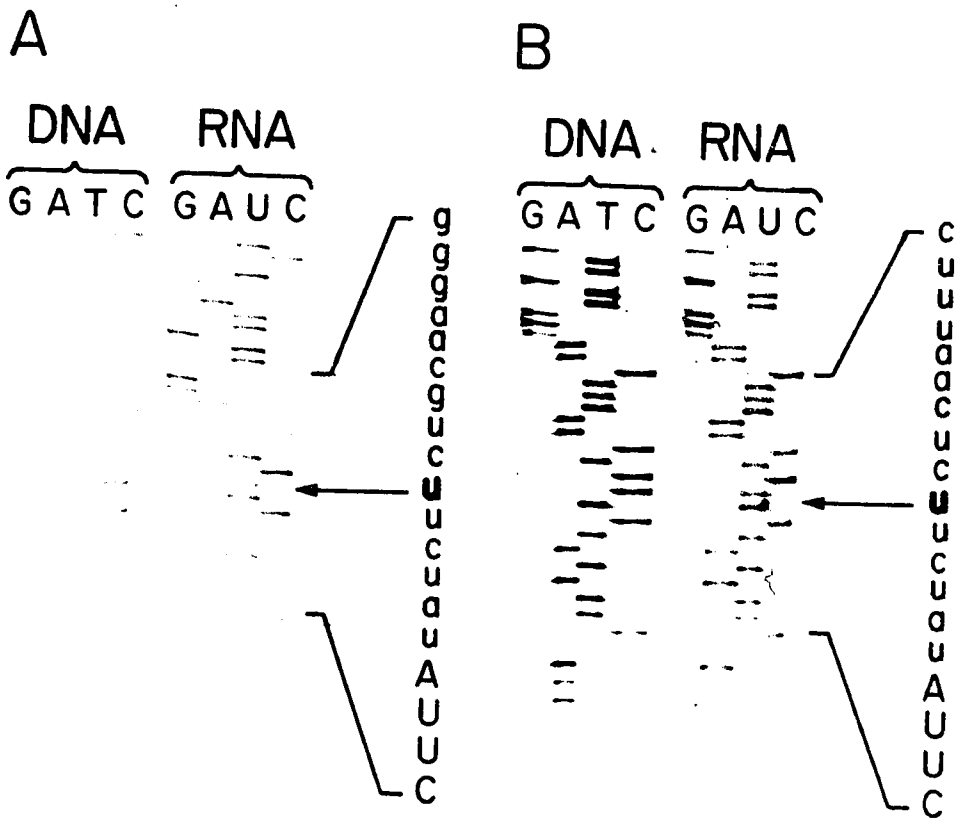
3.6.2. Analysis of the *nad1*a/b discontinuous intron.

3.6.2.1. Domain VI of the *nad1*a/b discontinuous intron.

The discontinuous intron between *nad1*a and *nad1*b is atypical of group II members in several respects. One of these concerns the domain VI which normally forms a helical structure, with a bulging A, that is immediately flanked by the domain V on one side and the 3' exon on the other. In the *nad1*a/b intron in wheat, domain VI cannot be folded in the typical helical structure. In the *Oenothera* counterpart, it has been suggested that RNA editing can establish the correct helical structure of the *nad1*a/b intron domain VI (Wissinger et al. 1991).

To address the issue of whether editing occurs within domain VI of the wheat *nad1*a/b intron and whether it might in fact be required for the correct RNA folding for splicing, reverse transcriptase-PCR experiments were carried out on precursor transcripts using oligomer D1 for primer extension

Figure 15. Analysis of RNA editing in domains V and VI of the *nad1*a/b discontinuous intron in wheat and pea. Sequencing gel autoradiographs of wheat (**A**) and pea (**B**) genomic DNAs (left) and cDNAs (right) are shown with intronic nucleotide positions in lower case. The horizontal arrow and a U in bold indicate the position of the single RNA editing site in domain VI. Pea sequences were determined directly from the PCR-amplification products of mtDNA and cDNA, whereas wheat sequences are derived from cloned DNA and cDNA.


A
DNA
RNA
G A T C
G A U C
g g g a a c g u c u c u a a A U U C
B
DNA
RNA
G A T C
G A U C
c u u a a c u c u c u a a A U U C

on wheat mtRNA and oligomer B3 as reverse primer for PCR amplification. The cDNAs were therefore derived from transcripts containing domains V and VI of the *nad1*a/b intron plus *nad1*b-*nad1*c-*nad1*d and possibly *nad1*e. Four PCR-amplified cDNA products were cloned and sequenced (see footnote on page 57). These experiments led to the identification, in each cDNA clone, of a single C-to-U editing site located 6 nt from the 3' splice site within domain VI (Figure 15A). Unexpectedly, this editing site does not appear to contribute in stabilizing any potential helix.

To further investigate this issue, the comparable region in the dicot pea was examined. Primer extension on RNA with reverse transcriptase was carried out using oligomer C1, and both cDNA and DNA were amplified using oligomers B3 and C1. These products were directly sequenced with Sequenase using oligomer B1 which is internal to these DNA fragments. The result indicates that the same position as in wheat, and only that position, was edited (Figure 15B). As in wheat, editing at this position in pea does not seem to contribute in establishing a domain VI helical structure.

There are several nucleotide substitutions and deletions/insertions among the *nad1*a/b intron sequences of different plants and some of these are illustrated by the alignments shown in Figure 16. It is perhaps not surprising that wheat and pea transcripts have the same C-to-U editing site at position -6, but it is unexpected that this position differs from the two sites (-7 and -11) reported in *Oenothera* (Figure 16B, underlined positions). It is worth noting that the RNA editing site at position -7 in *Oenothera* was observed in only one of ten cDNA clones (Wissinger et al. 1991). These observations contrast with typical RNA editing in coding

Figure 16. Alignment of *nad1*a/b intron sequences from various plants. The domains V (**A**) and VI (**B**) of the *nad1*a/b intron in wheat are aligned with those of maize (Bland et al. 1986), pea (this work), *Oenothera* (Wissinger et al. 1991), petunia (Conklin et al. 1991) and tobacco (Bland et al. 1986). The positions that have been found to be edited in wheat, pea and *Oenothera* are underlined, and the asterisk denotes the bulging A proposed for lariat formation in *Oenothera*. Exon sequences are shown in lower case letters. Pea sequence data end in the domain V, where oligomer B3 (which was used for PCR amplification) is located. **C** Alignment of the 5' splice site of the *nad1*a/b intron in wheat, *Oenothera* and petunia.

A *nad1a/b domain V*

```
WHEAT       ...GAGCCGTAAGCGCGGTGGGGG--TGACAGAGGATGTGCTCGTACGGTTC
MAIZE       ...GAGCCGTAAGCGCGGTGGGGG--TGACAGAGGATGTGCTCGTACGGTTC
PEA         ........GTAAGCGCGGTGGGGGGGTTACAGAGGACGTGCTCGTACGGTTC
OENOTHERA   ...GAGCCGTAAGCGCGGTGGGGGT-TGACAGAGGACGTGCTCGTACGGTTC
PETUNIA     ...GAGCCGTAAGCGCGGTGGGGGG-TGACAGAGGACGTGCTCGTACGGTTC
TOBACCO     ...GAGCCGTAAGCGCGGTGGGGGG-TGACAGAGGACGTGCTCGTACGGTTC
```

B *nad1a/b domain VI*

```
WHEAT       ATAGTTGGGTATGATATTCTCGTGGATTGTTGGGAACG-----TCCTCTATattcg...
MAIZE       ATAGTTGGGTATGATATTCTCGTGGATTGTTGGGAACG-----TCCTCTATattcg...
PEA         AAAGAAGGGTATGATCAACTCG----TTGTTGGGAACTTTAACTCCTCTATattcg...
OENOTHERA   ATAGAAGGGTATGATCAACTCGTTGATTGTTGGGAACTTTCACTCCTCTATattcg...
PETUNIA     ATAGAAGGATATGAAAAACTAGTTATTTGTTGGGAACTTTCACTCCTCTATattcg...
TOBACCO     ATAGAAGGATATGAAAAAGTACTTATTTGTTGGGAACTTTCACTCCTCTATattcg...
                                                      *
```

C *nad1a/b 5' splice site*

```
WHEAT       ...tagtaAGACGGGGGGCGGCCGTTCGGTCGCCTATG...
OENOTHERA   ...tagtaATTA-GGGGGCGGCCGTTCGGTCGCCTATG...
PETUNIA     ...tagtaATTA-GGGGGCGGCCGTTCGGTCGCCTATG...
```

regions where RNA sequences are more similar than DNA sequences.

Moreover, in the monocots wheat and maize, a 5-bp deletion is seen at the position of the bulging A proposed in *Oenothera* (Figure 16B; Wissinger et al. 1991). The pea sequence also shows a number of other differences including a nucleotide substitution at position -11 that is edited in *Oenothera*, and a deletion 4 nt farther upstream. These observations indicate that nucleotide substitutions and deletion/insertions have occurred among plants at positions that are usually conserved within the domain VI of the group II introns. Finally, the secondary structure equivalent to the one proposed in *Oenothera* for domain VI of the *nad1*a/b intron cannot be formed in wheat and pea.

3.6.2.2. Domain V of the *nad1*a/b discontinuous intron.

The wheat (this work) and maize (Bland et al. 1986) *nad1*a/b introns have a large loop of 17 nt in domain V (Figure 14A); this is unusual compared to the purine-rich tetraloop normally found in group II introns (Figure 14B, C and D). The loop of the *nad1*a/b domain V contains 19 nt in pea (this work) and 18 nt in *Oenothera* (Wissinger et al. 1991), petunia (Conklin et al. 1991) and tobacco (Bland et al. 1986), which suggests that its size may not be crucial although its sequence appears to be conserved (Figure 16). The size of domain V (34 nt) is almost universal among group II introns; the domain V is shorter by one or two nt in only four of the 68 group II introns, and two group II-like introns in *Euglena* chloroplast have up to 5 additional nt in the loop of domain V (Michel et al. 1989).

3.6.2.3. The 5' splice site of the *nad1*a/b discontinuous intron.

Another nearly universal motif of group II introns is the GTGCG or GCGCG consensus sequence located at the 5' splice site (Michel et al. 1989). In the wheat *nad1*a/b intron, the sequence at the 5' splice site is GACGG or AGACG because of a 2-nt redundancy; the latter is more likely since it corresponds to the 3' splice site consensus sequence TAT. Surprisingly, three nucleotide substitutions and one deletion/insertion are seen at the 5' splice site of the *nad1*a/b intron in comparisons between wheat and the dicots *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991; Figure 16C). These positions are normally involved in long-range interactions within group II introns (Jacquier and Jacquesson-Breuleux, 1991). It is therefore unexpected to observe the lack of the consensus sequence and, even more surprisingly, several nucleotide substitutions and a deletion/insertion between sequences from different plants. However, immediately downstream, a perfectly conserved block of 25 nt is present and may well play an important role in the *nad1*a/b intron structure (Figure 16C).

3.6.2.4. High rate of nucleotide substitution in the *nad1*a/b intron.

In order to examine the degree of sequence conservation within the four *nad1* introns among different plants, wheat and *Oenothera* exon-flanking sequences were compared (Table 3). Because the secondary structures are as yet incompletely known in wheat, except for the *nad1*b/c intron, only a block of 300 bp at the 5' end and the well-conserved domains V and VI at the 3' end of each intron were arbitrarily included. The first

Table 3. Sequence similarity levels between wheat and *Oenothera nad1* introns.

	5' 300 bp[a] (% identity)	domains V and VI[b] (% identity)	transversions	transitions	ratio
nad1a/b	82	88	50	10	5.0
nad1b/c	96	98	4	9	0.4
nad1c/d	93	97	7	10	0.7
nad1d/e	93	98	10	7	1.4

[a] The regions of 300 bp at the 5' end of the wheat *nad1* introns were compared (% identity) with *Oenothera* counterpart sequences: Figure 3A (519-818), Figure 3B (699-998), Figure 3B (2313-2612) and Figure 3C (261-560). [b] Sequences beginning immediately upstream of each domain V to the 3' end of the wheat *nad1* introns were compared (% identity) with *Oenothera* counterpart: Figure 3B (523-613), Figure 3B (2068-2120), Figure 3C (123-201) and Figure 3D (2665-2771).

300 bp sequence of group II introns usually contains at least domain I which includes multiple well-characterized and conserved helical structures and exon binding sites (Michel et al. 1989).

The *nad1*a/b intron sequences are only 82 and 88% identical at their 5' and 3' regions, respectively, between wheat and *Oenothera*, whereas the other *nad1* introns are 93-98% similar in those regions (Table 3). It appears that the low level of conservation in the *nad1*a/b intron is mostly due to a larger number of transversions (50) compared to other *nad1* introns (4-10). Consequently, the ratio of transversions to transitions for the *nad1*a/b intron (5.0) is much higher than that of other introns (0.4-1.4), although these numbers are based on a limited amount of data. Interestingly, these transversions could not be corrected at the RNA level by C-to-U RNA editing events. The reasons behind these peculiar rates of nucleotide substitutions are obscure, but are possibly attributable to various constraints on certain genomic or RNA sequences.

The wheat *nad1*a/b intron sequence was tentatively folded into a structure similar to that proposed for the *Oenothera* counterpart (Wissinger et al. 1991). Numerous nucleotide substitutions and insertion/deletions are located within helical regions and exon binding sites, and these changes appear to destabilize the equivalent domain I structure in wheat (Appendix IVA). Moreover, an insertion of approximately 300 bp has occurred 0.4 kb downstream of *nad1*a in wheat, which would be located within domain I of the structure proposed in *Oenothera*. Nevertheless, farther downstream, a block of more than 200 bp is at least 92% identical between wheat and

Oenothera (Appendix IVB). This region could potentially contain important sequences for the *nad1*a/b intron structure.

3.6.3. Positions of discontinuities in the core structure of *nad1* group II introns.

The presence of group II-like intron structures in the flanking regions of distantly-encoded wheat *nad1* gene pieces is analogous to the chloroplast *rps12* and *psaA* genes. In each case, a *trans*-splicing model has been proposed in which group II intron moieties interact specifically in splicing (Kohchi et al. 1988; Goldschmidt-Clermont et al. 1991). One prediction of the chloroplast *trans*-splicing models is that discontinuities within group II introns are located in non-helical regions which would not affect the stability of intron core structures. The simplest model for *nad1* gene expression would involve one *cis*- and three *trans*-splicing events. In this section, the *nad1* transcript analysis and the comparison of sequences from different plants are used to analyze the *nad1* intron structures.

Because RNA folding is believed to be crucial in the splicing of *nad1* segments, the *nad1*-flanking sequences were analyzed for potential complementary stretches between intron halves at the RNA level. This analysis has led to the identification of several regions of complementarity depending on the constraints that are imposed in the search (such as conserved positions, distances to the exons, number of mismatches; data not shown). The most thermodynamically-stable structures that could be found between the sequences immediately upstream of domain V and those within the 5' intron pieces are presented (Figure 14A, C, D; Figure 14B as in Michel et al. 1989 for the watermelon group IIA intron -

from Chapdelaine and Bonen, 1991). These hypothetical RNA interactions between intron pieces have been selected for illustrative purposes because (1) domain V, unlike domains I-III, is well-defined and enables the accurate positioning of one half of the domain IV helix and (2) it has been shown that *trans*-splicing can still proceed *in vitro* with a discontinuity within loop IV (Jarrell et al. 1988). Therefore, some of these interactions may not represent those that take place *in vivo*, but rather show the type of structure that is thought to be involved in *trans*-splicing. In fact, subsequent comparisons between wheat and *Oenothera nad1* sequences indicate that the 5' half of domain IV shown for the *nad1*a/b intron (Figure 14A) is absent in *Oenothera*. Another conserved complementary sequence is found 122 bp downstream of *nad1*a (not shown), but it is unlikely to represent the functional one because this region of 122 bp is too short to contain domains I-IV. The domains IV shown for the *nad1*c/d and *nad1*d/e introns (Figure 14C and D) are conserved among plants but flanking sequences are not.

Based on transcript analysis of *nad1*-flanking regions (section 3.5) and on sequence conservation between plants, one can predict within which domain of group II intron core structure the positions of discontinuities are most likely to have occurred. In the 3' half of the *nad1*a/b intron, the region of 946 bp between the *rps13* gene and *nad1*b includes several conserved blocks between plants, among which a stretch of approximately 50 nt immediately preceding domain V is likely to be involved in domain IV structure. In the 5' half of the *nad1*a/b intron, long stable transcripts of 1575 nt downstream of *nad1*a could easily include the domains I-IV. In contrast, the S1 nuclease sensitive site located 270 bp

downstream of *nad1*a would delimit a region that is rather short to accommodate domains I-IV. Therefore, the position of the discontinuity in the *nad1*a/b intron could theoretically be located in any of the domains I-IV.

A similar analysis can be made for the *nad1*c/d intron. Upstream of *nad1*d, the breakpoint in homology between wheat and broad bean sequences (Wahleithner et al. 1990) is located only 57 nt upstream of domain V (Figure 3C, arrowhead), which strongly suggests that a DNA rearrangement has occurred within loop IV of the *nad1*c/d intron. Downstream of *nad1*c, transcripts covering the region of approximately 1170 bp could potentially include domains I-IV.

In the case of the *nad1*d/e intron, the sequence downstream of *nad1*d is conserved between wheat and broad bean for 539 bp (Figure 3C), but for only 260 bp downstream of the EBS1 site proposed by Wahleithner et al. (1990; see Figure 3C, position 518-521). This region is rather short to accommodate the rest of domain I and domains II-IV. Therefore, it might be argued that a DNA rearrangement has occurred within the loop of domain I, II or III (Figure 14D, inset) rather than the loop of domain IV.

In summary, it appears that discontinuities in *nad1* group II introns can be located in different domains. While the *nad1*c/d intron appears to be discontinuous in domain IV, the discontinuities in the *nad1*a/b and *nad1*d/e introns could well be located in upstream domains. In the chloroplast *trans*-splicing models, the discontinuity in the *rps12* intron has been proposed to be located in domain III (Kohchi et al. 1988), those of the *psaA*-1 intron (including the *tscA* RNA) in domains I and IV, and that of the *psaA*-2 intron in domain IV (Goldschmidt-Clermont et al. 1991). However, discontinuous

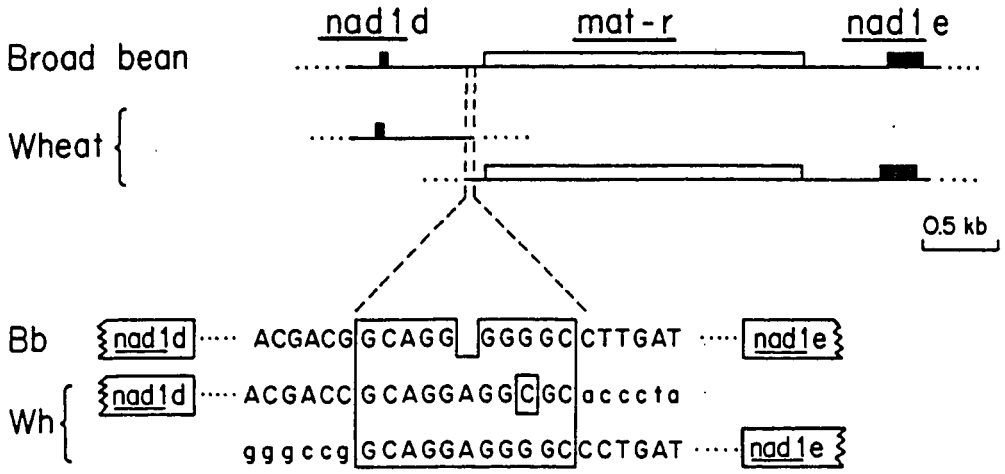
nad1 intron structures are as yet only partially characterized in wheat, and more structural elements must be identified in order to establish precisely where discontinuities are located.

3.7. COMPARISON OF *nad1* GENE ORGANIZATIONS IN DIFFERENT PLANTS.

The organization of the wheat *nad1*d-*nad1*e region differs from that reported for the three dicots broad bean (Wahleithner et al. 1990), *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991). In broad bean and *Oenothera*, *nad1*d and *nad1*e are separated by a classical (continuous) group II intron of approximately 3.1 kb containing a maturase-like (*mat-r*) ORF initially identified by Wahleithner and colleagues (1990). Between wheat and these two dicots, sequences are conserved for 539 bp downstream of *nad1*d (Figure 3C, position 799) and 68 bp upstream of the *mat*-r gene that precedes *nad1*e (Figure 3D, position 147). Beyond these points, homology abruptly ends presumably because of a DNA rearrangement (Figure 17) which separates the wheat *nad1*d and *nad1*e segments by more than 12 kb (see section 3.1), but does not result in the loss of any intron sequences.

Interestingly, the breakpoint occurs at the position of a purine-rich stretch of 11 nucleotides in broad bean and *Oenothera* which is duplicated in wheat; one copy is located downstream of *nad1*d and a second copy upstream of the *mat-r-nad1*e region (Figure 17). Short direct repeats are often present at sites of mtDNA rearrangements in plants (André et al. 1992). In maize for example, the nonchromosomal stripe 3 mutant (NCS3) contains a purine-rich direct repeat of 12 nt which is also located within a group II intron (Hunt and

Figure 17. Comparison of the *nad1*d and *nad1*e exon organization in wheat and broad bean mitochondria. The schematic shows the *mat*-r gene (open bar) and *nad1* segments (solid bars) in wheat and broad bean (Wahleithner et al. 1990). The expanded region depicts in detail the purine-rich sequence (blocked) that is repeated at the breakpoint in wheat. Lower case letters represent non-conserved positions.


nad1d
mat-r
nad1e
Broad bean
Wheat
0.5 kb
Bb
nad1d
ACGACGGCAGGGGGGCCTTGAT
nad1e
Wh
nad1d
ACGACCGCAGGAGGCGCacccta
gggccgGCAGGAGGGGCCCTGAT
nad1e

Newton, 1991), approximately 600 bp upstream of the 3' splice site (likely to be in the domain IV). The presence of similar purine-rich stretches at sites of breakpoints in homology suggests that they are involved DNA rearrangement events, either as sites mediating homologous recombination or as repeats generated as a consequence of double-stranded DNA repair.

Thus, it appears that a mtDNA rearrangement within the *nad1*d/e group II intron has occurred after the divergence of monocots and dicots from a common ancestor some 200 Myr ago (Wolfe et al. 1989). Surprisingly, the *nad1*d/e intron is also discontinuous in petunia (a dicot), but the *mat-r* gene, instead of being located upstream of *nad1*e like in wheat, is located downstream of *nad1*d region (Conklin et al. 1991). The position of the discontinuities in wheat and petunia may occur in the same domain (most likely domain IV) but at different ends of the *mat-r* gene. They could alternatively be located within different domains of the *nad1*d/e intron. Moreover, the three different *nad1* gene organizations in wheat, *Oenothera* and petunia must be due to two independent DNA rearrangements; one having occurred after the divergence of monocots and another one after the divergence of petunia and *Oenothera* (two dicots) lineages.

CHAPTER 4. GENERAL DISCUSSION.

4.1. EXPRESSION OF THE *nad1* GENE IN WHEAT MITOCHONDRIA.

The unusual organization of the *nad1* gene in the wheat mitochondrial genome has raised questions as to whether it is functional, and if so how it is expressed and how such an unusual structure arose. There are several convincing lines of evidence that support the idea that this gene is functional. (1) The wheat *nad1* sequences together are capable of encoding a polypeptide that is 45 and 84% identical, at the amino acid level, to its human and liverwort counterparts, respectively. (2) The five *nad1* exons are actively transcribed. (3) They are correctly linked and extensively edited at the RNA level, and RNA editing increases the predicted amino acid sequence similarity with non-plant organisms. (4) Structures reminiscent of group II introns flank *nad1* coding sequences. (5) Similar organizations of the *nad1* gene are seen in other plants. All these observations favour the hypothesis that the NAD1 protein is produced from this gene in wheat mitochondria. The similarity of the anticipated polypeptide with mitochondrial counterparts of non-plant organisms is higher than with the chloroplast one, which favours the hypothesis that the *nad1* gene is endogenous to the ancestral mitochondrion rather than having been transferred recently from the chloroplast.

4.2. PATTERNS OF RNA EDITING IN PLANT MITOCHONDRIA.

One interesting aspect of RNA editing in the *nad1* gene is the asymmetric distribution of the editing sites along the *nad1* coding sequences and how it might have arisen. The study of certain other mitochondrial genes in wheat, namely *nad4* and

nad5 (Lamattina and Grienenberger, 1991; De Souza et al. 1991), as well as *nad2* and *nad5* in *Oenothera* (Binder et al. 1992; Knoop et al. 1991), has also revealed asymmetric distributions of RNA editing sites. Interestingly, it has been observed that *nad5*b has no editing sites in wheat compared to at least 15 in *Oenothera*. This raises the possibility that the asymmetric distributions of RNA editing sites in mitochondrial genes could be due to the integration of edited reverse transcripts in genomic DNA. In the *nad1* gene, such events could have occurred in *nad1*a prior to the monocot/dicot divergence. This might explain how the low frequency of RNA editing sites in wheat, *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991) has occurred.

Plant mitochondrial RNA editing is known to occur at different sites in different species. In this respect, it is not surprising that the *nad1* gene would contain only approximately half of the sites that are shared by wheat, *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991). One wonders, however, how the asymmetric distribution has been maintained in all three plants through evolution. These observations may suggest that sites are acquired or lost non-randomly, and that certain regions may be particularly susceptible for editing. This is also consistent with the view that the recognition of RNA editing sites in plant mitochondria requires signals provided by sequences in the vicinity of those sites. Such a model could involve a mechanism similar to either the mammalian or trypanosomal ones, which are in fact very different. In mammalian apolipoprotein B transcripts, a stretch of 11 nt beginning 5 bp downstream of the C-to-U editing site is crucial for editing (Hodges and Scott, 1992). In trypanosome mitochondria,

guide RNAs are complementary to sequences located downstream of sites to be edited (Blum et al. 1990).

One of the editing sites found in the wheat *nad1* gene, which is also present in *Oenothera* and petunia counterparts, is located at the same position in the chloroplast *ndhA* counterpart in maize (Maier et al. 1992). These RNA editing sites might have occurred at the same position only by coincidence, but it is tempting to suggest that they reflect similarities in mechanisms between plant chloroplasts and mitochondria and that common nuclear factors could be used by both editing machineries. It may also reflect certain constraints for maintaining RNA editing and that evolutionary pressure may exist at the nucleotide sequence level (as well as at the protein level), perhaps due to RNA structure or recognition motifs required for editing.

The presence of partially edited transcripts in plant mitochondria could potentially lead to the production of heterogenous and non-functional polypeptides. Because the initiation codon is generated by RNA editing in wheat *nad1* mRNAs, as for several chloroplast mRNAs (Hoch et al. 1991), RNA editing seems obligatory for initiation of *nad1* translation. It has also been observed that the degree of editing in *nad3-rps12* transcripts of the polysomal fraction (Gualberto et al. 1991) and in spliced transcripts (Yang and Mulligan, 1991) is higher than in precursors. It is therefore likely that, in spite of the fact that transcription and translation are not physically separated in the mitochondria, incompletely edited transcripts are unavailable for translation.

4.3. A *TRANS*-SPLICING MODEL FOR *nad1* GENE EXPRESSION.

The five *nad1* exons are scattered at four distant genomic sites in the wheat mitochondrial genome. The assembly of the *nad1* gene pieces at the RNA level must require an unusual splicing process to generate functional mRNA molecules. A number of observations support the view that the wheat *nad1* coding segments are expressed through *trans*-splicing.

Firstly, the longest introns yet identified in any organellar gene are less than 4 kb in length, whereas the wheat *nad1* coding segments are separated by distances of at least 20, 7 and 12 kb. Moreover, the juxtaposition of other genes upstream of internal *nad1* pieces (e.g. *atp6/rps13* genes preceding *nad1*b) would place those genes within introns, an unorthodox organization in organellar genomes. The physical map of the wheat mitochondrial genome (Appendix II, Lejeune and Quétier, 1988) predicts that additional genes are located between *nad1* gene segments.

Secondly, the wheat *nad1* coding segments are flanked by group II-like intronic features. Only *nad1*b and *nad1*c are separated by a typical continuous 1.4 kb group II intron. *Nad1*c/d and *nad1*d/e intron structures, although not well defined, are also clearly related to group II introns. The *nad1*a/b intron falls in a special category since only the stem of domain V allows the classification of this intron as a member of group II (discussed in more detail in section 4.4). This unusual gene organization is reminiscent of that of two *trans*-split chloroplast genes, namely, the *rps12* gene in plants (Fukuzawa et al. 1986; Zaita et al. 1987; Hildebrand et al. 1988) and the *psaA* gene in *C. reinhardtii* (Kück et al. 1987). The *nad1* gene organization is also similar to those of the *nad5* and *nad2* genes in *Oenothera* (Knoop et al. 1991;

Binder et al. 1992) and the *nad5* gene in wheat (De Souza et al. 1991). The *nad2* and *nad5* genes require one and two *trans*-splicing events, respectively, for their expression.

Thirdly, the abundant wheat *nad1* precursor transcripts extend only 1-2 kb beyond the individual *nad1* coding regions, downstream of *nad1*a, *nad1*c and *nad1*d, and upstream of *nad1*b. In addition, there is no evidence for long precursor transcripts which could contain all *nad1* exons. Consequently, Northern blot hybridizations and S1 nuclease analysis of *nad1* steady state transcripts suggest that the *nad1* coding regions are located on different precursor transcripts. In *Oenothera*, the analysis of *nad2* transcripts led to similar conclusions (Binder et al. 1992). The wheat *nad1* transcripts are long enough to encompass group II intron "halves" of normal sizes (1-3 kb) that would flank the *nad1* coding segments.

Fourthly, the *nad1* gene structure differs radically among wheat, *Oenothera* (Wissinger et al. 1991) and petunia (Conklin et al. 1991) with respect to the *nad1*d/e intron. This situation is most easily explained by the occurrence of DNA rearrangements upstream of the *mat-r* gene in wheat and downstream of the *mat-r* gene in petunia. In wheat, the *mat-r* gene would be located in the 3' half of the *nad1*d/e intron, whereas in petunia it would be located in the 5' half.

Although those arguments strongly support a *trans*-splicing model, four unusual *cis*-splicing models have been considered theoretically by Choquet et al. (1988), and are discussed here in the context of the plant mitochondrial *nad1* gene. One which invokes long primary transcripts is excluded because, in petunia, the *nad1*d and *nad1*b-*nad1*c regions are located on opposite strands and 95 kb apart (Conklin et al. 1991). In a second model, the possibility is raised that low

level DNA molecules may exist. However, such molecules have not been detected in wheat either by Southern blot hybridizations or by PCR experiments. The two other *cis*-splicing models suggest that intron pieces may be covalently linked at the RNA level by priming (precursors from the upstream region being used to initiate transcription of the downstream region) or by end-to-end ligation of precursor RNAs from the different regions. Again, such RNA molecules are not detected by PCR; however, because of their size and their relatively low abundance, such splicing intermediates may be difficult to detect. Perhaps the strongest argument against these two latter *cis*-splicing models is that the splicing of *nad5*b to *nad5*c occurs prior to processing of *nad1*e-*nad5*c precursor transcripts. Therefore, *trans*-splicing is the only model that is consistent with the experimental results that have been obtained in different plants.

4.4. THE *nad1*a/b INTRON STRUCTURE.

Of the four *nad1* intervening sequences, the *nad1*a/b intron is the most unusual with respect to its degree of divergence among plants and to the apparent absence of group II intron features. The domain V stem is in fact the only structure that can be recognized as a feature of group II introns. This raises the question as to whether (1) this intron is a degenerate group II intron (i.e. with some functions now being provided *in trans*), (2) this intron became specialized or more interestingly (3) this intron could be a member of a yet un defined class of introns. The identification of phylogenetically conserved regions of this intron will help to distinguish among these possibilities.

The absence of a bulging A at the position -7 or -8 from the 3' splice site in the *nad1*a/b intron in all plants is somewhat puzzling. Starting from the 3' splice site, the first possible nucleotide that could serve as bulging A is at position -11 in monocots, and -10 in dicots. The 5-bp deletion/insertion in this region between monocots and dicots is also unexpected and reflects the poor conservation that has been described at the 3' splice site of the *nad1*a/b intron. Interestingly, in pea, a 4-nt repeat (AACT, Figure 16C) could have been involved in this deletion/insertion event.

The bulging A, which is directly involved in the long-range interaction leading to lariat formation, is among the best conserved positions of group II introns, including *Euglena* chloroplast group II-like introns (Michel et al. 1989). In fact, in the list of 68 group II introns presented by Michel et al. (1989), there are only three exceptions to this rule: the $tRNA^{Val}$ chloroplast intron (in which a G is seen instead of A), the chloroplast ORF203 (intron 2) and the *nad1*b/c intron in watermelon mitochondria. Domain VI of the *nad1*b/c intron in wheat is identical to that of watermelon and both plants have Cs at positions -7 and -8. In the *nad1*a/b intron, there is neither A nor G at these positions of any plants. It should be remembered, however, that lariat formation is dispensable *in vitro*. It normally serves to stabilize the tertiary structure of the intron-downstream exon splicing intermediate, and results in the acceleration of the second transesterification step (Jacquier and Jacquesson-Breuleux, 1991).

Another unexpected result from the study of the *nad1*a/b intron concerns the identification of an S1 nuclease sensitive site located 270 bp downstream of *nad1*a. The stable

transcripts observed in this region and those of >0.6 kb upstream of *nad1*b are rather short (although not too short) to accommodate a group II intron of normal size. It should also be kept in mind that *nad1*a and *nad1*b flanking sequences may include only part of the *nad1*a/b intron, and that other short RNAs (such as the *tscA* gene product, see below) might be necessary to complete the intron structure.

However, it seems doubtful that transcripts containing the 5' half of the *nad1*a/b intron end only 270 bp downstream of *nad1*a because (1) the first 300 bp downstream of *nad1*a are not well conserved (Appendix IV), and (2) other highly conserved sequences are located farther downstream which are likely to contain important sequences for the *nad1*a/b intron (Appendix IA and IVB). This S1 nuclease sensitive site could be due to relaxation of a stretch of approximately 20 Ts that is located at this position, or to RNA editing in this region. An alternative hypothesis is that a structure different from other group II introns exists; one in which non-linear (e.g. lariat) structures would prevent hybrid formation in this region of the *nad1*a/b intron. These hypotheses could be tested directly by using uniformly labelled probes that overlap this region: S1 products that cover only the region downstream of the S1 sensitive site should be observed.

4.5. EXPRESSION AND PROCESSING OF *nad1* TRANSCRIPTS.

Since multiple *trans*-split gene structures have been conserved over long evolutionary periods in both chloroplasts and mitochondria, they do not seem to interfere with normal gene functions. However, certain conditions must be fulfilled for DNA rearrangements within introns not to be lethal. One is that translocated segments must be placed under appropriate

transcriptional control. In this regard, the *nad1*e-*nad5*c arrangement in wheat differs from ORFX-*nad5*c in *Oenothera*, and *nad1*d-*nad1*e-*nad5*a-*nad5*b in *Arabidopsis* (Knoop et al. 1991). The different organizations of the *nad1* and *nad5* gene segments among angiosperms illustrate that gene pieces are often translocated upstream or downstream of actively transcribed regions. The fact that breakpoints in homology are located within approximately 25 bp of transcript termini downstream of *nad1*c and *nad1*e also suggests that there are functional constraints for the conservation of the sequences between those sites and the coding regions.

It has been debated in the literature whether single and double stem-loop structures are involved in transcription termination or in RNA processing in plant mitochondria (reviewed by Gray et al. 1992). Based on the thermodynamic stability and the conservation between plants, the most convincing double stem-loop structure was found in the vicinity of the transcript termini downstream of *nad1*e. Transcripts extend further downstream of each of the four major 3' termini that were detected in this work. This favours the view that *nad1* transcript termini are generated by RNA processing. It will be interesting to determine the factors that are involved in the stability of transcript termini that are located within introns. To clarify these questions, more transcript analysis will be necessary both *in vitro* and *in vivo*.

The presence of multiple genes that each require *trans*-splicing for their expression adds to the complexity of RNA processing in plant mitochondria. The cotranscription of *nad1*e and *nad5*c highlights this point in two different ways. (1) The *nad1*e-*nad5*c precursor transcripts contain parts of two

different mRNAs. (2) Splicing of the *nad5*b/c and *nad5*c/d introns may occur before cleavage between *nad1*e and *nad5*c, which suggests that *nad1*e does not interfere with the folding of *nad5*b/c and *nad5*c/d introns. It would be interesting to examine if these processing and splicing events occur in a predetermined order as suggested by De Souza and colleagues (1992) for the *nad5* gene, and if a particular precursor transcript can be involved in the production of both *nad1* and *nad5* mRNAs. In the *nad1* gene, transcripts in which the *nad1*b/c and *nad1*c/d introns had been removed but not the *nad1*a/b intron, have been detected. The relatively low level of transcripts arising from the *nad1*d and *nad1*e regions compared to the *nad1*a and *nad1*b-*nad1*c regions may reflect that splicing rates of the first two *nad1* introns are slow, or that splicing of the *nad1*c/d and *nad1*d/e introns must occur before that of the *nad1*a/b and *nad1*b/c introns.

4.6. FLUIDITY OF THE *nad1* GENE STRUCTURE.

The presence of three different structures of the *nad1*d-*nad1*e region(s) among plants strongly supports the idea that *trans*-splicing in the intron between *nad1*d and *nad1*e is the result of recent DNA rearrangements within a previously continuous group II intron, rather than being an ancient form of intron structure. It is worth mentioning that *trans*-split gene structures could, alternatively, arise from the duplication of certain portions of the gene and from the subsequent inactivation (or loss) of the original *cis*-copy by DNA rearrangement. Such a model could explain the presence of the ORF36 downstream of *nad1*c in watermelon (Stern et al. 1986) which is likely to correspond to a duplicated portion of *nad1*e in that plant. One would expect that the *cis*-copy would

offer a selective advantage unless *trans*-splicing is not too deleterious.

In fact, little is known about the positive or negative impact that *trans*-splicing could have on efficiency of gene expression. One argument that favours the view that *trans*-splicing is not too unfavourable is that most *trans*-split gene structures have occurred early in plant evolution. Except for the ones observed in the *nad1*d/e introns of wheat and petunia, other plant mitochondrial discontinuous introns in *nad1* (this work, Wissinger et al. 1991; Conklin et al. 1991), *nad2* (Binder et al. 1992) and *nad5* (Knoop et al. 1991, De Souza et al. 1991) genes seem to have occurred before the monocot/dicot divergence, approximately 200 Myr ago. The *trans*-split chloroplast *rps12* gene organization is also a very stable one, being present in both liverwort and tobacco, two land plants that have had separate lineages for some 350-400 Myr (Wolfe et al. 1989). Similarly, the fragmented *psaA* gene structure observed in *C. reinhardtii* is also present in widely divergent *Chlamydomonas* species (Palmer, 1991).

The intronless structure of the *nad1* gene in liverwort (Oda et al. 1992) could either result from intron losses in liverwort or from intron acquisitions in angiosperm mitochondria. An overview of the distribution of mitochondrial introns in different organisms (Table 4) supports the hypothesis that such acquisitions or losses of introns have been frequent during mtDNA evolution, and that angiosperm mitochondrial introns, which are all members of group II, are generally located at different sites from those of other organisms, including liverwort. For example, the *coxI* gene contains six group I and three group II introns in liverwort (Oda et al. 1992), sixteen introns in *Podospora* (Cummings et

Table 4. Distribution of introns in mitochondrial protein-coding genes of *Podospora*, liverwort and angiosperms, as well as in chloroplast homologues.

	Podospora anserina[a]	*Marchantia polymorpha*[b]	Angiosperms[c]	Chloroplast homologues[d]
Cytochrome oxidase				
coxI	16	9	0	-
coxII	2	2	0-2[e,g]	-
coxIII	0	2	0	-
Cytochrome b				
cob	2	3	0	+[j]
ATP synthase				
atp6	1	0	0	1
atp8	0	-	?	0
atp9	-	1	0	0
atpA	-	2	0	0
NADH dehydrogenase				
nad1	4	0	4[i]	1
nad2	0	1	4[i]	1
nad3	1	1	0	0
nad4	1	1	1-3[e]	0
nad4L	2	0	?	0
nad5	3	1	4[i]	0
nad6	0	0	0	0
nad7	-	2[f]	4	0
Ribosomal proteins				
rps3	-	0	1[g,h]	0
rps7, rps19	-	0	0[g]	0
rps12	-	0	0[g]	2[i]
rps13	-	0	0[g]	-
rps14	-	1	0[g]	0
rpl16	-	0	0[g]	0-1[e]
rps1, rps10, rpl5, rpl6	-	0	?	-
rps2, rps4, rps8, rps11	-	0	?	0
rpl2	-	1	?	0-1[e]

[a] Cummings et al. (1990); [b] Oda et al. (1992) [c] review by Bonen (1991); see also Lippok et al. (1992); Binder et al. (1992); Lamattina and Grienenberger (1991); Gass et al. (1992); Knoop et al. (1991); Linda Bonen, unpublished data. [d] *Marchantia polymorpha* (Ohyama et al. 1986) and tobacco (Shinozaki et al. 1986), reviewed by Palmer, (1991). [e] Presence of optional introns. [f] Pseudogene. [g] Genes not present in all plant examined. [h] The *rps3* intron is located in a fused upstream ORF (Hunt and Newton, 1991). [i] Genes for which *trans*-splicing models have been proposed. [j] Homology to apocytochrome *b6* and subunit 4 of the cytochrome *b/f* complex (Heinemeyer et al. 1984).

al. 1989), and has no introns in angiosperms. Conversely, while genes for NADH dehydrogenase subunits are intron-rich in angiosperms, they generally contain a small number of introns in liverwort. This variation in the distribution of mitochondrial introns among land plants contrasts with the distribution of chloroplast introns which is almost identical between liverwort and angiosperms (Palmer, 1991), and supports the view that these introns became widespread recently in mitochondrial genomes (Palmer and Logsdon, 1991).

The organization of the *nad1* gene in the wheat mitochondrial genome shows that DNA rearrangements alter not only gene order, but also gene structure. In *C. reinhardtii* mitochondria, genes for rRNAs also show radical intragenic rearrangements that have considerably altered gene structure (Boer and Gray, 1988b). It may be significant that both *C. reinhardtii* and plants have mitochondrial genes with homology to reverse transcriptases (Boer and Gray, 1988a; Brennicke and Schuster, 1987; Wahleithner et al. 1990). Although reverse transcriptase activity has not been demonstrated directly in plant mitochondria, it has been implicated in the transfer of an edited copy of the *coxII* gene to the nucleus in mung bean (Nugent and Palmer, 1991). Such activity may also have a role in the frequent acquisition or loss of RNA editing sites in plant mitochondria. It has been suggested that DNA rearrangements could occur by the integration of reverse transcribed sequences and subsequent recombination between duplicated sequences in plant mitochondrial genomes (André et al. 1992). Reverse transcriptase activity may be a key to part of the puzzle of plant mtDNA evolution and thus indirectly have implications for gene expression.

4.7. *TRANS*-INTERACTIONS IN GROUP II INTRONS.

One fundamental problem for the processing of *trans*-split genes is that the RNA pieces have to find their correct partners. The specific recognition of intron pieces is further complicated by the presence of two additional discontinuous introns in the wheat *nad5* (De Souza et al. 1991) and (probably) *nad2* (Binder et al. 1992) genes. There is no evidence for the presence of stable transcripts containing mis-matched *nad1* segments, which suggests that intron pieces can interact precisely and specifically for splicing or that the mis-matched transcripts are rapidly degraded.

There are several types of interactions that can be envisioned for the proper recognition of intron moieties. In the models proposed for the chloroplast *psaA* (Goldschmidt-Clermont et al. 1991) and *rps12* genes (Kohchi et al. 1988), base-pairing between intron halves within the discontinuous domains III or IV is viewed as a strong *trans*-interaction since, in each case, long complementary sequences are seen. Such structures could be present in *nad1* introns but have not been identified because core intron structures are as yet unknown, but no long stretches of perfect complementarity could be found.

It is worth pointing out that other interactions are likely to occur between intron halves, and that base pairing within the discontinuous domain could play a relatively minor role. For example, in the *cis*-split *nad1*b/c intron the domain IV helix appears to be only 3 bp long (Michel et al. 1989). In *in vitro trans*-splicing studies, the domain IV can be removed completely without abolishing splicing, but domain V is crucial (Jarrell et al. 1988). Although the role of domain V in the tertiary folding of group II introns is yet unclear, it

is likely that it plays a key role in bringing 5' and 3' splice sites in close proximity (Michel et al. 1989; Jacquier and Jacquesson-Breuleux, 1991). It has been shown recently that a binding site for domain V is located within domain I (Koch et al. 1992). Such interaction could potentially act as a primary factor of recognition between the two intron halves.

Trans-splicing might be expected to differ from *cis*-splicing in requiring either additional specialized secondary structural interactions or extra machinery (such as small RNAs and/or proteins) for the correct recognition and alignment of intron pieces (particularly in a milieu where multiple *trans*-splicing events occur). Indeed, in the case of the *C. reinhardtii trans*-split *psaA* gene, at least fourteen nuclear loci are required for the processing of *psaA* introns (Choquet et al. 1988; Herrin and Schmidt, 1988). Moreover, the chloroplast *tscA* gene product appears to be a short RNA that would be directly involved in the formation of a typical group II intron structure (Goldschmidt-Clermont et al. 1991). These studies on the chloroplast *psaA* gene suggest that all intron structures are not necessarily transcribed from the vicinity of coding regions. This possibility has to be kept in mind in the analysis of the wheat mitochondrial *nad1* introns.

4.8. INCIDENCE OF *TRANS*-SPLIT GENE STRUCTURES.

One may also question why the phenomenon of *trans*-splicing has not been observed more frequently in nature. Of more than 68 group II introns characterized in a wide variety of organisms (Michel et al. 1989), only two chloroplast genes had been identified to be *trans*-split prior to the present study. In plant mitochondria, in addition to the three discontinuous introns in the *nad1* gene (this work, Wissinger

et al. 1991, Conklin et al. 1991), there are two other examples in the *nad5* gene (Knoop et al. 1991, De Souza et al. 1991) and one in the *nad2* gene (Binder et al. 1992). This large number of discontinuous introns (six out of 22 introns) may simply reflect the recombinogenic nature of plant mtDNA. However, *trans*-split genes appear to be absent from fungal mtDNA in which rearrangements are also frequent. Therefore, the numerous DNA rearrangements in plant mitochondria cannot alone be held responsible for the high ratio of discontinuous introns.

Trans-splicing may have a selective advantage by providing a more refined or efficient control of gene expression. In trypanosomes, this view has been considered for nuclear *trans*-spliced leaders but this seems to be a somewhat complicated process to provide mRNAs that arise from multicistronic transcripts with a 5' cap (Laird, 1989). This model appears even more unlikely in plant mitochondria or chloroplasts since *trans*-splicing is involved only in isolated cases. Moreover, for the wheat mitochondrial *nad1* gene, the excess of certain precursor transcripts makes the issue of efficiency less plausible.

Another hypothesis is that particular subclasses of introns (as yet undefined) could be more amenable for *trans*-splicing because they require different *trans*-acting factors for example. Interestingly, the four *nad1* introns share as many as 16 nt positions in common out of 34 nt in domain V. Oda and colleagues (1992) noticed that the liverwort mitochondrial group II introns can be grouped into four families, two of which are closely related. This is consistent with the view that most plant mitochondrial introns have a

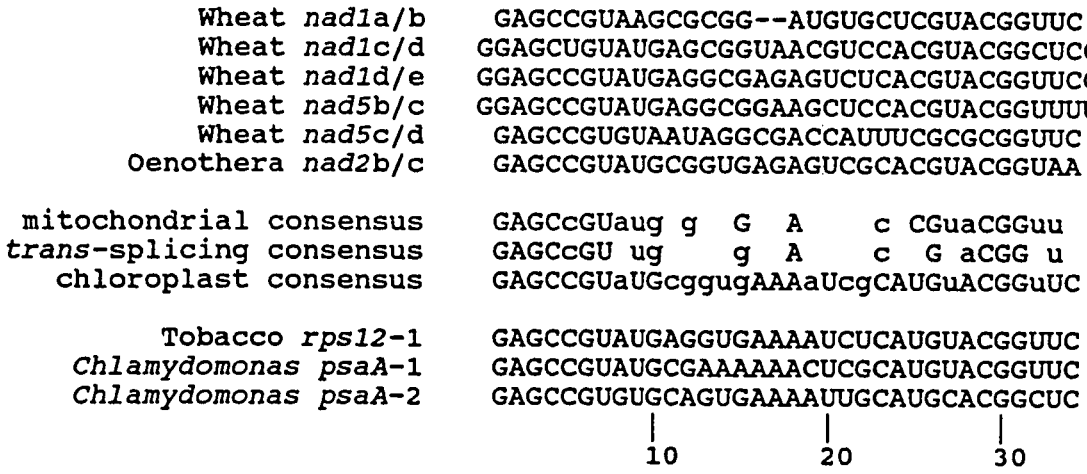
recent origin (discussed above), and have possibly evolved from a small number of ancestors.

To test further the idea that group II introns involved in *trans*-splicing may share some specific features, an alignment of domain V sequences of the nine discontinuous group II introns found so far in mitochondria and chloroplasts is presented (Figure 18). In spite of the fact that these introns belong to different subgroups and that the *nad1*a/b intron is atypical in several respects, there are 11 invariant positions among the nine introns and 7 other positions are found in eight of these introns. It may be significant that one position of the group II consensus (Michel et al. 1989) is absent in four mitochondrial discontinuous introns (position 16), and that as many as 5 invariant nucleotides in discontinuous introns are not part of the general consensus. Moreover, the chloroplast *rps12*-1 domain V is nearly identical to the wheat *nad1*d/e intron (only 4 mis-matches), and is also closely related to the wheat *nad5*b/c and *Oenothera nad2*b/c introns (7 mis-matches each). In fact, the domain V of tobacco *rps12*-1 intron in the chloroplast is as similar to that of mitochondrial discontinuous intron as it is to its liverwort chloroplast counterpart (6 mis-matches, Kohchi et al. 1988). However, these observations must be interpreted cautiously since domain V is highly conserved, which may lead to convergence of sequences.

4.9. **FUTURE DIRECTIONS.**

With the identification of at least 22 group II introns in plant mitochondrial genes (Table 4), introns appear to contribute substantially to genome sizes. However, the analysis of sequences flanking known genes supports the idea

Figure 18. Alignment of domain V sequences of discontinuous introns. The sequences shown above are the consensus ones from the mitochondrial *nad1* (this work), *nad5* (De Souza et al. 1991) and *nad2* (Binder et al. 1992) introns. The sequences shown below are from the chloroplast *rps12* (Zaita et al. 1987) and *psaA* (Goldschmidt-Clermont et al. 1991) introns. Dashes represent the loop of 17 nt in *nad1*a/b domain V. Lower case letters depict the consensus positions in 8 out of 9 introns. The numbers indicate the position in the domain V structure.


Wheat nad1a/b GAGCCGUAAGCGCGG--AUGUGCUCGUACGGUUC
Wheat nad1c/d GGAGCUGUAUGAGCGGUAACGUCCACGUACGGCUCC
Wheat nad1d/e GGAGCCGUAUGAGGCGAGAGUCUCACGUACGGUUCC
Wheat nad5b/c GGAGCCGUAUGAGGCGGAAGCUCCACGUACGGUUUU
Wheat nad5c/d GAGCCGUGUAAUAGGCGACCAUUUCGCGCGGUUC
Oenothera nad2b/c GAGCCGUAUGCGGUGAGAGUCGCACGUACGGUAA
mitochondrial consensus GAGCcGUaug g G A c CGuaCGGuu
trans-splicing consensus GAGCcGU ug g A c G aCGG u
chloroplast consensus GAGCCGUaUGcggugAAAaUcgCAUGuACGGuUC
Tobacco rps12-1 GAGCCGUAUGAGGUGAAAAUCUCAUGUACGGUUC
Chlamydomonas psaA-1 GAGCCGUAUGCGAAAAAACUCGCAUGUACGGUUC
Chlamydomonas psaA-2 GAGCCGUGUGCAGUGAAAAUUGCAUGCACGGCUC
10
20
30

that the presence of non-coding spacer DNA is the main cause of the large genome sizes and of the size variation (Gray, 1989a). It is possible that the identification of additional genes is hindered by the presence of short *trans*-split exons (such as *nad5*c of 22 bp). Nevertheless, probes from liverwort mtDNA (Oda et al. 1992) will be helpful in the isolation of additional genes in angiosperm mitochondria.

Because wheat *nad1* introns are as yet only partially characterized at the RNA level and because they have several unusual structures, it will be interesting to find continuous counterparts of *nad1*a/b and *nad1*c/d introns to determine when the different *nad1* gene organizations occurred during evolution, and to establish the complete secondary structures of these introns. A continuous form of the *nad1*a/b intron might help to clarify why this intron is so poorly conserved. Since the *nad1* gene has no intron in *C. reinhardtii* (Boer and Gray, 1988a) and liverwort (Oda et al. 1992), gymnosperms or early-diverging land plants might be potential candidates.

It is notable that 19 of the 22 (86%) introns are located in genes for NADH dehydrogenase subunits, which represent only 35% of the plant mitochondrial genes identified so far (see Table 4). It is possible that introns could play a role in plant mitochondrial gene regulation. This is an attractive model considering that plant mitochondria contain rotenone-resistant NADH dehydrogenase and cyanide-resistant terminal oxidase activities which are not present in animal mitochondria. Major variations in these activities have been observed during early stages of germination in several plants (Douce and Neuburger, 1989). In this regard, it is also interesting that the *nad1* and *nad5* genes have different transcript profiles after 24 hours and 6 days of germination

(Bonen, 1987; Bird and Bonen, unpublished). The use of probes from the four regions of the *nad1* gene at different stages of development may represent an interesting approach. The study of the *mat-r* gene expression during wheat development is also of interest. Using S1 nuclease mapping and anchored PCR methods, it may be possible to determine how *mat-r* transcripts are processed and how its translation is initiated.

A more detailed S1 nuclease analysis of the *nad1* gene may also help us to understand better certain regulatory aspects of plant mitochondrial gene expression. First, upstream of the *nad1*a, *nad1*d and *nad1*e coding regions, it will be important to examine whether *nad1* stable transcripts arise from primary transcripts or processing events. Second, in the *nad1*b upstream region, complex processing events are expected. The preliminary S1 nuclease mapping results for the *rps13-nad1*b interval should be pursued considering that Northern blot hybridizations suggest that the *rps13* gene is silent after 24 hours and 6 days of germination (Bonen, 1987). Third, the short abundant transcript observed in the *nad1*e region should be investigated further to identify its 5' and 3' termini, and to examine its relative abundance at different stages of development. Fourth, the ratio of precursor to mature transcripts appears to be much higher for the *nad1*a and *nad1*b-*nad1*c regions than for *nad1*d and *nad1*e regions. It would be of interest to determine whether this is due to different promoter strengths, or to different splicing rates.

The identification of RNA editing sites within *cis*- and *trans*-splicing intron structures raises fundamental questions about the temporal relationship between editing and splicing processes. It has also been suggested that RNA editing could play a role in the regulation of the splicing process

(Wissinger et al. 1991; Binder et al. 1992). However, the contribution of RNA editing to plant mitochondrial intron structure remains difficult to define. Of the six sites that occur in the domain VI region (this work; Wissinger et al. 1991; Conklin et al. 1991; Knoop et al. 1991), two appear to weaken the helical structure, three have no clear contribution to the stability of the helix one way or the other, and only one stabilizes the structure but is present in only one of ten clones (appendix V). It is worth noting that they are all located between positions -6 and -11 from the 3' splice site.

The population of precursor transcripts may not be representative of the molecules that take part in splicing, although these stable precursor transcripts are extensively edited within the coding regions. To address these issues, the analysis of excised introns would allow us to identify the editing sites (if any) that are absolutely required for splicing. Because the analysis of such splicing by-products is difficult *in vivo*, one interesting avenue would be to study the self-splicing properties of plant mitochondrial introns *in vitro*. Although *in vitro* self-splicing properties of several group II introns of other systems have been extensively studied (Cech, 1986), preliminary work would be necessary to determine *in vitro* self-splicing conditions of certain plant mitochondrial introns. Eventually, the disruption of continuous introns as well as the reconstruction of discontinuous introns could be analyzed. It would also be possible to determine how the presence or absence of editing sites affects splicing rates.

4.10. CONCLUDING REMARKS.

This work clearly demonstrates that short gene segments can be dispersed around the wheat mitochondrial genome and yet still comprise functional coding entities that are expressed through complex RNA processing. The presence of group II introns is intimately associated with this phenomenon and could play an important evolutionary role in the creation of novel genes from pre-existing coding domains or the conversion of long genes into shorter autonomous ones. This is an attractive model if one considers that group II intron ancestors may have spread rapidly at certain times during evolution. The ability of such introns to undergo *trans*-splicing could provide a selective advantage in the acquisition of novel gene functions.

During the course of this work, evidence has emerged that the *trans*-split *nad1* gene structure is not unique in plant mitochondria; the *nad5* and *nad2* genes also require *trans*-splicing involving group II intron structures for their expression (Knoop et al. 1991; De Souza et al. 1991; Binder et al. 1991). The identification of group II introns that can be divided into halves or pieces supports the idea that the group II introns represent a mosaic of modules that can assemble into functional entities. Complex *trans*-interactions between RNA molecules are also required for the assembly of rRNAs in *C. reinhardtii* mitochondria (Boer and Gray, 1988b). This illustrates that mosaic RNA structures are functional in contemporary organelles (Butow and Perlman, 1991). The growing evidence that group II introns can be viewed as an assembly of independent modules, that normally act in *cis*, adds to their similarities with nuclear pre-mRNA introns (Jacquier, 1990).

Our notion of gene expression has started from the simple model in which DNA is transcribed into RNA, and RNA is translated into proteins. The picture has become more complex with the discovery of intervening sequences and alternative splicing; the primary transcription product is no longer the RNA that is translated. *Trans*-splicing extends by a further step the complexity of gene expression by requiring multiple promoters and independently-produced transcripts. The phenomenon of RNA editing has also increased the importance of RNA maturation in the process of gene expression. RNA (or cDNA) sequences are now necessary to deduce amino acid sequences and this has strong practical implications. The wheat mitochondrial *nad1* gene with three introns involved in *trans*-splicing, one in *cis*-splicing, 23 RNA editing sites (including one in the initiation codon), as well as RNA editing within at least one intron depicts strikingly the complexity of gene expression.

REFERENCES

Agabian N (1990) *Trans* splicing of nuclear pre-mRNAs. Cell 61:1157-1160

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-465

André C, Levy A, Walbot V (1992) Small repeated sequences and the structure of plant mitochondrial genomes. Trends Genet 8:128-132

Araya A, Domec C, Begu D, Litvak S (1992) An *in vitro* system for the editing of ATP synthase subunit 9 mRNA using wheat mitochondrial extracts. Proc Natl Acad Sci USA 89:1040-1044

Attardi G, Schatz G (1988) Biogenesis of mitochondria. Ann. Rev Cell Biol 4:289-333

Augustin S, Müller MW, Schweyen RJ (1990) Reverse self-splicing of group II intron RNAs *in vitro*. Nature 343:383-386

Bachl J, Schmelzer C (1990) Effect of deletions at structural domains of group II intron bI1 on self-splicing *in vitro*. J Mol Biol 212:113-125

Bendich AJ, Smith SB (1990) Moving pictures and pulsed-field gel electrophoresis show linear DNA molecules from chlorplasts and mitochondria. Curr Genet 17:421-425

Bennoun P (1982) Evidence for a respiratory chain in the chloroplast. Proc Natl Acad Sci USA 79:4352-4356

Bibb MJ, Van Etten RA, Wright CT, Mark W, Walberg MW, Clayton DA (1981) Sequence, gene organization of Mouse mitochondrial DNA. Cell 26:167-180 rem p 78

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

Bland MM, Levings CS, Matzinger DF (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

Blum B, Sturm NR, Simpson AM, Simpson L (1991) Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. Cell 65:543-550

Blum B, Bakalara N, Simpson L (1990) A model for RNA editing in kinetoplastid mitochondria: "Guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell 60:189-198

Boer PH, Gray MW (1988a) Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are

linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. EMBO J 7:3501-3508

Boer PH, Gray MW (1988b) Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. Cell 55, 399-411

Bonen L (1991) The mitochondrial genome: so simple yet so complex. Curr Opinion Genet Devel 1:515-522

Bonen L, Bird S (1988) Sequence analysis of the wheat mitochondrial atp6 gene reveals a fused upstream reading frame and markedly divergent N termini among plant ATP6 proteins. Gene 73:47-56

Bonen L (1987) The mitochondrial S13 ribosomal protein gene is silent in wheat embryos and seedlings. Nucleic Acids Res 15:10393-10404

Bonen L, Gray MW (1980) Organization and expression of the mitochondrial genome in plants I. The genes for wheat mitochondrial ribosomal and transfer RNA: evidence for an unusual arrangement. Nucleic Acids Res 8:319-335

Bonnard G, Gualberto JM, Lamattina L, Grienenberger J-M (1992) RNA editing in plant mitochondria. Crit Rev Plant Sci 10:503-524

Brears T, Londsdale DM (1988) The sugar beet mitochondrial genome: a complex organisation generated by homologous recombination. Mol Gen Genet 214:514-522

Brennike A, Schuster W (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

Brown GG, Auchincloss AH, Covello PS, Gray MW, Menassa R, Singh M (1991) Characterization of transcription initiation sites on the soybean mitochondrial genome allows identification of a transcription-associated sequence motif. Mol Gen Genet 228:345-355

Burger G, Citterich MH, Nelson MA, Werner S, Macino G (1985) RNA processing in *Neurospora crassa* mitichondria: transfer RNAs punctuate a large precursor transcript. EMBO J 4:197-204

Burger G, Werner SJ (1985) The mitochondrial URF1 gene in *Neurospora crassa* has an intron that contains a novel type of URF. J Mol Biol 186:231-242

Burke JM (1988) Molecular genetics of group I introns: RNA structures and protein factors required for splicing-a review. Gene 73:273-294

Butow RA, Perlman PS (1991) Introns in pieces. Curr Biology 1:331-333

Cattano R (1991) Different types of messenger RNA editing. Ann Rev Genet 25:71-88

Cavalier-Smith T (1991) Intron phylogeny: a new hypothesis. Trends Genet 7:145-148

Cech TR (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell 44:207-210

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

Chasan R (1991) Splices and edits: RNA processing in plants. Plant Cell 3:1045-1047

Choquet Y, Goldschmidt-Clermont M, Girard-Bascou J, Kück U, Bennoun P, 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

Clayton DA (1991) Replication and transcription of vertebrate mitochondrial DNA. Ann Rev Cell Biol 7:453-478

Conklin PL, Wilson RK, Hanson MR (1991) Multiple *trans*-splicing events are required to produce a mature *nad1* transcript in a plant mitochondrion. Genes Develop 5:1407-1415

Covello PS, Gray MW (1991) Sequence analysis of wheat mitochondrial transcripts capped *in vitro*: definitive identification of transcription initiation sites. Curr Genet 20:245-251

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

Covello PS, Gray MW (1989) RNA editing in plant mitochondria. Nature 341:662-666

Cummings DJ, MacNally KL, Domenico JM, Matsuura ET (1990) The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. Curr Genet 17:375-402

Cummings DJ, Domenico JM, Michel F (1988) DNA sequence and organization of the mitochondrial ND1 gene from *Podospora anserina*: analysis of alternate splice sites. Curr Genet 14:253-264

Dawson AJ, Jones VP and Leaver CJ (1984) The apocytochrome b gene in maize mitochondria does not contain introns and is preceded by a potential ribosome binding site. EMBO J 3:2107-2113

De Souza AP, Jubier M-F, 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

Delahodde A, Goguel V, Becam AM, Creusot F, Perea J, Banroques J, Jacq C (1989) Site-specific DNA endonuclease and RNA maturase activities of two

homologous intron-encoded proteins from yeast mitochondria. Cell 56:431-441

dePamphilis CW, Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337-339

Douce R, Neuburger M (1989) The uniqueness of plant mitochondria. Ann Rev Plant Physiol Plant Mol Biol 40:371-414

Dujon B (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations-a review. Gene 82:91-114

Feagin JE, Abraham JM, Stuart K (1988) Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell 53:413-422

Finnegan PM, Brown GG (1990) Transcriptional and post-transcriptional regulation of RNA levels in maize mitochondria. Plant Cell 2:71-83

Folkerts O, Hanson MR (1989) Three copies of a single recombination repeat occur on the 443 kb mastercircle of the *Petunia hybrida* 3704 mitochondrial genome. Nucleic Acids Res 18:7345-7357

Forde BG, Leaver CJ (1980) Nuclear and cytoplasmic genes controlling synthesis of variant mitochondrial polypeptides in male-sterile maize. Proc Natl Acad Sci USA 77:418-422

Fukuzawa H, Kohchi T, Shirai H, Ohyama K, Umesono K, Inokuchi H, 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

Gass DA, Makaroff CA, Palmer JD (1992) Variable intron content of the NADH dehydrogenase subunit 4 gene of plant mitochondria. Curr Genet 21:423-430

Geliebter J, Zeff RA, Melvold RW, Nathenson SG (1986) Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: K^{bm9} and K^{bm6}. Proc Natl Acad Sci USA 83:3371-3375

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

Goldschmidt-Clermont M, Girard-Bascou J, Choquet Y, Rochaix J-D (1990) *Trans*-splicing mutants of *Chlamydomonas reinhardii*. Mol Gen Genet 223:417-425

Graham A, Steven J, McKechnie D, Harris WJ (1986) Direct DNA sequencing using Avian Myeloblastosis Virus and Moloney-Murine Leukemia Virus reverse transcriptase. Focus 8(2):4-5

Gray MW, Hanic-Joyce PJ, Covello PS (1992) Transcription, processing and editing in plant mitochondria. Ann Rev Plant Physiol Plant Mol Biol 43:(In press)

Gray MW (1991) Origin and evolution of plastid genomes and genes. In: Cell and somatic cell culture genetics of plants, Vol 7A, The molecular biology of plastids. Bogorad L, Vasil IK eds. Academic Press pp 303-330

Gray MW (1989a) Origin and evolution of mitochondrial DNA. Ann Rev Cell Biol 5:25-50

Gray MW (1989b) The evolutionary origins of organelles. Trends Genet 5:294-299

Gray MW, Cedegren R, Abel Y, Sankoff D (1989) On the evolutionary origin of the plant mitochondrion and its genome. Proc Natl Acad Sci USA 86:2267-2271

Gruissem W, Barkan A, Deng X-w, Stern D (1988) Transcriptional and post-transcriptional control of plastid mRNA levels in higher plants. Trends Genet 4:258-263

Gualberto JM, Bonnard G, Lamattina L, Grienengerger J-M (1991) Expression of the wheat mitochondrial *nad3-rps12* transcription unit: correlation between editing and mRNA maturation. Plant Cell 3:1109-1120

Gualberto JM, Weil J-H, Grienenberger J-M (1990) Editing of the wheat *coxIII* transcript: evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. Nucleic Acids Res 18:3771-3776

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

Gualberto JM, Wintz H, Weil J-H, 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 are co-transcribed. Mol Gen Genet 215:118-127

Hanson MR (1991) Plant mitochondrial mutations and male sterility. Ann Rev Genet 25:461-486

Harris ME, Hajuk SL (1992) Kinetoplastid RNA editing: *in vitro* formation of cytochrome b gRNA-mRNA chimeras from synthetic substrate RNAs. Cell 68:1091-1099

Heinemeyer W, Alt J, Herrmann RG (1984) Nucleocide sequence of the clustered genes for apocytochrome *b6* and subunit 4 of the cytochrome *b/f* complex in spinach plastid chromosome. Curr Genet 8:543-549

Herrin DL, Schmidt GW (1988) *Trans*-splicing of transcripts for the chloroplast *psaA1* gene. *In vivo* requirement for nuclear gene products. J Biol Chem 262:14601-14604

Hiesel R, Wissinger B, Schuster W, Brennicke A (1989) RNA editing in plant mitochondria. Science 246:1632-1634

Hildebrand M, Hallick RB, Passavant CW, Bourque DP (1988) *Trans*-splicing in chloroplasts: the *rps12* loci of *Nicotiana tabacum*. Proc Natl Acad Sci USA 85:372-376

Hintz WE, Mohan M, Anderson JB, Horgen PA (1985) The mitochondrial DNA of *Agaricus*: heterogeneity in *A. bitorquis* and homogeneity in *A. brunnescens*. Curr Genet 9:127-132

Hoch B, Maier RM, Appel K, Igloi GL, Kössel H (1991) Editing of a chloroplast mRNA by creation of an initiation codon. Nature 353:178-180

Hodges P, Scott J (1992) Apolipoprotein B mRNA editing: a new tier for the control of gene expression. Trends Biochem Sci 17:77-81

Hsiao K-c (1991) A fast and simple procedure for sequencing double stranded DNA with sequenase. Nucleic Acids Res 19:2787

Hunt MD, Newton KJ (1991) The NCS3 mutation: genetic evidence for the expression of ribosomal protein genes in *Zea mays* mitochondria. EMBO J 10:1045-1052

Jacquier A, Jacquesson-Breuleux N (1991) Splice site selection and role of the lariat in a group II intron. J Mol Biol 219:415-428

Jacquier A (1990) Self-splicing group II and nuclear pre-mRNA introns: how similar are they? Trends Biochem Sci 15:351-354

Jarrell KA, Dietrich RC, Perlman PS (1988) Group II intron domain 5 facilitates a *trans*-splicing reaction. Mol Cell Biol 8:2361-2366

Joyce PBM, Gray MW (1989) Chloroplast-like transfer RNA genes expressed in wheat mitochondria. Nucleic Acids Res 17:14-5461-5476

Joyce PBM, Spencer DF, Gray MW (1988) Multiple sequence rearrangements accompanying the duplication of a $tRNA^{pro}$ gene in wheat mitochondrial DNA. Plant Mol Biol 11:833-843

Kemmerer EC, Kao T-h, Deng G-r, Wu R (1989) Isolation and nucleotide sequence of the pea cytochrome oxidase subunit I gene. Plant Mol Biol 13:121-124

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

Koch JL, Boulanger SC, Dib-Hajj SD, Hebbar SK, Perlman PS (1992) Group II introns deleted for multiple substructures retain self-splicing activity. Mol Cell Biol 12:1950-1958

Kohchi T, Umesono K, Ogura Y, Komine Y, Nakahigashi K, Komano T, Yamada Y, Ozeki H, Ohyama K (1988) A nicked group II intron and *trans*-splicing in

liverwort, *Marchantia polymorpha*, chloroplasts. Nucleic Acids Res 16:10025-10036

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

Kunkel TA, Eckert KA (1989) Fidelity of DNA polymerases used in polymerase chain reactions. Curr Comm Mol Biol: Polymerase Chain Reaction. Cold Spring Harbor Laboratory Press. (ed. Erlich HA, Sibbs R, Kazazian HH)

Laird PW (1989) *Trans* splicing in trypanosomes - archaism or adaptation? Trends Genet 5:204-208

Lamattina L, Grienenberger J-M (1991) RNA editing of the transcript coding for subunit 4 of NADH dehydrogenase in wheat mitochondria: uneven distribution of the editing sites among the four exons. Nucleic Acids Res 19:3275-3282

Lambowitz AM, Perlman PS (1990) Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. Trends Biochem Sci 15:440-444

Lambowitz AM (1989) Infectious introns. Cell 56-323-326

Lang BF, Ahne F, Bonen L (1985) The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*: The cytochrome b gene has an intron closely related to the first two introns in the *Saccharomyces cerevisiae cox*1 gene. J Mol Biol 184:353-366

Lejeune B, Quétier F (1988) Structure de l'ADN mitochondrial du blé. In Variabilité génétique cytoplasmique et stérilité mâle cytoplasmique. (Paris:INRA) pp 201-208

Levings CS, Brown GG (1989) Molecular biology of plant mitochondria. Cell 56:171-179

Levy AA, André CP, Walbot V (1991) Analysis of a 120-kilobase mitochondrial chromosome in maize. Genetics 128:417-424

Lippok B, Brennicke A, Wissinger B (1992) The *coxII* gene in carrot mitochondria contains two introns. Mol Gen Genet 232:322-327

Liu AW, Narayanan KK, André CP, Kaleikau EK, Walbot V (1992) Co-transcription of *orf25* and *coxIII* in rice mitochondria. Curr Genet 21:507-513

Lonsdale DM (1989) The plant mitochondrial genome. In The Biochemistry of Plants: A Comprehensive Treatise. Vol. 15, A Marcus, ed. (New York: Academic Press), pp 229-295

Lonsdale DM (1984) A review of structure and organisation of the mitochondrial genome in higher plants. Plant Mol Biol 9:201-206

Lonsdale DM, Hodge TP, Fauron CM-R (1984) The physical map and organization of the mitochondrial genome from the fertile cytoplasm of maize. Nucleic Acids Res 12:9249-9261

Mahendran R, Spottswood MR, Miller DL (1991) RNA editing by cytidine insertion in mitochondria of *Physarum polyceophalum*. Nature 349:434-438

Maier RM, Hoch B, Zeltz P, Kössel H (1992) Internal editing of the maize chloroplast ndhA transcript restores codons for conserved amino acids. Plant Cell 4:609-616

Makaroff CA, Palmer JD (1987) Extensive mitochondrial specific transcription of *Brassica campestris* mitochondrial genome. Nucleic Acids Res 15:5141-5156

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory)

Marres CAM, De Vries S, Grivell LA (1991) Isolation and inactivation of the nuclear gene encoding the rotenone-insensitive internal NADH:ubiquinone oxidoreductase of mitochondria from *Saccharomyces cerevisiae*. Eur J Biochem 195:857-862

Matsubayashi T, Wakasugi T, Shinozaki K, Yamagushi-Shinozaki K, Zaita N, Hidaka T, Meng BY, Ohto C, Tanaka M, Kato A, Maruyana T, Sugiura M (1987) Six chloroplast genes (*ndhA-F*) homologous to human mitochondrial genes encoding components of the respiratory chain NADH dehydrogenase are actively expressed: Determination of the splice sites in *ndhA* and *ndhB* pre-mRNAs. Mol Gen Genet 210:385-393

Meng BY, Matsubayashi T, Wakasugi T, Shinosaki K, Sugiura M, Hirai A, Mikami T, Kishima Y, Kinoshita T (1986) Ubiquity of the genes for components of a NADH dehydrogenase in higher plant chloroplast genomes. Plant Sci 47:181-184

Michaelis G, Vahrenholz C, Pratje E (1990) Mitochondrial DNA of *Chlamydomonas reinhardtii*: The gene for apocytochrome *b* and the complete functional map of the 15.8 kb DNA. Mol Gen Genet 223:211-216

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

Michel F, Jacquier A (1987) Long range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. Cold Spring Harbor Symposia on Quantitative Biology, Vol. LII. Cold Spring Harbor Laboratory. pp 201-212

Michel F, Lang BF (1985) Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature 316:641-643

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

Mishra NC (1991) Genetics and Molecular biology of *Neurospora crassa*. Adv Genet 29:1-62

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

Mulligan RM, Lau GT, Walbot V (1988) Numerous transcription initiation sites exist for the maize mitochondrial genes for subunit 9 of the ATP synthase and subunit 3 of cytochrome oxidase. Proc Natl Acad Sci USA 85:7998-8002

Nagley P (1989) Trafficking in small mitochondrial RNA molecules. Trends Genet 5:67-69

Neupert W, Schatz G (1981) How proteins are transported into mitochondria? Trends Biochem Sci 6:1-4

Newton KJ (1988) Plant mitochondrial genomes: organization, expression and variation. Ann Rev Plant Physiol Plant Mol Biol 39:503-532

Nixon PJ, Gounaris K, Coomber SA, Nunter CN, Dyer TA, Barber J (1989) *psbG* is not a photosystem two gene but may be an *ndh* gene. J Biol Chem 264:14129-14135

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

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

Ohyama K Fukuzawa H Kohchi T Shirai H Sano T Sano S Umesono K Shiki Y Takeuchi M Chang Z Aota S Inokuchi H Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast. Nature 322:572-574

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Cell and somatic cell culture genetics of plants, Vol 7A, The molecular biology of plastids. Bogorad L, Vasil IK eds. Academic Press pp 5-53

Palmer JD, Logsdon JM (1991) The recent origins of introns. Curr Opinions Genet Dev 1:470-477

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol 28:87-97

Palmer JD, Herbon LA (1987) Unicircular structure of the *Brassica hirta* mitochondrial genome. Curr Genet 11:565-570

Palmer JD, Shields CR (1984) Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307:437-440

Perlman PS, Butow RA (1989) Mobile introns and intron-encoded proteins. Science 246:1106-1109

Pfanner N, Neupert W (1990) The mitochondrial protein import apparatus. Ann Rev Biochem 59:331-353

Pritchard AE, Seilhamer JJ, Mahalingam R, Sable CL, Venuti SE, Cummings DJ (1990) Nucleotide sequence of the mitochondrial genome of *Paramecium*. Nucleic Acids Res 18:173-180

Rapp WD, Stern DB (1992) A conserved 11 nucleotide sequence contains an essential promoter element of the maize mitochondrial *atp1* gene. EMBO J 11:1065-1073

Reinhold-Hurek B, Shub DA (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. Nature 357:173-176

Rodgers JH (1990) The role of introns in evolution. Febs Lett 268:339-343

Ruby SW, Abelson J (1991) Pre-mRNA splicing in yeast. Trends Genet 7:78-85

Runeberg-Ross P, Grienenberger J-M, Guillemault P, Marechal L, Gruber V, Weil J-H (1987) Localization, sequence and expression of the gene coding for $tRNA^{pro}$ (UGG) in plant mitochondria. Plant Mol Biol 9:237-246

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, 2nd edition

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

Scazzocchio C (1987) The natural history of fungal mitochondrial genomes. In: Evolutionary Biology of Fungi. Rayner ADM, Brasier CM, Moore D Eds. Cambridge University press. pp 53-73

Schmidt U, Riederer B, Mörl M, Schmelzer C, Stahl U (1990) Self-splicing of the mobile group II intron of the filamentous fungus *Podospora anserina* (COl 11) *in vitro*. EMBO J 9:2289-2298

Schuster W, Wissinger B, Hiesel R, Unseld M, Gerold E, Knoop V, Marchfelder A, Binder S, Schobel W, Scheike R, Grönger P, Ternes R, Brennicke A (1991a) Between DNA and protein - RNA editing in plant mitochondria. Physiologia Plantarum 81:437-445

Schuster W, Ternes R, Knoop V, Hiesel R, Wissinger B, Brennicke A (1991b) Distribution of RNA editing sites in *Oenothera* mitochondrial mRNAs and rRNAs. Curr Genet 20:397-404

Schuster W, Hiesel R, Isaac PG, Leaver CJ, Brennicke A (1986) Transcript termini of messenger RNAs in higher plant mitochondria. Nucleic Acids Res 14:5943-5954

Sharp PA (1987) *Trans* splicing: Variation on a familar theme? Cell 50:147-148

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043-2049

Small I, Maréchal-Drouard L, Masson J, Pelletier G, Cosset A, Weil J-H, Dietrich A (1992) *In vivo* import of a normal or mutagenized heterologous transfer RNA into the mitochondria of transgenic plants: towards novel ways of influencing mitochondrial gene expression? EMBO J 11:1291-1296

Small I, Suffolk R, Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell 58:69-76

Sommer B, Köhler M, Sprengel R, Seeburg PH (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell 67:11-19

Stern DB, Bang AG, Thompson WF (1986) The watermelon mitochondrial URF-1 gene: evidence for a complex structure. Curr Genet 10:857-869

Stern DB, Palmer JD (1984) Extensive and widespread homologies between mitochondrial DNA and chloroplast DNA in plants. Proc Natl Acad Sci USA 81:1946-1950

Sutton CA, Conklin PL, Pruitt KD, Hanson MR (1991) Editing of pre-mRNAs can occur before *cis* and *trans*-splicing in *Petunia* mitochondria. Mol Cell Biol 11:4274-4277

Tessier L-H, Keller M, Chan RL, Fournier R, Weil J-H, Imbault P (1991) Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. EMBO J 10:2621-2625

van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. Cell 44:225-234

Van de Peer Y, Neefs J-M, De Watcher R (1990) Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms, and mitochondrial origins. J Mol Evol 30:463-476

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

Wahleithner JA, Wolstenholme DR (1988) Ribosomal protein S14 genes in broad bean mitochondrial DNA. Nucleic Acids Res 16:14-6897-6913

Walbot V (1991) RNA editing fixes problems in plant mitochondrial transcripts. Trends Genet 7:37-39

Ward BL, Anderson RS, Bendich AJ (1981) The mitochondrial genome is large and variable in a family of plants (*Cucurbitaceae*). Cell 25:793-801

Weiss H, Friedrich T, Hofhaus G, Preis D (1991) The respiratory-chain NADH dehydrogenase (complex I) of mitochondria. Eur J Biochem 19:7563-576

Wilson AJ, Chourey PS (1984) A rapid inexpensive method for the isolation of restrictable mitochondrial DNA from various plant sources. Plant Cell Rep 3:237-239

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

Wissinger B, Schuster W, Brennicke A (1990) Species-specific RNA editing patterns in the mitochondrial *rps13* transcripts of *Oenothera* and *Daucus* Mol Gen Genet 224:389-395

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

Wolf K, Del Giudice L (1988) The variable mitochondrial genome of ascomycetes: organisation, mutational alterations and expression. Adv Genet 25:185-308

Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201-6205

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial chloroplast and nuclear DNAs. Proc Natl Acad Sci USA 84:9054-9058

Woodson SA, Cech TR (1989) Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell 57:335-345

Wu M, Nie ZQ, Yang J (1989) The 18-kD protein that binds to the chloroplast DNA replicative origin is an iron-sulfur protein related to a subunit of NADH dehydrogenase. Plant Cell 1:551-557

Yang AJ, Mulligan RM (1991) RNA editing intermediates of *cox2* transcripts in maize mitochondria. Mol Cell Biol 11:4278-4281

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

APPENDIX I. Sequence analysis of the four *nad1* coding regions. Most sequences presented here have been obtained from only one strand.

A. Sequence analysis of the *nad1a* region. The restriction sites for HindIII and BamHI are underlined. The blocked sequence is also shown in Figure 3. Runeberg-Ross and colleagues (1987) have sequenced a 0.7-kb HindIII fragment that is located immediately upstream of this HindIII restriction site.

HindIII
AAGCTTATTT TGCAATCCAT TATCTTTGGT ATTCTTCTAT ATTTAGATCA TATTATATAG-60
CTATTTCACA GGTTTTATTT ACATACCTCA GTCATGTTAT TTGTCTAGCC TTCAATCAGA-120
ATAGCCTCTC CACCGGTTCC GAAGATTCTT TCGGCATCAA TGTGCTTTTG GAATCTTCCT-180
ATGAAACAAA TAACATTCTG GAACAGTTTA GGACAGAACA GGCCTCAGTA GAAAGCGAGC-240
TTTTTTCGCG TATAGCCACC TCGAAGCCCA ATTAGCGCAC GGGCTGCCCC CTCAACTCAA-300
CCCTGACGAG TACGCAAACT TGGTCAGGGA GCATTTAGAC GGTGCAGTTA GTATTGACCA-360
CTACCGCCAA GTCCATAATT CAAAAATTGA TGAAGTCCAC ATAATGGAGC TCAAAAGCCG-420
ATTACCAAAT CTACTTTTTT AGCATCTTGA AACCGCACAT AAGAAATATT CTGAATGAAT-480
CACCTTACAA TAACATACGA GAGACAGCTT TCTATTTTGT CGAACAACAA GTCGAACCCA-540
ATGAACAATT CCATTCAAAG GAATGTTCTA GAGTTTTTTC CACTATATTC GCGATATTAA-600
AAAAAATGGA AGAAATTCTT CACGATTTCT AGAGTTCCGG GATTACTTCA CTGATATCGA-660
TTTTCGCCTA AGCACGGCCT CTAGGTTAGT TCAAAGTGTA ATACAACGAA CTATCGATAC-720
ACCGAAAACG AGTCAAGATG GATACATAAA ATGGAGACAT CGTGGATTAC CCAAATTGGG-780
ACGACGATAG GACATTTCAC TTTCTTTCGC AATATTTCTG ATCATGACTC AACCACTAGG-840
AAAGGGGATT GCTTACTCTC AGCCACGTTT TCGCTTATGC TTAGTCAAGT GAGGATTCCA-900
TTCGAAGTAA CGTAACAGGA GCACCATCTT CAAAACTTCT CGGGATCCGG AGTTTTTCGA-960
TAAAAGGTCC CATCCTTCAA TATCATGATT GGGTCAACCA GGCCAGATCA TAAGTGAAAT-1020
AGTTTGATCG GGTCGACCAG GTCAGGCCCG ATCATGAGTG AATAGAAAAT CGAAAACGTA-1080
CATAGCTGTT CCAGCGGAAA TACTTTGTTT AATTCTACCA CTTCTACTAG GAGTAGCCTT-1140
TTTAGTGCTA GCTGAACGTA AAGTAATGGC TTTTGTGCAA CGTCGAAAGG GTCCTGATGT-1200
AGTGGGATCG TTCGGATTGT TACAACCTCT AGCAGATGGT TTGAAATTGA TTCTAAAAGA-1260
ACCTATTTCA CCAAGTAGTG CTAATTTCTC CCTTTTTAGA ATGGCTCCAG TGGCTACATT-1320
TATGTTAAGT CTGGTCGCTT GGGCCGTTGT ACCTTTTGAT TATGGTATGG TATTGTCAGA-1380
TCCGAACATA GGGCTACTTT ATTTGTTTGC CATATCTTCG CTAGGTGTTT ATGGAATAAT-1440
TATAGCAGGT TGGTCTAGTA AGACGGGGGG CGGCCGTTCG GTCGCCTATG ATATACGGAC-1500
CAATTGGTCA AAAATGGGTT TGTGCCGCAG GTGTTGAACG ATCTACTCTA CACAGGTGTG-1560
GGCTTACAGG GCTAGGGCTC ATAAACCCTT TCTTTCATTC ATCAAGGGGT CGGTCACTTT-1620
TCCGGGCCGG GATCGAGAAG TGGAAGTCAT AAAAAAGAGA TGTTTCTCTT CGCACCTCAT-1680
ATCAAGAGGA AAGGTAGCTT GTCAAGCTGG TCTCACTATT GGAAATTCTA GCTTTCTTTT-1740
TTTTTTTTTT TTCACCGGCT CTTCTTCTTT GTCAACGCTA CTAAGGACCT GACGGTTCGC-1800
CTACTTGATG GATGAAAGAA CATGAGAAAG GGTTCTAAAA TAAAAGGCTA GAAAGTCTAC-1860
TACAGTACAG TCAAAGTCAG CGGCTGAGTT TGCCTAGCCT CGCCTACTCA TTTTTGTAGG-1920

CGAGCGAGTT AGCGAAGAGG CTTAGGGATA AGAGAGTGTC GGTGCGTACG TAGCCTTCAT-1980
TCTACTACGC TACACAAAGT CACTTAGCTC GACGTTAATT AAGAGAGTAA AGGAGGAATA-2040
CCCTACATAG AAGAACCACA GTTCGCTTAC AAAGGTATAA CCTTTAGCTC CCCGGGGCTA-2100
AGCTGCTTCG CACCACTGAT AGTTAATTAA GATTCACTTT CAAAGGCGCT ACTTTGTTTC-2160
GCAAGCCTTT GTCATTGTCA GTCAACTAAG GTTGGCCCTG GCCCCCTTGC GAATCCGTAA-2220
ATCAGAGAGC ATTGCCGCAA AAAAAGGATG GTCCCCTATG CATTTCATCT TTCCGGAACG-2280
CAAAGAATTA AAGTACCTGA CCCCCATCAT GGTGAACCCT CTCCCTTGTT GATCGGGATG-2340
AGGTAGATGC CTCCCAGCCG GGGGGGCCGA TCGAATCGGA GTTTCCTTAG GTAGCCACCG-2400
ACCTACAGTT CTCCTTAAAC TTCTGTGCTT GGTGGAAAAG AAGCGAACAA AGGTACGCTC-2460
GCTTGCTGTC TTGTTCTCTG CCGCGGACTG GGATCGCTCG CCAGCTAGGC CCTCTAGAAA-2520
AAAGTTGGAG CAACATTGTA TGAGAACATA TTACCCATCT TCGGGGACAA GGGGCGGAAC-2580
GACCTCTCGA TCTACTTACT GCAGCCCAGG ACGGGCGTCG TCGTCTAGGC GTGACTTCTT-2640
GTTTTGATCT CCCCTACGCC TAGGACGTTG TCTGGGCCAA GAGCCATAGG TTAGTTGCTG-2700
TTTCCATTTG GTTCTTCTTT CTCGTTGTTG ATACCGGCAA GACCCAGCCA GATGATGATG-2760
TCTGCTGGTT GGTAGTGAGA GGACTCTTAG TACCCGCAAA AAAAGGGCGC TGGTGAAAAA-2820
AAACAATCAT TTGATTTAGT TTCTTGCTCT CCCTTAGCAG CGGGAAAGGA GTCTATCTAT-2880
CTGCCTAGCT TTGGTAGATT TCCCCCAACG CAATTCAAAA AGGAAATTCC ACGAATCCCT-2940
GAGGTGGATA AGTCCGACGA CTCAGCAGCA GTGCGGAATT GAGTTTCTGG TCCGGTGGAT-3000
CTGATCTATA GTTAGGGGGC AATACATACC AAAAGGTTCG AAGAAGGAAG GCGCAGTATA-3060
TAACCAACCC ACCCATAGCC GGTCTAACTG CTGAGAGAGA AAAGTGCTTT TCTTTCCCTA-3120
AGAGTCCTCG AGTACACACA GCTATTGCGG ATCC
BamHI

B. Sequence analysis of the *nadIc* downstream region. The sequence begins immediately downstream of the sequence data shown in Figure 3B and is numbered accordingly. The restriction sites PvuII and XbaI are underlined.

PvuII
CAGCTGGCTGGTCAATCTCAG TGATCTTATC GGCCGGCAAA GCGGAGACGG ACGACCACGG-2984
TCCCGACCTT ACCAGCACCG TAGTTTACTA ATCAATGAAG CCCCTCAACC TACTATCTTT-3044
TCTTACAAAC TCAACCAAGC CTAACCAAAC CCCATGCTCA CGACATTGTT CTTGATATTG-3104
ATTGAGTTGG AGGGAGATCA GAGACTACCA CGGAATCCTT CCATAGTCAC CCCTGGAGCC-3164
TTCGTCACCA AAGCGTCACG ACAGTTTCCA AGACCCCTCA TATAATCTCC AGTGGGGATC-3224
AGTCGGAGCA CGTAGGAATG CCGACCACTA CATAAGCCAC GTCTGGCTGA GGCGAGCAGC-3284
AAGCCGGGGG AGAGTATTTT GAAGTACAAG AACGTTGGGG TGGGACGCTA GTACTTAAAC-3344
TAGGGTTGCC TTCTTAGCGC TAATCTTGCG TTTTTCAGCA GTTAGTTGTA GCGCGCCTTC-3404
CCTCCTTCTC TCATTTCGGA AAGATTTGGC AGTATTTTCT TATGATGACC TGGTCGAGAG-3464
AGTACGAAAC ATCGGTGTAA AAGATTGAGT GGGATGGTTA TAGTAAGGGG CGAACCAAGT-3524
GCTTCGCGGA AGAAGCGAAT CAATCAGAAT GGAGACAAGG AAAGGAATCG AGGCCCGGCC-3584
CGGTAGCGAT GCCTCTCACG CGGATGGGAG TTACTTTTCT TTATCACCGT AAGCGTTACG-3644
AAATGGACCC GCATTCCCCT TTGCTTTTCT TTTGCAACTA AGCTTAGGTC ATACATTTCC-3704
AAGCTAATTA GAATACAACC ATGAAGGTGG GGTTCGAATC GCCATCACAA CTTATTTATT-3764
CTGGTAAGTA GGATTTGGAA TGTAGTCCCT TTTTATTGGA TCCGCCGCTC CGTTAAGTGA-3824
AATTCAAAGT AGGTGTGGGT CTAGGTAGAG GAAGTCATAA ATCAAGTATC AAGTATTGTT-3884
ATAAATATCG TATATAGTAG TTGCCAAACT CACAAGCATC ACACACTTCG ATCTACTTTA-3944
TTGAATAACT TAGCTAGATA GAACCGACTC AAGCTTTTAA ATGAAAGAAG CAATAATGGC-4004
ATAATTTCTG TCTCAAGACC TCTTGCAGCT ACAGCCAATG CAGCACTCCT TCATCAACAT-4064
ACAACACAGC CCATTAGCNC CTTGGCGCCA GTCGAGTATA AAGAGAGACT AGAACCTACC-4124
GGGAGACCCT TCNCGGAGAG TCCCTATCGT TGTTGCACAT TACACACAGA GTTCGGATTA-4184
ATTGTGAGGA AGGAAGGGTA TGTGTGTTGA AGGAGTGGTG CGAAGACATA TTTNNNTCAG-4244
TGAGTATAAA GGGTAGACTA CGGGAGCCTG GAGGACTCCT AGTAACACAA CTGTTCGCCT-4304
TTCCGAAACG ATTCTTATCA GGACCCTCCT TTATATAAAC CACTCGTAGA TATATACCTC-4364
CGTGTGGCCT CAGTGATACA TCGATAGATA AGAAGGGAAA TGCGTCATCT AGA
XbaI

C. Sequence analysis of the *nad1*d region. The *nad1*d exon is in bold and the sequence shown in Figure 3C is blocked. The restriction sites EcoRI and HindIII are underlined.

EcoRI

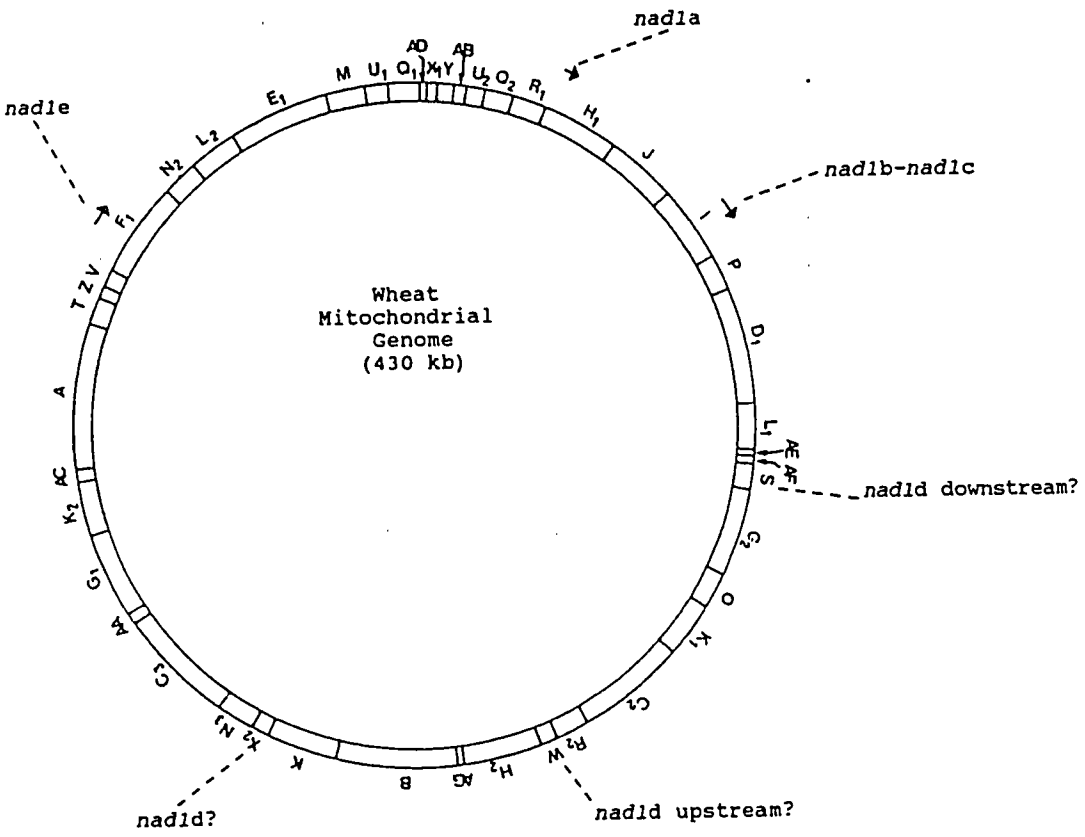
GAATTCCATT CGGTACCTTT GGGGTTTTCT TTTTCCTTCT GGATGACGAC GTCAACCGCG-60
TCTCCTTATT GTTATTGCGA TGGTGACGAG GAACAAAGGG CCGGGGAGAT CATCGAAATT-120
ACCTCTAGCC CGGAGCCAGA GGGGGTCGTC GACACAACGC CAACCCTGGT GGTGGGGCAC-180
TTCGAAGGTC TGGAGACCCT CCTTGAGATT CCCCCTTTAC CGGAAACGCC ACCCGCAGTC-240
CTTCCTTAGC TCCAGAAGAA AGGGAGCAAG CCCCCTCTGT TGGCCTAAAC CTCGATCCAT-300
CCCAGAGTAG CCGCCGGCTG GACCAGGACG GGGGGCAAAC TCCTCAAAAC GAAGTTCTAC-360
TTTGTCTTTG GAAGAATTGG AAGAATCGGA CTCTAAGTTT AAAGTAGGAG TAGAGTCTTT-420
AGAGGCTCGC TTAGAAGGAA TAAAGAGAAA AAAGCAACTG CAAGCGTTCG GGGCTTCTCC-480
TTTTGAGACC GCGAAAGCAG AGCTGGTTAC GTCCATTGGG GAAAATTCCC AGTTTCGTGT-540
GGGACTGCTT TTGTCTCTGG ATGTCAAATG CCCTTGACGG GCAAACGGGA GGCTCTCGAA-600
AAGACGATAG AGGCCATCTT GAAATGCCAA GGGAAGCCCT AGCGTCTCTA CTCGACCTAA-660
ATGCGATCTC GAGGACCCCG AAAAACTCGA GACCGCTCTC CGTCCATATT TAGGACGGTT-720
AGGGCGATTT TTTAAGAAAT AAAAACCTTC TGTCGATTTA TCCACACTTC CATGACTTCT-780
AGAGGAAAGC TAACTGCTTG CTGGCTGGGA GCTGTATGAG CGGTAACGTC CACGTACGGC-840
TCCGTGAGAA GGTGGACGGA AATGGCCTTG TTGTACCTCA CTCCCG**TCTT CAATGGGGTC**-900
TGCTCTTTTT TTTTTAGGAG AGTATGCCAA TATGATCTTA ATGAGGTGCG GGGCTTTGCA-960
TCTGACATTC GTTGGGCTTT CCTCTGCGGG AGCCCGCGTC CCGGCCTTTT TGTGCAATAA-1020
ACCCCTCCGG CCGAAGACTA GTGATAGGTG GTCCCGCGGA GCTTTCGGAG AAGGGTAGCC-1080
TAGTGTGTAA GCACAGCAAT GAACCGCGGC GAACCCTCAG ACGACCCTTC TAAGATAAGG-1140
GGGGAGATCC TCAGTAGTGG TGACCCTTTG ACTCTTCCAC TGACTTATAT ATGTACCGAA-1200
TGCTCATACG GGAAAGTGAA CTCCTGGGTC TGGAACCTGG GGGGGTTGCT CCGATAAAAA-1320
ATCCTTTCTT TCTCGTCCAC TCTAGGGGGT GCGGACACAC CTGCGCGGAT TACAGGTGAC-1380
GGTTACAAGA ATGGCGGGGA AGTGAAGAGT ACCCGACGAC ATTCAGGGAT GAATGTAGAC-1440
CCATCGGGCG GGGATAATCA TTCCGGTCCT GGGAGAGGTG GCGACACCAG TCTGAGCTGA-1500
GGGAAGCCAG CCAATTGAGT CACTGAAACG ACCGCAGGAG GCGCACCCTA AGCCCAGCTT-1560
CTCGGAGCTG CTATTGCGAA ATACGGCTTC CTTAATATCC AAATTTCAAC AAGGGGGCTG-1620
GGCTGCCCGC CTTCATAGAA AGGCGCCCGG GCCTAGATTA GATGTTCGCC TTTTTATAGA-1680
ATAGAAAGAG AAGGTAAAGG GGCCCCCTTT GGAGATTCTT TCAAGAATGC ATGGCTTCCT-1740
CCTATTCCGC TTCAACAGAA GCCTTAAAAG CCTGCTGCTA TCGGCAGAAG GGTTATCCAT-1800
TTCATAGGCG AGGGTNNNAA GAAGCCGAAC CTCTGATCGA CCCGGACACG TCAAAAATTC-1860
TCGGCGCCGA GAGAAAGACG AGATTCAGTA AGTAATTCAG TGAGTGCTTT CTTTTATAAG-1920
AAGAGCGTCG GCTTGGAAGA AGAAAATGAA ATACGAACAA CCGCGCTGGT TGTAATAGAT-1980
CGACTTGAAT GCGTCTTCTT GCTCCAGCAT TCAAGTTCCA TTTCAAGGGA GGACGACGTA-2040
CCATGATACT TTCTGTTTTG TCGAGATCTG CTTTGGTCTC TGGTTTGATG GTTGTACGTG-2100
CTAAAAATCC GGTACATTCC GTTTTGTTTC CCATCCTAGT CTTTTGCGAC ACTTCTGGTT-2160
TACTTATTTT GTTAGGTCTC GACTTCTCCG CTATGATCTC CCCAGTAGTT CATATAGGAG-2220
CTATTGCCGT TTCATTCCTA TTCGTGGTTA TGATGTTCAA TATTCAAATA GCGGAGATTC-2280

ACGAAGAAGT ATTGCGCTAT TTACCAGTGA GTGGTATTAT TGGACTGATC TTTTGGTGGG-2340
AAATGTTCTT CATTTTAGAT AATGAAACCA TTCCATTACT ACCAACCCAC AGAAATACGA-2400
CCTCTCTGAG ATATACGGTT TATGCCGGAA AGGTACGAAG TTGGACTAAT TTGGAAACAT-2460
TGGGAATTTA CTTTATACCT ACTATTCCGT CTGGTTTTTG GTTTCTAGTC TGATTTTATT-2520
AGTTGCTATG ATTGGGGCTA TAGTACTTAC TATGCATAGG ACTACAAAGG TGAAAAGACA-2580
GGATGTATTC CGACGAAATG CCTTGGATTC TAGGAGCCAT ATAATGAACA GGACTATTTC-2640
TCCTTTTGGC CATAGCCATA GAAGAAGCTT
HindIII

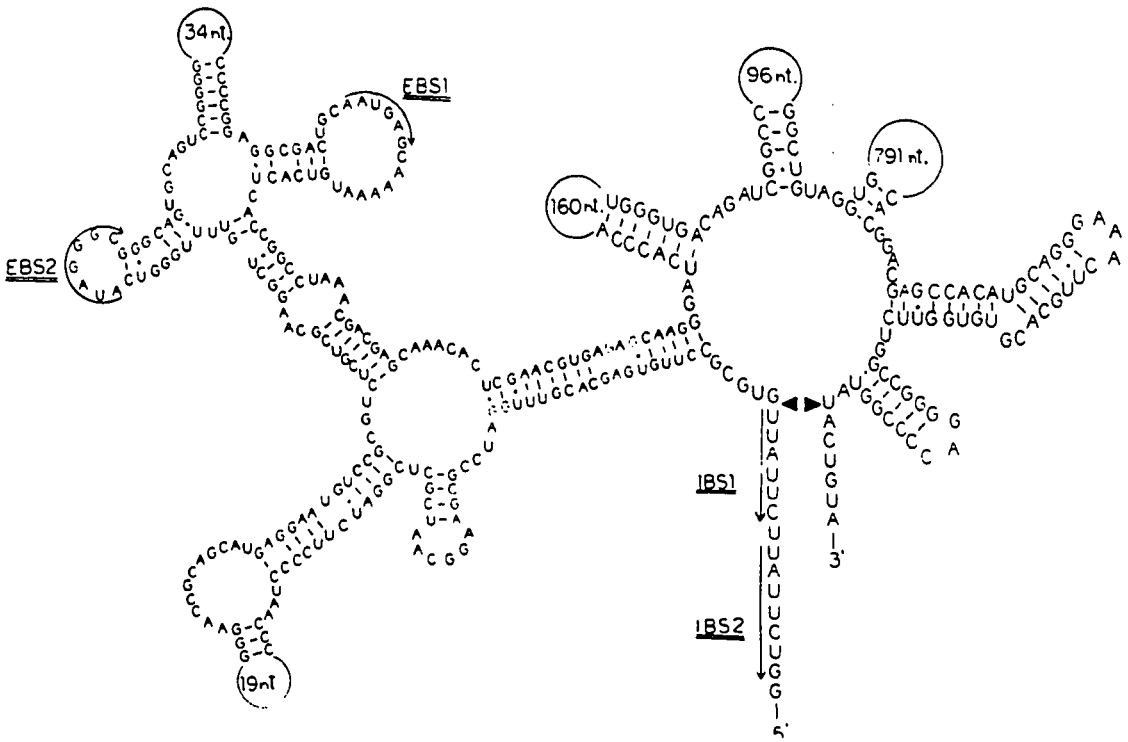
D. Sequence analysis of the *nad1e-nad5c* region. The sequence begins immediately downstream of the sequence data shown in Figure 3D and is numbered accordingly. The exon *nad5c* is in bold. The restriction sites XbaI and BamHI are underlined.

XbaI
TCTGCGGTTA TAAGTAACGT AGCCAGCCTA ATTTCTAGAA TAGGGCTGCG GGCAGGAGGG-3180
CTTTGTTTGT GCAATCTTGC TCGCAACACC CTCTCCTCCG GCGAACCTGC GATCTCCGCC-3240
TCTAACTGAA TGCTCAACTG AGGTCGTGTA CAACAACCTT AACCGCACAA GCCTGGGCTG-3300
GGCTATCCCC ATCCCTAGAG GAGCCGTATG AGGCGAACGT CCACGTACGG TTTTGAAGCC-3360
GAGCCTTTCC AGCAATGGGG CCTAGGGACC **GATATGATGA TTGGTTTAGG TAGGGCGGCC**-3420
GGCTACTATG GGCACTGTAG GGATTAGTGC GTGAGACCGG ATCCACAAAT GCGCATGGAC-3480
TCACCCTTGA CTTGAGAATG GAAGAAGGGA AAGACAATAG CATGTCACAA GAGCGAGGCA-3540
GAGAGGGCTC CGCCCTTACA GCGAGAGGGA GCCCTCGCGG AGCCGGGCTT TGGAGATGAG-3600
GCCTTTTGCG AAGCCAAGTC AATTTCGGGC CACCAAACCC TGCAACTGAG GAAAAGGCCC-3660
TATGGAGTAA AGGGAAGCGT GTACGTTGTC ACACTCCCCG CCCTCCAAAG GTGCCTAGAG-3720
GACGGGCCAG ACGCAGCAGA GCGACGCCCC AGGAGCGGAT TCCCCACGAG CAGGGGGACA-3780
GGAGACGGCC ATCTCGAGGC ACATCACGAC CTACAGGCAA CACCGGCGAG ACCCAGGAAG-3840
GTAACCCGAT TGGGAGTCAG AGGATCC
BamHI

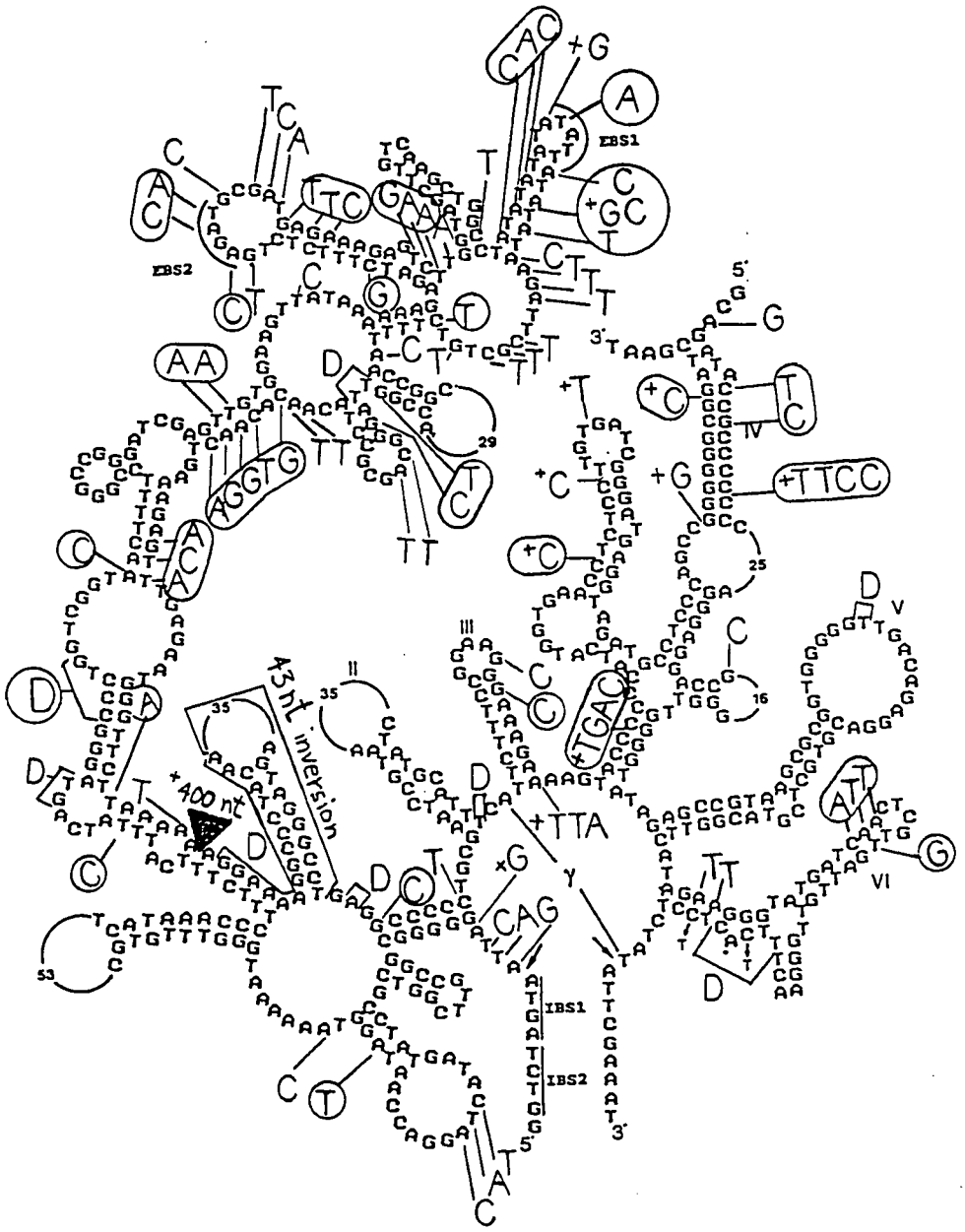
APPENDIX II. Physical map of the wheat mitochondrial genome derived from Lejeune and Quétier (1987). Letters represent the SalI restriction fragments that constitute the "master circle". The positions of *nad1* sequences and their orientations and shown (arrows).



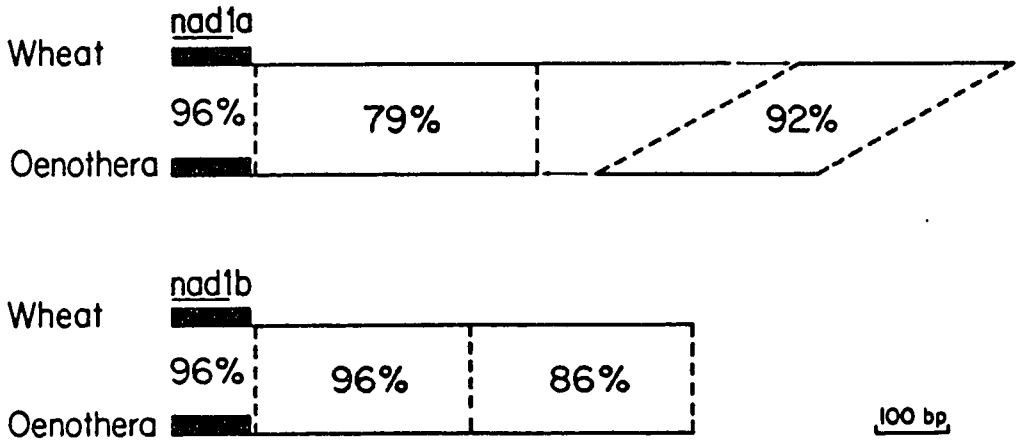
APPENDIX III. Secondary structure model of the wheat mitochondrial *nad1*b/c intron. The group II intron structure is based on the one proposed for the watermelon counterpart (Michel et al. 1989). Intron/exon borders are indicated by solid triangles. Intron binding sequences (IBS1 and IBS2) and exon binding sequences (EBS1 and EBS2) are shown by arrows.



APPENDIX IV. A. Secondary structure model of the *Oenothera nad1*a/b intron proposed by Wissinger and colleagues (1991). Nucleotide substitutions and deletions that are observed in wheat are indicated; those that destabilize helices are circled.



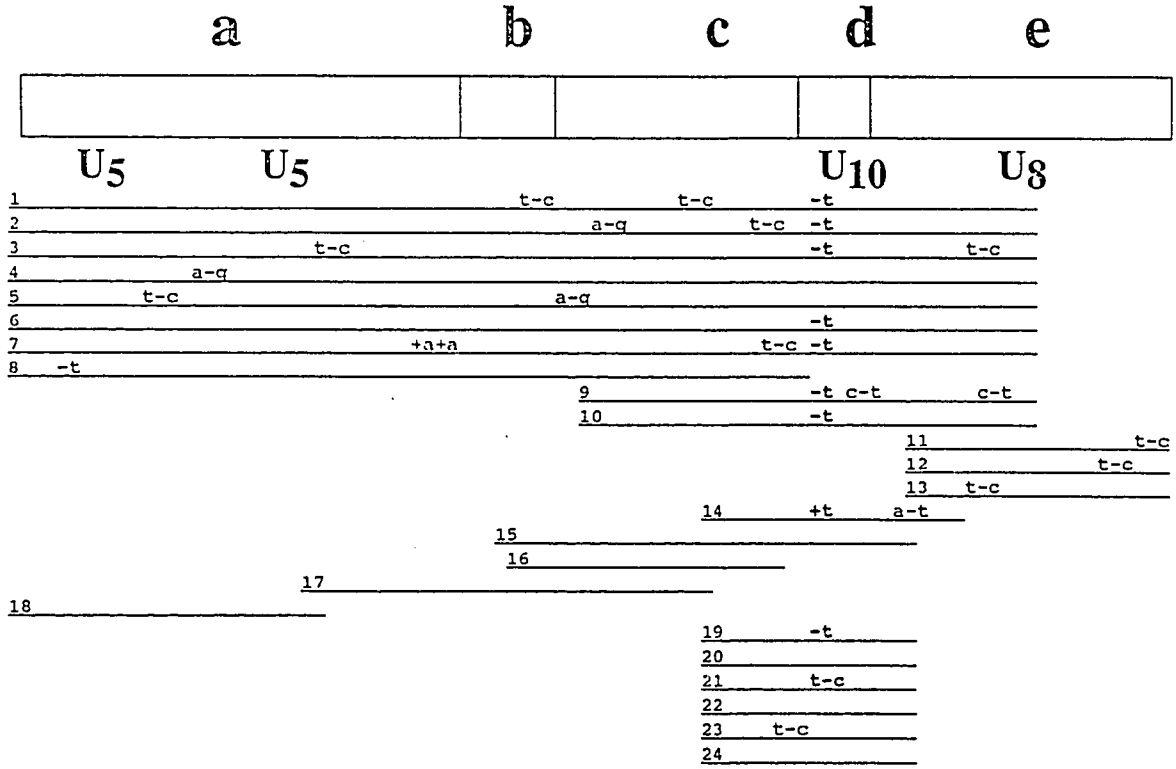
B. Schematic representation of the alignment of sequences from *nad1a* region between wheat and *Oenothera*. The arrows depict an inversion of 43 bp.



APPENDIX V. Compilation of RNA editing sites found within plant mitochondrial introns.

Intron	Plant	Domain	contribution to stability	frequency	Reference
*nad1*a/b	wheat	VI(-6)	?	4/4	this work
"	pea	VI(-6)	?	high	this work
"	Oenothera	VI(-7)	+	1/10	Wissinger et al.1991
"	"	VI(-11)	-	10/10	"
*nad1*d/e	petunia	VI(-10)	-	?	Conklin et al. 1991
*nad5*a/b	Oenothera	VI(-10)	-	?	Knoop et al. 1991
*nad5*d/e	"	unmodelled	?	?	"
"	"	unmodelled	?	?	"
*nad2*b/c	"	I(+7)	+	high	Binder et al. 1992
"	"	IV	+	high	"
"	"	unmodelled	?	high	"

APPENDIX VI. Rare nucleotide changes observed in wheat mitochondrial *nad1* coding sequences. The five exons (*nad1*a-e) in wheat (Chapdelaine and Bonen 1991) are shown and U-stretches of at least 5 nucleotides are indicated. The different cDNA clones analyzed are shown below with the type of changes (numbered as in the *nad1* ORF, see Table 2) being: T-to-C (136, 285, 452, 582, 658, 663, 678, 824, 827, 941, 963); C-to-T (702, 835); A-to-G (186, 474, 522); +A (361, 364); A-to-T (765); -T (64). Open circles represent sites that are not edited in certain clones. Changes in the U_{10}-stretch are in bold.



APPENDIX VII

Cell, Vol. 65, 465–472, May 3, 1991, Copyright © 1991 by Cell Press

The Wheat Mitochondrial Gene for Subunit I of the NADH Dehydrogenase Complex: A *Trans*-Splicing Model for This Gene-in-Pieces

Yvan Chapdelaine and Linda Bonen
Department of Biology
University of Ottawa
Ottawa
Canada K1N 6N5

Summary

The *nad1* gene encoding subunit I of the respiratory chain NADH dehydrogenase is fragmented into five unique-copy coding segments that are scattered over at least 40 kb and interspersed with other genes in the wheat mitochondrial genome. The *nad1* segments are flanked by sequences with group II intron features, and transcript analysis demonstrates the presence of correctly spliced mRNAs. RNA editing occurs at sites asymmetrically distributed along the wheat *nad1* coding region, and the initiation codon is created by RNA editing. The unusual organization of the wheat *nad1* gene is attributed to mitochondrial DNA rearrangements within introns, and a *trans*-splicing model involving secondary structural interactions between group II–like intron pieces is proposed for its expression.

Introduction

Plant mitochondrial DNA (mtDNA) rearrangements are known to occur frequently during evolution (reviewed by Gray, 1989; Palmer, 1990), but are expected to be confined to regions that do not affect individual gene structure or expression. Such intergenic spacer regions are plentiful in plant mtDNA because, although these genomes are very large compared with those of other organisms (i.e., 15- to 200-fold larger than animal mtDNA), they contain relatively few extra genes (reviewed by Levings and Brown, 1989; Lonsdale, 1989). Somewhat surprisingly, comparative analysis of mtDNA sequences from closely related plants indicates that DNA rearrangements can occur very close to or even within coding regions. Consequently, genes that are otherwise almost identical can differ in predicted protein sequence at their termini and have unrelated regulatory sequences. Although the mechanisms underlying DNA reorganization in plant mitochondria are not well understood, most genomes contain long recombinationally active repeated elements (Lonsdale, 1989; Small et al., 1989) as well as short stretches within intergenic regions that are homologous to bits and pieces of genes.

One such short gene segment of approximately 30 codons has been identified at a single genomic site downstream of the *rps13* gene in the mitochondria of tobacco, maize (Bland et al., 1986), and wheat (Bonen, 1987). It shows homology to an internal region of the gene for subunit I of the respiratory chain NADH dehydrogenase (NADH–ubiquinone oxidoreductase), designated as *nad1* in plant mitochondria. Unexpectedly, sequences corresponding to the 125–130 amino-terminal *nad1* codons could not be identified within the stretch of approximately 1 kb up to the *rps13* gene. Several *nad1* homologous segments, including the one mentioned above, have also been identified in watermelon mitochondria (Stern et al., 1986), and two segments homologous to 3′-terminal *nad1* coding sequences have recently been characterized in broad bean (Wahleithner et al., 1990). In no case, however, do the reported fragments comprise a full-length *nad1* gene capable of encoding a polypeptide of the expected 320–330 amino acids. It has been suggested that these plant *nad1* homologous segments might be pseudogenes (see Makaroff and Palmer, 1987; Lonsdale, 1989), with the functional NAD1 polypeptide being either nuclear encoded or a product of the evolutionarily related chloroplast *ndhA* gene (Matsubayashi et al., 1987).

We have considered an alternative possibility, that is, that the entire *nad1* gene is indeed present in wheat mtDNA, but that coding segments are dispersed so that transcripts must undergo unusual *cis* or *trans* splicing to yield functional mRNA molecules. It should be noted that although plants have exceptionally large mitochondrial genomes, only a few genes have been reported to be split (reviewed in Lonsdale, 1989), and their introns have been classified as group II because of distinctive secondary structural features (Michel et al., 1982, 1989). Organellar group I or II introns typically range in length from 0.6 kb to as much as 4 kb. Some are known to be self splicing in vitro (Cech, 1986), and some behave as mobile genetic elements (Perlman and Butow, 1989; Lambowitz, 1989). Of particular interest here is the unusual organization of two chloroplast genes whose exons are widely separated in the genome and independently transcribed: the *rps12* gene in plants (Fukuzawa et al., 1986; Zaita et al., 1987; Hildebrand et al., 1988) and the *psaA* gene in Chlamydomonas (Kück et al., 1987). It is believed that secondary structural interactions between group II intron halves flanking their coding segments are crucial for mRNA maturation (Choquet et al., 1988; Kohchi et al., 1988).

Results

Identification of the Five Wheat *nad1* Gene Segments

We have identified and characterized five *nad1* coding segments, designated *nad1*a–*nad1*e, that together are capable of encoding a polypeptide of 325 amino acids in wheat mitochondria (Figure 1; Figure 2). These *nad1* sequences are however scattered over four distant genomic sites. Only the second and third exons (namely, *nad1*b and *nad1*c) are close together; they are separated by a group II intron of 1422 bp. These two exons were initially identified by DNA sequence analysis of the region beginning 947 bp downstream of the *rps13* gene (Bonen, 1987).

To search for 5′ *nad1* coding sequences, we used a synthetic oligomer (oligo B2) that is complementary to *nad1*b as primer on wheat mtRNA template in reverse

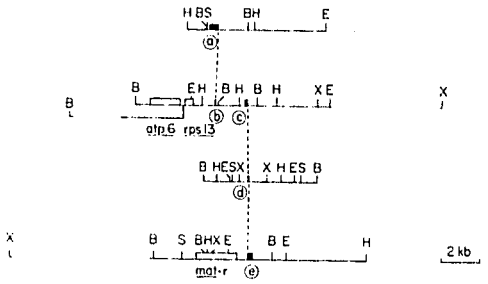


Figure 1. Restriction Maps of the Four Wheat Mitochondrial Genomic Regions Containing *nad1* Gene Segments

The schematic shows the *nad1* coding regions (closed bars and circled letters), the *atp6*, *rps13*, and *mat-r* genes (open bars). Only the EcoRI (E), BamHI (B), and HindIII (H) sites flanking the *nad1* exons are shown, in addition to those XhoI (X) and SalI (S) sites used to establish minimum distances between exons. The dotted lines represent uncloned regions.

transcriptase dideoxy sequencing experiments. We observed approximately equimolar amounts of two sequences (data not shown), one corresponding to an unspliced RNA and the second to a processed RNA species. This information was used for the synthesis of another oligomer (oligo A) that served as a hybridization probe to retrieve the putative 5′ *nad1* coding region from a plasmid clone bank. DNA sequence analysis confirmed the presence of a reading frame for the amino-terminal portion of an NAD1 polypeptide (Figure 2A). This segment displays 50% amino acid identity with its human mitochondrial ND1 counterpart (Anderson et al., 1981) and 98% with that of Oenothera (Wissinger et al., 1991) after adjustments are made for RNA editing (see below). Our restriction mapping and Southern blot analysis have established that this *nad1* segment (as well as each of the other four segments) is present as a single genomic copy (data not shown) and that *nad1*a is a minimum of 20 kb away from *nad1*b. Moreover, at least two genes, namely *atp6* and *rps13*, are located between them (Figure 1; Bonen, 1987; Bonen and Bird, 1988).

The 3′ *nad1* coding segments (designated *nad1*d and *nad1*e; Figures 2C and 2D) were identified by heterologous hybridization experiments using synthetic oligomers (oligos D and E2) that were designed from published plant *nad1* homologous segments, namely, broad bean *nad1* subterminal exons (Wahleithner et al., 1990) and the watermelon ORF36 (Stern et al., 1986), respectively. Surprisingly, these regions are also very far apart from each other and from the mid-region of the gene. The *nad1*c/d and *nad1*d/e segments are estimated to be separated by at least 7 kb and 12 kb, respectively (Figure 1). Consequently, the five wheat *nad1* coding segments are scattered over a minimum distance of 40 kb. In fact, they may well be much farther apart, encoded on opposite DNA strands, or even on different molecular forms of the genome (see Lonsdale, 1989). When this limited mapping information is correlated with the 430 kb SalI restriction map of wheat mtDNA (Lejeune and Quétier, 1988), much greater distances (and additional mitochondrial genes) are anticipated between *nad1* segments.

The five *nad1* exons are predicted to encode 129, 27, 64, 20, and 85 amino acids, respectively (Figure 2), taking into account exon/intron borders and RNA editing events (discussed below). The positions of the discontinuities in the wheat *nad1* coding sequence do not precisely correspond to the locations of introns identified in homologous genes from other organisms; however, the third (group I) intron in the Podospora mitochondrial ND1 gene (Cummings et al., 1988) is close to the wheat *nad1*a/b junction, and the single chloroplast *ndhA* (group II) intron (Matsubayashi et al., 1987) is near the site of the wheat *nad1*-b/c intron.

The region preceding the wheat *nad1*e exon contains an open reading frame of 678 codons (Figure 2D) that is very closely related to the one (maturase-related gene [*mat-r*]) previously identified in broad bean (Wahleithner et al., 1990) at the same genomic site (compare 86% amino acid identity versus 90% for the wheat/broad bean *nad1*d and *nad1*e exons). The pattern of nucleotide substitutions observed between these two *mat-r* genes supports the view that this sequence is under functional constraint. Both genes are related to fungal mitochondrial group II intronic ORFs (see Lang et al., 1985) and possess six of the nine "invariant" amino acid residues characteristic of reverse transcriptases (Michel and Lang, 1985; Wahleithner et al., 1990).

The *nad1* Initiation Codon Is Created by RNA Editing, and Other Sites Are Asymmetrically Distributed

The wheat NAD1 sequence is homologous along its full length to its counterparts from other organisms; however, the gene lacks an ATG triplet in the vicinity of the predicted start codon. Initiation of translation may involve a nonconventional initiator, additional splicing events to generate the extreme 5′ coding sequences possessing an AUG codon, or RNA editing to create an AUG codon. To distinguish among these possibilities, we used a synthetic oligomer (oligo A2) in direct cDNA sequencing experiments. The results, shown in Figure 3A, demonstrate the presence of RNA editing that transforms the ACG codon at position 134–136 (Figure 2A) to AUG in approximately 50% of the RNA molecules. The conversion of C to U is typical of RNA editing in plant mitochondria (Gualberto et al., 1989; Covello and Gray, 1989; Hiesel et al., 1989); however, this is the first report of an RNA editing event that would be obligatory for protein synthesis to proceed. Interestingly, the *mat-r* gene, which also lacks an initiation codon in both wheat (Figure 2D) and broad bean (Wahleithner et al., 1990), does not contain any codon near its 5′ terminus that could in this manner be converted to AUG. As yet, it is not known how translation of this reading frame is initiated.

To further examine the nature and extent of RNA editing within the wheat *nad1* coding region, we determined the sequences of cloned PCR-amplified cDNAs derived from transcripts in which all five exons had been spliced (see Figure 3B, lane 5; oligomers A3 and E as primers). Positions at which the wheat *nad1* genomic and cDNA se-

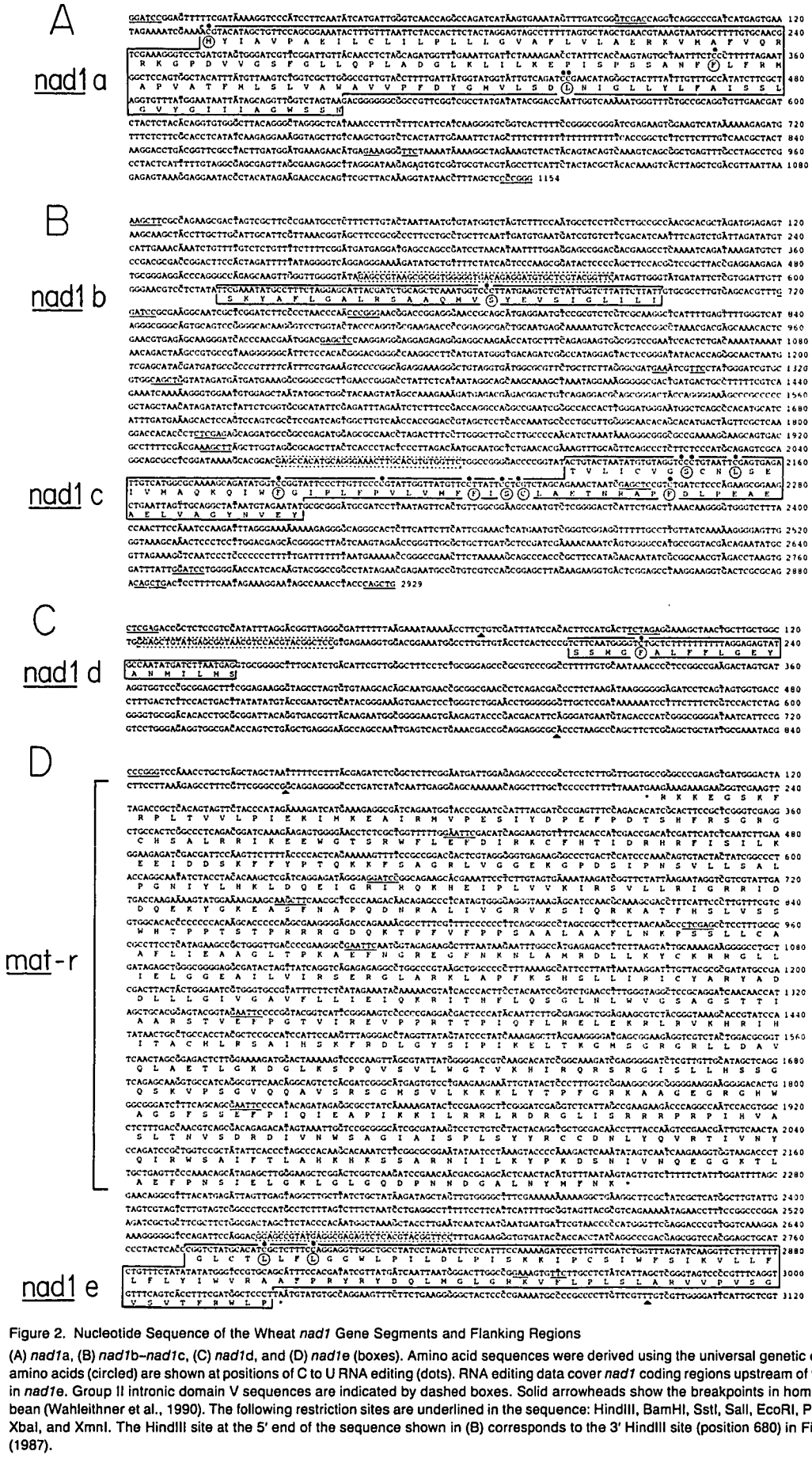


Figure 2. Nucleotide Sequence of the Wheat *nad1* Gene Segments and Flanking Regions

(A) *nad1*a, (B) *nad1*b–*nad1*c, (C) *nad1*d, and (D) *nad1*e (boxes). Amino acid sequences were derived using the universal genetic code, but altered amino acids (circled) are shown at positions of C to U RNA editing (dots). RNA editing data cover *nad1* coding regions upstream of the open triangle in *nad1*e. Group II intronic domain V sequences are indicated by dashed boxes. Solid arrowheads show the breakpoints in homology with broad bean (Wahleithner et al., 1990). The following restriction sites are underlined in the sequence: HindIII, BamHI, SstI, SalI, EcoRI, PvuII, XhoI, SmaI, XbaI, and XmnI. The HindIII site at the 5′ end of the sequence shown in (B) corresponds to the 3′ HindIII site (position 680) in Figure 2 of Bonen (1987).

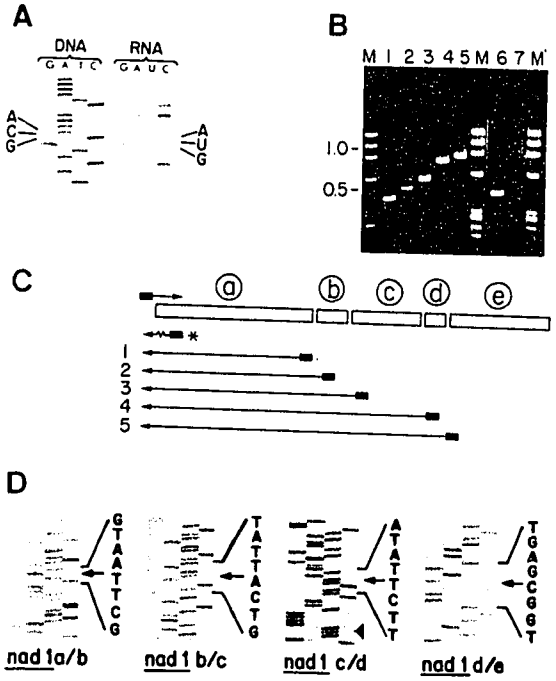


Figure 3. Direct cDNA Sequencing and Characterization of PCR-Amplified *nad1* cDNAs

(A) cDNA sequencing gel autoradiographs of *nad1* RNA sequences at the initiation codon site, using oligo A2 (asterisk) as depicted in schematic (C). (B) PCR-amplified products of five *nad1* cDNAs (lanes 1–5) synthesized using nested, exon-specific primers (oligos A, B, C, D, and E) with oligomer A3 shown in schematic (C). Wheat mtDNA was used as template with oligomer sets A/A3 (lane 6) and B/A3 (lane 7). (D) Sequence analysis of cloned PCR products at the four *nad1* exon junctions (arrows). The arrowhead indicates an edited U at position 215 within *nad1*d (Figure 2C). Lane order: G, A, T, C.

quences differ are shown by dots in Figure 2, and altered amino acids are circled. Each exon contains at least one editing site, and all consist of C to U conversions. Fourteen of the 17 observed editing positions result in amino acid substitutions, the majority of which increase the similarity of the predicted wheat NAD1 protein with those of other organisms. Of particular interest are the large number of RNA editing sites within *nad1*c compared with the other exons; 13% of the *nad1*c codons are edited, whereas only 2%–5% of the codons are altered in *nad1*a, *nad1*b, or *nad1*d, as if the latter have undergone RNA-mediated gene conversion events more recently than has *nad1*c. Alternatively, this asymmetric distribution of editing sites may relate to constraints imposed by the complex RNA splicing events required for expression of this gene.

A comparison of the wheat and Oenothera *nad1* RNA sequences (this paper; Wissinger et al., 1991) indicates that editing occurs at both homologous and unique sites in the two plants and that the two resulting proteins are predicted to be almost identical. The Oenothera *nad1* transcript is somewhat more heavily edited than that of wheat, and there are seven sites at which it is converted to the sequence found in the wheat mtDNA. In the converse comparison, five such sites are seen in wheat, and interestingly, four of them are located within *nad1*c, again highlighting the clustering of sites within this exon.

Transcripts of *nad1* Coding Segments Are Correctly Spliced

Direct cDNA sequence analysis of wheat *nad1* transcripts, as used in the identification of *nad1*a, indicates that both precursor and processed RNAs are present. This has been confirmed by Northern blot analysis in which complex RNA species, from 2 kb to >3.5 kb in apparent size, are seen (Bonen, 1987; data not shown). Some of the abundant precursors contain individual coding segments and only immediate flanking regions. For example, transcripts containing sequences upstream of *nad1*b do not extend into the *rps13* gene region (Bonen, 1987).

To determine the precise splice junctions, PCR-amplified reverse transcripts were analyzed. Oligomers A, B, C, D, and E, which are specific to the respective *nad1* exons, were used as primers on wheat mtRNA, and the resulting cDNAs were PCR amplified using oligomer A3 as the second primer. The products, shown in Figure 3B (lanes 1–5), correspond approximately to the predicted sizes of 422, 494, 578, 762, and 825 bp, respectively. Sequence analysis of cloned cDNAs allowed the assignment of precise exon boundaries (Figure 3D) and confirmed that the correct reading frame for NAD1 protein synthesis is generated. It should be noted that a 2 nucleotide redundancy occurs at the *nad1*a/b border (see below). The PCR-generated results were corroborated by direct cDNA sequencing across splice junctions (data not shown).

When wheat mtDNA, rather than RNA, was used with oligomers A and A3 in PCR experiments, the predicted product (approximately 422 bp) was seen (Figure 3B, lane 6). In contrast, when oligomers B and A3 were used as primers, no product was seen (Figure 3B, lane 7). These data, in conjunction with our Southern hybridization data, are consistent with the view that *nad1*a and *nad1*b are not closely linked in the wheat mitochondrial genome.

nad1 Coding Sequences Are Flanked by Group II–Like Intronic Structures

Previous studies of watermelon and broad bean *nad1* homologous segments (Stern et al., 1986; Wahleithner et al., 1990) had established the presence of group II introns between the segments corresponding to *nad1*b/c and *nad1*d/e, respectively. Closely related sequences are seen in wheat except that an mtDNA rearrangement has disrupted the intron between *nad1*d and *nad1*e (see below). When sequences flanking the wheat *nad1* coding segments were examined in detail, in three of the four cases (Figures 4B–4D), features were found that are clearly diagnostic of group II introns (Michel et al., 1989): namely, a conserved block at the 5′ splice site (GTGCG or GCGCG) and the helical domain V structure near the 3′ splice site (Figure 2, dashed blocks; Figure 4).

The discontinuous intron between *nad1*a and *nad1*b, however, is atypical for many reasons: First, domain V has 17 nucleotides in the loop rather than the usual 4 (GAAA) purines (Figure 4A). Second, the 5′ splice site lacks the

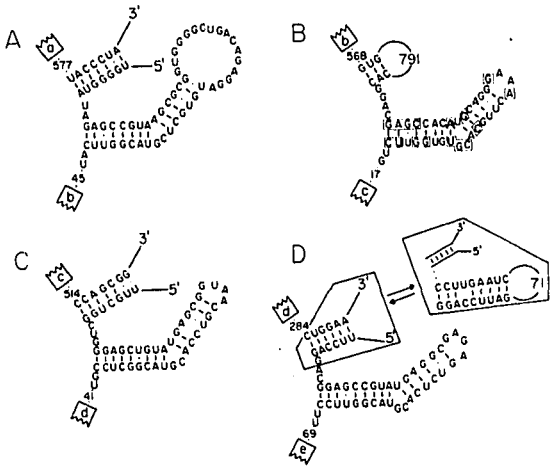


Figure 4. Secondary Structural Models Illustrating Putative Splicing of Wheat *nad1* Exons

(A, C, and D) *Trans* splicing. (B) *Cis* splicing. The only features of the group II core structure (Michel et al., 1989) shown are the well-defined domain V (hairpin) and the upstream domain IV, with a potential discontinuity within the looped region (A, C, and D) or within domains I, II, or III (inset in [D]). The structure in (B) is shown as proposed for watermelon (Michel et al., 1989), and positions conserved among the four wheat *nad1* domain V sequences are boxed. The numbers indicate distances to the exons.

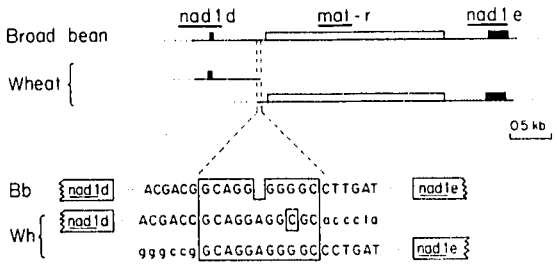


Figure 5. Comparison of the nad1d and nad1e Exon Organization in Wheat and Broad Bean Mitochondria

The schematic shows the *mat-r* gene (open bar) and *nad1* segments (closed bars) in wheat and broad bean (Wahleithner et al., 1990). The expanded region depicts in detail the purine-rich sequence (blocked) that is repeated at the breakpoint in wheat. Lower case letters represent nonconserved positions.

consensus sequence (compare GACGG or AGACG due to dinucleotide redundancy at the border, Figure 2A; Figure 3D). Third, domain VI lacks the bulged adenosine that is normally found 7 or 8 nucleotides upstream of the 3′ splice site and is involved in lariat structure (van der Veen et al., 1986; Schmelzer and Schweyen, 1986). Fourth, domain VI cannot be folded into the conventional helical structure. Interestingly, these latter two points are predicted to hold for the maize mtDNA counterpart (Bland et al., 1986), because both monocots lack a 5 nucleotide stretch (containing a potential bulged adenosine) that is found in the dicots, tobacco (Bland et al., 1986), and Oenothera (Wissinger et al., 1991), and the monocot/dicot sequences differ within domain VI at numerous positions (in ways that cannot be corrected by C to U RNA editing nor result in compensatory base changes). In addition, the wheat and Oenothera sequences differ both at their 5′ splice sites (with neither conforming to the classical group II consensus sequence) and within EBS1 and EBS2, the proposed exon-binding sites (Wissinger et al., 1991). We conclude that although domain V, a hallmark of group II introns, is present in the wheat *nad1a/b* intervening region, many other features are absent or altered.

Because RNA folding is believed to be crucial in the splicing of these discontinuous transcripts, we have searched for potential base pairing between the flanking intron halves. In Figure 4, we present the most thermodynamically stable structures that could be found between the sequences immediately upstream of domain V and those within the 5′ intron pieces (Figures 4A, 4C, and 4D; Figure 4B as in Michel et al., 1989 for the watermelon group IIA intron). These hypothetical RNA interactions between intron pieces have been selected for illustrative purposes because domain V, unlike domains I–III, is well defined and enables the accurate positioning of one-half of the domain IV helix, and it has been shown that *trans* splicing can still proceed in vitro with a discontinuity within loop IV (Jarrell et al., 1988). Because the breakpoint in homology between the wheat and broad bean sequence upstream of *nad1*d is located only 57 nucleotides from domain V (Figure 2C, arrowhead), a DNA rearrangement within loop IV seems likely. In the case of the *nad1d/e* discontinuity, on the other hand, because the sequence downstream of *nad1*d is conserved between wheat and broad bean for only 539 bp (Figure 2C; Figure 5), which is rather short to accommodate domains I–IV, it might be argued that the DNA rearrangement involves loop II or III (Figure 4D, inset) rather than loop IV. It should be noted that in vitro, the presence of domain V is crucial for *trans* splicing, whereas domain IV can be removed without abolishing splicing (Jarrell et al., 1988).

A Short Direct Repeat Is Located at the Site of the Wheat *nad1d/e* Intron Discontinuity

The structure of the 3′ end of the wheat *nad1* gene differs radically from that recently reported for broad bean, in which exons *nad1*d and *nad1e* are separated by a classical group II intron of 3143 bp containing a *mat-r*-like open reading frame (Wahleithner et al., 1990). The wheat and broad bean sequences are conserved for 539 bp downstream of *nad1*d (Figure 2C, position 799) and 68 bp upstream of the *mat-r* gene that precedes *nad1*e (Figure 2D, position 147). Beyond these points, homology abruptly ends because of a genomic rearrangement (Figure 5) that separates the wheat *nad1*d and *nad1*e segments by at least 12 kb (see above), but does not result in the loss of any intron sequences. Interestingly, the breakpoint occurs at the position of a purine-rich stretch of 11 nucleotides in broad bean, so that there are two copies of this sequence present in wheat, one downstream of *nad1*d and the second upstream of *mat-r/nad1*e (Figure 5).

Discussion

The unusual organization of the wheat *nad1* gene raises fundamental questions about how gene pieces that are physically very far apart in the mitochondrial genome can give rise to functional messenger RNA molecules. A number of observations suggest that the wheat *nad1* coding segments, which are scattered over four distant genomic sites, are expressed through *trans* splicing. First, the longest introns yet identified in any organellar gene are less than 4 kb in length, whereas the wheat *nad1* coding segments are a minimum of 20, 7, and 12 kb apart (Figure 1). Moreover, the juxtaposition of other genes upstream of internal *nad1* pieces (e.g., *atp6/rps13* genes preceding *nad1*b) would place those genes within introns, an unorthodox organization in organellar genomes. Their proximity to upstream genes could, on the other hand, provide transcriptional signals for expression. Second, the abundant wheat *nad1* precursor transcripts do not extend far beyond the individual coding regions (see Bonen, 1987). Third, in broad bean mitochondria, there is a group II intron separating *nad1*d and *nad1*e, whereas in wheat the intron is broken into two pieces that are at least 12 kb apart. Finally, the wheat *nad1* coding segments are flanked by group II–like intron features, reminiscent of the unusual organization of two *trans*-split chloroplast genes, namely, the *rps12* gene in plants (Fukuzawa et al., 1986; Zaita et al., 1987; Hildebrand et al., 1988) and the *psaA* gene in Chlamydomonas (Kück et al., 1987). *Trans* splicing of these two chloroplast genes, which is believed to involve interactions between group II intron pieces, is considered distinct from another well-studied form of *trans* splicing, that is, the addition of 5′ leaders to protist and nematode nuclear-encoded mRNAs (Agabian, 1990). However, similarities among organellar group II introns, nuclear pre-mRNA introns, and nuclear *trans*-split introns are emerging (Jacquier, 1990), and the phenomenon of *trans* splicing may be more widespread than previously thought (Dandekar and Sibbald, 1990).

We propose that expression of the wheat *nad1* gene involves three *trans*-splicing and one *cis*-splicing events mediated through RNA folding of group II–like sequences. To illustrate the concept simply, we have shown intron halves base pairing to form domain IV (Figure 4). *Trans* splicing might be expected to differ from *cis* splicing in requiring either additional specialized secondary structural interactions or extra machinery (such as small RNAs and/or proteins) for the correct recognition and alignment of intron pieces (particularly in a milieu where multiple *trans*-splicing events occur), and this might explain the absence of certain group II features and unexpected sequence differences (that cannot be compensated for by C to U RNA editing) seen between wheat and Oenothera (compare 5′ splice site and domain VI of *nad1*a/b). Indeed, in the case of the Chlamydomonas *trans*-split *psaA* gene, a product of the chloroplast *tscA* genetic locus has been determined to be essential for the splicing of exon 1 to 2 but not exon 2 to 3 (Roitgrund and Mets, 1990; Goldschmidt-Clermont et al., 1990), and a number of nuclear gene products are also required for processing (Choquet et al., 1988; Herrin and Schmidt, 1988). Transcript analysis to precisely define the ends of plant mitochondrial *nad1* intron pieces and additional RNA editing information as well as functional in vitro studies similar to those performed on fungal mitochondrial group II introns (see Jarrell et al., 1988) will be important in elucidating key RNA interactions.

One fundamental problem in the processing of *trans-split genes is that the RNA pieces* have to find their correct partners. We find no evidence for the presence of transcripts containing mismatched *nad1* segments (compare single PCR-generated products, Figure 3B); however, we do see much higher levels of *nad1* precursor RNAs than those of *cis*-split genes such as *coxII* (Bonen et al., 1984). This presumably reflects different efficiencies of RNA processing. Interestingly, our preliminary data suggest that the wheat *nad1* and *nad5* genes follow developmentally specific RNA processing pathways (Bonen, 1987; S. Bird and L. B., unpublished data).

In addition to investigating the mechanisms by which wheat *nad1* mRNAs are generated, it will also be of interest to learn how and when such events have taken place. From a comparison of the *nad1*d/e regions in broad bean and wheat, it is apparent that an mtDNA rearrangement within that group II intron has occurred since the divergence of monocots and dicots from a common ancestor some 200 million years ago (Wolfe et al., 1989). The presence of almost identical purine-rich stretches at the breakpoints (Figure 5) suggests their involvement in the DNA rearrangement processes, either as sites mediating homologous recombination or as repeats generated as a consequence of double-stranded DNA repair. Upstream of *nad1*d, the wheat and broad bean sequences share homology for only 136 bp, and downstream of *nad1*e the breakpoint in homology occurs 66 bp after the termination codon (Figure 5). This suggests that additional rearrangement events have occurred within these regions of plant *nad1* genes. In contrast, sequences preceding the *nad1*b segment are conserved for at least 0.9 kb in all plants examined (Stern et al., 1986; Bland et al., 1986; Bonen, 1987; Wissinger et al., 1991), indicating a long evolutionary history for this genomic organization. The *trans*-split chloroplast *rps12* gene organization is similarly a very stable one, being present in both liverwort and tobacco, two land plants that have had separate lineages for some 350–400 million years (Wolfe et al., 1989).

One may also question why this phenomenon has not been observed more frequently in nature. Of more than 70 group II introns characterized in a wide variety of genetic systems (Michel et al., 1989), only two chloroplast genes had been identified to be *trans* split prior to the present studies. Interestingly, the ribosomal RNA genes in Chlamydomonas mitochondria (Boer and Gray, 1988a) also show radical intragenic rearrangements that necessitate long-range, noncovalent RNA interactions (although not splicing) for their expression. It may be significant that both of these organisms have mitochondrial genes with homology to reverse transcriptases (Boer and Gray, 1988b; Wahleithner et al., 1990; this paper). It must be

emphasized that even though plant mitochondria are exceptional among organelles in the extent of recombinational activity and in the proportion of noncoding DNA, intragenic rearrangements would be expected to be deleterious unless the translocated segment is placed under appropriate transcriptional control and the RNA molecules can associate correctly to undergo proper processing.

This work clearly demonstrates that short gene segments can be dispersed around the wheat mitochondrial genome and yet still comprise functional coding entities that are expressed through complex RNA processing. The presence of group II introns is intimately associated with this phenomenon and could play an important evolutionary role in the creation of novel genes from pre-existing coding domains or the conversion of long genes into shorter autonomous ones. Plant mitochondria, long known for the fluidity of their genome organization, must now be recognized for a remarkable plasticity in individual gene structure.

Experimental Procedures

Nucleic Acid Isolation and Cloning

MtDNA and RNA were isolated from wheat embryos (Triticum aestivum var. Thatcher and Frederick) that had been germinated in the dark for 24 hr (DNA) or 3 days (RNA) as described previously (Bonen and Gray, 1980; Wilson and Chourey, 1984). Wheat mtDNA restriction fragments were recovered from agarose gels using GeneClean (BIO/CAN Scientific) and ligated into pUC plasmid vectors for transformation of E. coli strain TB1. Clones were retrieved from banks using colony hybridization methods (Maniatis et al., 1982). DNA restriction fragments of interest were subcloned into appropriate M13 vectors (Maniatis et al., 1982) and sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977) using [α-^{35}S]dATP or [α-^{32}P]dATP (Amersham and NEN/Du Pont) with either Sequenase (USB) or Klenow polymerase (Amersham). Both strands were sequenced, and dITP was used in some cases. The sequence data were analyzed using Microgenie (Beckman) programs and GenBank/National Biomedical Research Foundation data banks.

Reverse Transcriptase Sequencing

Synthetic oligomers were labeled at their 5′ termini with T4 polynucleotide kinase (Pharmacia) and [γ-^{32}P]ATP (NEN/Du Pont). After purification on Sephadex G-50 spin columns, the oligonucleotides were annealed with approximately 25 μg of wheat mtRNA and extended by the dideoxy chain termination method using avian myeloblastosis virus reverse transcriptase (Life Sciences), following the procedure two as described by Geliebter et al. (1986). Products were resolved on 7% acrylamide sequencing gels.

Oligonucleotide Synthesis

The following oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and have been designated according to the exon that they specify.

A: CCTGCTATAATTATTCCATAAACGC	(Figure 2A, positions 507–483)
A2: AAGGCTACTCCTAGTAGAAG	(Figure 2A, 198–179)
A3: TCGGGTCGACCAGGTCAGGC	(Figure 2A, 86–105)
B: ACTTCATAAGGGACCATTTG	(Figure 2B, 676–657)
B2: CGTAATGCTCCTAGAAAGGC	(Figure 2B, 646–627)
C: ATCTGCTTTTGCGCCATGAC	(Figure 2B, 2182–2163)
D: TAAGATCATATTGGCATACT	(Figure 2C, 255–236)
E: AATGGCAAGATCTAGGATAG	(Figure 2D, 2829–2810)
E2: GAGCTAATGATAGAGGCAAGAACAC	(Figure 2D, 2980–2956)

The oligomer E2 sequence was derived from a conserved region within the watermelon mitochondrial ORF36 (Stern et al., 1986) and oligomer D from the broad bean *nad1* subterminal exon (Wahleithner et al., 1990).

PCR Amplification of cDNAs

cDNAs were prepared using the method described above for reverse transcriptase dideoxy sequencing, except that the dideoxynucleotide mixes were omitted. The products were then incubated an additional hour with 2 μl of 2 mM dNTPs. Purified cDNA pellets were dissolved in 100 μl of TE (10 mM Tris–HCl [pH 7.5], 1 mM EDTA). Each amplification tube contains 600 ng of each primer (approximately 100 pmol), 10 μl of amplification buffer (10 mM Tris–HCl [pH 9.0], 500 mM KCl, 15 mM $MgCl_2$, 0.1% gelatin, 1% Triton), 10 μl of dNTPs (2 mM), 2 U of Taq DNA polymerase (BIO/CAN), and 10 μl of DNA template in a final volume of 100 μl. Thirty cycles were performed on a Perkin Elmer Cetus DNA thermal cycler, following the recommended protocols of the manufacturer: 1 min at 94°C, 2 min at the appropriate annealing temperature (Geliebter et al., 1986), and 3 min at 72°C. The SalI and BglII restriction sites within oligomers A3 and E, respectively were used for cloning the amplification products into M13 vectors.

Acknowledgments

We thank P. H. Boer for helpful discussions and D. Wolstenholme for providing a preprint of the broad bean *nad1d/e* paper. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. L. B. is the recipient of an NSERC University Research Fellowship, and Y. C. holds a graduate scholarship from le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received December 20, 1990; revised February 27, 1991.

References

Agabian, N. (1990). *Trans* splicing of nuclear pre-mRNAs. Cell *61*, 1157–1160.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457–465.

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

Boer, P. H., and Gray, M. W. (1988a). Scrambled ribosomal RNA gene pieces in Chlamydomonas reinhardtii mitochondrial DNA. Cell *55*, 399–411.

Boer, P. H., and Gray, M. W. (1988b). Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in Chlamydomonas reinhardtii mitochondrial DNA. EMBO J. *7*, 3501–3508.

Bonen, L. (1987). The mitochondrial S13 ribosomal protein gene is silent in wheat embryos and seedlings. Nucl. Acids Res. *15*, 10393–10404.

Bonen, L., and Bird, S. (1988). Sequence analysis of the wheat mitochondrial *atp6* gene reveals a fused upstream reading frame and markedly divergent N termini among plant ATP6 proteins. Gene *73*, 47–56.

Bonen, L., and Gray, M. W. (1980). Organization and expression of the mitochondrial genome in plants I. The genes for wheat mitochondrial ribosomal and transfer RNA: evidence for an unusual arrangement. Nucl. Acids Res. *8*, 319–335.

Bonen, L., Boer, P. H., and Gray, M. W. (1984). The wheat cytochrome oxidase subunit II gene has an intron insert and three radical amino acid changes relative to maize. EMBO J. *3*, 2531–2536.

Cech, T. R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell *44*, 207–210.

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.

Covello, P. S., and Gray, M. W. (1989). RNA editing in plant mitochondria. Nature *341*, 662–666.

Cummings, D. J., Domenico, J. M., and Michel, F. (1988). DNA sequence and organization of the mitochondrial ND1 gene from Podospora anserina: analysis of alternate splice sites. Curr. Genet. *14*, 253–264.

Dandekar, T., and Sibbald, P. R. (1990). *Trans*-splicing is predicted to occur in a wide range of organisms including vertebrates. Nucl. Acids Res. *18*, 4719–4725.

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.

Geliebter, J., Zeff, R. A., Melvold, R. W., and Nathenson, S. G. (1986). Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: K^{bm9} and K^{bm6}. Proc. Natl. Acad. Sci. USA *83*, 3371–3375.

Goldschmidt-Clermont, M., Girard-Bascou, J., Choquet, Y., and Rochaix, J.-D. (1990). *Trans*-splicing mutants of Chlamydomonas reinhardii. Mol. Gen. Genet. *223*, 417–425.

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

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

Herrin, D. L., and Schmidt, G. W. (1988). *Trans*-splicing of transcripts for the chloroplast *psaA1* gene. In vivo requirement for nuclear gene products. J. Biol. Chem. *262*, 14601–14604.

Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A. (1989). RNA editing in plant mitochondria. Science *246*, 1632–1634.

Hildebrand, M., Hallick, R. B., Passavant, C. W., and Bourque, D. P. (1988). *Trans*-splicing in chloroplasts: the *rps12* loci of Nicotiana tabacum. Proc. Natl. Acad. Sci. USA *85*, 372–376.

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

Jarrell, 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, chloroplasts. 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 reinhardii: evidence for in vivo *trans*-splicing. EMBO J. *6*, 2185–2195.

Lambowitz, A. M. (1989). Infectious introns. Cell *56*, 323–326.

Lang, B. F., Ahne, F., and Bonen, L. (1985). The mitochondrial genome of the fission yeast Schizosaccharomyces pombe: the cytochrome b gene has an intron closely related to the first two introns in the Saccharomyces cerevisiae *cox1* gene. J. Mol. Biol. *184*, 353–366.

Lejeune, B., and Quétier, F. (1988). Structure de l'ADN mitochondrial du blé. In Variabilité Génétique Cytoplasmique et Stérilité Mâle Cytoplasmique (Paris: INRA), pp. 201–208.

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

Lonsdale, D. M. (1989). The plant mitochondrial genome. In The Biochemistry of Plants: A Comprehensive Treatise, Vol. 15, A. Marcus, ed. (New York: Academic Press), pp. 229–295.

Makaroff, C. A., and Palmer, J. D. (1987). Extensive mitochondrial specific transcription of Brassica campestris mitochondrial genome. Nucl. Acids Res. *15*, 5141–5156.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Matsubayashi, T., Wakasugi, T., Shinozaki, K., Yamagushi-Shinozaki, K., Zaita, N., Hidaka, T., Meng, B. Y., Ohto, C., Tanaka, M., Kato, A., Maruyana, T., and Sugiura, M. (1987). Six chloroplast genes (*ndhA-F*) homologous to human mitochondrial genes encoding components of the respiratory chain NADH dehydrogenase are actively expressed: determination of the splice sites in *ndhA* and *ndhB* pre-mRNAs. Mol. Gen. Genet. *210*, 385–393.

Michel, F., and Lang, B. F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature *316*, 641–643.

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

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

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

Perlman, P. S., and Butow, R. A. (1989). Mobile introns and intron-encoded proteins. Science *246*, 1106–1109.

Roitgrund, C., and Mets, L. J. (1990). Localization of two novel chloroplast genome functions: *trans*-splicing of RNA and protochlorophyllide reduction. Curr. Genet. *17*, 147–153.

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.

Small, I., Suffolk, R., and Leaver, C. J. (1989). Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell *58*, 69–76.

Stern, D. B., Bang, A. G., and Thompson, W. F. (1986). The watermelon mitochondrial URF-1 gene: evidence for a complex structure. Curr. Genet. *10*, 857–869.

van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., and Grivell, L. A. (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. Cell *44*, 225–234.

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.

Wilson, A. J., and Chourey, P. S. (1984). A rapid inexpensive method for the isolation of restrictable mitochondrial DNA from various plant sources. Plant Cell Rep. *3*, 237–239.

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

Wolfe, K. H., Gouy, M., Yang, Y.-W., Sharp, P. M., and Li, W.-H. (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA *86*, 6201–6205.

Zaita, N., Torazawa, 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.

EMBL Accession Numbers

The accession numbers for the sequences reported in Figure 2 of this paper are X57965–X57968.