INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Université d'Ottawa • University of Ottawa

RNA SPLICING AND EDITING OF GROUP II INTRONS IN FLOWERING PLANT MITOCHONDRIA

Catherine Carrillo

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

ABSTRACT

Within plant mitochondria, multiple processing events are required following transcription to produce functional mature messenger RNAs. These events include the conversion of certain cytosine residues to uridine by the process of RNA editing, and the removal of introns from some mitochondrial transcripts. Virtually all of the plant mitochondrial introns are classified as "group II" based on their structure which can be folded into six helical domains (D1-D6). However, most plant mitochondrial introns have aberrant features relative to conventional group II introns. The focus of this study is on examining the features and splicing of unusual group II introns observed in plant mitochondria.

Plant mitochondrial group II introns contain a number of mispairs within the helical regions of D5 and D6 which are normally among the most highly conserved structures of classical group II introns. In introns within the wheat gene encoding subunit 7 of the NADH dehydrogenase (*nad7*), six AC mispairs in these domains were predicted to be corrected to Watson-Crick AU pairs by RNA editing to restore helical structures. The editing status of these regions was determined by direct sequencing of RT-PCR products from partially-spliced and excised intron molecules. Editing was observed at only two of the six predicted sites and it was not conserved among angiosperms; for example, for *nad7* intron 4 the AC mispair which was edited in wheat remained unedited at the homologous site in tobacco, and was only partially edited in the population of excised introns in *Arabidopsis*. Editing of AC mispairs within intron sequences was unexpectedly low and did not restore the normally highly conserved D5 helical structures.

The 3' regions of several group II introns within the mitochondrial *nad1*, *nad4* and *nad7* genes show unexpected sequence divergence among flowering plants. For example, the *nad7* intron 3 linker joining domains 5 and 6 was longer than in conventional group II introns and it varied among plants, ranging from 4 to 11 nucleotides. Other examples of this occur within non-conventional D6 structures which lack the branch site adenosine normally involved in lariat formation. The D6 structures of *nad1* intron 1 and *nad4* intron 2 are highly variable in sequence among plants and, for *nad1* intron 1, editing actually reduces similarity among plants. These observations suggest that the core structures and sequences of certain mitochondrial introns in flowering plants are under reduced or different evolutionary constraint compared to conventional group II introns.

The temporal relationship between editing and splicing was examined for exon editing sites at splice junctions since some of these are involved in long range interactions with the intron which are required for splicing. Editing at such non-silent exon sites in *nad7* exons c and e, and *nad4* exon c, within three nucleotides downstream or one nucleotide upstream from the splice junction, was not observed in intron-containing transcripts, in contrast to full editing in mature mRNAs. These results were unexpected since editing is usually an early event in transcript processing. These results have implications for models regarding editing site recognition in plant mitochondria.

If splicing of plant mitochondrial introns proceeds by typical group II intron mechanisms, excised introns would be in a lariat form whereas if splicing occurred by first step hydrolysis the excised introns would be in a linear form. I have examined the physical forms of excised plant mitochondrial group II introns using a combination of northern, RT-PCR and S1 nuclease protection analyses. Steady state levels of *nad7* intron-containing precursors and excised intron RNAs were higher in 24 hr germinating embryos than in 6 day etiolated seedlings, suggesting different RNA processing pathways or RNA stabilities.Moreover, excised intron molecules have differing electrophoretic mobilities, indicating the presence of several forms of this intron RNA. Nevertheless, RT-PCR, northern and S1 nuclease protection analyses demonstrate the presence of an abundant lariat form of excised *nad7* introns 3 and 4 in 24 hr wheat mt RNA whereas linear forms of excised intron were not detected by these methods. For wheat *nad7* intron 4, an RT-PCR product from a putative "circular" form of excised intron was also observed; however, this form of the intron was not detected by northern analyses. Interestingly, an abundant species representing a putative splicing intermediate was observed in 6 day mtRNA as well as in 24 hr embryos. No appreciable difference in intron editing levels were seen between the two developmental stages. The techniques applied to the characterization of these introns will be useful for examining more unusual plant mitochondrial introns, such as those lacking branch site adenosines.

The reduced evolutionary constraint of plant mitochondrial introns may indicate that the role of intronic domains has been taken over by trans-acting proteins and/or RNAs. The study of such factors and their regulation should provide insight into the complexity of post-transcriptional events occurring in plant mitochondria.

RÉSUMÉ

Dans les mitochondries des plantes, plusieurs évènements sont requis pour la production des ARN messagers (ARNm). Ces évènements comprennent la conversion de cytosine a uridine par le processus d'édition d'ARN et, dans quelques cas, l'enlèvement des séquences interposées (introns) des transcrits. Presque tout les introns mitochondriaux des plantes sont désignés 'groupe II', à cause de leur structure qui peut être pliée en six domaines hélicoïdaux (D1-D6). La plupart des introns mitochondriaux ont des caractéristiques anormales lorsque comparées aux introns classique du groupe II. Le but de ce projet est d'examiner les traits et l'épissage des introns observés dans les mitochondries des plantes.

Les introns dans les mitochondries de plantes contiennent plusieurs nucléotides mal-oppariés à l'intérieur des régions hélicoïdales de D5 et D6, qui sont normalement les structures les plus conservées des introns du groupe II. Dans les introns du gène *nad7* qui code pour la sous-unité 7 du complexe de la déshydrogénase NADH du blé, six AC mal-oppariés dans ces régions devraient être corrigés à AU par l'édition d'ARN afin de rétablir les structures hélicoïdales. Le niveau d'édition était déterminé par le séquençage directe des produits TR-RPC dérivé de transcrits partiellement épissés et des introns excisés. L'édition a été observé à seulement deux des six sites prédits et elle n'était pas conservée chez les gènes de différentes espèces d'angiospermes; par exemple, pour l'intron 4 de *nad7*, un site d'édition dans le blé n'était pas édité dans le tabac et était seulement partiellement édité dans la population d'introns excisés dans *Arabidopsis*. L'édition des AC mal-oppariés dans les séquences d'introns était rare et n'a pas reconstitué la structure secondaire de D5 qui est généralement bien préservé.

Les régions 3' de plusieurs introns du groupe II dans les gènes mitochondriaux *nad1*, *nad4* et *nad7* ne sont pas bien conservées parmi les plantes à fleur. Par exemple, dans l'intron 3 de *nad7* la région qui joint D5 à D6 était plus longue que dans les introns typiques du groupe II et varie en longueur de 4 à 11 nucléotides chez différentes espèces d'angiospermes. D'autres exemples de ce phénomène ont été trouvé dans les structures D6 anormales à qui il manquait une base adénosine au site de branchement qui permet normalement la formation du lariat. La structure D6 de l'intron 1 de *nad1* et l'intron 2 de *nad4* était très variable en séquence parmi les plantes et, pour l'intron 1 de *nad1*, l'édition d'ARN a réduit la similarité entre les plantes. Ces observations suggèrent que la structure de base et les séquences de certains introns mitochondriaux soient sujettes à contraintes réduites ou différentes comparés aux introns typiques du groupe II.

La relation temporale entre l'édition et l'épissage a été examiné pour l'édition des exons aux jonctions d'épissage parce que certaines de ces régions sont impliquée dans des intéractions à distance avec les introns qui sont requis pour l'épissage. L'édition d'ARN a certaines positions non-silencieuces dans les exons c et e de *nad7* et l'exon c de *nad4* n'était pas observé dans les transcrits contenant des introns comparés aux ARNm qui étaient complètement édités. Ces résultats étaient imprévus parce que l'édition est normalement un événement qui ce passe en même temps que l'épissage des ARNm. Ces observations ont des implications pour des modèles de l'édition d'ARN dans les mitochondries des plantes.

Si l'épissage des introns mitochondriaux des plantes procède par le mécanisme typique des introns du groupe II, les introns excisés auriont une forme lariat tandis que si l'épissage se produit avec une hydrolyse en première étape, les introns excisés auriont une forme linéaire. J'ai examiné les formes physiques des introns excisés des mitochondries de blé en utilisant une combinaison d'hybridation d'ARN, de TR-RPC et de protection à la nucléase S1. Les niveaux des introns excisés et de transcrits contenant des introns était plus haut dans les embryons de 24 heures comparés aux jeunes plantes de six jours. Ceci suggère qu'il y a une variation dans la modification et la stabilité des ARNm entre les deux niveaux de développement. Ces différences ne correspondaient pas aux différences d'édition d'ARN dans les introns. De plus, les introns excisés ont des mobilités electrophorétique différentes, ce qui indique la présence de plusieurs formes de ces ARN introniques. Cependant, les expériences d'hybridations d'ARN et de protection à la nucléase S1 démontrent la présence d'une forme lariat d'intron abondant pour les introns excisés 3 et 4 de *nad7* dans les embryons de 24 hr de blé tandis que la forme linéaire était sous la limite de détection de ces méthodes. Une forme 'circulaire' de l'intron excisé a aussi été observée par TR-RPC pour l'intron 4 de *nad7*, mais cette forme de l'intron n'a pas été détecté par des expériences d'hybridation d'ARN. La détection de transcrits abondants représentants des intermédiaires d'épissage était aussi observé dans le blé pour les deux stages de développement. Les techniques appliquées à la caractérisation de ces introns seront utiles pour examiner des introns mitochondriaux des plantes qui sont irrégulier, comme ceux qui n'ont pas d'adénosine aux site de branchement.

La contraintes évolutionnaires réduites observées chez les introns mitochondriaux des plantes à fleur peut indiquer que le rôle des domaines introniques est assumé par des protéines et/ou des ARN agissant en trans. L'étude de ces facteurs et leur régulation pourrait clarifier la complexité des événements qui ont lieu après la transcription dans les mitochondries des plantes.

ACKNOWLEDGEMENTS

First and foremost, I gratefully acknowledge my supervisor, Dr. Linda Bonen, for her assistance with technical and theoretical aspects of this project and for providing me with the opportunity to do this research. I greatly appreciate all of the time she devoted to this work. I would also like to thank the members of my advisory committee, Dr. G. Drouin, Dr. C. Wyndham, Dr. V. Mezl, for their time and helpful advice.

I would like to acknowledge the contribution of fourth year and summer students Patricia Clark, Jason Szeto, Sean McDonald and Nima Niknejad. I would also like to acknowledge Karen Williams, for preliminary work done on this project. Finally, I have greatly appreciated the help, support and friendship of my fellow graduate students, particularly, Magid Fallahi, Shehrebanoo Malik, Lia Luus and Selvi Subramanian..

Thanks also to all of my friends and family for their support. In particular, I would like to thank my husband, Randy, for his patience and support, my sister, Lisa, for always being there for me, and my mother, for her strength and encouragement.

This work is dedicated to the memory of my father.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES:

ABBREVIATIONS

A	adenosine
AMV	avian myeloblastosis virus
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
cp	chloroplast
crs	chloroplast RNA splicing genes
C	cytidine
D5	domain 5
D6	domain 6
ddH_2O	double distilled water
dH_2O	single distilled water
ddNTP	dideoxyribonucleotide triphosphate
dNTPs	all four (G, A, T, C) deoxyribonucleotide triphosphates
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
G	guanosine
kb	kilobases
MFE	minimum free energy
mRNA	messenger RNA
mt	mitochondrial
µl	microlitre
µg	microgram
ml	millilitre
mg	milligram
$MgCl_2$	magnesium chloride
mrs	mitochondrial RNA splicing genes
NAD	nicotinamide adenine dinucleotide
nad	genes for NADH dehydrogenase subunits
nt	nucleotide
nts	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
rRNA	ribosomal RNA
R	purine
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
T	thymidine
Tris	tris (hydroxymethyl) aminomethane
tRNA	transfer RNA
U	uridine
Y	pyrimidine

CHAPTER 1:
LITERATURE REVIEW

The central tenet of molecular biology is that DNA is transcribed to RNA which is translated into protein. This simplified view became more complicated in 1977 with the discovery of extraneous sequences which interrupted some coding sequences in eukaryotic nuclear genes. These regions, called introns, are commonly found within nuclear genes of higher eukaryotes but are also found, to a lesser degree, in lower eukaryotes, archaebacteria, prokaryotes, bacteriophage and in the organelles of plants, protists and fungi. Such introns are removed at the RNA level by the process of splicing. While this process seems somewhat wasteful, in that it requires the expenditure of additional energy by the cell, in some cases the presence of introns can prove to be useful to the organism. For example, alternative splicing can be used to generate different proteins from the same DNA sequence, and there is some debate over whether introns have been involved in enabling processes such as exon shuffling throughout evolution.

In the following sections I will be examining the processes which are involved in generating mature messenger RNAs (mRNA) in plant mitochondria with emphasis on splicing. The mitochondrion is an essential component of eukaryotic cells due to its role in energy metabolism. While some of the proteins required for mitochondrial function are encoded by the nucleus, other essential proteins (and RNAs) are encoded and expressed within the mitochondrion. Expression of mitochondrial genes in plants is of particular interest due to the complexity of events which are involved and also due to the requirement for coordination with nuclear gene expression for regulation of mitochondrial activity. Mitochondrial gene expression may also be affected by chloroplast activity and thus is potentially linked to photosynthetic gene expression.

1.1 Classes of introns

Introns have been found in all categories of genes and can be divided into four major classes, nuclear pre-mRNA, tRNA, group I and group II introns. Additional classes have also been reported in the literature (i.e., group III introns in *Euglena gracillis*, tRNA-like introns in archaebacteria) and are reviewed elsewhere (Michel and Ferat, 1995; Abelson *et al.*, 1998). For each class of intron there is a unique biochemical mechanism for splicing.

However, there are some common themes in splicing. For example, virtually all of the introns that have been well characterized to date rely on proteins and/or RNAs for splicing *in vivo*. Interestingly, it appears that the actual splicing reaction for most introns may well be catalyzed by an RNA component rather than a protein enzyme (see below). In general, this is where the similarities in splicing mechanisms end among the introns. In the following sections I will examine splicing among the different classes of introns. Group II introns will be discussed more extensively, since virtually all plant mitochondrial introns identified to date are members of this class.

1.1.1 Catalytic activity of intronic RNAs

Some classes of introns (group I, group II and tRNA introns) can be folded into well defined structures and certain members of the group I and II intron classes have been shown to self-splice *in vitro*, albeit under non-physiological conditions. The catalytic functions required for splicing reside within the RNA molecule itself. This activity is reminiscent of that of protein enzymes, and for this reason, RNAs which can catalyze biochemical reactions have been called "ribozymes" (reviewed by Scott, 1999). Group I introns within *Tetrahymena thermophila* nuclear rDNA genes were among the first catalytic RNAs identified (Kruger *et al.*, 1982). Their discovery, and the identification of other naturally occurring ribozymes such as group II introns (see below) was particularly exciting as it demonstrated that protein enzymes were not solely responsible for the catalysis of reactions within a cell.

1.1.2 Nuclear pre-mRNA spliceosomal introns

Spliceosomal introns within nuclear genes are widespread in nature, occurring in organisms as diverse as yeast, amoeba, plants and vertebrates (reviewed in Logsdon, 1998). Generally introns are found at a higher density in later diverging organisms: densities range from 0.03 introns per kb of coding sequence in *Entamoeba* to 5.6 in vertebrates (Logsdon, 1998). In plants, intron density is approximately 4.3 introns per kb of coding sequence. Spliceosomal introns are not highly conserved in sequence; however, they have four characteristic short elements: the 5' splice site (GT), the 3' splice site (AG), a short sequence upstream of the 3' splice containing an adenosine which is directly involved in the splicing reaction (described below) and a polypyrimidine tract. Plant spliceosomal introns are shorter

than those in vertebrates and most are within 80-139 nt, although introns 2 kb up to 7 kb have also been observed in a small number of cases (Simpson and Filipowicz, 1996). An important distinguishing feature of plant spliceosomal introns is that they are particularly AU rich. Interestingly, *in vitro* studies have shown that this varies among plants, where dicots require about 60% AU for splicing and monocots require only 30% AU (Goodall and Filipowicz, 1991).

Nuclear pre-mRNA splicing has been extensively studied in yeast due to the availability of an *in vitro* assay. While such an assay has not yet been developed for plant nuclear splicing, evidence suggests that similar mechanisms are used (reviewed in Brown and Simpson, 1998). Splicing of nuclear pre-mRNA introns takes place in the nucleus of the cell where a number of proteins and RNAs assemble on the transcript into a complex called a spliceosome. The spliceosome is comprised of ribonucleoprotein particles (snRNPs) including U1, U2, U5 and U4/U6 which are named based on the small nuclear RNAs (snRNAs) present in the complex. A number of other proteins factors are also involved in splicing, structure, assembly and regulation of the spliceosome. These include RNA binding DEAD-box proteins such as RNA helicases and RNA dependent ATP-ases.

Intron excision is an early processing step and splicing of the nascent RNA occurs during transcription. Splicing occurs by transesterification in two cleavage-ligation steps (Figure 1.1A). In the first step, the phosphodiester bond joining the first nucleotide of the intron with the upstream exon is cleaved with the concomitant ligation of the 5' phosphate of the intron to the 2' hydroxyl of a branchpoint adenosine 18-40 nts upstream of the 3' splice site. In the second splicing step, the free 3' hydroxyl group of the upstream exon performs a nucleophilic attack on the phosphodiester bond between the intron and the downstream exon. The end result is that the intron is released in a lariat form, and the two exons are ligated together. The 2'-5' bond of the lariat intron is subsequently disrupted with a debranching enzyme, so that the linearized intron can be rapidly degraded. Both steps in the splicing reaction are believed to be catalyzed by the RNA, rather than the protein components of the spliceosomal complex (Gaur *et al.*, 1997).

1.1.3 Nuclear tRNA introns

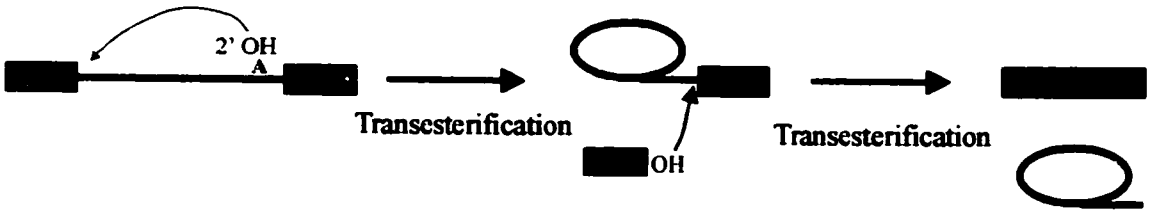
Eukaryotic nuclear genes encoding tRNAs contain another class of introns which are

Figure 1.1. Splicing of nuclear pre-mRNA introns and tRNA introns.

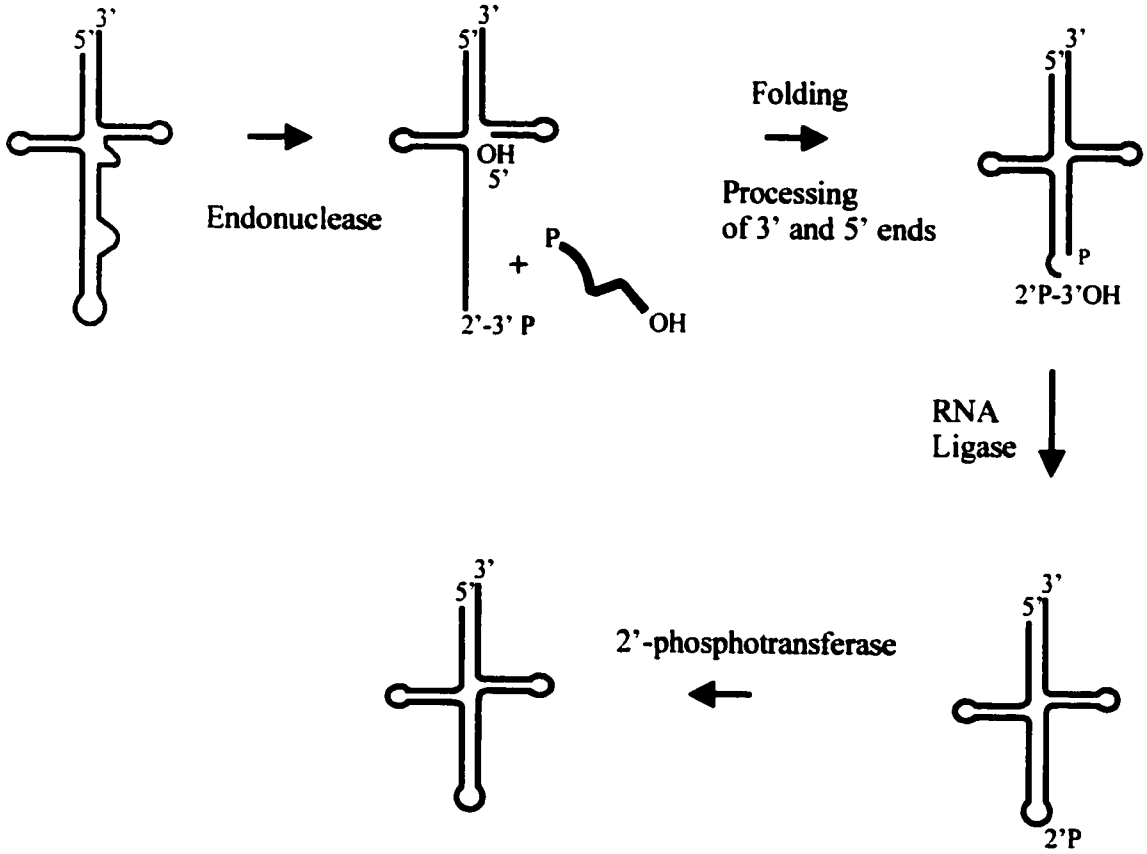
[A] Splicing of nuclear pre-mRNA introns. Green arrows indicate nucleophilic attack in the course of the two transesterification steps of spliceosomal splicing. Introns are shown as red lines and exons are shown as grey and blue boxes.

[B] Mechanism of tRNA splicing. Black lines represent tRNA and red lines represent introns. Splicing requires three enzymes: an endonuclease, an RNA ligase and a 2' phosphotransferase. Adapted from Lewin, 1994.

A. nuclear pre-mRNA Splicing



B. tRNA Splicing



unrelated to spliceosomal introns (reviewed in Abelson *et al*, 1998). These introns were first identified in yeast where they occur in approximately 20% of tRNA genes (Abelson *et al*, 1998). Related introns have also been identified in a number of other organisms including plants, amphibians and mammals. It should be noted that some eubacterial and organellar tRNA genes also contain introns, but these are unrelated introns belonging to the group I or group II class (cf. Sections 1.1.4, 1.1.5). Archaebacterial tRNA introns were originally thought to be unrelated as well, however recent evidence suggests this is not the case (see below). Introns within eukaryotic nuclear tRNAs are relatively small (14-60 nts in yeast, 12-25 nts in plants) and are invariably located 1 nt downstream of the tRNA anticodon. The intron sequence is not highly conserved; however, introns include a sequence complementary to the anticodon, which is necessary to generate a splicing-competent structure within the tRNA (Figure 1.1B).

Splicing of tRNA introns involves separate cleavage and ligation steps rather than two transesterification steps as in the other major classes of eukaryotic introns (Figure 1.1B). This system has been well characterized in yeast, due to the availability of an *in vitro* splicing system (Abelson *et al.*, 1998). The first step of splicing involves an endonuclease which cleaves the two phosphodiester bonds flanking the intron, resulting in the excision of a linear intron and the generation of two tRNA pieces. The 5' section of the tRNA has a terminal 2',3'-cyclic phosphate group and the 3' section has a hydroxyl at the 5' terminus. These products are rather unique since typically RNA processing enzymes cleave on the other side of the phosphate bond. Before ligation of the two tRNA halves can take place, the cyclic phosphate group is opened with a cyclic phosphodiesterase, generating a 2' phosphate, 3' hydroxyl end on the first tRNA half and the 5' hydroxyl of the second tRNA piece is phosphorylated with a kinase. The two ends to be ligated are in close proximity due to the higher order structure of the tRNA, and the 3' hydroxyl is ligated by tRNA ligase to the 5' phosphate forming a phosphodiester bond. In the final step of splicing, a NAD-dependent phosphotransferase removes the 2'-phosphate (from the 3' end of the first tRNA piece) remaining at the splice junction.

Thus, the catalysis of tRNA splicing involves three enzymes: an endonuclease, a tRNA ligase and a phosphotransferase. Genes for all of these enzymes have been isolated

for yeast (*Saccharomyces cerevisiae*), and their activity has been demonstrated in plants and vertebrates (reviewed in Spinelli *et al.*, 1998). Interestingly, *in vitro* experiments with an *Arabidopsis* tRNA intron indicate that the RNA component is capable of self-cleavage under certain conditions (Weber *et al.*, 1996). The endonuclease may be involved in improving the efficiency of this reaction *in vivo*. Homologues of the endonuclease have been identified in archaebacteria where they perform a similar function in splicing (Abelson *et al.*, 1998). Recently, a homologue of the 2'-phosphotransferase has been identified in *Escherichia coli*, even though tRNA introns requiring this enzyme have not been identified in eubacteria (Spinelli *et al.*, 1998). Note that mutation of this gene does not affect the growth of *E. coli* and this gene is not observed in other eubacteria examined suggesting that it does not encode an essential protein (Spinelli *et al.*, 1998). The presence of the 2'-phosphotransferase over large evolutionary times may indicate an ancient origin of tRNA introns or that this gene has been secondarily acquired to perform an alternate cellular function.

1.1.4 Group I introns

In 1982 the introns found within the yeast mitochondrial genome were classified into two groups, group I and group II, based on sequence and secondary structural characteristics (Michel *et al.*, 1982). Subsequently, introns from other organisms belonging to each group were identified. These two groups appear to be unrelated and no introns falling into an intermediate category have been identified; however, some general similarities between the two classes exist. For example, members of both classes are capable of self splicing *in vitro*, splicing occurs via two transesterification steps and, in some cases, introns encode ORFs.

Group I introns have been found in nuclear ribosomal RNA genes of protists, in protein coding and tRNA genes within eubacteria and in organelles of fungi and plants and in bacteriophage (reviewed in Lambowitz and Belfort, 1993; Saldanha *et al.*, 1993). These introns range in size from about 200 to 3000 nt depending on the lengths of the regions outside core structures and whether or not they contain open reading frames (ORFs) (Lambowitz and Belfort,1993).

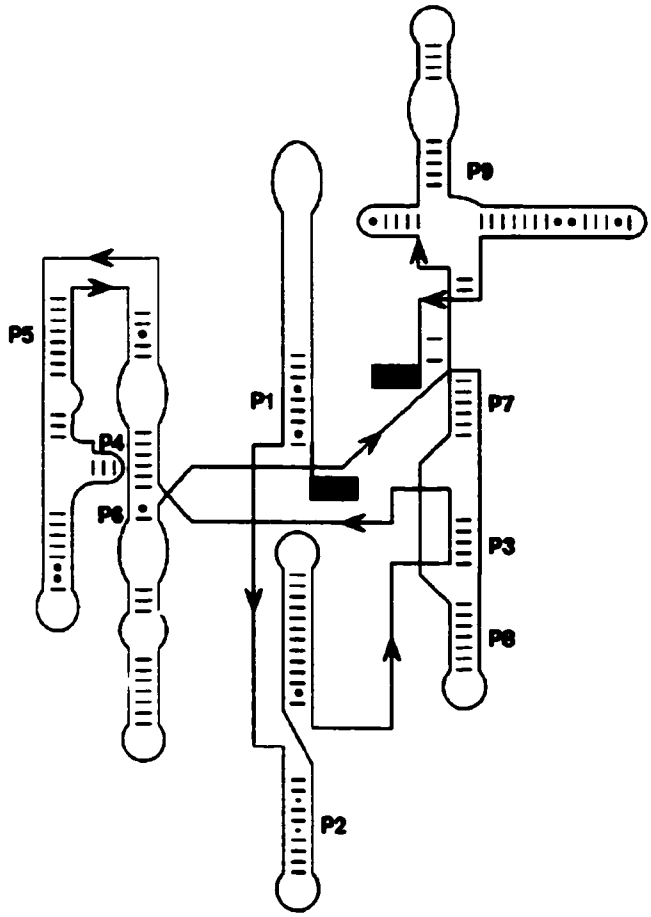
The first self-splicing intron characterized was a group I intron in the pre-rRNA of *Tetrahymena* (Kruger *et al.*, 1982). Several other group I introns were subsequently shown to catalyze their own excision *in vitro* in the absence of any host proteins or RNAs. Group

Figure 1.2. Group I introns.

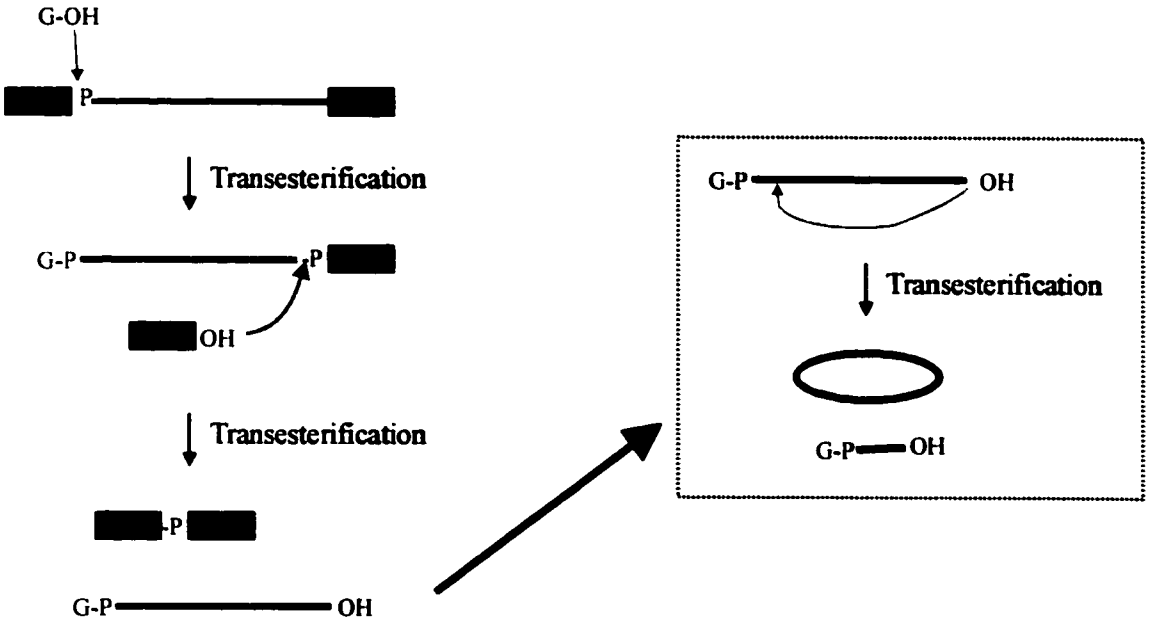
[A] Secondary structural model for group I introns adapted from Carola and Eckstein (1999). Helical domains are labeled P1-P9. Exons are shown as blue and grey boxes.

[B] Group I splicing through two transesterification steps releasing a linear intron. Green arrows depict the nucleophilic attack in each splicing step, introns are shown as red lines and exons are shown as grey and blue boxes. A third transesterification step occurring after intron excision is shown in the box (see text).

A. Group I intron structure



B. Group I Intron Splicing



I introns can be folded into a characteristic secondary structure consisting of nine base pairing regions (P1-P9, Figure 1.2A). The core intron elements required for self-splicing are P4-P6 and P3-P9, as defined by mutational studies on self-splicing introns *in vitro* (reviewed in Saldanha *et al.*, 1993) and X-ray crystallography studies of this intron (reviewed in Carola and Eckstein, 1999). A tertiary interaction between the 3' end of the intron and a region between P7 and P9 is also necessary to generate a splicing competent structure.

Splicing of group I introns involves two transesterification steps which are different from those of the nuclear pre-mRNA and group II introns (Figure 1.2B; Saldanha *et al.*, 1993). A guanine nucleotide (in the form of GMP, GDP or GTP) is necessary to initiate the reaction. The free 3' hydroxyl of the guanosine attacks the 5' splice junction (presented within the P1 structure), resulting in the release of the upstream exon. In the second step, the 3' splice site is cleaved and the two exons are ligated together while the intron is released in a linear form. No free exons have been observed, indicating the close coupling of the two steps. In some cases, the linear excised intron is circularized in a secondary reaction. The 3' terminal G hydroxyl attacks at a position near the 5' end of the intron, forming a circular intron and releasing a short linear fragment representing the original 5' end of the intron.

While group I introns are capable of self-splicing *in vitro*, cellular factors are known to be involved *in vivo* (reviewed in Saldanha *et al.*, 1993; Grivell, 1995). In yeast mitochondria, several cytochrome b group I introns encode proteins which are capable of assisting in the splicing of the intron in which they are found. These proteins are called maturases or homing endonucleases due to their additional function in intron mobility. It has been suggested that the splicing function of intron-encoded ORFs evolved secondarily due to their ability to bind the intron in mobility reactions (Saldanha *et al.*, 1993). Other proteins involved in splicing are nuclear-encoded factors which may be derived from bifunctional proteins which are capable of binding RNA. For example, *CYT18*, which aids in the splicing of a group I intron in *Neurospora*, is in fact a mitochondrial tyrosyl tRNA synthetase with an extra domain to perform its role in splicing (Kämper *et al.*, 1992). This RNA binding protein can assist in splicing by stabilizing group I intron structures (Caprara *et al.*, 1996). Other nuclear-encoded factors with less direct effects on group I splicing have also been identified and are reviewed in Grivell (1995).

Group I intron spread has been largely through horizontal transfer. Many of these introns are mobile genetic elements: they can move into intron-less alleles by recognition of a specific target sequence. To this end, the endonuclease-related ORFs encoded within these introns, called homing endonucleases, assist in this mobility by creating double-stranded breaks in the target sequence. Such intron mobility has been reported in plant mitochondria where group I introns have been recently identified. The discovery of a group I intron in the coxI gene of *Peperomia* prompted the mass screening of 335 diverse genera of land plants and group I introns were found in cox I genes in a wide range of angiosperms (Cho *et. al.* 1998). Interestingly, the distribution of the intron suggests at least 32 independent intron acquisitions, indicating the recent origin of this intron at these locations.

1.1.5 Group II introns

Membership of the group II intron class includes over 100 introns from fungal and plant organelles as well as a few in cyanobacteria, proteobacteria and gram-positive bacteria (Holländer and Kück, 1999A; Michel and Ferat, 1995; Michel *et al.*, 1989). Introns with group II characteristics have also been identified in *Euglena gracilis* chloroplasts, however many of these introns lack certain entire domains of the classical group II introns (Michel and Ferat, 1995). Fungal group II introns range in size from about 900 to 2500 nts, depending on whether or not they encode ORFs (Copertino and Hallick, 1993), and about 600 nts are required for the core, catalytic activity of the group II ribozyme (Michel and Ferat, 1995). Several fungal and algal mitochondrial introns and some bacterial group II introns are capable of self-splicing *in vitro* under non-physiological conditions (reviewed in Perlman and Podar, 1996). This self splicing ability was first demonstrated for the yeast mitochondrial introns aI5γ, within the cytochrome I gene, and bI1, within the cytochrome oxidase b gene (van der Veen *et al.*, 1986; Schmelzer and Schweyen, 1986; Peebles *et al.*, 1987). An interesting feature of group II splicing is that the intron need not be in one piece for splicing to occur: *trans*-splicing has been observed with introns in pieces *in vivo* and *in vitro* (Michel and Ferat, 1995; Perlman and Podar, 1996)

The secondary structure of group II introns was first determined based on only two intron sequences from yeast mitochondria (Michel *et al.*, 1982). Surprisingly, this model was supported by comparative analysis when more sequences became available and also by

biochemical studies, albeit with some minor modifications (Michel *et al.*, 1989; Michel and Ferat, 1995). Group II intron structure consists of six helical domains (D1-D6) which are characteristically depicted as radiating from a central wheel (Figure 1.3A) (reviewed by Michel and Ferat, 1995; Michel *et al.*, 1989). Of these, D5 is the most highly conserved in both sequence and secondary structure and accordingly, it can be used as a diagnostic feature of this class of introns (Knoop *et al.*, 1994B). Group II introns can be further divided into two major subclasses (group IIA and IIB) based on differences in structural characteristics (Michel *et al.*, 1989). These include differences in the location of the bulging adenosine within D6, being located 7 nts (IIA) or 8 nts (IIB) upstream from the splice junction, and the terminal three nucleotides of the intron, which are YAY(IIA) or RAY (IIB).

1.1.5.1 Group II splicing: transesterification vs. hydrolysis reaction

In general, splicing of group II introns proceeds through two transesterification steps which are similar to splicing of nuclear pre-mRNA introns (cf. Figure 1.1A). In the first step, the 2' hydroxyl group of the bulging adenosine of D6 acts as a nucleophile, attacking the phosphodiester bond joining 5' end of the intron to the upstream exon (Figure 1.3B). This results in the formation of a 2'-5' bond between the first nucleotide of the intron and the bulging adenosine, and the release of the upstream exon. Note the intron-downstream exon splicing intermediate (Figure 1.3B). In the next step, the 3' hydroxyl of the released exon attacks the phosphodiester bond joining the last nucleotide of the intron with the downstream exon, resulting in the release of the lariat intron and the concomitant ligation of the two exons.

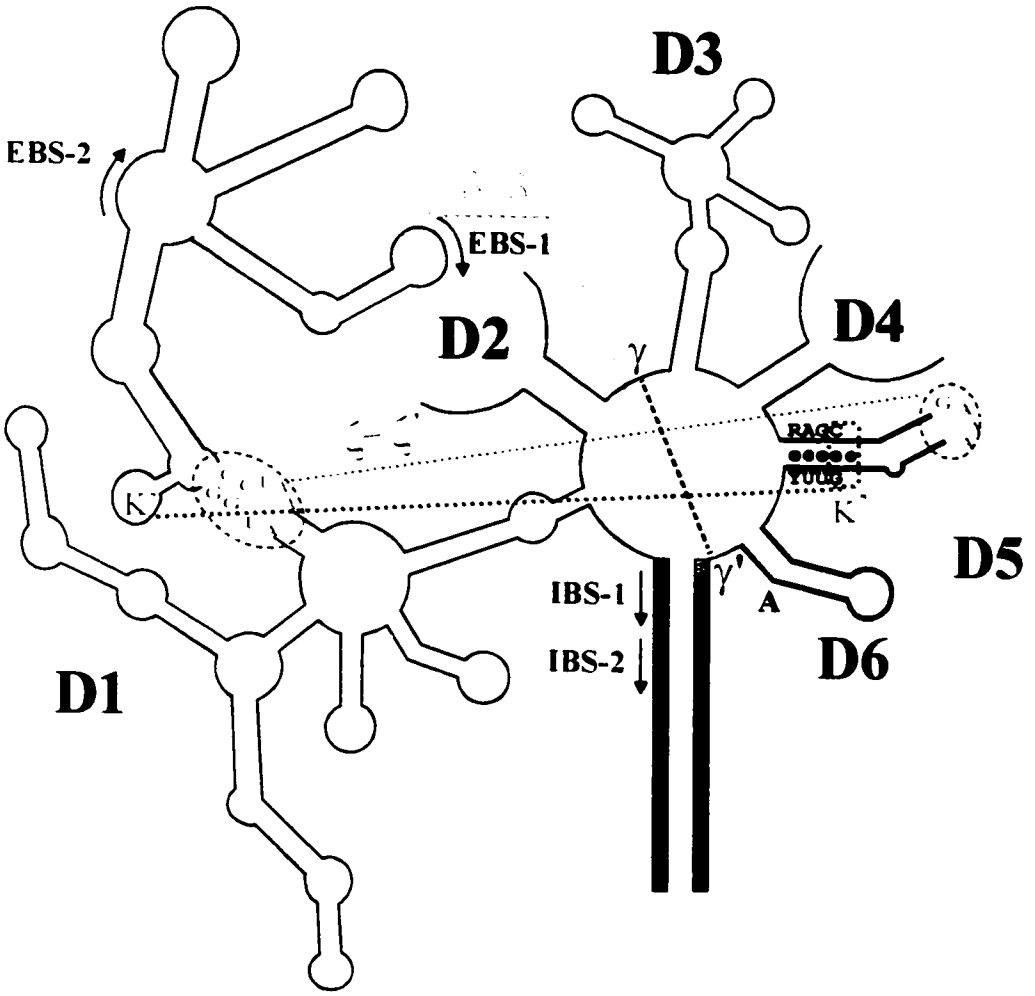
An alternative splicing mechanism has also been observed in the course of *in vitro* splicing experiments with yeast autocatalytic introns (reviewed in Perlman and Podar, 1996). In this mechanism, the first step of splicing occurs by the hydrolysis of the bond between the upstream exon and the first intron nucleotide. The second step of splicing occurs via the typical transesterification step and splicing is accurate, but the intron is excised in a linear form. This mechanism of splicing has been observed *in vitro* with both yeast branch site mutants (van der Veen *et al.*, 1987) and in *trans*-splicing reactions (Jacquier and Rosbash, 1986). In fact, both pathways occur in parallel under all reaction conditions *in vitro*, although one pathway can dominate under varying conditions (Daniels *et al.*, 1996). Recently, splicing

Figure 1.3. Group II introns.

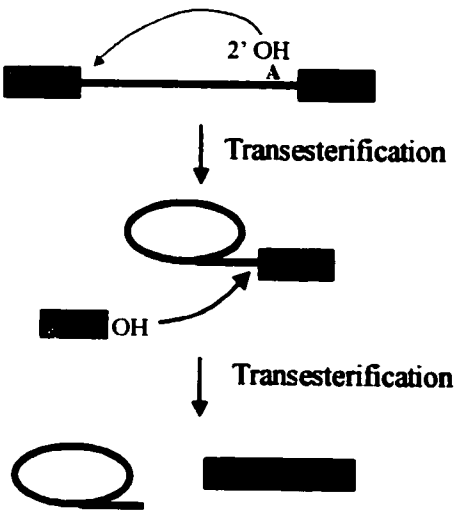
[A] Secondary structural model for group II introns. The six helical domains are labeled D1-D6 and some tertiary interactions are indicated by dotted lines. Domains 5 and 6 are highlighted in red. Intron binding sequences (IBS) and exon binding sequences (EBS) are indicated by arrows. Grey and blue boxes represent exon sequences. Adapted from Michel and Ferat (1995).

[B&C] Typical group II splicing via two transesterification steps resulting in excised intron lariats [B] or alternate splicing by first step hydrolysis and second step transesterification resulting in the excision of the intron in a linear form [C]. Introns are shown as red lines, green arrows depict nucleophilic attack and exons are shown as blue and grey boxes.

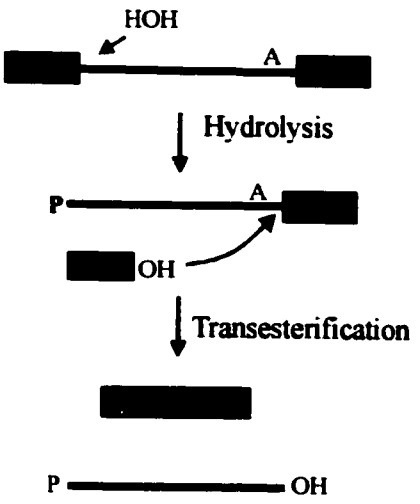
A. Group II intron structure



B. Group II intron splicing



C. Group II intron splicing via hydrolysis



via this alternative mechanism has been demonstrated *in vivo* in the mitochondria of yeast strains with branch site mutations (Podar *et al.*, 1998). However, the hydrolytic mechanism of splicing is inefficient: the catalytic rates measured under optimal conditions for both splicing steps were faster by three orders of magnitude in the branching pathway than in the hydrolytic pathway (Dème *et al.*, 1999). It is notable that splicing by both pathways results in accurate exon ligation. To date, splicing of naturally occurring introns via the hydrolytic pathway has not been demonstrated.

1.1.5.2 Structure of group II introns

Mutational analysis, mainly of the yeast self-splicing mitochondrial introns aI5γ and bI1, has aided in establishing the importance of each group II intron domain for *in vitro* splicing. D1 and D5 are the most highly conserved structures as they form the catalytic core of self-splicing introns, while D6 includes a bulging adenosine which is directly involved in the splicing reaction. Note that the location of the bulging adenosine, 7 to 8 nt upstream of the 3' end of the intron is one of the most highly conserved features of group II introns from a wide variety of organisms. Domains 2 and 4 are less conserved and do not seem to contain elements that are directly involved in the splicing reactions, although these helices are important for proper intron structure.

Several long range interactions have been identified in group II introns and those involving flanking exon sequences and D5 or D6 are shown in Figure 1.3A. Note that all of the interactions shown are supported by some *in vitro* evidence where mutation of nucleotides involved in the interactions disrupt splicing to some degree and compensatory changes restore splicing (Perlman and Podar, 1996). Interactions such as the γ-γ', the ζ-ζ' (Michel *et al.*, 1989) and the κ- κ'(Boudvillain and Pyle, 1998) assist with stabilizing tertiary intron structures by bringing distal regions of the intron closer together. Other long range interactions are between the intron and flanking exon sequences. The exon binding sequence 1 (EBS1) is found in D1 in all group II introns. This element is usually 6 nts in length and is complementary to the intron binding sequence 1 (IBS1) which comprises the last 6 nts of the 5' exon (Michel *et al.*, 1989). Similarly EBS2 binds to the complementary IBS2 sequence located 0-3 nt upstream of IBS1. This interaction appears to be less important as it varies in location within D1 and is occasionally missing in group II introns. The EBS-IBS interactions

are particularly important for group II intron splicing since loss of affinity for the 5' exon before the second step of splicing results in the accumulation of splicing intermediates (Michel *et al.*, 1989). The only other known tertiary interaction involving flanking exon sequence is the δ - δ', which involves the first nucleotide of the 3' exon.

Domain 5 has been the focus of many mutational studies with autocatalytic yeast introns due to its highly conserved nature and its importance in catalysis. D5 is essential for splicing *in vitro* and *in vivo*. It is highly conserved in both sequence and length, with a 34 nt helix containing a 2 nt bulge and closed by a purine tetraloop (Figure 1.3A). Note that there is now evidence of an alternate 9 bp-bulge-5 bp conformation (Costa *et al.*, 1998) compared with the 8 bp-bulge-6 bp version shown in Figure 1.3A. The most conserved bases in this structure, numbered from the 5' end of the helix, are A2, G3 and C4 which pair with U33, U32 and G31, respectively. The bases A2, G3 and C4 have been shown by mutational analysis to be crucial for splicing (Peebles *et al.*, 1995; Boulanger *et al.*, 1995) and the terminal tetraloop, which is usually GAAA, interacts with a motif within domain 1 (Costa and Michel, 1995). This interaction appears to increase the affinity of the group II intron for the upstream exon by 100 times (Costa and Michel, 1999). The nucleotide sequence of D5 is conservative to such an extent that the identification of group II introns in data bank searches has been possible (Knoop *et al.*, 1994B).

Domain 6, although less well-conserved and having a shorter helix, almost always possesses a bulging adenosine which is located 7 to 8 nt from the 3' splice site and which is involved in lariat formation. However, *in vitro* studies of the branch site suggest that it is not absolutely required as introns lacking a bulging adenosine can splice by first step hydrolysis under high salt conditions *in vitro* (van der Veen *et al.*, 1987; Chu *et al.*, 1998) and under physiological conditions *in vivo* (Podar *et al.*, 1998). Furthermore, when a nucleotide which pairs with the branch site nucleotide is introduced so that the adenosine is no longer bulged, branching is not prevented although lariat formation is greatly reduced (Chu *et al.*, 1998). Taken together, these data indicate that, while the branch site is not absolutely required, its presence increases the efficiency of splicing.

1.1.5.3 Group II intron encoded proteins

Genetic evidence for yeast mitochondrial group II introns indicates that even introns

capable of self-splicing *in vitro* require additional factors for efficient splicing *in vivo* (reviewed in Grivell, 1995). In some cases these factors are encoded by ORFs which are found within about a quarter of known group II introns (Michel and Ferat, 1995). The encoded proteins can facilitate RNA splicing (maturases) and/or function in intron mobility (Lambowitz and Belfort, 1993). Group II intron-encoded proteins consist of up to three domains with reverse transcriptase function (RT domain), maturase function (domain X) and/or a zinc-finger like domain which is related to endonucleases (Lambowitz and Belfort, 1993). Phylogenetic comparisons indicate that the reverse transcriptase (RT) domains of the group II-intron encoded ORFs most closely resemble those of the non-LTR class of retroid elements (Mohr *et al.*, 1993). This may indicate a retroviral origin of group II introns where the intron-encoded ORFs were originally involved in transposition. Most of these protein-coding sequences are found within D4, the least conserved intron domain and often the ORFs extend upstream through D1-D3 to the 5' exon so that they can be translated as chimeric proteins which include the upstream exon (Michel and Ferat, 1995). These chimeric proteins are subsequently processed so that the smaller maturase protein is released.

The maturase activity of group II-encoded ORFs was originally demonstrated in yeast by genetic analysis (Moran *et al.*, 1994; van der Veen *et al.*, 1988) and is thought to be conferred by a region of the protein called domain X. The activity of these maturases in facilitating splicing *in vitro* has been demonstrated with the *Lactococcus lactis* group II intron L1.ltrB (Saldanha *et al.*, 1999). In this case, the maturase has been shown to assist with splicing by binding to its own sequence within D4 (Wank *et al.*, 1999).

The yeast intron-encoded maturases also have functions in intron mobility due to their RT and endonuclease activities (Guo *et al.*, 1997). A ribonucleoprotein complex containing the excised aI2 intron RNA and aI2-encoded protein catalyzes a double-stranded site-specific break in the recipient DNA. This reaction involves a second ribozymic activity of the intron RNA as the RNA actually catalyzes the cleavage of the sense strand of the DNA. The aI1- and aI2- encoded RTs preferentially reverse transcribe the intron RNAs, using excised intron or unspliced precursor RNA as templates. The properties of intron-encoded proteins allow introns to move into an intronless allele and, at a much lower frequency, to a new site. Interestingly, the ability of the protein to bind to the intron RNA

to act as an endonuclease also allows it to stabilize the folding of an intron RNA for splicing.

In seed plants only two maturases have been identified: *mat-k* within the chloroplast *trnK* intron (Vogel *et al.*, 1997B) and *mat-r* in the fourth intron of *nad1* in mitochondria (Farré and Araya, 1999; Thomson *et al.*, 1994; Chapdelaine and Bonen, 1991; Wissinger *et al.*, 1991). These maturases were identified by sequence homology to domain X (*mat-r*, *mat-k*) and the RT domain (*mat-r*) of other group II intron encoded proteins. Interestingly, the *mat-r* ORF is transcribed from an intron internal promoter in wheat mitochondria rather than being translated as a chimeric protein as described above (Farré and Araya, 1999). Evidence for a role for the plant mitochondrial maturase is somewhat indirect. However, conservation of RT and X domains of the protein, which is improved by RNA editing within mRNAs (Bégu *et al.*, 1998; Thomson *et al.*, 1994), suggests a role for these proteins *in vivo*. Interestingly, a reverse transcriptase activity which could perhaps be attributed to this protein has been found in potato mitochondria (Moenne *et al.*, 1996). Evidence is accumulating for *mat-k* activity in chloroplast splicing. For example, its maturase activity is specific to the intron in which it is encoded: the *mat-k* gene product was able to bind its own precursor *in vitro*, but not to the intron-containing *rps16* transcript (Liere and Link, 1995). Furthermore, lack of translation in ribosome deficient plastids resulted in a reduction of *trnK* splicing indicating that a plastid-encoded protein is required for splicing of this intron (Vogel *et al.*, 1999). Similarities of maturases in plant organelles to other group II intron-encoded ORFs is likely to correspond to similar roles of these proteins in plants.

1.1.5.4 Nuclear-encoded accessory factors involved in group II splicing

Accessory factors involved in group II splicing are also encoded by nuclear genes. These factors may assist in splicing directly by binding to the intron and assisting in folding, or they may have indirect effects. For example, the *MRS3* and *MRS4* genes encode mitochondrial carrier proteins and they may be able to suppress group II splicing defects by altering the ionic balance within the mitochondrion (Grivell, 1995). Self-splicing introns require high salt concentrations for splicing *in vitro*. In contrast, the yeast *MRS2* and *MSS116* genes encode proteins which interact directly with group II introns in splicing, perhaps by improving the structure of these introns so that splicing is more efficient (Schmidt *et al.*, 1998; Niemer *et al.*, 1995). Overexpression of these proteins suppresses splicing defects in

yeast. Interestingly, the *MSS116* protein has a DEAD-box motif which is characteristic of RNA helicases which function in spliceosomal splicing. Similarly, nuclear encoded factors are involved in splicing of introns in the *psaA* gene of the alga, *Chlamydomonas* (Goldschmidt–Clermont *et al.*, 1990; 1991), where *trans*-splicing mutants belonging to at least 13 nuclear complementation groups were found. These factors could be divided into 3 classes based on their specificity. Class A aids in splicing of *psaA* intron 2, class C is involved with intron 1 splicing and class B aids in the splicing of both introns. One of the class A factors, *maa2*, has recently been characterized and is a protein related to pseudouridine synthase (Perron *et al.*, 1999).

While accessory factors involved in splicing of angiosperm mitochondrial group II introns have not yet been identified, recently, there has been some evidence of the involvement of nuclear-encoded accessory factors in splicing of group II introns within plastids (Hess *et al.*, 1994; Jenkins *et al.*, 1997; Vogel *et al.*, 1999). Two maize nuclear mutants, chloroplast RNA splicing 1 and 2 (*crs1* and *crs2*) have now been characterized. The *crs1* is defective in splicing only the *atpF* intron whereas the *crs2* mutants are defective in splicing of all but one of the group IIB introns (Jenkins *et al.*, 1997; Vogel *et al.*, 1999). Furthermore, the study of mutants of barley and maize which lacked plastid ribosomes demonstrated that a plastid encoded factor was required by all group IIA and to a lesser degree by some group IIB introns within chloroplasts (Hess *et al.*, 1994; Jenkins *et al.*, 1997; Vogel *et al.*, 1999). From these data it is evident that only the second intron of *ycf3* in maize is not affected in any of the mutants examined in these studies and the *atpF* intron required both a nuclear factor and a plastid-encoded factor. As more mutants are discovered it is likely that more than one factor will be found to be involved in splicing of each of these chloroplast introns.

1.1.5.5 Origin of group II introns

The similarities in the splicing mechanism and boundary sequences between group II introns and spliceosomal introns has led to speculation that group II introns are in fact the evolutionary predecessors of the nuclear pre-mRNA introns (Sharp, 1985; Cech, 1986; Jacquier, 1990; Weiner, 1993; Sharp, 1991). The proposal is that nuclear pre-mRNA introns have evolved from group II introns due to progressive loss of self-splicing capability and an

increasing reliance on host factors (Jacquier, 1990); for example, with the role of certain domains being taken over by *trans*-acting snRNAs in a spliceosome complex. Once these proteins and RNAs had been recruited to aid in the splicing reaction, degeneration of intron structures would take place as splicing became increasingly reliant on generalized, *trans*-acting splicing factors.

Since there is very little sequence similarity between group II introns and spliceosomal introns, an alternate theory has been put forth: common splicing mechanisms observed between group II introns and nuclear pre-mRNA introns are simply due to chemical determinism or convergent evolution (Weiner, 1993). Perhaps splicing via two transesterification steps and the formation of a lariat with the 2' hydroxyl of an adenosine residue is simply one of the most efficient manners of carrying out intron excision. Splicing of group I introns also occurs via two transesterification steps, albeit with different splicing intermediates.

Whether or not a relationship exists between group II introns and spliceosomal introns remains an open question. However, evidence for a relationship between the two is mounting. Hetzer *et al.* (1997) demonstrated that the role of the group II intron subdomain ID3 could be functionally substituted with the spliceosomal snRNA U5. Furthermore, when branch site adenosines were mutated in group II introns and nuclear pre-mRNA introns, effects on the splicing reaction were very similar (Gaur *et al.*, 1997). Mechanistic similarities between group II intron and nuclear pre-mRNA splicing have also been described by Sontheimer *et al.* (1999). The accumulation of evidence supporting the evolution of spliceosomal introns from group II introns is consistent with the idea that these eukaryotic nuclear introns were acquired relatively late in the course of evolution (reviewed in Logsdon, 1998).

The fact that nuclear pre-mRNA introns are not generally found in amitochondriate organisms led to the proposal by Cavalier-Smith (1991) that these introns arose from group II introns introduced into the eukaryotic nuclear genome in the course of the acquisition of organelles. This theory, called the "mitochondrial seed hypothesis" has been supported with the relatively recent identification of introns within eubacteria, particularly the α-proteobacteria, and the absence of introns in archaea and in amitochondriate protists such as

Giardia and *Trichomonas* (reviewed in Logsdon, 1998). Furthermore, group II introns are commonly found in organelles of plants, fungi and algae, although not in animal mitochondria where they may have been secondarily lost due to streamlining. A problem with this theory is that mitochondrial type genes have now been found in the nuclear genomes of *Giardia* and *Trichomonas*, suggesting that these organisms once had mitochondria which were secondarily lost (Logsdon, 1998). However, another explanation for these data is that these genes have been secondarily acquired from the uptake of DNA found in the environment of the protists (Doolittle, 1998). A more complete analysis of closely related protists with and without mitochondria may resolve this issue.

1.1.5.6 Group II introns in flowering plant mitochondria

To date, a total of 25 group II introns, including both group IIA and group IIB types, have been identified within mitochondrial genes of flowering plants (Figure 1.4), and most are located within genes encoding NADH dehydrogenase subunits (*nad1*, *nad2*, *nad4*, *nad5*, *nad7*) or ribosomal proteins (*rps3*, *rps10*, *rpl2*). All but one of these, namely *nad2* intron 3, differ in position from the 25 *cis*-spliced group II introns in the mitochondrial genes of the liverwort, Marchantia polymorpha (Oda *et al.*, 1992) and minor differences are seen in the presence or absence of particular introns among flowering plants (notably the *cox2* introns 1 and 2, *nad4* introns 2 and 3 and *nad7* intron 3; Figure 1.4). Plant mitochondrial *cis*-splicing introns typically range in length from about 0.49 (i.e., *coxII* intron 1, Albertazzi *et al.*, 1998) to 3.4 kb (i.e., *nad4* intron 2, Lamattina and Grienenberger, 1991) and variation in size among plants occurs mostly within domain 4. Only one intron (*nad1* intron 4) contains a maturase-type gene (*mat-r*) with homology to reverse transcriptases.

Several of the plant mitochondrial introns have features which deviate from conventional group II intron structures (cf. Section 6.2). For example, some are lacking the bulging adenosine normally involved in group II intron splicing (i.e., *nad1* introns 1 and 2; Chapdelaine and Bonen, 1991) and some have mispairs within helical regions (i.e., *nad7* intron 3; Bonen *et al.*, 1994). None of the plant mitochondrial introns have been shown to self-splice *in vitro* (Malek and Knoop, 1998).

Frequent genomic rearrangements within plant mitochondria result in the disruption of some genes within intron sequences. The result of this is that adjacent exons occur in

Figure 1.4. Plant mitochondrial introns. The 25 plant mitochondrial group II introns and the single group I intron are listed. Group II introns are categorized as *cis*-splicing, *trans*-splicing or variable and those that are absent in some species are indicated. Accession numbers and sequences of D5 and D6 of the group II introns are listed in Appendices I and II.

Gene	Intron	cis- vs. trans-splicing	variability in presence
nad1	1	trans	no
	2	cis	no
	3	trans	no
	4	Variable among plants	no
nad2	1	cis	no
	2	trans	no
	3	cis	no
	4	cis	no
nad4	1	cis	no
	2	cis	yes
	3	cis	yes
nad5	1	cis	no
	2	trans	no
	3	trans	no
	4	cis	no
nad7	1	cis	no
	2	cis	no
	3	cis	yes
	4	cis	no
cox2	1	cis	yes
	2	cis	yes
rpl2		cis	no
rps3		cis	no
rps10		cis	no
orf454		cis	no
cox1		group I intron	yes

distant regions of the genome and must re-associate at the RNA level for splicing. In other words, *trans*-splicing can occur between two or more RNA molecules, as long as the essential group II intron domains are present. Six of the plant mitochondrial introns within *nad1*, *nad2* and *nad5* genes are *trans*-splicing (reviewed in Bonen, 1993; Figure 1.4). Interestingly the third intron of the *nad5* in *Oenothera* is actually split into three pieces (Knoop *et al.*, 1997). Similarly, *trans*-splicing events are required for excision of the chloroplast *rps12* intron in flowering plants and for the two *psaA* introns in *Chlamydomonas* plastids (reviewed in Lambowitz and Belfort, 1993). With the exception of the *nad1* intron 4 (Figure 1.4) transition from *cis*- to *trans*-splicing introns predates the divergence of monocots and dicots (120-200 mya) as *cis*-splicing homologues for all of the plant mitochondrial *trans*-splicing introns identified to date have been identified in early branching land plants (Malek and Knoop, 1998). Intron fragmentation usually occurs within D3 or D4, which are among the more variable structures of group II introns. In fact, D4 can be omitted without affecting splicing in *in vitro* reactions of yeast autocatalytic introns (Bachl and Schmelzer, 1990). Little is known of the machinery or the mechanism for *trans*-splicing but it is believed that the intron pieces recognize their correct partners through base pairing in the interrupted domain, other tertiary interactions described for group II introns, and/or RNA-protein interactions. *Trans*-splicing is highly specific: if splicing errors occur, they are beyond the detection limit of RT-PCR (Knoop and Brennicke, 1993). Regulation of such *trans*-splicing genes would be expected to be particularly complex, since the pieces of the intron are independently transcribed.

1.2 Expression of plant mitochondrial genes

Plant mitochondria have evolved complex mechanisms for genome maintenance, gene decoding and gene regulation. The mitochondrial genome encodes less than 10% of the information necessary for its biogenesis and function; however these genes are vital to the mitochondrial function in respiration (reviewed in Bonen and Gray, 1997). A majority of the essential mitochondrial proteins are nuclear-encoded, so that they are translated in the cytosol then imported into the mitochondrion. Regulation of mitochondrial gene expression must therefore be coordinated with that of the nucleus.

The study of plant mitochondria is particularly beneficial due to the agronomical

importance of cytoplasmic male sterility (CMS) which causes defective pollen development that can usually be attributed to aberrations within mitochondria (reviewed in Hanson, 1991). Production of hybrid seeds which are more robust and have superior disease resistance is often dependent on male sterility in crops where manual emasculation is not feasible.

1.2.1 Organization and gene content of the plant mitochondrial genome

Flowering plant mitochondrial genomes are variable in size, ranging from 208 kbp in *Brassica hirta* to 2400 kbp in watermelon (reviewed in Mackenzie, 1999). These genomes are relatively large compared to the streamlined vertebrate mitochondrial genomes which are between 14 and 42 kbp (Bonen and Gray, 1997). Much of this size variation is due to extraneous DNA integration and also to the presence of introns in the plant mitochondrial genome. For example, the completely sequenced Arabidopsis mitochondrial genome of 367 kbp has an estimated 60% noncoding sequence (Unseld *et al.*, 1997). In plant mitochondrial genomes, there is little highly repetitive sequence as determined by renaturation kinetic measurements (Mackenzie, 1999). While gene density in plant mitochondria is low compared to their animal counterparts, gene content is almost double. The *Arabidopsis* mitochondrial genome encodes 57 genes including tRNAs, rRNAs, subunits of components of the electron transport chain, ribosomal proteins and proteins involved in cytochrome c biogenesis (Unseld *et al.*, 1997). Interestingly, gene content varies among even closely related plants due to the ongoing transfer of genes to the nucleus (Bonen and Gray, 1997).

Generally, plant mitochondrial genomes evolve slowly at the sequence level; however, they change quickly in structure as they undergo frequent rearrangements due to recombination at repeat sequences (reviewed in Bonen and Gray, 1997). Even between closely related species, changes in gene order are observed. Such rearrangements generally occur through homologous recombination involving repeated regions. An interesting implication of this is that higher plant mitochondrial genomes do not exist as a single circular molecule (as in vertebrate mitochondria), but are divided into a number of subgenomic circles.

1.2.2 Post-transcriptional processing of plant mitochondrial transcripts

Post-transcriptional processing of plant mitochondrial transcripts can be quite complex as it can involve group II splicing (cf. Section 1.1.5), the modification of specific

cytosine residues within coding regions to uridine by a process called RNA editing (cf. Section 1.3) and processing of transcript termini (reviewed in Bonen and Gray, 1997). Due to the size of the mitochondrial genome, most genes are separated by at least several kilobases of noncoding sequence. However, some genes that are located close together are co-transcribed. Populations of RNA transcripts arising from a given mitochondrial gene-coding region can be quite heterogeneous. This variation is due to multiple transcription initiation and termination sites and to post-transcriptional events such as cleavage of transcript ends and splicing. Since transcriptional modulation appears not to be the primary means of gene regulation in plant mitochondria, post-transcriptional events can be an important level of control of gene expression (reviewed in Binder *et al.*, 1996).

Frequently both 5' and 3' end processing is required to generate mRNAs, tRNAs and rRNAs (Binder *et al.*, 1996). This end processing involves endonucleolytic cleavage at specific sites which are presumably recognized by specific endonucleases. Processing of tRNAs has been well characterized and involves two separate endonucleases: an RNase P-like enzyme for 5' end processing and a 3' endonuclease. The 3' ends of mRNAs can sometimes fold into stem-loop structures which may be recognized by endonucleases, but are also thought to bind to proteins which confer stability to the messages. Mitochondrial mRNAs are different from those in the cytosol in that they are not capped or polyadenylated in the course of their maturation. In this respect plant mitochondrial gene expression resembles prokaryotic expression. One caveat to this is that polyadenylation has been recently observed in the mitochondria of sunflowers (Gagliardi and Leaver, 1999) and in maize (Lupold *et al.*, 1999); however, such adenylation targets the transcript for degradation and is not involved in translation. The involvement of endonucleases, which are likely to be nuclear-encoded, in post-transcriptional processing of plant mitochondrial genes enables an additional level of control for mitochondrial gene expression.

1.2.3 Developmental- and tissue-specific regulation of plant mitochondrial gene expression

Plant mitochondrial activity is influenced by factors such as age, tissue location and interaction with other organelles such as chloroplasts and peroxisomes (reviewed by Whitehouse and Moore, 1995). For example, in areas of rapid growth which contain

immature plastids, such as the meristematic regions, an important function of mitochondria is energy production. In contrast, in mature leaf cells where most of the cell's energy is produced by photosynthesis, the role of the mitochondrion is primarily the production of carbon skeletons for cellular biosynthetic synthesis via the TCA cycle. Accordingly, differences in transcript levels have been observed in different tissue types. In *Petunia*, differences in levels of mitochondrial-encoded proteins have been observed among different regions of developing anthers (Conley and Hanson, 1994). In wheat mitochondria, the levels of *coxI*, *coxII*, *cob* and *atpA* transcripts were found to decrease from the basal meristem to the tip of leaves (Topping and Leaver, 1990). Tissue specific regulation of plant mitochondrial genes has also been observed in sunflowers with the CMS phenotype. Amounts of transcripts for four mitochondrial encoded genes, *atpA*, *atp9*, *cob* and *rrn26* significantly increase in young meiotic cells, while nuclear-encoded genes for mitochondrial proteins showed little difference in their transcription rates (Smart *et al.*, 1994). Differences have also been observed in intron-containing transcript levels between different stages of development. Northern blots of RNA preparations from 24 hour wheat embryos and 6 day etiolated seedlings using *nad1* and *nad5* intron probes demonstrate stage specific differences in ratios of partially spliced transcripts to excised introns (Chapdelaine and Bonen, unpublished results).

Differences in transcript levels observed may be due to differences in rates of transcription or to post-transcriptional regulation. Post-transcriptional events such as processing of transcript ends, endonucleolytic cleavage and splicing may be involved in regulating the levels of mRNA. Differences in RNA stability conferred by factors bound to 3' ends of transcripts may also be involved. In addition, post-translational events, including rapid proteolytic cleavage and translational control factors, could regulate levels of mitochondrial proteins.

1.2.4 Structure and expression of plant mitochondrial NADH dehydrogenase genes

An important group of genes encoded by plant mitochondria are the *nad* genes which code for nine (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7* and *nad9*) of the approximately 30 polypeptides (Combettes and Grienenberger, 1999) required to form complex I (NADH dehydrogenase). The activity of this dehydrogenase is required by the

electron transport chain for the function of plant mitochondria in respiration. The remaining polypeptides (approximately 20) which make up this complex are nuclear-encoded and must be imported from the cytoplasm. Interestingly, 19 out of the 25 group II introns identified to date in plant mitochondria occur within five of the *nad* genes. This disproportionate presence of introns raises the possibility that the mitochondrial *nad* genes are regulated at the level of splicing.

The plant mitochondrial electron transport chain contains enzyme complexes similar to those found in other organisms. However, in addition to complex I, plants encode several unique NAD(P)H dehydrogenases which are also involved in the reduction of ubiquinone. One of these NAD(P)H dehydrogenases competes directly with complex I inasmuch as it also oxidizes matrix NADH. This enzyme is characterized by its ability to function in the presence of rotenone, a complex I inhibitor, and is therefore called a rotenone insensitive NADH dehydrogenase. However, while this dehydrogenase can participate in electron transfer within the inner mitochondrial membrane, reactions are not coupled to proton translocation; thus, the energy resulting from their oxidation reactions is lost in the form of heat instead of being harnessed by oxidative phoshorylation. Shunting of electrons through such an alternative pathway results in the uncoupling of electron transport to ATP synthesis, and therefore loss of the energy generating capacity of the mitochondrion. On the surface, this process appears to be very wasteful; however, if the mitochondrion was performing a role in synthesis, rather than for ATP production, then this secondary electron transport chain enzyme would be useful in the plant's life cycle. The composition of the alternative dehydrogenase has not yet been elucidated. This complex may be composed of some subunits of complex I or it may be composed of some novel nuclear encoded ones. Alternatively, complex I may have a rotenone-insensitive activity (Soole and Menz, 1995). Two nuclear genes encoding homologues to rotenone-insensitive NADH dehydrogenase proteins of yeast and bacteria have recently been identified in potato; however, their function *in vivo* has not yet been demonstrated (Rasmusson *et al.*, 1999).

Complex I may not be essential at certain stages of development. In fact, there is evidence that disruption of complex I genes is not lethal in plants. For example, Pla *et al.* (1995) described a tobacco cytoplasmic male sterile (CMS) mutant in which the last two

exons of the *nad7* gene were deleted. Although this mutant had no detectable *nad7* polypeptide, the plant was able to survive, albeit with a decreased growth rate, altered leaf shape and male sterility. The remaining dehydrogenase activity was also much less sensitive to rotenone, indicating the breakdown of complex I activity. Interestingly, despite the absence of functional complex I, basal mitochondrial functions were maintained.

1.3 RNA Editing

RNA editing is a term used to describe any changes to a precursor mRNA, during or after transcription, which cause the sequence of the mature mRNA to be different from that of the coding sequence of DNA. This process has been described in rRNAs, tRNAs and mRNAs in a variety of organisms, including organelles of land plants and unicellular organisms, in the nucleus of animals, and in viruses (reviewed by Smith *et al.*, 1997). Editing is accomplished through a number of different mechanisms which can be divided into two general types. In one type, nucleotides are added to, or deleted from the RNA to generate the correct reading frame and sequence. Such editing is observed in the kinetoplasts (mitochondria) of trypanosomes where the addition and deletion of uridine residues alters the sequence of the precursor RNAs. The second type of RNA editing involves base modification, where the reading frame is not altered but specific residues within the precursor mRNA are changed. Examples of this type of editing are observed in plant organellar precursor mRNAs and mammalian apolipoprotein B mRNAs, where specific cytidines are edited to uridines. Similarly, A-to-I editing is observed in hepatitis delta virus and for the expression of glutamate receptor genes in vertebrates. It is likely that the number of examples of RNA editing will continue to increase as more organisms are examined.

In the editing systems which have been characterized to date, editing machinery involves both protein and usually RNA components, with the protein moiety being responsible for the catalysis of the reaction (Smith *et al.*, 1997). The editing site is specified by a number of different means. In some cases (i.e., apolipoprotein B), the editing machinery requires a mooring sequence downstream of the editing site, in other cases the editing site is delineated by regions of double-stranded RNA (i.e., glutamate receptor transcripts in vertebrates), while in other systems (i.e., trypanosomes) editing sites are specified by base pairing of a *trans*-acting RNA called a 'guide RNA'. In plant mitochondria, the mechanism

of editing has not yet been determined (discussed below).

1.3.1 RNA editing in plant mitochondria

RNA editing has been observed in the mitochondria of all groups of land plants studied to date, except for marchantiid liverworts (Steinhauser *et al.*, 1999). Usually flowering plant mitochondrial editing involves the conversion of specific cytosine (C) residues to uridine (U), however, rarely U-to-C changes are observed (e.g., *coxIII* in Gualberto *et al.*, 1990) and in lower plants both types of editing are frequent (Yoshinaga *et al.*, 1996; Hiesel *et al.*, 1994). It is estimated that more than 1000 editing sites occur within protein coding regions of transcripts in plant organelles, compared to 25-30 in plastids (Maier, 1997; Mulligan *et al.*, 1999); however, in *Arabidopsis* mitochondria, only 441 editing sites were found within transcribed regions (Giegé and Brennicke, 1999). For the most part, RNA editing causes changes in the encoded amino acid and results in proteins which are more similar to their homologues in other plants. However, a significant number of edits (11.6% in *Arabidopsis*) also occur at silent sites (Giegé and Brennicke, 1999). RNA editing is an early processing event: fully-spliced transcripts are usually fully edited whereas unspliced ones may show partial editing (Sutton *et al.*, 1991; Yang and Mulligan, 1991). Interestingly, RNA editing has also been observed in a small number of cases in noncoding regions, rRNAs, tRNAs and in introns (Smith *et al.*, 1997).

It seems very inefficient to encode the information needed for editing machinery that would have been unnecessary had the genetic information been encoded correctly in the DNA of the organism. Why has this process arisen and been maintained among widely divergent plants? One possibility is that RNA editing arose by chance and then got locked in due to the accumulation of mutations which were not detrimental because of the availability of editing machinery (Covello and Gray, 1993). Another possibility is that RNA editing provides a method of regulation. In a small number of cases, RNA editing alters the sequence of an mRNA precursor so that sequences that are involved in translation initiation are corrected and translation can proceed. For example, in *nad1* transcripts of wheat (Chapdelaine and Bonen, 1991), *rps10* transcripts of *Solanum tuberosum* (Araya *et al.*, 1994) and in *coxII* transcripts of radish (Dong *et al.*, 1998) a functional AUG start codon is generated by RNA editing of an ACG triplet. Similarly, in *atp6*, *atp9* and *rps10* genes,

editing creates stop codons within transcripts (Hanson *et al.*, 1996A). While the idea that editing at such sites could be involved in regulating gene expression is appealing, if this were the case, one would expect editing events such as these to occur more frequently, or in groups of co-regulated genes.

On the other hand, there may be some forms of regulation of the RNA editing machinery that affect how efficient the editing of the complete precursor mRNA will be as opposed to the editing at a single site. Tissue and developmental specificity of RNA editing has been observed in the *nad3* transcripts in maize mitochondria (Grosskopf and Mulligan, 1996). Because most editing sites are at non-silent codon positions within exon sequences, proteins made from unedited mRNAs would be non-functional or may have an altered function. However, there does not appear to be a mechanism for preventing incompletely edited transcripts from being translated. Partially edited transcripts of *atp6*, *nad9* and *rps12* were found to be associated with translational apparatus (polysomes) (Lu and Hanson, 1994; 1996; Hanson, 1996B; Lu *et al.*, 1996). Polysomes appear to be rather indiscriminate, since intron containing transcripts have also been found to be associated with polysomes in maize (Yang and Mulligan, 1993). Furthermore, proteins generated from the incorrect *rps12* transcripts were detected *in vivo* (Lu *et al.*, 1996; Phreaner *et al.*,1996). Perhaps the observed proteins which are generated based on unedited sequences perform a secondary function in the mitochondrion which has not yet been determined. In contrast, proteins generated from the unedited *atp6* and *atp9* transcripts could not be detected *in vivo* (Lu and Hanson, 1994; Grohmann *et al.*, 1994). Perhaps such non-functional proteins are rapidly degraded *in vivo*. If this is the case, then levels of such proteins could be regulated through RNA editing.

1.3.2 Mechanism of RNA editing in plant organelles

Currently the machinery required for editing in plant organelles has not yet been identified; however, recent studies have provided insight into how editing is accomplished in plant organelles. Editing in chloroplasts and mitochondria appears to occur by similar processes. There are three likely possibilities for the mechanism of editing: either nucleotide replacement by an insertion/deletion mechanism, base replacement or deamination of the cytidine to produce uridine (reviewed by Mulligan *et al.*, 1999). To date, evidence supports deamination of the cytidine as being the mechanism of RNA editing. Labeling studies, using

^{3}H on the cytidine base or ^{32}P at the α phosphate have supported this hypothesis (Blanc *et al.*, 1996; Yu *et al.*, 1998). The protein machinery involved in editing is thought to be encoded by the nucleus. In a barley mutant with extremely low levels of chloroplast ribosomes, editing could still take place in the absence of plastid encoded proteins. However, Karcher and Bock (1998) showed that treatment of plants with prokaryotic translational inhibitors impeded plastid RNA editing at certain sites suggesting that some site-specific factor(s), but none of the generalized editing machinery, are proteins encoded within plastids.

While general editing factors appear to be similar, factors conferring specificity differ between mitochondria and plastids, and also among different plants. For example, when the second exon of the *Petunia* coxII gene was transformed into tobacco chloroplasts, none of the seven editing positions were edited (Sutton *et al.*, 1995). This may reflect the low level of editing in chloroplast sequences compared to those in mitochondria. However, an *rpoB* fragment that was found to be transferred from chloroplasts to mitochondria in rice is also not edited even though it contains several editing sites (Zeltz *et al.*, 1996). Editing within a gene is also not conserved among plants. For example, carrot and *Oenothera rps13* have different editing sites, even though primary sequence is conserved in the region immediately flanking the editing site (Wissinger *et al.*, 1990).

No consensus sequences for delineating editing positions have been identified. However, sequence comparisons have indicated a bias against a purine (especially guanine) in the position upstream of the editing site (Mulligan *et al.*, 1999). In fact, 96.3% of *Arabidopsis* editing positions are preceded by pyrimidines (Giegé and Brennicke, 1999). Recently plastid transformation studies have provided new insight into the specificity of RNA editing. For a *psbL* site a short sequence upstream of the editing position (16 nt) is important for editing while only 5 nt downstream of the site is required (Chaudhuri and Maliga, 1996). However, the *ndhB* editing sites differed in the length of sequence required for delineating an editing site (Bock *et al.*, 1996; 1997). Such studies have not been undertaken in mitochondria as a plant mitochondrial transformation system has not been developed, but frequent duplication/ rearrangements within the mitochondrion have provided some insight into the sequence requirements for editing. For example, the study of *rps12* editing sites within a chimeric pseudogene, *rps12b*, created by recombination between *rps12*

and *rps3* regions indicate that the sequence at least 7 nucleotides upstream of an editing site was required for editing (Mulligan *et al.*, 1999). In a chimeric rice *atp6* gene, an editing site was not lost despite a rearrangement at the flanking 3' nucleotide (Kubo and Kadowaki, 1997). Editing in mitochondria appears to be specified primarily by upstream sequences as it is in plastids.

The *trans*-factors involved in recognizing an editing position are still undetermined. Guide RNAs (as in kinetoplasts) have been proposed to perform this role. Searches for *cox 3* antisense transcripts in wheat using gel blots have not been successful (Gualberto *et al.*, 1990). Low stringency RNase protection assays using sense *cox2* transcripts also failed to detect any antisense RNAs (Mulligan *et al.*, 1999). While a putative guide RNA was identified in tobacco chloroplast DNA sequence, mutagenesis of this RNA did not abolish editing at the expected site (Bock and Maliga, 1995). It will be interesting to discover what factors are actually involved in conferring specificity of editing at the numerous plant organellar editing sites.

1.3.3 RNA editing in mitochondrial group II introns

Plant mitochondrial RNA editing is not restricted to protein coding sequences. There have been a few examples of RNA editing within structural RNAs such as tRNA and group II intron transcripts and in flanking non-translated sequences. In the potato and bean $tRNA^{Phe}$(GAA) a genomically encoded C corresponding to a position in the stem of this tRNA is found to be a U in the tRNA, correcting an AC mismatch which would have affected the stability of this stem (Maréchal-Drouard *et al.*, 1993). RNA editing at 16 sites in group II introns has been reported in the literature (Appendix III) and a role for this editing in improving helical structures for splicing has been proposed (Wissinger *et al.*, 1991).

RNA editing which improves base pairing within stems of group II introns has been observed in plants at a total of six sites within helical positions in D1 of *rps 10* (Zanlungo *et al.*, 1995), D6 in *nad1* introns 3 and 4 (Wissinger *et al.*, 1991; Farré and Araya, 1999), D1 and D4 of *nad2* intron 2 (Binder *et al.*, 1992) and D6 of *nad5* intron 2 (Knoop and Brennicke, 1993). In these structures, RNA editing may be required to improve base-pairing within helical regions (Schuster and Brennicke, 1994; Wissinger *et al.*,1992). This model is conceptually appealing because intron editing would then serve an important biological role,

as it does in coding regions.

Not all editing events in introns improve pairing within core helical structures. In five cases RNA editing events in D6 of *nad1* introns 1 and 4 (Conklin *et al.*, 1991; Wissinger *et al.*, 1991; Chapdelaine and Bonen, unpublished) and *nad5* intron 1 (Knoop *et al.*, 1991; Giegé and Brennicke, 1999) do not appear to improve secondary structure. Five other sites have been reported in *nad2* introns 1 and 2 (Binder *et al.*, 1992), *nad5* introns 1 and 4 (Giegé and Brennicke, 1999; Knoop *et al.*, 1991) but the effects of these changes are not clear as the structure of these regions has not been determined. It should be noted that a number of RNA editing sites have also been observed in the *mat-r* gene which is located within the non-core part of *nad1* intron 4 (Farré and Araya, 1999) and they follow editing patterns seen in other plant mitochondrial coding sequences. The question remains as to whether editing of group II introns is an essential event, as it is in coding sequences, or a chance occurrence, due to the presence of elements that would predispose certain positions to editing.

1.4 Objectives

The aim of this project is to study the expression of intron-containing plant mitochondrial genes. Most plant mitochondrial introns occur within the *nad* genes which encode polypeptides that form part of the NADH dehydrogenase in the electron transport chain. Expression of the *nad* genes is particularly complex due to the requirement for C-to-U RNA editing and the presence of *cis*- and *trans*-spliced introns within 5 out of 9 known mitochondrial *nad* genes. Furthermore, stage-specific differences in steady state levels of intron-containing *nad* transcripts suggests that these genes may be developmentally regulated. The goal of my project is to examine transcript processing within intron-containing *nad* genes, with particular emphasis on splicing due to the aberrant nature of the introns within these genes. The following questions will be addressed:

1. Does RNA editing play a role in splicing? Editing events which improve base pairing in highly conserved regions of the group II introns, such as domains 5 and 6, may help restore the classical structure of the intron. This improved structure may be required for excision of the intron from the precursor mRNA or may simply enhance the efficiency of intron excision. By looking at the degree of RNA editing within core structures of plant mitochondrial introns, the conservation of such editing among plants, and the temporal

relationship between editing and splicing, this question can be addressed.

2. Are seed plant mitochondrial group II introns evolving in a different direction than group II introns in other organisms? The unusual features of plant mitochondrial introns, along with the observation of intron RNA editing suggests that splicing of these introns may depend on accessory factors; perhaps more so than group II introns in other organisms which are capable of self-splicing. Such aberrant intron structures may reflect degeneration after their role has been taken over by a *trans*-acting factor in the cell. If this is the case, then intron structures would become highly variable among plants, due to the lack of evolutionary constraint. Alternatively, unusual structures may be conserved as they take on new roles, such as protein-binding for the production of a spliceosomal-like complex. The degeneration of plant mitochondrial introns may reflect an increased reliance on host factors, in much the same manner as has been proposed for the acquisition of generalized splicing machinery in the evolution of nuclear spliceosomal introns. These possibilities can be addressed by examining intron sequence conservation among plants at the DNA and at the RNA level.

3. How do aberrant group II introns splice? Generally, splicing of group II introns proceeds via two transesterification steps and introns are excised in a lariat form. The excision of plant mitochondrial introns in such a manner has not yet been demonstrated and the presence of non-classical features (i.e., lack of bulging adenosine in D6 in some cases) may preclude splicing by typical mechanisms. For example, in yeast, mutants lacking a bulging adenosine in D6 were excised in a linear form, by first step hydrolysis *in vivo* (Podar *et al.*, 1998). Alternatively, another nucleotide within D6 could provide the 2' hydroxyl for the first transesterification step in the splicing reaction, resulting in near-normal splicing with lariat formation. These possibilities will be addressed by looking at the physical form of excised plant mitochondrial introns to gain insight into the mechanism of plant mitochondrial splicing.

4. Are developmentally-specific differences in the steady state levels of intron-containing transcripts in some of the *nad* genes due to regulation at the level of splicing? Northern blots of RNA preparations from 24 hour embryos and 6 day etiolated seedlings probed with *nad1* and *nad5* intron probes have clearly demonstrated that there are stage specific differences in the levels of RNA intermediates (Chapdelaine, 1992). These differences cannot be explained by changes in levels of transcripts alone, as differences in

apparent sizes are observed as well. The possibility of the regulation of complex I genes is of particular interest since plant mitochondria are unique in that they possess an additional NADH dehydrogenase on the inside of the inner mitochondrial membrane which could potentially perform the role of complex I under different developmental conditions. The fact that most mitochondrial introns occur within complex I genes may indicate that splicing is involved in regulating the expression of these genes. Alternatively, their expression may be regulated by transcription and RNA turnover rates. The examination of intron editing and sequence variation, as well as developmental differences among intron-containing mitochondrial transcripts should provide insight into the origin and role of plant mitochondrial introns.

CHAPTER 2:
MATERIALS AND METHODS

2.1 Isolation of plant mitochondrial DNA

Mitochondrial nucleic acids were isolated from wheat (*Triticum aestivum* var. Frederick), soybean (*Glycine max* var. Maple Arrow), rice (*Oryza sativa* var. v20B and v26B), pea (*Pisum sativum* var. Thomas Claxton), *Arabidopsis thaliana* (var. Columbia) and tobacco (*Nicotiana tabacum* subsp. petite Havana). For wheat, rice, soybean and pea, seeds were surface sterilized with diluted Javex in dH_2O (1:10) for 10 minutes, followed by soaking in 10 mM HCl for 5 minutes. After each treatment, seeds were rinsed with sterile water 3 to 5 times. Planting was done in approximately 3 cm of sterile water-soaked vermiculite and seedlings were grown for 6 days in the dark. For wheat, rice and soybean, embryos were dissected from surface sterilized seeds in order to remove the endosperm and then grown on 2 layers of 3 MM filter paper saturated with 1% glucose for 24 hours in the dark. For *Arabidopsis* and tobacco, mitochondrial nucleic acids were isolated from greened leaf tissue after 4 to 6 weeks of growth in soil since germination of these seeds required a longer period of time and was dependent on an external source of nutrients which was present in the soil and not in vermiculite.

The methods used for isolation of mitochondrial DNA and RNA were adapted from Bonen (1987). Plant tissue was homogenized in 0.5 ml/g wet weight of cold buffer I (0.44 M sucrose, 50 mM Tris pH 8.0, 3 mM EDTA, 1mM β-mercaptoethanol, 0.1% BSA) using a cold mortar and pestle. This homogenate was filtered through four layers of cheesecloth then through one layer of Miracloth (Calbiochem). The cellular fraction was pelleted by two five minute spins at 500 x g in an SS34 rotor, the supernatant was transferred to a new tube and the mitochondrial fraction was then pelleted by a 20 minute spin at 12 000 x g in the SS34 rotor. For the greened leaf tissue from *Arabidopsis* and tobacco, this pellet also contained chloroplasts; however, in tissue from etiolated seedlings and embryos only minor amounts of proplastids were presumed to be present.

Mitochondrial DNA isolation was performed using methods adapted from Wilson and Chourey (1984). The crude mitochondrial pellet was resuspended in 600 µl buffer II (50 mM Tris pH 8.0, 20 mM EDTA) then mitochondria were lysed in 1.2 ml of buffer III (200

mM Tris pH 8.0, 100 µM EDTA, 200 mM NaCl, 2% SDS, 200 mM β-mercaptoethanol) then incubated at 65°C for 20 minutes. Following this, 600 µl of 5 M KOAc was added and samples were incubated on ice for 30 minutes before spinning for 3 minutes. The supernatant was removed to a new tube and mtDNA was precipitated in 120 µl of 5 M NH_4OAc and 1.2 ml of isopropanol at -20°C for 30 minutes. DNA was pelleted by centrifugation then resuspended in 1.4 ml of buffer II. DNA precipitation was repeated with 150 µl of 3M NaOAc and 1 ml isopropanol as above. The DNA pellet was resuspended in 50 µl of TE buffer (10 mM Tris pH 7.5, 1.0 mM EDTA) and stored at -20°C.

2.2 Isolation of plant mitochondrial RNA

For mitochondrial RNA isolation, crude mitochondrial pellets were obtained by differential centrifugation as described above. These pellets were resuspended in 200 µl of buffer IV (10 mM Tris pH 7.5, 50 mM KCl, 10 mM $MgCl_2$), then disrupted with 200 µl of buffer IV containing 8% Triton-X 100. After mixing well, 600 µl of detergent mix (2% tri-isopropylnaphtalene sulfanate,12% sodium p-aminosalicylate, 0.1M NaCl, 20 mM Tris pH7.4) was added, then samples were extracted once with phenol (0.7 ml phenol containing 0.1% 8-hydroxyquinoline and saturated with TE buffer), then once with phenol:chloroform: isoamyl alcohol (24:24:1) according to standard procedures (Sambrook *et al.*, 1989). Samples were then either LiCl precipitated (see below) or precipitated by 2.0 vol of 95% EtOH and 0.1 vol of 5M NaCl before being stored at -20°C.

2.2.1 LiCl and DNAse treatment of mtRNA

To remove DNA, RNA preparations were precipitated overnight at 4°C in 2M LiCl following standard procedures (Sambrook *et al.*, 1989). Samples were centrifuged 20 minutes at 4°C followed by one wash with 500 µl of 4M LiCl and two washes with 70% EtOH. The RNA pellet was then resuspended in the desired amount of low TE. In some cases this precipitation was repeated two more times.

Some samples were subsequently treated with 7.5 units of FPLC pure, RNAse-free DNAse I (Pharmacia) in 1x DNAse buffer (40 mM Tris (pH 7.5), 6 mM $MgCl_2$) containing 20 units of RNAse inhibitor for 20 minutes at 33°C. Samples were extracted once with an equal volume of phenol saturated with TE (pH7.5) and once with phenol-chloroform (Sambrook *et al.*, 1989) before ethanol precipitation with 2.0 vol of 95% EtOH and 0.1 vol

5M NaCl.

For RT-PCR experiments, LiCl precipitation immediately following RNA isolation was usually sufficient for removing most of the contaminating DNA. DNAse treatment following this precipitation helped reduce levels of contaminating DNA further; however, low levels of DNA could still be detected by RT-PCR following these treatments (cf. Figure 3.4CD).

2.2.2 Gel fractionation and purification of mtRNA

Approximately 10 µg of wheat mtRNA which had been LiCl precipitated three times was run on a 1.25 % agarose gel containing 6.6 % formaldehyde in MOPS buffer as described in the protocol provided with the RNAid kit (Bio101). The region of the gel containing the fraction of mitochondrial RNA of interest was determined by staining a parallel lane containing RNA size markers (Gibco-BRL) with ethidium bromide. RNA was isolated from the gel slices by melting the slices, binding the RNA to silica beads then eluting the RNA into ddH_2O using the RNAid kit (BIO 101). RNA fractions were subsequently treated with DNase I, as above, in order to remove any remaining DNA contamination.

2.3 Oligonucleotides used for PCR and hybridization

The following oligonucleotide primers used in RT-PCR, primer extension and as oligomer probes for northern analyses were designed from the genomic sequence of the wheat mitochondrial *nad7* gene (Bonen *et al.*, 1994). The positions and orientations of the primers are numbered according to the sequence available in GenBank and EMBL databases (X75036).

Primer	Sequence (5' to 3')	Position in *nad7*	Description	Strand
LB39	TAGGATCCTGATCGAGCAAG	6639-6620	3' UTR	Antisense
LB40	GCAGTCGACTGAGTTCTGAA	6493-6512	intron 4	Sense
LB41	CCACCAGATCTTAAGGAAAG	4226-4245	intron 3	Sense
LB42	GCCGGTGCGTGGTGAACGGA	5640-5621	intron 4	Antisense
LB43	CTGGACAAGCTTTAGGGGAA	552-571	5' UTR	Sense
LB46	ACTTATCTTCAAGCTTTACC	1594-1623	exon 2	Sense

Primer	Sequence (5' to 3')	Position in *nad7*	Description	Strand
LB47	GCACAGCAAGCAAAGGATTG	3361-3380	exon 3	Sense
LB48	CTGGTACCTACTGGTACTTC	4488-4469	exon 4	Antisense
LB53	ACGCGAATTCGCTTCCGAGG	1408-1427	intron 1	Sense
LB54	CTAGGATCCCCACTCATATT	2705-2724	intron 2	Sense
LB60	CCACCACTTCACTTTTGCAC	4390-4371	intron 3	Antisense
LB64	ATGCATGCTTTTGTAGGGTC	3493-3474	intron 3	Antisense
LB87	TCTGACTCTATTGGTCATAG	4164-4183	intron 3	Sense
LB91	GGAGTGGCACAAGATCTGCC	3220-3239	exon 3	Sense
LB92	ATCGGCTTTGATCATGCCAC	4592-4573	exon 4	Antisense
LB94	CACTTTCATTCGACGTTCCG	3973-3954	intron 3	Antisense
LB133	CGTGTCAGCTTAGTTATC	4977-4958	intron 4	Antisense
LB151	CATGTCGCACTGGGTAG	3419-3410, 4399-4393	intron 3 lariat	Antisense
LB152	CACGGCGCACGATTCCA	4660-4644	exon 4/ intron 4	Antisense
LB159	AACCTCGGGATGCTTTACTC	4752-4733	intron 4	Antisense
LB165	GGCGCACGTAGAGTAGA	4657-4651, 6349-6340	intron 4 circle	Antisense
LB172	GCGCACTAGACCGATGA	4656-4651, 6343-6333	intron 4 lariat	Antisense
LB178	GTAGTCATTTGGCCGGCTAT	6328-6309	intron 4	Antisense
LB181	TTACTGACCAGAAAGACACC	6460-6441	exon 5	Antisense
LB182	ACGGCGCACGTAGAGTAGAC	4659-4651, 6349-6339	intron 4 circle	Antisense
LB183	ACGGCGCACTAGACCGATGA	4659-4651, 6343-6333	intron 4 lariat	Antisense

The following oligonucleotides were designed based on wheat *nad1* (Chapdelaine and Bonen, 1991) and *nad4* sequences (Lamattina and Grienenberger, 1991). The positions and orientations of the primers are numbered according to the sequence available in GenBank and EMBL databases.

Primer	Sequence (5' to 3')	Location	Accession number	Strand
LB18	ACTTCATAAGGGACCATTTG	*nad1* exon 2	X57967 676-657	Antisense
LB19	ATCTGCTTTTGCGCCATGAC	*nad1* exon 3	X57967 2182-2163	Antisense
LB115	AAGGCTACTCCTAGTAGAAG	*nad1* exon 1	X57968 198-179	Antisense
LB65	GTAGATCTTTCAACACCTGC	*nad1* intron 1	X57968 604-585	Antisense
LB95	GACTTCCACTTATCGATCCC	*nad1* intron 1	X57968	Antisense
LB116	TCGGGTCGACCAGGTCAGGC	*nad1* 5' UTR	X57968 86-105	Sense
LB128	TTGGGTTGGGGTCTCGAGCCG	*nad1* intron 1	X57967 508-527	Sense
LB175	CCATGCTGAGTCACAGGCAG	*nad1* intron 2	X57967	Sense
LB176	GATCCCTTGCTCTCACGTTC	*nad1* intron 2	X57967	Antisense
LB90	TTTCTTTGCTGATTCCTCTC	*nad4* exon 4	X57164	Antisense
LB93	GTTCCGGTTCACTGATAAGG	*nad4* intron 2	X57164	Sense
LB168	CTATCAAGAACAGATCAGGC	*nad4* intron 2	X57164	Antisense

2.4 Polymerase Chain Reaction (PCR)

PCR amplifications were carried out using either mtDNA (wheat, rice, soybean, pea) or total DNA (*Arabidopsis*, tobacco). Approximately 10 ng (mtDNA) or 50-100 ng (total DNA) of template was used per 50 µl reaction and amplification was carried out using 1.25 U Taq DNA polymerase (Promega) in the manufacturer's buffer containing 0.2 mM dNTPs, 4.0 mM $MgCl_2$ and the appropriate primer pairs (125 ng each primer). Amplification conditions were typically 35 cycles of 94°C for 40 seconds denaturation, 42-55°C for 60 seconds ramping and 30 seconds annealing and 72°C for 90 seconds. This was followed by 1 cycle of extension at 72°C for 5 minutes. The annealing temperature (Ta) ranged from 42-

55°C depending on the GC content of the oligomers used (Ta = 4°C (# GC) + 2°C (# AT) - 5°C).

2.5 Reverse Transcriptase-PCR (RT-PCR)

2.5.1 Complementary DNA (cDNA) synthesis

For generating PCR products from spliced mRNA, partially spliced precursors (in which at least one intron had been removed) and excised introns, a complementary DNA synthesis step was done before PCR. cDNAs were generated from 4-10 µg mitochondrial RNA using 15 units of avian myeloblastosis virus (AMV) reverse transcriptase (Promega) and 25 ng of a gene specific oligomer. Synthesis was performed in 1x RT buffer (24 mM Tris (pH 8.3), 16 mM $MgCl_2$, 8 mM dithiothreitol (DTT), 0.4 mM dNTPs) containing 10 units RNasin and 4 mM sodium pyrophosphate according to the manufacturer's protocol (Krug and Berger, 1987). Synthesis was carried out at an increased extension temperature of 50°C instead of 42°C to reduce the secondary structure of RNA templates. The reaction was stopped by adding 25 µl of TE. PCR was carried out as described above using approximately 4% of the cDNA synthesis reaction mixture.

2.5.2 RT-PCR across lariat branch sites

RT-PCR across lariat branch sites was done as described by Vogel *et al.* (1997A). Approximately 1 µg of mtRNA (wheat, rice, soybean, pea) or 4 µg of total RNA (*Arabidopsis*, tobacco) purified by LiCl precipitation was pelleted and washed with 70% EtOH. The pellet was resuspended in 12 µl low TE containing 25 ng of an antisense oligomer mapping to the 5' end of the intron of interest (in some cases several oligomers were used in a single cDNA reaction). Samples were denatured at 70°C for 10 minutes then quick cooled on ice for 5 minutes. Reverse transcription with 200 units of Superscript II RT was carried out in the manufacturer's buffer (50 mM Tris (pH8.3), 7.5 mM KCl, 3 mM $MgCl_2$) and 10 mM DTT, 0.5 mM dNTPs and 20 units RNAse inhibitor (RNAsin). Reactions were incubated for 2 hours at 49°C before denaturing the enzyme by incubating at 70°C for 15 minutes then adding 30 µl of TE. The samples were stored at -20°C. Typically, approximately 5% of the cDNA synthesis reaction was used for PCR amplification. To examine the branchpoint region of excised intron RNA molecules, an antisense oligomer mapping close to the 5' end of an intron was used in combination with a sense oligomer

mapping to the 3' end of the same intron for PCR amplification (cf. Section 4.2).

2.6 Gel purification of DNA fragments

To purify the RT-PCR products for cloning or direct sequencing, products were run on an agarose gel in TBE buffer and then eluted using Gene Clean II (Bio101) according to the manufacturer's instructions or filter tips (BIO-RAD) according to Dean and Greenwald (1995). In the latter procedure, the ends of the tips were cut so that they fit into a 1.5 ml microfuge tube, and the gel slice containing the desired PCR product was placed at the top of the filter tip. Samples were spun at 4000 rpm for 10 minutes, then the eluate was ethanol precipitated according to standard procedures (Sambrook *et al.*, 1989) or the filtrate was used directly. For sequencing, the purity of this filtrate was improved by incubating the agarose gel in ddH_2O overnight at 4°C before isolating DNA fragments.

2.7 Cloning DNA fragments and RT-PCR products

PCR products were prepared for cloning by creating blunt ends (Sambrook *et al.*, 1989) since Taq polymerase adds an extra adenosine residue to the 3' end of the DNA strand. For the blunting reaction of approximately 1 µg of gel purified PCR product, adenosine residues were removed using 10 units of DNA polymerase (large fragment Klenow), which has a 3' exonuclease activity, and phosphate groups were added to the 5' end of DNA strands using 10 units of T4 polynucleotide kinase in 100 µl of 1x reaction buffer (1mM ATP, 50 mM Tris, pH 7.5, 10 mM MgCl2, 1 mM DTT, 20 µM dNTPs, 0.5 mg/ml BSA). The reaction was incubated for 60 minutes at 37°C. Fragments were subsequently purified by Gene Clean II (BIO/CAN Scientific) or EtOH precipitation before ligation into dephosphorylated SmaI pUC vector (Pharmacia) as described below.

For PCR products which contained restriction sites or mtDNA fragments cloned into pUC vectors, DNA was prepared for cloning by restriction enzyme digestion following standard protocols (Sambrook *et al.*, 1989). DNA fragments were separated by agarose gel electrophoresis and subsequently eluted from the gel using filter tips or Gene Clean II (see above). Fragments were ligated into pUC18, pGEM or M13 vectors which had been digested with the appropriate enzymes. Approximately 200 ng of DNA and 50 ng of M13 vector or 15 ng pUC or pGEM vector in a total volume of 15 µl were ligated using 1 unit of T4 DNA ligase in 1x ligation buffer (50 mM Tris (pH 7.5), 10 mM $MgCl_2$). Constructs were

transformed into *Escherichia coli* JM101 for M13 clones, JM109 for pGEM clones or TB1 for pUC clones. Plasmid DNA was isolated by the boiling method (Sambrook *et al.*, 1989) and M13 DNA was isolated by the polyethylene glycol method (Sambrook *et al.*, 1989).

2.8 Sequencing of cloned DNAs and RT-PCR products

Samples were sequenced by dideoxy chain-termination (Sanger *et al.*, 1977) using Sequenase version 2.0 (US Biochemical) and α-^{35}S-dATP according to the manufacturer's recommendations. Products were separated on 4 mm thick, 7% polyacrylamide gels containing 7 M urea in 0.5x TBE buffer (25 mM Tris pH 8.0, 1 mM EDTA, 10 mM boric acid) using standard procedures.

PCR products were sequenced directly using a modified protocol in which 75 ng of primer was annealed to 200 ng template in the presence of 0.5 % NP-40 and 1x reaction buffer (provided in the Sequenase kit) by boiling for 3 minutes then quick cooling in ethanol cooled to -20°C. This was followed by gradual warming to room temperature in a heat block cooled to -20°C. The labeling cocktail was also modified to include four times more labeling mix, approximately 0.5 % NP-40, Mn buffer and 1.5 times the suggested amount of α-^{35}S-dATP. The primer used in the sequencing reaction was one of the primers used in the PCR reaction or a primer mapping to an internal region of the PCR product.

2.9 Preparation of radiolabeled probes for hybridization

2.9.1 Labeling 5' ends of oligomers with ^{32}P

Probes for primer extension and for northern or Southern hybridization were prepared by end labeling oligomers ranging from 17-20 nt in length (see section 2.2). Oligomers (100 ng) were incubated at 37°C for 60 minutes with 50 µCi of γ-^{32}P-ATP (3000 Ci/mM, Amersham) and 7 units of T4 polynucleotide kinase in 1x kinase buffer (50 mM Tris, pH 9.5, 10 mM $MgCl_2$, 5 mM DTT) in a final volume of 15 µl. The reaction was stopped by adding 35 µl TE (10 mM Tris (pH7.5), 1 mM EDTA), then labeled oligonucleotides were spun through a Sephadex G-50 column in a 1 ml syringe. The eluate with the highest specific activity (usually the 2nd or 3rd as judged by measurement with a Geiger counter) was used in hybridization experiments. For oligomers which were to be used for primer extension experiments, the latter purification step was omitted.

2.9.2 ^{32}P-labeling by second strand synthesis using M13 templates

Hybridization probes were radioactively-labeled by strand-specific primer extension for M13 cloned DNAs. Approximately 50 ng of single stranded M13 DNA was used in a reaction containing 1.5 ng forward sequencing primer, 15 μCi of α-^{32}P-dATP (3000 Ci/mM, Amersham) in 1x reaction buffer (10 mM Tris (pH 7.0), 10 mM $MgCl_2$, 10 mM NaCl) in a volume of 5 μl. This mixture was incubated at 65°C for 3 minutes then cooled to room temperature over 30 minutes. The labeling was carried out at room temperature for 3 to 4 hours in a volume of 10 μl containing 1.7 μM dNTPs (G, T, C) and 2 units of Klenow. After adding 40 μl of TE, the unincorporated label was removed by spinning the reaction through a Sephadex G-50 column. Typically incorporation of label was approximately 50% (determined by monitoring column and eluate) and probes with at least 25% incorporation of label were used in subsequent hybridization experiments after denaturing by boiling for 2 minutes and quick cooling on ice.

2.9.3 ^{32}P-labeling DNA by PCR

PCR was used to generate uniformly labeled probes for S1 nuclease mapping experiments. Approximately 0.2 ng of plasmid DNA containing the insert of interest, which had been linearized by restriction enzyme digestion, was used in a 25 μl reaction containing 1x PCR buffer (Promega), 4.0 μM $MgCl_2$, 10 μM dNTPs, 125 ng of each specific oligomers (usually universal and reverse sequencing oligomers) and 15 μCi of α-^{32}P-dATP (3000 Ci/mM, Amersham). Amplification conditions were 20 cycles of 94°C for 1 minute, 47°C for 90 seconds and 72°C for 2 minutes, followed by 1 cycle of 72°C for 5 minutes. The unincorporated label was removed by spinning the reaction through a Sephadex G-50 column. Probes with at least 10% incorporation of label (as determined by monitoring column and eluate) were used in subsequent S1 nuclease experiments.

2.10 Northern and Southern hybridization

Northern hybridization analyses were performed using standard procedures (Sambrook *et al.*, 1989). Approximately 4 μg of mtRNA was denatured in a volume of 10 μl containing 1x MOPS buffer pH 7.0 (40 mM 3-[N-morpholino]propane-sulphonic acid (MOPS), 10mM sodium acetate and 1 mM EDTA, pH 8.0.), 6.6% formaldehyde, 50% formamide and 40 ng of ethidium bromide by heating at 65°C for 5 minutes with frequent

vortexing. RNA samples were then quick cooled on ice and sterile RNA loading buffer (50% glycerol, 1mM EDTA and bromophenol blue) was added.

Denaturing agarose gels containing 7.2% (v/v) formaldehyde, 1.2 % agarose, in 1x MOPS buffer were pre-run for 1 hour in a submarine gel system at 60 V (24 mA) before samples were loaded. Electrophoresis was performed at 60V for the first hour and increased to 100 V for 3 hours. The RNA was transferred by capillary action from the gel to a nylon membrane (HyBond-N, Amersham) overnight in 20x SSC buffer (6 M NaCl, 0.6 M sodium citrate, pH 7.0). Blots were air dried before cross-linking by UV light (254 nm) for 5 minutes.

Southern blots of cloned DNAs were prepared by standard procedures (Sambrook *et al.*, 1989). Samples were separated on 1 - 1.5% agarose gels in 1X TBE. Gels were stained with ethidium bromide and photographed before denaturation in 1.5 M NaCl, 0.5 M NaOH for 30 minutes and neutralization in 3 M NaOAc (pH 5.5) for 30 minutes. DNA was transferred to a nylon membrane (HyBond-N, Amersham) as described above.

Hybridization conditions used varied depending on probes. For M13 and PCR probes, hybridization was carried out in 50% deionised formamide, 5 x Denhardt's solution, 5 x SSC , 1% sodium dodecyl sulphate (SDS), 50 mM $NaPO_4$, pH 7.0 and 250 µg/ml denatured sheared herring sperm DNA (Sambrook *et al.*, 1989). For oligomer probes, hybridization was generally carried out in 5% deionized formamide, 5x SSC, 0.1% SDS and 50 µg/ml yeast tRNA. In order to increase the stringency of some oligomer hybridizations, the concentration of formamide was increased up to 20%. Hybridizations were carried out in 25 ml falcon tubes in hybridization ovens at varying temperatures. For M13 probes, temperatures were generally 42°C for homologous probes and 37°C for heterologous probes. For oligomer hybridizations, temperatures ranged from 37°C to 51°C depending on the melting temperature of the oligomer and the stringency required for the hybridization. In both cases, membranes were prehybridized for at least one hour in 10-15 ml hybridization buffer. This buffer was removed prior to the addition of the probe and 1-3 ml of fresh buffer was added. Hybridization took place overnight.

Following hybridization, membranes were rinsed in wash solution (2x SSC or 0.2x SSC for higher stringency, 0.1% SDS), then washed in 20 ml of the same solution for 20

minutes at the appropriate incubation temperature. This step was repeated 1-2 times. Membranes were exposed to XAR film (Kodak) at -80°C for the appropriate length of time.

2.11 *In vitro* synthesis of RNA

RT-PCR and PCR products derived from wheat *nad7* introns 3 and 4 and cloned into pGEM vectors were used as templates to generate transcripts which were used as controls for RNAse H and S1 experiments. Approximately 1 µg of plasmid DNA, digested with the appropriate restriction enzyme, was used as template for synthesis using either SP6 or T7 RNA polymerase (depending on whether sense or antisense RNA was being transcribed). DNA template was incubated in 1x SP6/T7 Buffer (Promega), 10 mM DTT, 0.5 mM NTPs, 50 ng/µl BSA, 15 units of the appropriate enzyme and 1 unit of RNAsin for 30 minutes at 37°C. The reaction was stopped by adding 10 U of DNAse 1 (HPLC pure, Pharmacia) and incubating for another 10 minutes at 37°C followed by phenol:chloroform extraction then ethanol precipitation by standard procedures (Sambrook *et al.*, 1989).

2.12 S1 nuclease protection assays

In order to characterize *nad7* intron-containing transcripts, S1 nuclease protection assays were done according to standard procedures (Sambrook *et al.*, 1989). Approximately 10 µg of RNA was dissolved in 5 µl of ddH_2O and 3 µl PIPES buffer (0.2 M piperazine-N,N'-bis (2 ethanesulfonic acid) disodium salt, pH 6.5, 5 mM EDTA, 2 M NaCl) and vacuum dried. This pellet was then resuspended in 12 µl deionized formamide and 3 µl PCR-labelled probe (section 2.9.3). Samples were then denatured at 85°C for approximately 15 minutes, after which they were slow cooled to 47°C, followed by overnight incubation at this temperature. The DNA/RNA hybrids were then cooled to 43°C and 300 µl of S1 buffer (0.25 M NaCl, 30 mM NaOAc pH 5.5, 1 mM $ZnSO_4$) containing 20 µg/ml carrier salmon sperm DNA was added. The reaction mixture was divided into four tubes containing 0, 50, 200 and 400 units (or 0, 10, 50 and 200 units) of S1 nuclease then incubated at 30°C for 30 minutes. Reactions were stopped with 25 µl of stop solution (4 M ammonium acetate, 50 mM EDTA, 50 µg/ml carrier tRNA) and 240 µl of 95% EtOH. After 1 hour precipitation at -20°C, the protected RNA/DNA hybrids were pelleted by centrifugation for 20 minutes at 4°C, then resuspended in 6 µl of TE. One third of this sample was electrophoresed through a 7% polyacrylamide gel along with a DNA sequencing ladder which was used as a size marker.

2.13 RNA ligation

In an attempt to circularize putative linear forms of excised introns, in order to identify these by RT-PCR, approximately 1 µg of wheat mtRNA from 24 hr embryos was incubated in 1x RNA ligation buffer (50 mM HEPES, pH 8.3, 10 mM $MgCl_2$, 5 mM DTT, 2 mM ATP, 50 µg/ml BSA), 1 unit of RNA ligase and 1 unit of RNAse inhibitor in a total volume of 100 µl. Samples were incubated at 17°C for at least 10 hrs then ethanol precipitated by standard procedures before use in RT-PCR reactions (cf. Section 2.5.2).

2.14 Random cleavage of RNA with Mg^{2+}

Treatment with the divalent cation Mg^{2+} at a slightly alkaline pH was used in order to randomly cleave RNA (Sanger *et al.*, 1979), thereby linearizing excised lariat introns. Approximately 40 µg of wheat mtRNA was pelleted and resuspended in 25 mM glycine pH 9.0 and 5 mM $MgCl_2$ in a total volume of 200 µl. Samples were incubated at 37°C and 40 µl aliquots were removed at various time points (i.e. 0, 30, 90, 150, 210 minutes). The reaction was stopped by adding 100 µl of 95% EtOH and 4 µl of 5 M NaCl and placing the samples on ice. The RNA was pelleted after overnight precipitation then characterized by northern hybridization analyses.

2.15 Specific cleavage of RNA using Ribonuclease H

Ribonuclease H (RNAse H) from *E. coli* is an endoribonuclease that specifically hydrolyzes the phosphodiester bonds of RNA hybridized to DNA. It will not degrade single-stranded nucleic acids, or double-stranded RNA. This enzyme can therefore be used to digest RNA at specific positions following hybridization to oligonucleotides. Approximately 10 µg of RNA was pelleted then resuspended in 12 µl low TE containing 25 ng of oligomer. This mixture was then incubated at 65°C for 5 minutes with vortexing to denature the RNA, followed by incubation at 37°C for 30 minutes for primer annealing. Digestion of hybrids took place in 1x RNAse H buffer (50 mM Tris pH 8.3, 70 mM KCl, 25 mM $MgCl_2$, 250 ng/µl BSA) with 0.5 units of RNase H for 30 minutes at 37°C. The reaction was stopped by the addition of 70 µl TE followed by a phenol:chloroform extraction and ethanol precipitation by standard procedures (Sambrook *et al.*, 1989). Digested RNA was characterized by northern hybridization or by primer extension analyses.

2.16 Primer extension analysis

Primer extension experiments were carried out to map the 5' ends of excised introns RNAs or to determine levels of linear, lariat or circular forms of these transcripts. Approximately 10 µg of mtRNA or RNase H digested mtRNA (cf. Section 2.15) in 5 µl TE was used with 25 ng of ^{32}P end labeled antisense oligomers (cf. 2.9.1) in a total volume of 10 µl. Samples were denatured by incubation at 70°C for 5 minutes then quick cooling on ice. Extension of the hybrids took place in 1x First Strand Buffer (Gibco-BRL) and 10 µM DTT, 0.1 mM dNTP containing 1 unit RNAse inhibitor and 100 units Superscript II RT for 2 hours at 42-49°C. Elevated extension temperatures were used in order to reduce the effects of intrinsic RNA secondary structure on extension. Samples were ethanol precipitated by standard procedures and resuspended in 12 µl Low TE. Approximately one third of the sample was electrophoresed through a 7% denaturing polyacrylamide gel. A DNA sequencing ladder which was extended from the same primer (whenever possible) was used to estimate sizes of extension products.

2.17 Intron secondary structure prediction

Domain 5 and 6 sequences from various plant introns were folded using 'mfold' version 3.1 (Zuker *et al.*, 1999; Mathews *et al.*, 1999), courtesy of Dr. Michael Zuker's homepage (http://mfold2.wustl.edu/~mfold/rna/form1.cgi) and free energies were calculated based on a temperature of 37°C. Domain 5 structures were constrained according to the either the 9 bp + 5 bp (Figure 6.1B) or the 8 bp + 6 bp pair model (Figure 6.1A) Domain 6 structures were constrained to have pairings at the base of the helix (*nad1* intron 1, *nad4* intron 2) or to have a bulging adenosine 7 to 8 nts from the 3' end of the intron. When several structures with similar folding energies were obtained, the structure which was supported by the most plant sequences was shown.

CHAPTER 3:
RNA EDITING STATUS OF WHEAT MITOCHONDRIAL INTRON DOMAINS AND FLANKING EXON SEQUENCES.

3.1 Rationale

In plant mitochondria, C-to-U type editing is a frequent occurrence at non-silent codon positions within protein coding sequences where it changes the amino acid specified to one that improves similarity with homologous proteins from other organisms. In contrast, only a small number of RNA editing sites have been described in plant mitochondrial introns; however, not many intron sequences have been examined at the RNA level. Plant mitochondrial introns have features which deviate from conventional group II structures, including mispairs within helical structures. Does RNA editing play a role in correcting AC mispairs within helical regions of plant mitochondrial group II introns? In some cases, intron editing has been observed to improve helical structure by converting AC mispairs into Watson-Crick AU pairs, whereas in other cases structure is not improved by editing (see section 1.3.3). The model that RNA editing may be required to improve base-pairing to improve secondary structure for splicing (cf. Wissinger *et al.*, 1991) is conceptually appealing as intron editing would then serve an important biological role, as it does in coding regions. The improved structures may be required for excision of the intron from the precursor mRNA or may simply enhance the efficiency of intron excision.

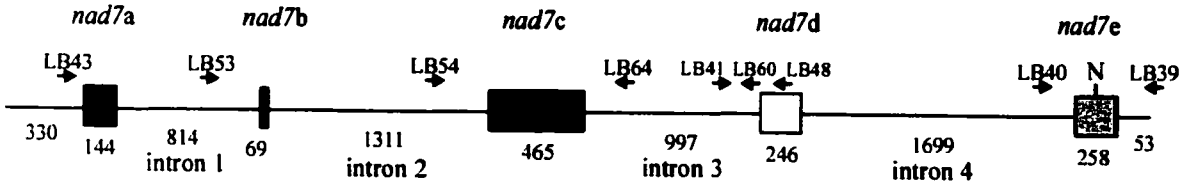
Based on DNA sequence analysis, six AC mispairs are observed in helical regions of D5 and D6 within the four *nad7* group II introns in wheat mitochondria (Bonen *et al.*, 1994; Figure 3.1B, arrows). In the present study, RT-PCR sequencing analysis was used to examine the RNA editing status of these intronic RNAs in wheat embryos and seedlings. I have also looked at exon sequences flanking the introns since these regions are known to be involved in long-range interactions with group II introns through base pairing (Michel and Ferat,1995). This chapter is essentially as published (Carrillo and Bonen, 1997) with the additional figures (3.2A, 3.3, 3.6A&B, 3.7) and additional information regarding editing at splice junctions (section 3.4).

Figure 3.1. Group II introns in the wheat mitochondrial *nad7* gene.

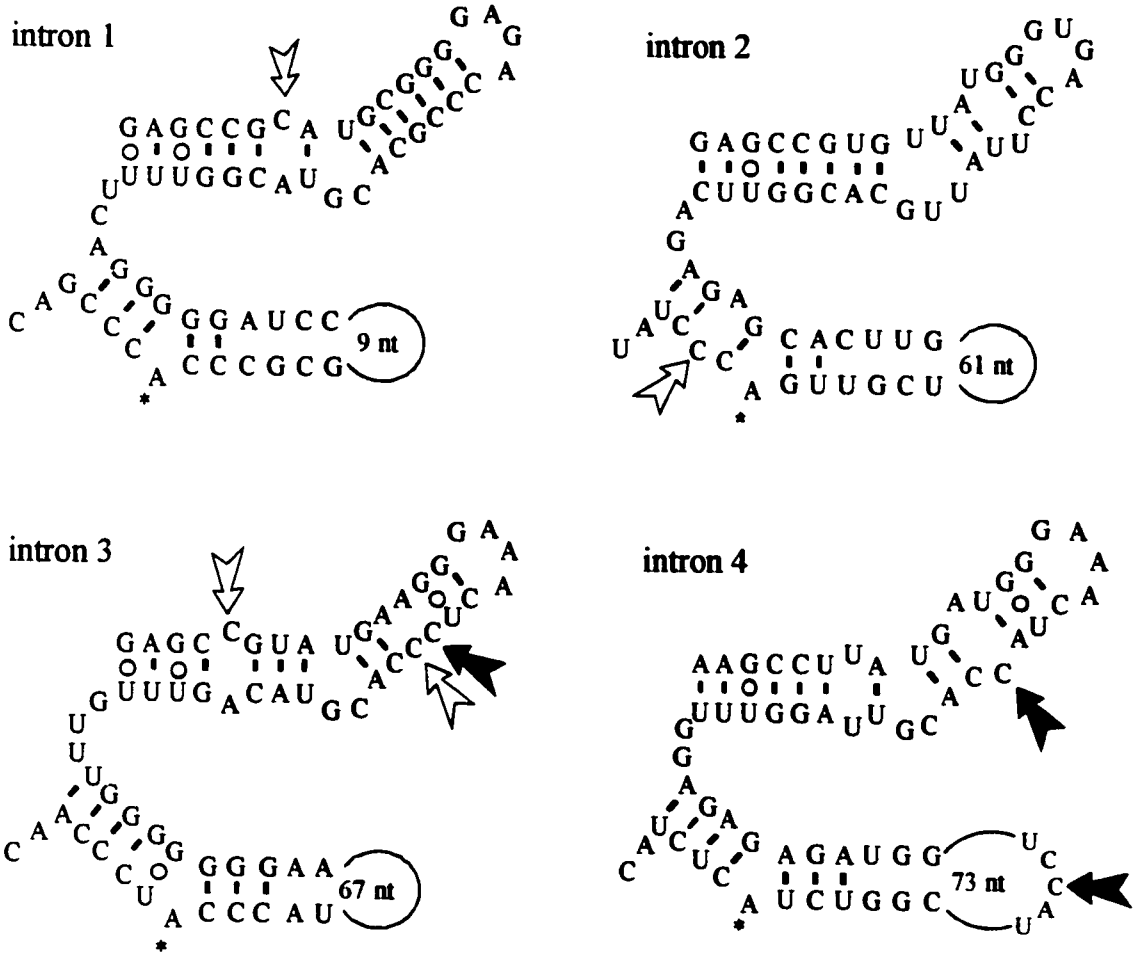
[A] Structure of the wheat *nad7* gene with lengths of exons (boxes) and introns (lines) shown in bp. Numbered arrows represent orientation and approximate locations of primers used in RT-PCR experiments (cf. Section 2.3). The NheI restriction site (N) shown within exon e is not present in DNA but is created in cDNA due to RNA editing.

[B] Secondary structural models of the domain 5 and 6 sequences of the wheat *nad7* introns 1-4 (updated from Bonen *et al.*, 1994). The predicted RNA editing sites within the six AC mispairs are shown by arrows and the RNA editing sites which were experimentally observed are depicted by solid red arrows. Bulging adenosines within domain 6 are shown by asterisks.

A.


nad7a
nad7b
nad7c
nad7d
nad7e
LB43
LB53
LB54
LB64
LB41
LB60
LB48
LB40
N
LB39
330
144
814
intron 1
69
1311
intron 2
465
997
intron 3
246
1699
intron 4
258
53

B.


intron 1
9 nt
intron 2
61 nt
intron 3
67 nt
intron 4
73 nt

3.2 Developmental differences in levels of *nad7* intron-containing transcripts

Prior to carrying out the RNA editing analysis, levels of intron-containing transcripts were determined by northern analyses using wheat mitochondrial RNA from 24-hr germinating embryos and 6-day etiolated seedlings. The gel loadings were standardized for levels of mitochondrial ribosomal RNA (Figure 3.2B, 18S rRNA) and blots were hybridized with M13 probes (Figure 3.2A, introns 1 and 2, exon e, red lines) or oligomer probes (introns 3 and 4, arrows). Markedly higher levels of intron-containing precursor transcripts and excised intron RNAs were observed in 24-hr germinating wheat embryos than in 6-day etiolated seedlings (Figure 3.2B), relative to rather equivalent levels of the 1.6 kb *nad7* mRNA (Figure 3.2B, *nad7*e). In addition, the RNA profiles from these two stages show qualitative differences, most notably in the region of excised intron 4 RNA species (Figure 3.2B, intron 4, asterisk). These differences will be discussed further in Chapter 5. There are also differences between the introns in the complexity and abundance of precursors and in the levels of excised introns relative to precursors. For example, excised intron is the major species of intron 3, whereas the levels of the approximately seven intron 4-containing transcripts are relatively equal. Similarly, the levels of intron 1 and 2 containing precursors are high relative to excised introns. Perhaps the less complex *nad7* intron 3 profile indicates that this intron is excised early relative to the other three introns.

To assess the levels of intron 3-containing precursors and excised intron RNA relative to mRNA, I performed S1 nuclease analysis using a probe derived from a region of *nad7* which included 103 nt of exon c and 85 nt of intron 3 (Figure 3.3A). Protected fragments representing input probe (Figure 3.3B, 235 nt), intron 3-containing precursor RNAs (188 nt), mature messenger RNA (103 nt) and excised intron (85 nt) were detected in 24 hour embryos; whereas the major signal in 6 day seedlings corresponded to the mature mRNA. The signal from the excised intron (85 nt) was very weak since the protected fragment was 45 % shorter than the protected precursor RNA fragment (188 nt) and would therefore be expected to have a correspondingly lower intensity if the levels of precursor were equivalent to those of the excised intron, assuming that the nucleic acid sequence contained roughly 25% of adenosine. Additionally, binding of the probe to excised introns may be somewhat inhibited by the higher order structure of the excised intron so that the

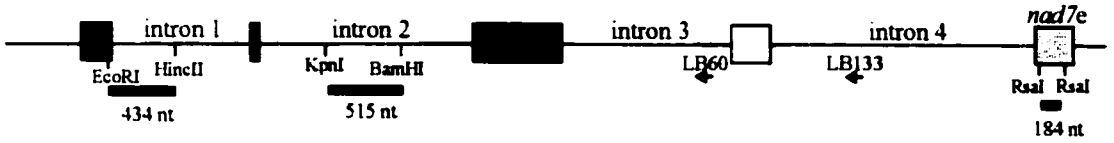
Figure 3.2. Northern blot analysis of wheat mitochondrial *nad7* transcripts using RNA isolated from 24 hr wheat embryos and 6-day etiolated seedlings.

[A] Schematic of the wheat *nad7* gene showing positions of M13 probes (red lines) and oligomer probes (black arrows). Restriction sites used for cloning into M13 vectors are shown.

[B] RNA blots were hybridized with wheat mitochondrial probes for 18S rRNA, *nad7* introns 1, 2, 3 and 4, and *nad7* exon e. Asterisks indicate the positions of excised intron RNAs. Arrows indicate the positions of the 26S and 18S mitochondrial ribosomal RNAs and bars denote the positions of RNA size markers in kb (Gibco-BRL).

Adapted from Carrillo and Bonen (1997).

A.


intron 1
intron 2
intron 3
intron 4
nad7e
EcoRI
HincII
KpnI
BamHI
LB60
LB133
RsaI
RsaI
434 nt
515 nt
184 nt

B.

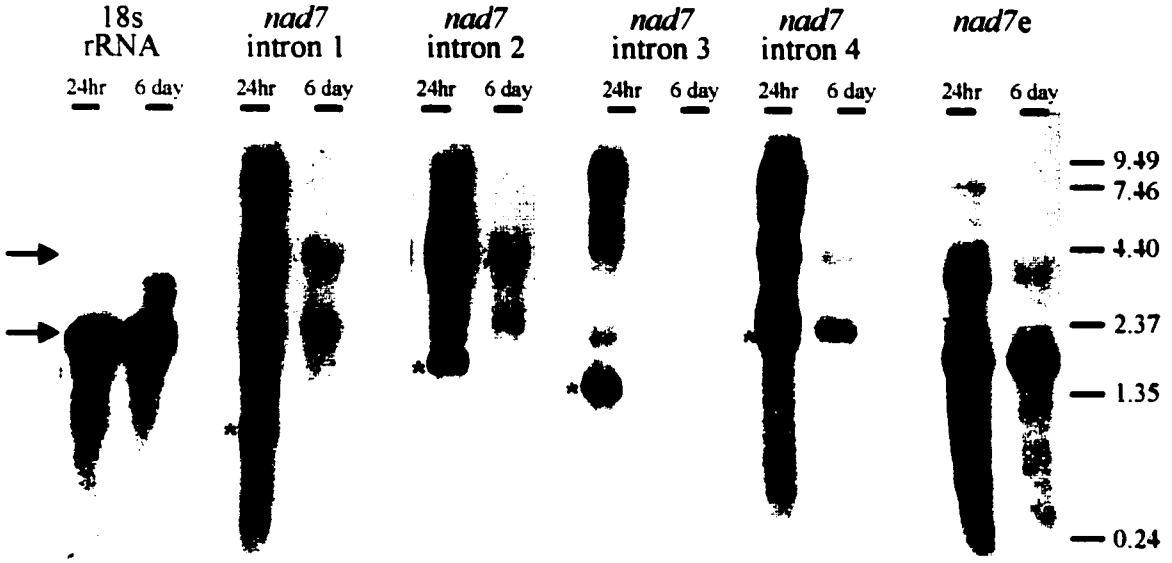
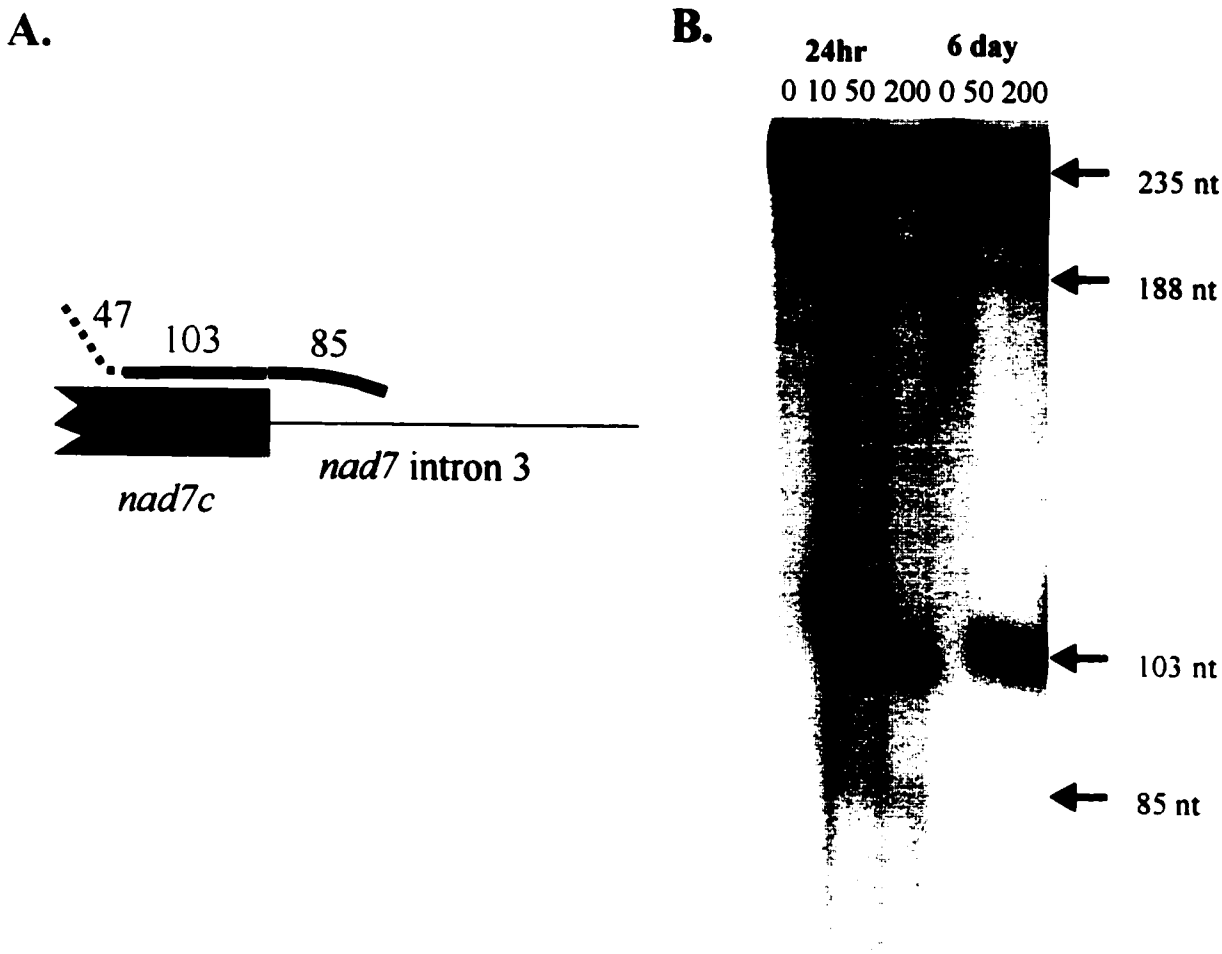

18s rRNA
nad7 intron 1
nad7 intron 2
nad7 intron 3
nad7 intron 4
nad7e
24hr
6 day
9.49
7.46
4.40
2.37
1.35
0.24

Figure 3.3. S1 nuclease analysis of *nad7* intron 3-containing transcripts.

[A] Schematic of the region used for S1 nuclease protection generated by PCR using the forward sequencing primer and primer LB64 (Figure 3.2A) with an M13 clone containing an *nad7* HincII-NarI fragment as template. Red represents the exon portion of the probe (103 nt) and blue represents the intron portion of the probe (85 nt). The dotted line represents M13 sequence (47 nt).

[B] Autoradiograph of protected products from the S1 nuclease protection assay run on a denaturing polyacrylamide gel. Lanes are numbered according to the number of units of nuclease used in each reaction (0, 10, 50, 100). Numbered arrows indicate sizes of products which were deduced based on a sequencing ladder run beside the S1 products.


A.
47
103
85
nad7c
nad7 intron 3
B.
24hr
6 day
0 10 50 200 0 50 200
235 nt
188 nt
103 nt
85 nt

level of excised intron is under-represented in this experiment. This experiment confirms that for *nad7* intron 3 the levels of intron-containing precursor RNAs were much lower in 6 day seedlings than in 24 hour embryos relative to levels of mRNA.

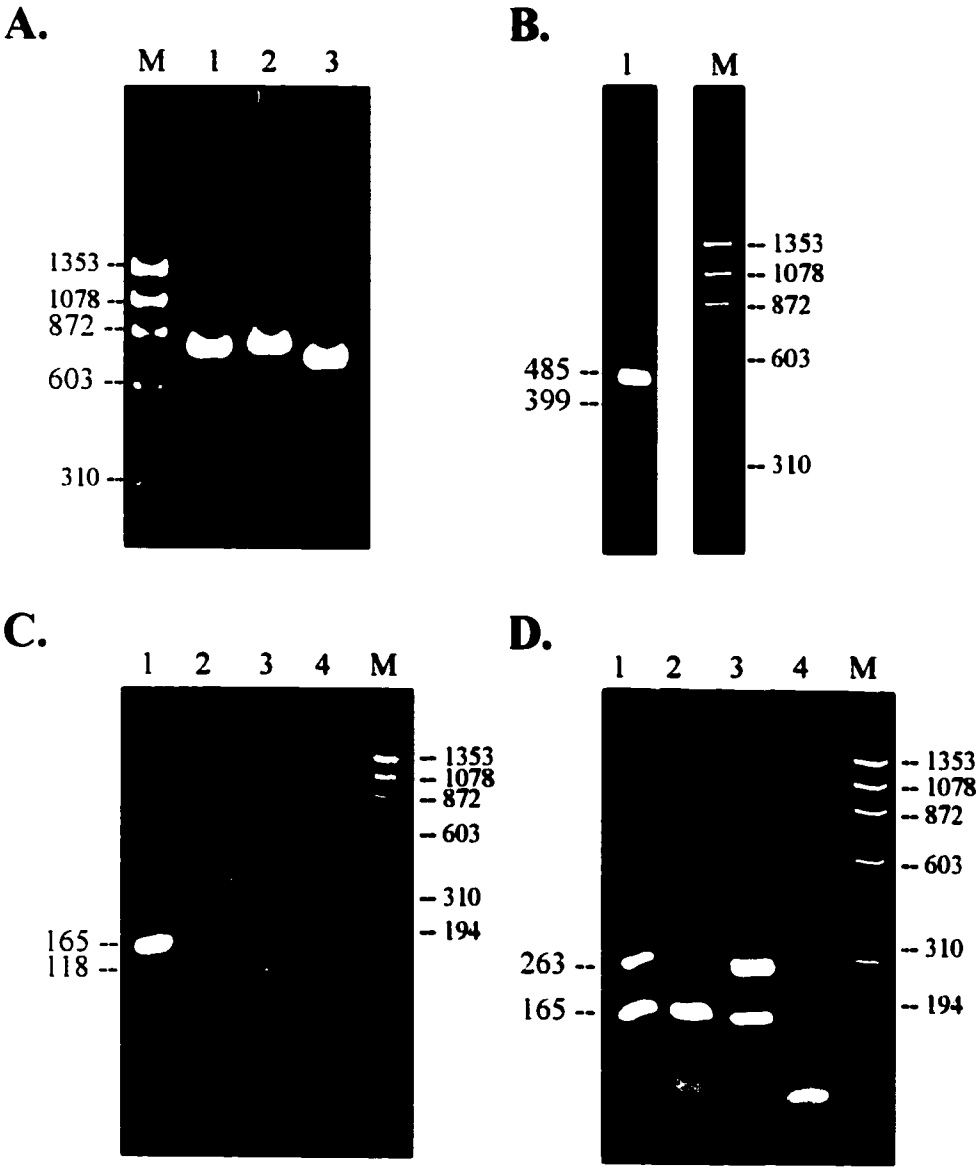
The developmentally-specific differences in the abundance and apparent complexity of *nad7* intronic transcripts may reflect differential RNA stabilities or RNA processing pathways. If splicing is involved in regulating transcript levels, this raises the possibility of a role for intron editing in regulation if such events are important to improve folding for splicing as has been proposed (Schuster and Brennicke, 1994; Wissinger *et al.*, 1992). The editing status of wheat *nad7* intronic mtRNAs was examined in both 24 hr and 6 day stages of development to determine if differences in levels of editing were observed.

3.3 Editing status of the domain 5 and 6 regions of wheat *nad7* introns

Based on their DNA sequence, the wheat *nad7* introns have weaker base pairing within the domain 5/6 helical regions (Figure 3.1B) than do conventional group II introns. There are a total of six AC mispairs in D5 and D6 helices (Figure 3.1B, arrows), which could be converted to AU pairs by RNA editing. Of particular note are two adjacent ones in intron 3. It has been estimated, based on NMR spectroscopy, that an AC mispair within a hairpin is about +2 kcal/mol less stable than an AU pair (Puglisi *et al.*,1990).

To determine the RNA sequences of these intronic regions, an RT-PCR method was used. A primer mapping downstream of the terminal *nad7* exon (primer LB39, Figure 3.1A) was used to generate cDNA from wheat RNA template from 24 hr embryos or 6 day seedlings and then partially-spliced cDNAs were PCR-amplified using LB48 (or LB39) and an additional oligomer located upstream of the respective D5 sequences of introns 1 (LB53), 2 (LB54), and 3 (LB41), as shown in Figure 3.1A. The observed products, in which all introns except the one of interest had been excised, were 773, 787, 715 bp in length, respectively (Figure 3.4A). Larger RT-PCR products in which no intron or only one intron was excised were not detected on agarose gels, however Southern hybridization experiments demonstrated their presence (data not shown). In the case of the intron 4 amplification product of 485 bp, an NheI restriction site within exon e (Figure 3.1A) which was created by RNA editing enabled the separation of products derived from edited cDNA molecules (Figure 3.4B, 399 bp) from those derived from unedited precursor RNA or residual DNA

Figure 3.4. RT-PCR amplification products from wheat embryo *nad7* RNA template.
[A] RT-PCR products from partially spliced *nad7* template generated using primer pairs LB53 and LB48 (lane 1), LB54 and LB48 (lane 2) and LB41 and LB39 (lane 3). See Figure 3.1A for primer locations.
[B] Restriction profile of RT-PCR product generated from primer pair, LB40 and LB39, after digestion with NheI, which converts the 485 bp product into 399 bp + 86 bp ones derived from RNA-edited template.
[C] RT-PCR product from excised intron 3 RNA template using LB41 and LB60 to give a product of 160 bp. The PCR reaction included a second set of primers, LB115 and LB116, (Chapdelaine and Bonen, 1991) which yield a product of 118 bp from *nad1*a, to assess residual DNA contamination. The templates were unfractionated mtRNA (lane 1), gel-purified excised intron 3 (lane 2), wheat mtDNA (lane 3) and a negative control with no DNA template (lane 4). [D] RT-PCR products generated from gel-fractionated RNA template using primers LB41 and LB48 (263 bp product) as well as primers LB41 and LB60 (165 bp product). The templates were gel-purified RNA from the high MW precursor fraction (lane 1), gel-purified RNA from the excised intron 3 fraction (lane 2), wheat mt DNA (lane 3) and a negative control with no template (lane 4) .
[A, B, C & D] Size markers (M) were ΦX174 DNA restricted with HaeIII. Expected sizes of products are shown in red. Adapted from Carrillo and Bonen (1997).


A.
M 1 2 3
1353
1078
872
603
310
B.
1 M
485
399
1353
1078
872
603
310
C.
1 2 3 4 M
165
118
1353
1078
872
603
310
194
D.
1 2 3 4 M
263
165
1353
1078
872
603
310
194

(485 bp) and this population was examined further. RT-PCR products were either cloned into M13 vectors for sequencing or were sequenced directly. It should be noted that similar results were later observed using RT-PCR products from *nad7* introns 3 and 4 generated exclusively from excised lariat introns (cf. Section 4.2).

For introns 1 and 2, no editing was seen within D5 or D6 sequences by direct sequencing of RT-PCR products, despite apparent full editing in downstream exon positions. In addition, none of the 10 clones of RT-PCR products examined for each of these introns was edited within the intron sequence. D5 and D6 of intron 3 RNA differed from the DNA sequence at only one position, namely C22 was changed to U22 (Figure 3.1B, red arrow), so that an AC mispair was converted to an AU pair. At this site, editing was seen in 8 out of 10 clones examined for wheat embryo RNA, and 6 out of 10 clones for seedling RNA (Figure 3.6A, red dots). In contrast, at the positions of the other two AC mismatches (Fig. 3.1B open arrows), none of the 20 clones showed editing. In the case of intron 4, editing was seen at the position homologous to the edited one in D5 of intron 3 (Figure 3.1B, red arrow) and also at a site within the loop of D6, namely 42 nucleotides upstream of the 3' splice site (Figure 3.6C, red arrow). In wheat embryo RNA, these two intron 4 sites showed editing in 11/12 and 10/12 clones examined, compared to 8/10 and 7/10 clones, respectively when seedling RNA was used as template (Figure 3.6B, red dots). Thus, the degree of editing is similar in the two developmental stages, but slightly higher in embryos than in seedlings. With respect to editing within *nad7* coding sequences contained in the RT-PCR products of partially-spliced templates, generally exon sites examined showed complete or nearly-complete editing (ranging from 64% to 90% in the clones examined, Figure 3.6A and B). It should be noted that in wheat *nad7* mRNA, exon editing positions appear to be fully edited except for four positions at silent sites which are edited in approximately 50% of the population of mRNA (Figure 3.7A, red bars vs. hatched red bars; Bonen *et al.*, 1994).

Because the partially-spliced transcripts showed incomplete intron editing at the positions of two AC mispairs and none at the other four predicted sites, I considered the possibility that editing is a rate-limiting step for splicing, and that once editing occurs, the correct structure enables rapid splicing. If so, the population of excised intron RNAs would be expected to be fully edited at those sites. To address this issue, gel-fractionated and

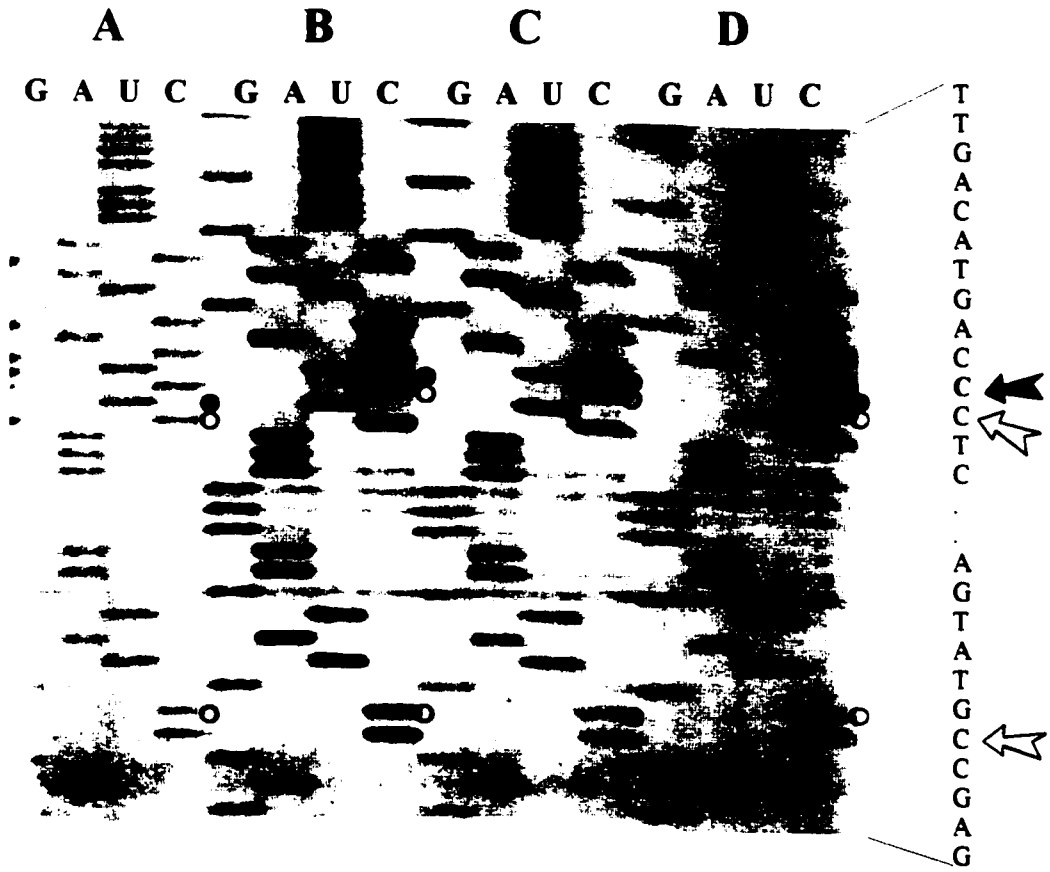
purified excised intron RNA was used as template in RT-PCR sequencing experiments. RNA from 24-hr embryos was selected because steady state levels of *nad7* intron-containing transcripts are much higher than in seedlings (Figure 3.2B). I concentrated on intron 3 because the excised form is the major intron 3 RNA species in embryos and this intron has three AC mispairs within D5, notably two adjacent ones. RT-PCR amplification using intron-specific primers LB41 and LB60 (Figure 3.1A) generated an amplification product of 165 bp (Figure 3.4C). To assess the amount of residual contaminating DNA, the PCR amplification step included an additional set of primers (LB115 and LB116) which generated a 118 bp product from *nad1a* template DNA (Chapdelaine and Bonen,1991). The levels of contaminating DNA were seen to be reduced, but not completely eliminated when the intron was gel purified (Figure 3.4C, lane 2). To ensure that the excised intron 3 fraction was not contaminated with other intron 3-containing transcripts, cDNAs were synthesized using LB48 and LB60 prior to PCR amplification with LB41. No products (263 bp in length) derived from intron 3/*nad7*d-containing transcripts were observed in the excised intron fraction (Figure 3.4D, lane 2), in contrast to their presence in the gel-purified precursor RNA fraction (lane 1) and in mtDNA (lane 3). Direct sequencing of the *nad7* intron 3 amplification products indicated that only the same single site shows editing (Figure 3.5C, red arrow) and that as for the partially-spliced template (Figure 3.5A) and unfractionated RNA template (Figure 3.5B), it appears to be incomplete in the population of RNA molecules. No editing was detected at the other two positions (Fig. 3.5C, open blue arrows).

In summary, C-to-U type RNA editing was observed at only three sites within the D5/D6 regions of the four wheat *nad7* introns. One occurs within the unmodelled D6 loop in intron 4, and the other two convert AC mismatches to AU pairs at identical positions within the D5 helix of introns 3 and 4. This was the first report of RNA editing within D5 of any plant mitochondrial intron. Although these two events improve helicity and so may be important for splicing, they do not generate the structure that is invariably present in this catalytic domain of ribozymic group II introns (Perlman and Podar, 1996). Moreover, four other AC mispairs within these helices show no editing, so Watson-Crick base-pairing at those sites must not be required for proper folding for splicing.

In the case of intron 3, the single editing event shifts the calculated free energy (cf.

Figure 3.5. RNA editing status of domain 5 of wheat *nad7* intron 3.

[A-D] Direct DNA sequencing of RT-PCR products generated from *nad7* intron 3 template from wheat embryos using primer LB41 for sequencing. Red arrow indicates the observed RNA editing site whereas open blue arrows indicate predicted but unedited sites (cf. Figure 3.1B, intron 3). Sequences of RT-PCR products are from [A] partially-spliced template (LB41 and LB39; Figure 3.4A, lane 3), [B] unfractionated RNA (LB41 and LB60; Figure 3.4C, lane 1), [C] gel-fractionated excised intron (LB41 and LB60; Figure3.4C, lane 2) and [D] wheat mitochondrial DNA (LB41 and LB60; Figure 3.4C, lane 3). Adapted from Carrillo and Bonen (1997).


A
B
C
D
G A U C G A U C G A U C G A U C
T T G A C A T G A C C C T C . . A G T A T G C C G A G

Section 2.17; Zuker *et al.*, 1999; Mathews *et al.*, 1999) of D5 from -6.3 to -8.8 kcal/mol, which is still markedly less stable than those of the ribozymic introns bI1 and aI5γ, which have values of -14.9 to -19.8 kcal/mol respectively. Editing at the other two AC mismatches (Figure 3.1B, open arrows) would have resulted in a free energy of -16.2 kcal/mol. Interestingly, this intron which has such a weak D5 structure in wheat mitochondria, is absent from the *nad7* gene of tobacco although the other three are present (Pla *et al.*,1995). For intron 4, the edited D5 helix is considerably more stable (free energy of -14.2 kcal/mol) than the unedited form (-9.6 kcal/mol). The D5 sequences of introns 1 and 2, which showed no editing, have much more stable helices, with free energies of -20.7 and -18.5 kcal/mol, respectively. Note that these structures are more stable than those of the self splicing introns despite the presence of a mispair in D5 in both introns. The higher stability is due to a higher number of GC pairs within these structures.

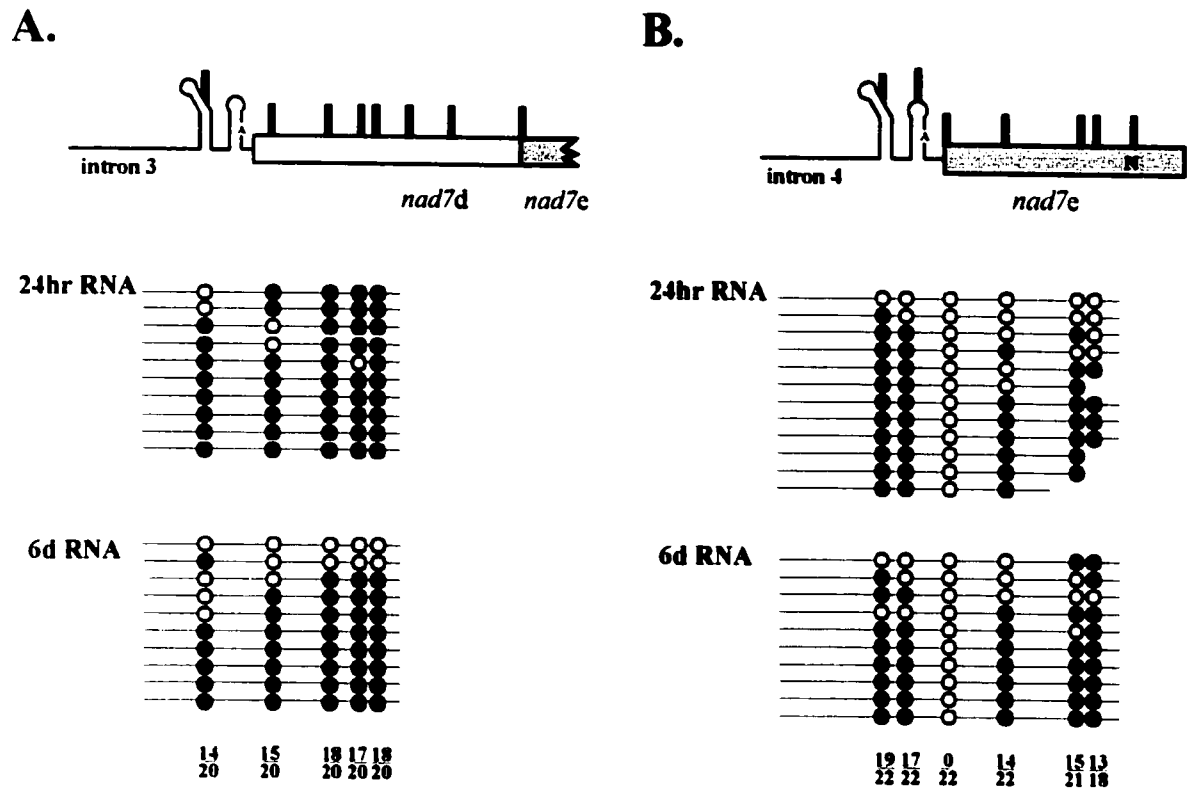
Similar patterns of intron editing were observed in 24-hr wheat embryos and 6-day seedlings, with the degree of editing being slightly higher in embryos. Thus, there appears to be no obvious correlation between editing status and the markedly different profiles of *nad7* intron-containing transcripts seen between these two stages of development. Differences in the completeness of RNA editing have been reported for maize *nad3* transcripts, with an increased level seen in 7-day seedlings compared to 3-day seedlings (Grosskopf and Mulligan,1996). Although it is not known what confers editing site selection in plant mitochondria, it is possible that the editing observed within the D5 helix, designated as base-pair position #11 (cf. Figure 6.1A) in introns 3 and 4, involves an internal guide RNA (discussed in section 6.3)

3.4 Editing at splice junctions in precursor and fully-spliced *nad7* and *nad4* RNA

RNA editing is typically an early post-transcriptional event in plant mitochondria (Mulligan *et al.*, 1999) so that transcripts in which at least one intron has been excised would be expected to be edited. However, editing does not appear to be an early event for exon sequences near splice junctions. The RT-PCR product used in the analysis of intron 4 (see above) contained four *nad7*e editing sites (Bonen *et al.*, 1994). One of these, located three nucleotides downstream from the intron-exon border appeared to be unedited as judged by direct sequence analysis (Figure 3.6C, open blue arrow) in contrast to its full editing in

Figure 3.6. RNA editing of wheat *nad7* intron versus exon sequences near splice sites. [A & B] Schematics representing editing status of *nad7* intron 3 [A] and intron 4 [B] RT-PCR products generated from 24hr and 6 day incompletely spliced mtRNAs and cloned into M13 vectors for sequencing. Red bars represent exon and intron editing positions. Red dots indicate positions within each clone which are edited and open dots indicate positions which are not edited. [B] Blue dots indicate an *nad7e* editing site which is not edited in intron 4-containing precursors.

[C& D] Direct cDNA sequencing of the *nad7* exon e region near the splice junction. [C] RT-PCR products generated from pre-mRNA (primers LB40 and LB39) restricted with NheI (Figure 3.3B, lane 1) and [D] RT-PCR products representing mature *nad7* mRNA (primers LB43 and LB39). Sequencing primers were primer LB40 for [C] and an internal primer mapping to *nad7*c (LB47) for [D]. Red arrows indicate observed C-to-U edits, whereas open blue arrow indicates an exon position not edited in precursor transcripts. Editing sites within the upstream intron 4 [C] or exon d [D] are also shown with red arrows. Adapted from Carrillo and Bonen (1997).


A.
intron 3
nad7d
nad7e
24hr RNA
6d RNA
14 15 18 17 18
20 20 20 20 20
B.
intron 4
nad7e
24hr RNA
6d RNA
19 17 0 14 15 13
22 22 22 22 21 18
C.
G A U C
u c u u u c c u a u a
34
c u c u a c
Intron 4
nad7e
C U C A A U U
D.
G A U C
A A U G U C U U A A U
79
U G G A A U C
nad7d
nad7e
C U U A A U

completely-spliced *nad7* mRNA (Fig. 3.6D, lower red arrow) and the nearly-complete editing seen at the intronic site within the loop of D6 (Figure 3.6C, red arrow). Moreover, none of the 22 clones examined from embryo or seedling RNAs showed editing at this position (Figure 3.6B, open blue dots). In contrast, the other three exon sites showed greater than 50% editing in these RNAs, which would include early unspliced precursors, but all of which were known to have the Nhe I site generated by editing (Figure 3.6B, red dots).

To determine if this temporal relationship between editing and splicing was observed at other editing positions close to splice junctions, I examined the editing status of other such non-silent exon editing sites in *nad7* and *nad4*. Exon c of the wheat *nad7* gene which has an editing position at the last nucleotide (Figure 3.7A, blue bar) was examined by direct sequencing of RT-PCR products generated using primers LB64 and LB43 (Figure 3.1A). Editing at this position was not detected by direct sequencing, compared to apparent full editing at an exon position 297 nt upstream of this site (Figure 3.7B, open blue arrow vs. red arrow). In the *nad7* gene there was also a first exon editing position 7 nts upstream from the splice junction (Figure 3.7A, box). However, selecting for RT-PCR products derived from RNA rather than DNA was not as easily achieved since there was no upstream intron and there was no restriction site created by editing. Direct sequences of RT-PCR products generated from RNA template which had been DNAse treated appeared not to be edited at any of the exon editing sites, indicating that the PCR products were derived from DNA or newly synthesized transcripts (data not shown). I have also examined editing of the first nucleotide of *nad4* exon c in incompletely-spliced RT-PCR products (LB90 and LB93). As observed for *nad7*c and *nad7*d exon sites, editing was not apparent by direct sequencing, compared to full editing of downstream exon d editing positions (Figure 3.7C). *Nad4d* was examined since there are no editing sites in *nad4*c other than at the first nucleotide position. These observations are consistent with editing being a relatively "late" RNA processing event at exon sites close to splice junctions such that editing does not occur until after splicing, or that it is the rate limiting step which triggers splicing.

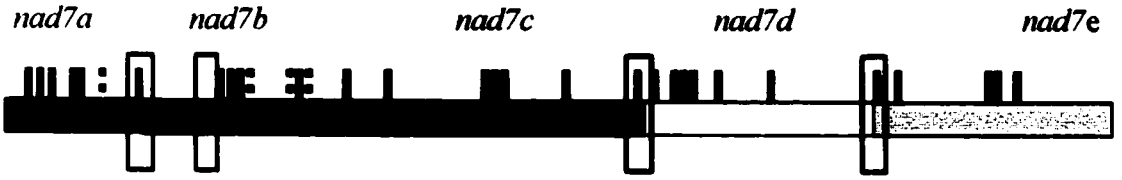
An interesting observation about *nad4* is that exon c has only one editing position compared to *nad4a* and *nad4d* where 21 editing sites are scattered throughout the exons (Lamattina and Grienenberger, 1991). This appears to have occurred due to gene conversion

Figure 3.7. RNA editing near splice junctions in wheat *nad7* and *nad4*.

[A] Schematic of *nad7* gene showing exon editing positions which are partially edited (hatched red bar) and fully edited (red bar) in mature mRNA. Exon editing sites within 10 nt of the splice junction are boxed. Blue bars represent exon editing positions which are unedited in intron-containing transcripts but fully edited in mRNA.

[B & C] Direct sequencing of *nad7* intron 3 containing RT-PCR products generated with primers LB46 and LB64 [B] or *nad4* intron 2 containing RT-PCR products generated with primers LB90 and LB93 [C]. Red arrows indicate positions which are edited and open blue arrows indicate exon positions not edited in precursor transcripts but fully edited in mRNA.

A.

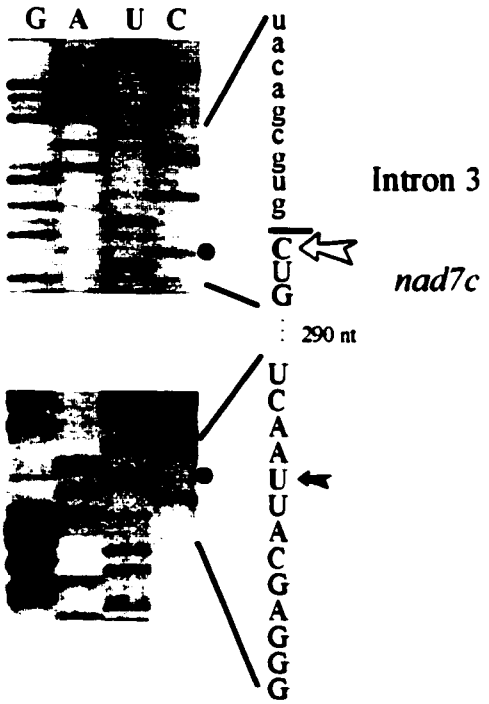

nad7a
nad7b
nad7c
nad7d
nad7e

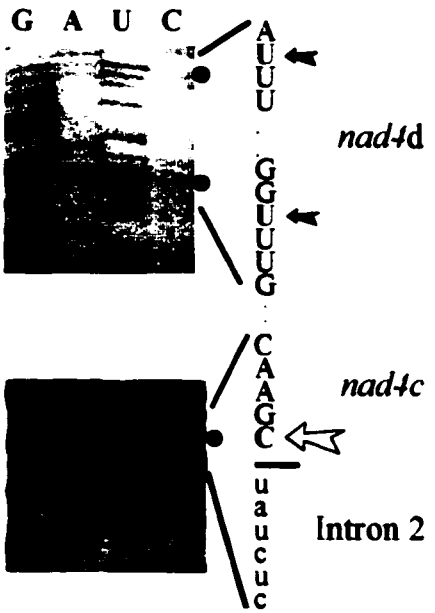
B.


nad7b
nad7c

C.


nad4c
nad4d


G A U C
uacagcgug
CUG
Intron 3
nad7c
290 nt
UCAAUUACGAGGG


G A U C
AUUU
GGUUUG
nad4d
CAAGC
nad4c
uaucuc
Intron 2

with an edited copy of this exon. However, the one editing position remaining is the first nucleotide of *nad4c* suggesting that the template for gene conversion was an intron 2-*nad4c* RNA transcript which was not edited due to the requirement for editing before splicing Alternatively, the editing site may have been gained after gene conversion.

The observation of exon positions which remain unedited in the population of intron-containing RNA molecules was unexpected since editing is normally an early post-transcriptional event and for *nad7* intron 3 and *nad4* intron 2 at least one intron had already been excised from within transcripts. This finding is interesting since exon sequences which flank group II introns are known to base pair with nucleotides within the intron in long range interactions required for splicing in ribozymic introns (cf. Section 1.1.5.2). The editing sites I have examined here fall into three categories: the editing position falls within the IBS1 (*nad7*c) which is a short region (5-7 nt) that complements the EBS1 within D1 of the intron, the editing site is at the first position of the 3' exon (*nad4*c) which is involved in the δ - δ' interaction with a single nucleotide flanking the EBS1 sequence and the editing position (*nad7*e) is outside any regions which are known to interact with the introns (Michel and Ferat, 1995). For *nad7c* and *nad4c*, I was unable to determine the intronic positions of the EBS1 or the δ elements so the potential effect of editing at these positions for splicing is unknown. In *Acorus calamus*, editing at a *coxII* exon 1 position 4 nt upstream from the splice junction is within the IBS1 site and in this case, editing appears to correct an AC mispair between the IBS1 and the EBS1 interactions (Albertazzi *et al.*, 1998). In contrast, editing at the same *coxII* position in *Petunia* changes a GC pair to a weaker GU pair within this interaction. Thus editing at either the *nad7*c or the *nad4*c sites could potentially affect splicing, whereas the editing status of the *nad7*e position may play a role in splicing but may be more likely to have implications for RNA editing which will be discussed in Section 6.1.1.

CHAPTER 4:
VARIATION IN SEQUENCE AND EDITING WITHIN CORE DOMAINS OF MITOCHONDRIAL GROUP II INTRONS AMONG PLANTS

4.1 Rationale

Mitochondrial genes in flowering plants are known to accumulate nucleotide substitutions more slowly than do chloroplast or nuclear genes, and for the most part intron sequences are very highly conserved as well (Yang *et al.*, 1999; Learn *et al.*, 1992). The latter is attributed in part to the distinctive folding constraints of group II introns (reviewed in Michel and Ferat, 1995) which are presumably required to maintain the ability to catalyze their splicing reaction. In plant mitochondrial introns, domains 5 and 6, normally among the most highly conserved structures in group II introns, have features which differ from conventional group II intron structures (cf. Section 6.2). For example, some introns have mispairings within helical regions, including some AC mispairs which are, in a small number of cases, corrected by RNA editing (i.e., *nad7* introns 3 and 4, cf. Chapter 3). However, other introns exhibit much more unusual features, including the absence of the D6 branch site normally required for splicing (i.e., *nad1* intron 1 and *nad4* intron 2). The question arises as to whether such aberrant introns are actively under functional constraint. To assess the conservation of RNA editing and structural features, D5 and D6 regions from various introns were examined in some closely related plants (wheat and rice within monocots; soybean and pea within dicots) as well as some more distantly related dicots (*Arabidopsis* and tobacco).

4.2 Variation in RNA editing status among plants at homologous domain 5 sites of *nad7* introns

In the analysis of the four introns in the wheat *nad7* gene (Carrillo and Bonen, 1997), RNA editing was observed at only two of the six AC mispairs within domain 5/6 core helices and in both cases these were within D5 at position C22 of introns 3 and 4. D5 is typically very highly conserved, so it might be expected that if such editing were important for correct intron folding for splicing, the comparable positions of these two *nad7* introns in other flowering plants would either show editing or have a genomically-encoded thymidine.

The D5/D6 sequences of *nad7* introns 3 and 4 from wheat, rice, *Arabidopsis*, pea, soybean and tobacco were obtained using an RT-PCR method designed so that amplification

products are derived exclusively from excised intron RNA template (Vogel *et al.*, 1997A) rather than from partially-spliced precursor transcripts as in Chapter 3. In this method, cDNA was extended from a primer mapping to the 5' end of the intron (Figure 4.1A, grey arrow) with Superscript II reverse transcriptase which is capable of reading through the 2'-5' bond at the branch site. Two primers mapping to the 5' and 3' ends of the intron (Figure 4.1 A, black arrows) were then used to amplify the region across the branch site. Note that linear excised intron template would not be amplified by this method. RT-PCR products were typically examined by direct sequencing; however, wheat products for *nad7* introns 3 and 4 were also cloned into plasmid (pGEM) vectors and individual clones were sequenced.

Intron editing was observed in D5 in excised introns of both *nad7* intron 3 (Figure 4.1B, red arrow) and intron 4 (data not shown) from wheat mitochondria confirming observations in which partially-spliced RNA template or gel-fractionated excised intron RNA had been used in RT-PCR experiments (Carrillo and Bonen,1997). In all cases, products which were consistent with a lariat form of excised intron for *nad7* intron 3 were observed in wheat mtRNA from 24 hr embryos (Figure 4.1C) and from 6 day etiolated seedlings (data not shown). The junction between the extreme 5' end of domain 1 and the bulging adenosine of D6 is shown for *nad7* intron 3 (Figure 4.1C, green arrow). It should be noted that the misincorporation of a thymidine instead of adenosine was observed at the nucleotide position corresponding to the branch point of these cDNA products (Figure 4.1B, green arrow), as was also seen by Vogel *et al.* (1997A) for chloroplast introns.

In the case of *nad7* intron 4, the entire DNA sequence of D5 except for positions #6 and #22, is identical among all the plants examined (Figure 4.2A). In rice, *Arabidopsis* (Unseld *et al.*, 1997; this work) and tobacco (Pla *et al.*, 1995; this work), there is a cytosine at position #22 (as in wheat) compared to thymidine in soybean and pea. This site C22 was seen to be edited to U22 in rice and wheat excised introns (Figure 4.2B, red dot), but it remained unedited in tobacco (open blue dot) and an approximately equal mixture of edited and unedited forms was seen in the excised intron RNA population of *Arabidopsis* (purple dot). Clearly, editing at this position cannot be essential for splicing in tobacco or *Arabidopsis*. At position #6 of D5, editing was observed in soybean, pea and tobacco (Figure 2B, red dots) to result in conventional AU pairing at the RNA level as in *Arabidopsis*, rice

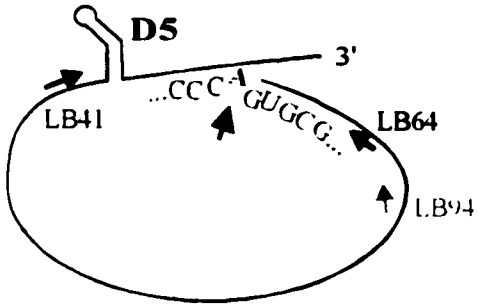
Figure 4.1. RNA editing in excised *nad7* intron 3 in wheat.

[A] Schematic of the strategy used for amplifying RT-PCR products from excised *nad7* intron 3. Black and grey arrows indicate primers used for RT-PCR (see text) and the green arrow indicates the branch site. Domain 1 sequence is shown in pink and D6 sequence is shown in blue.

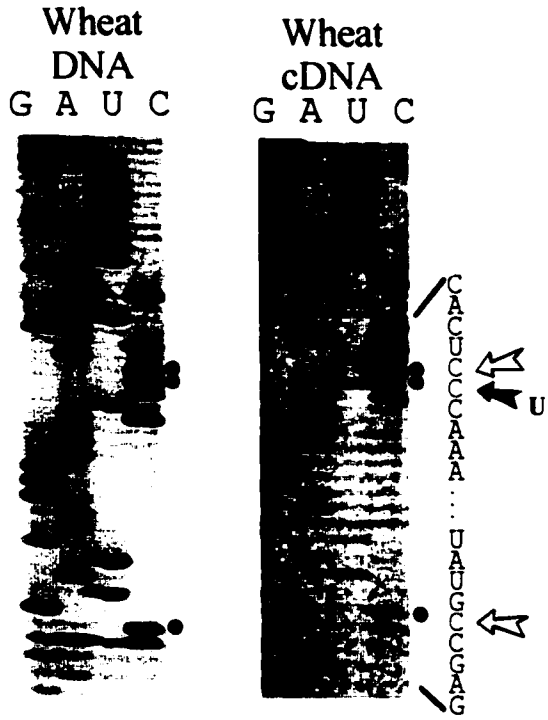
[B] RNA editing in excised *nad7* intron 3 in cDNA from 24 hr wheat mtRNA compared to wheat mtDNA sequence. The RNA editing position is indicated by a red arrow and AC mispairs which were not edited are indicated by open blue arrows.

[C] Direct sequencing of RT-PCR products (Primers LB87 and LB64) of *nad7* intron 3 generated using 24 hr wheat mtRNA showing the lariat branch site (green arrow)

A.


D5
3'
...CCCA
GUGCG...
LB41
LB64
LB94

B.


Wheat
DNA
G A U C
Wheat
cDNA
G A U C
CACUCCCAAA...UAUGCCGAG
U

C.

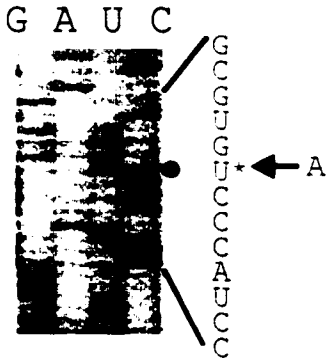
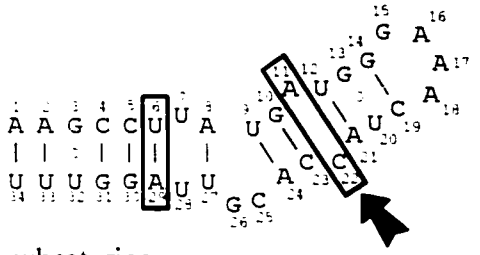

G A U C
GCGUGUCCCAUCC
A

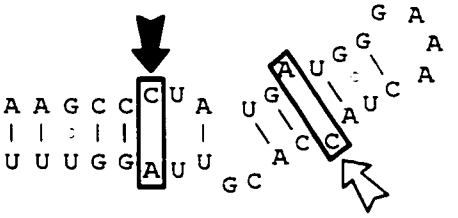
Figure 4.2. Variation in RNA editing status within domain 5 of *nad7* intron 4 among plants. [A] RNA secondary structural models of domain 5 sequences from wheat (X75036), rice, soybean, pea, *Arabidopsis* (Y08501) and tobacco. Nucleotide positions are numbered (1-34) and editing sites observed in at least one plant are shown with boxes. The degree of editing in excised intron 4 RNA populations is indicated by red (apparent full editing), purple (partial editing) or open blue (unedited) arrows.

[B] Direct sequencing of the D5 region from RT-PCR products (LB40 and LB159) generated from excised intron RNAs across branch junctions from rice, soybean, *Arabidopsis*, and tobacco. Editing status is shown by red dots for apparent full editing, purple dots for partial editing, open blue dots for absence of editing, and black dots for the genomically-encoded thymidine. Our analysis revealed several differences from the published tobacco (Nicotiana sylvestris, X86706) *nad7* intron 4 DNA sequence (Pla *et al.*, 1995, Appendix II)

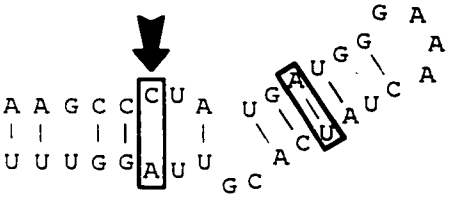
A.



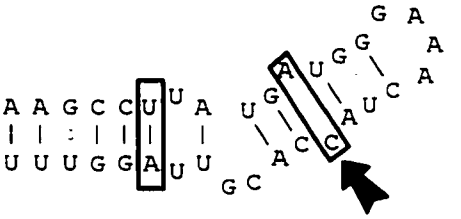
wheat, rice



tobacco

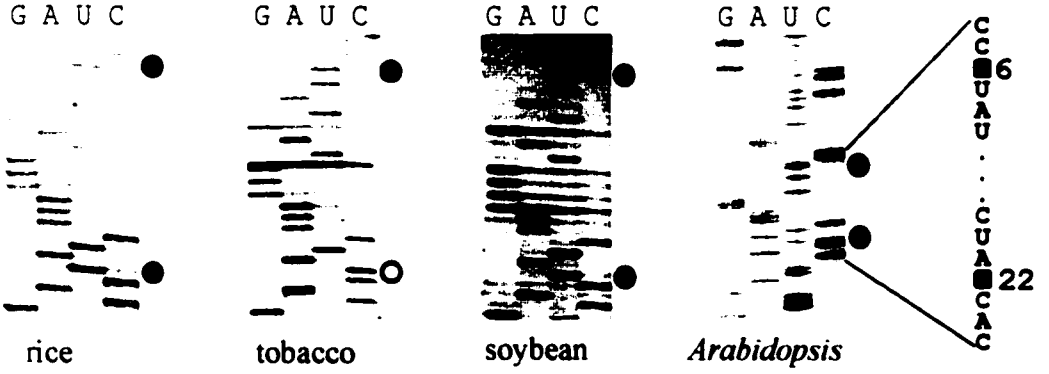


soybean, pea



Arabidopsis

B.



rice tobacco soybean *Arabidopsis*

and wheat, where there is a genomically-encoded thymidine. It should be noted that this U6-A29 base pair is adjacent to a pyrimidine-pyrimidine mispair (U7-U28), so that without editing the D5 helix would be particularly weak.

In wheat and rice, an additional editing position has been observed within the distal non-helical part of D6 of *nad7* intron 4 (Carrillo and Bonen,1997, Figure 4.3A, red arrow); however, it is not an evolutionarily conservative position in that there is adenosine at the comparable site in the dicot DNAs examined (Figure 4.3B, red arrow). Secondary structural modelling of this domain suggests that editing in wheat and rice may increase the stability of the D6 loop (Figure 4.3C, unedited vs. edited). The most energetically favourable structures of tobacco and *Arabidopsis* have similar stabilities to wheat, despite the destabilizing AA mispair at the position of the editing site in monocots (Figure 4.3C). Note that most nucleotide substitutions in these dicots do not disrupt the predicted D6 structure (Figure 4.3 A and B, highlighted in purple) so that the secondary structural model for wheat is supported by rice, tobacco and *Arabidopsis* sequences. For pea and soybean, the thermodynamic stabilities of D6 structures constrained to the model shown in Figure 4.3A are low relative to the other plants examined. This is due to an insertion/deletion event at a position adjacent to the edited AC mispair in monocots (Figure 4.3A, green triangle) which results in an internal loop. The lack of conservation of the *nad7* intron 4 D6 editing position among plants suggests that pairing at this position is not important for splicing.

Within *nad7* intron 3, position C22 showed editing in wheat and rice, and there is a genomically encoded thymidine in *Arabidopsis*, pea and soybean, so that in all cases U22 is observed at the RNA level (see Figure 4.4B, red arrow). Interestingly, the adjacent C21, which remained unedited and forms an AC mispair in wheat, rice and *Arabidopsis*, comprises an AA mispair in soybean and pea (see Figure 4.4A), further indicating that Watson-Crick pairing at this position is not important for splicing. As noted for position C6 in *nad7* intron 4, editing at position C22 in *nad7* intron 3 may be important for counteracting the destabilizing effect of otherwise having two adjacent unpaired sites. In contrast, the C5 site remains unedited so that there is an AC mispair here in all of these plants at the RNA level (see Figure 4.4A and B, open blue arrows). It should be noted that in tobacco, the *nad7* gene lacks intron 3 (Pla *et al.*, 1995).

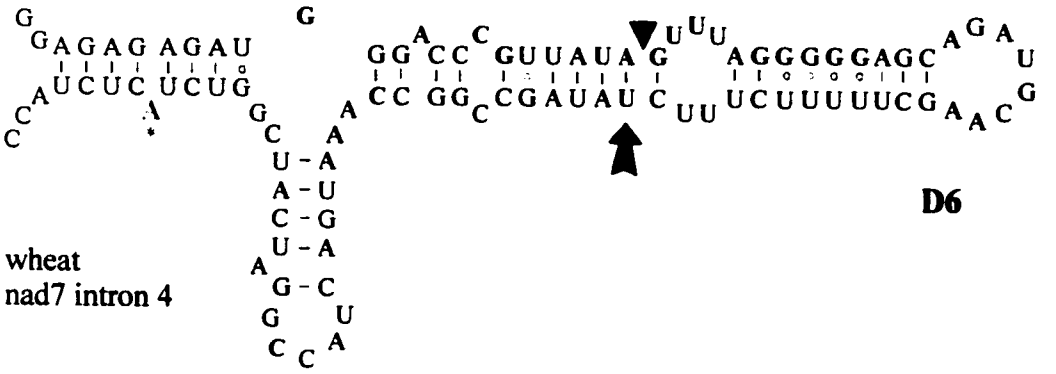
Figure 4.3. Sequence alignment and secondary structure of D6 of *nad7* intron 4 among plants.

[A] Secondary stuctural model for D6 of wheat *nad7* intron 4 with the editing position in monocots indicated by a red arrow. Structural models were predicted from highest thermodynamic stabilities (given in kcal/mol) using mfold version 3.0 by Zuker and Turner (Mathews *et al.*, 1999; Zuker *et al.*, 1999) and by comparing structures among plants. Regions which are variable among plants are highlighted to show changes which do not affect structure (purple) or which destabilize the predicted secondary structure (blue). The location of the insertion/deletion is shown with a green triangle. Note that all plants have D6 structures with conventional bulging adenosines (asterisk).

[B] Alignment of D6 sequences of *nad7* intron 4 from wheat (X75036), rice, pea, *Arabidopsis* (Y08501), tobacco and soybean. Dots indicate positions identical to the wheat sequence and dashes denote insertion/deletions. The observed editing in monocots is highlighted in red and the branchpoint adenosine is shown by an asterisk.

[C] Structural stabilities of D6 (shown in A) among different plants. Predicted minimum free energies are listed in kcal/mol for edited (rice and wheat) and unedited D6 structures. For dicots only the value of the unedited D6 structure is shown since the edited position in monocots is an uneditable A in dicots.

A.



B.

***nad7* intron 4**

```
wheat       GGAGAGAGATGGGACCCGTTATA-----GTTTAGGGGGAGCAGATGCAA-GCT
rice        ..............T..............A...................
tobacco     ..............G..........TAA.....A...............
Arabidopsis ..............G..........TAA...A...............
Soybean     ..........C.....TA....TATATTAAA.....A...............
Pea         ..........C.....TA...G.TTAAC.AA.....A...............

wheat       TTTT-CTTTCATAGCCGGCCAAATGACTACCGGATCATCGGTCTACTCTACC
rice        ..........|.....................................|.....
tobacco     ..........A.....................A...............|.....
Arabidopsis ....T.....A.....................A...............|.....
Soybean     ..........A.........A.........A...............|.....
Pea         ..........A.....................A...............|.....
```

C.

Plant	ΔG unedited (kcal/mol)	ΔG edited (kcal/mol)
wheat	-24.5	-29.9
rice	-19.7	-24.0
tobacco	-32.5	-
Arabidopsis	-29.1	-
soybean	-19.3	-
pea	-20.4	-

Because the only two sites edited within D5 helices of *nad7* introns in wheat were both at position #22, and differences in editing status were seen among plants, I considered the possibility that the editing seen in wheat might occur for reasons other than a requirement in splicing, perhaps because of accessibility of this structure to editing machinery. Therefore, I examined position C22 within the trans-spliced wheat *nad5* intron 2 (Pereira *et al.*, 1991; see Appendix VIII) which also has an AC mispair. However no editing was seen, even though a D6 editing position (9 nucleotides upstream of the 3' splice site) appeared fully edited as did downstream exon sites in partially-spliced RNA molecules from wheat (this work), soybean or pea (Carrillo and Clark, unpublished data). These observations are consistent with the view that A11-C22 mispairs within D5 do not impede splicing and that certain plant mitochondrial introns have weaker helical structures than do conventional group II introns.

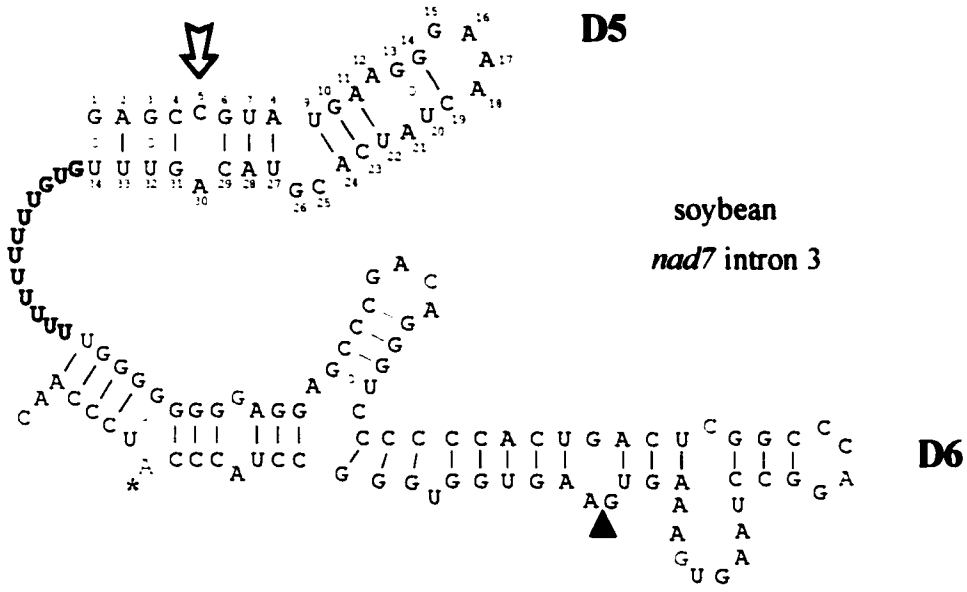
Since AC mispairs within D5 or D6 helices in many instances remain unedited or show plant-specific variation in editing status, these events must not be essential for proper folding for splicing. Interestingly, however, the observed editing sites are located primarily within regions of particularly weak helical structure (for example, having uneditable neighbouring mispairs) so that at such positions editing may serve an important role in improving stability. It is also possible that additional protein machinery assists in the stabilization of such weak RNA structures. Interestingly, plant-specific variation in intron editing is reminiscent of what has been observed at silent codon positions within exons, where this was interpreted as the absence of evolutionary constraint (Shields and Wolfe, 1997). Perhaps the lack of conservation of plant mitochondrial intron editing positions also indicates a lack of evolutionary constraint for these structures.

4.3 Length variation in the linker joining domains 5 and 6 of *nad7* intron 3 among plants

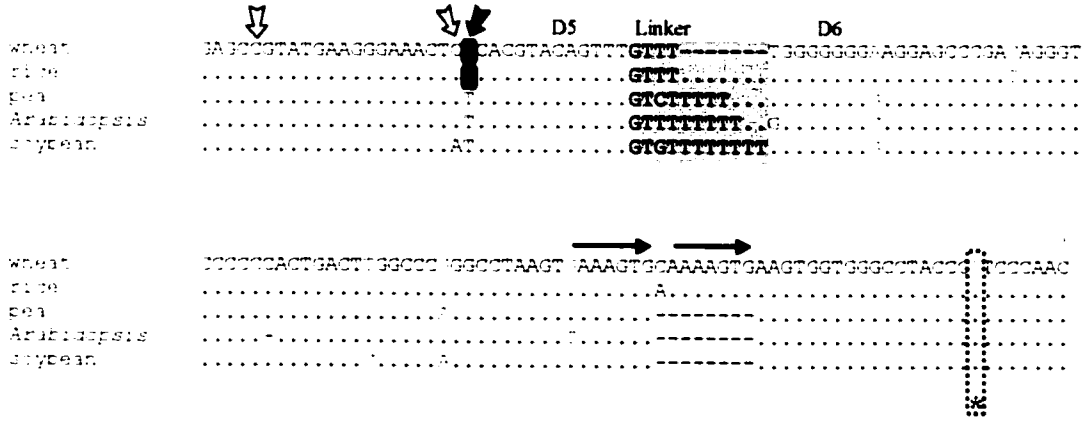
Intron 3 of *nad7* also shows differences among plants with respect to the number of nucleotides separating D5 and D6 (Figure 4.4). This joining segment, J(56), is uridine-rich and varies from 4 nt in wheat and rice, to 8 nt in pea, 9 nt in *Arabidopsis* and 11 nt in soybean (Figure 4.4B and C, yellow shading). Interestingly, despite a difference of 7 nt in the linker region, a lariat form of both soybean and rice excised introns was detected by the

Figure 4.4. Length variation of the J(56) sequence joining D5 and D6 of *nad7* intron 3.
[A] RNA secondary structural model of domains 5 and 6 from soybean, with J(56) shaded in yellow, an open blue arrow indicating an unedited AC mispair. Regions which are variable among plants are highlighted to show changes which do not affect structure (purple) or which destabilize (blue) the predicted secondary structure. The location of the insertion/deletion among plants is shown by a green triangle.
[B] Alignment of domain 5/6 sequences of *nad7* intron 3 from wheat (X75036), rice, pea, *Arabidopsis thaliana* (Y08501) and soybean. Dots indicate positions identical to the wheat sequence and dashes denote insertion/deletions. The J56 is shaded in yellow and observed editing is highlighted in red. Black arrows indicate the direct repeat present in the monocot but not dicot sequences. The branchpoint adenosine is shown by an asterisk.
[C&D] Direct sequencing of RT-PCR products generated from excised intron 3 RNAs from wheat and soybean, showing the linker region ([B], yellow shading) and branch junction ([C], green asterisk and arrow). Note a uridine instead of a cytosine at the position of the branch site.

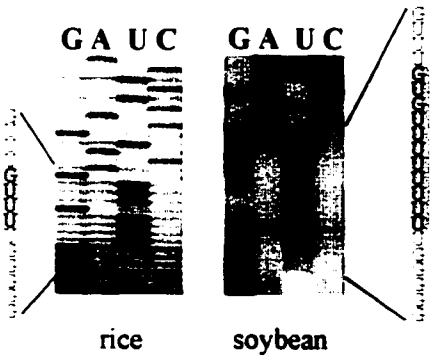
A.


D5
soybean
nad7 intron 3
D6

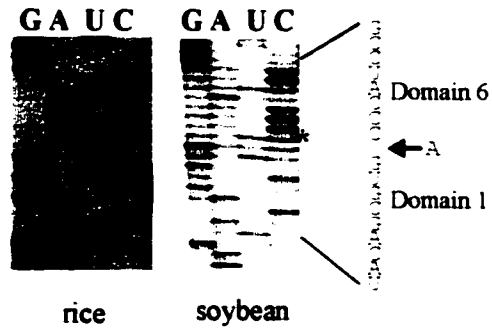
B.


D5
Linker
D6
wheat
rice
pea
Arabidopsis
soybean

C.


G A U C
G A U C
rice
soybean

D.


G A U C
G A U C
Domain 6
A
Domain 1
rice
soybean

RT-PCR methods used (Figure 4.4C and D). The observed variation may have been caused by slippage during DNA replication. Since results were reproducible for each plant examined, these results are not likely to reflect artefacts generated in the course of RT-PCR or PCR experiments. It is notable that the dicot sequences with longer J(56) linkers, have a shorter D6 due to an 8 nt deletion distal to the helical core within a stretch that forms a tandem repeat in the monocot sequences (Figure 4.4B, black arrows).

In conventional group II introns, the J(56) linker is highly conservative in length, typically being between 2 and 4 nt (Michel *et al.*, 1989) and among other higher plant mitochondrial introns, it is 2 to 3 nt in length (cf. Appendix V). The importance of this length in self-splicing introns has been demonstrated in the case of the yeast mitochondrial ribozymic aI5γ intron, where alterations were observed to reduce the efficiency of splicing (Boulanger *et al.*, 1996). Crosslinking studies using the yeast aI5γ intron have shown the close interaction between J(56) linker nucleotides and ones in the internal loop of D6 distal to the branch point adenosine. This is believed to be important for providing specificity and energy for the correct arrangement of the D5 and D6 helices and it has been proposed that they have a side-by-side orientation (Podar and Perlman, 1999). The variation in length of J(56) among closely related plants and its unusually long sequence was unexpected given its importance in self-splicing introns. Again this points to reduced or different structural constraints in plant mitochondria.

4.4 Domain 6 sequence variation among plants within the trans-splicing *nad1* intron 1.

When the wheat mitochondrial *nad1* gene was characterized (Chapdelaine and Bonen,1991), it had been noted that the first *trans*-spliced intron had a number of atypical features, including the absence of a bulged adenosine 7 or 8 nts upstream of the 3' splice site (Figure 4.5A). Moreover, this intronic region was poorly conserved between wheat and *Oenothera*. Two editing sites were observed in this region of the *Oenothera* intron (Wissinger *et al.*, 1991) and the secondary structural model presented was in part improved by these changes, leading to the proposal that RNA editing might be essential for proper folding for splicing. However, this structural model cannot be applied to the monocot sequences. In addition, this intron also lacks the classical GTGCG motif at its 5' end which is virtually invariant among classical group II introns and even more remarkably, the plant

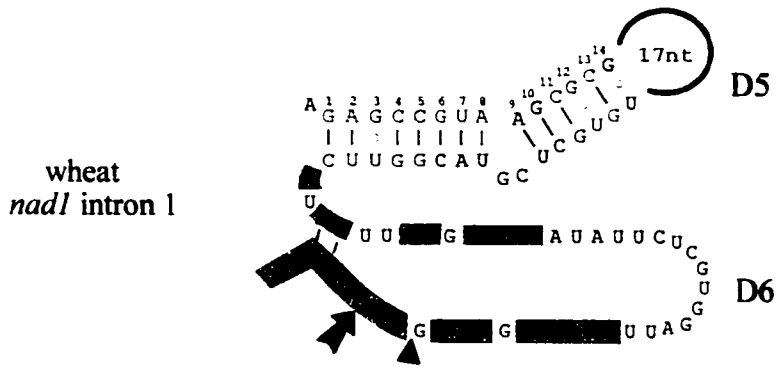
Figure 4.5. Variation in sequence and editing within domains 5/6 of the trans-splicing *nad1* intron 1 among plants.

[A] RNA secondary structural model for wheat, with shaded positions showing identical nucleotides within domain 6 for the flowering plants shown in panel B. Red arrow shows position of observed editing in wheat. The location of an insertion/deletion is shown with a green triangle.

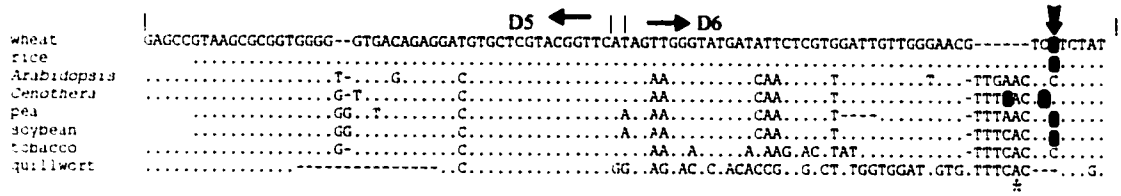
[B] Alignment of D5/D6 sequences of *nad1* intron 1 from wheat (X57967), rice, *Arabidopsis thaliana* (X98301), *Oenothera bertiana* (M63033), pea, soybean, tobacco (X04019), and the quillwort *Isoetes lacustris* (Y17812). The maize DNA sequence (M18339) is identical to that shown for wheat and rice, and the petunia sequence (X60401) shows 2 differences within the variable region compared to tobacco. Positions identical to the wheat sequence are indicated by dots, and insertion/deletions by dashes. The single editing site observed in rice, wheat, pea, and soybean (but not in tobacco or *Arabidopsis*) partially-spliced *nad1* transcripts is highlighted in red, and the two editing sites in *Oenothera* are those reported to show editing in 1/10 (purple) and 10/10 (red) cDNA clones (Wissinger *et al.*, 1991).

[C] Taxonomic grouping of the plants in panel B, shown with secondary structural models for domain 6 of *nad1* intron 1 as predicted from highest thermodynamic stabilities (given in kcal/mol) using "mfold" version 3.0 by Zuker and Turner (Mathews *et al.*, 1999; Zuker *et al.*, 1999). Note that only the *Isoetes* sequence (Malek and Knoop, 1998) shows the conventional bulging adenosine (asterisk). The presence or absence of editing is shown by red and open blue arrows, respectively.

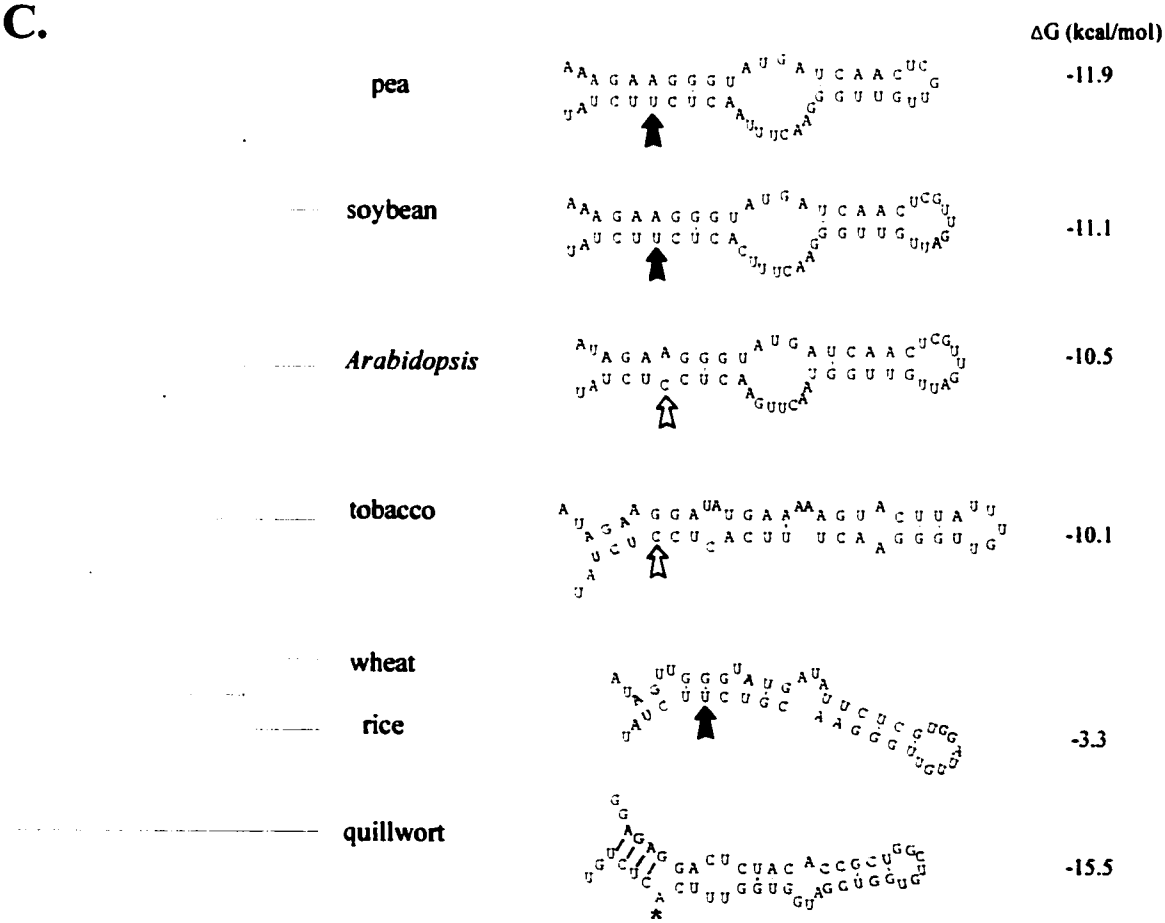
A.


wheat
nad1 intron 1
17nt
D5
D6

B.


D5
D6
wheat
rice
Arabidopsis
Oenothera
pea
soybean
tobacco
quillwort

C.


ΔG (kcal/mol)
pea
-11.9
soybean
-11.1
Arabidopsis
-10.5
tobacco
-10.1
wheat
rice
-3.3
quillwort
-15.5

intron sequences differ from each other in this region (cf. AGACGGGGG in wheat and ATTAGGGGG in *Oenothera*). Furthermore, in this intron D5 has an unusually large loop of 17-19 nt compared to the typical purine tetraloop (Figure 4.5A). Such features are not expected of sequences under functional constraint.

To further assess the degree of sequence variation and RNA editing status among plants, the RNA sequences of the D5/D6 region of *nad1* intron 1 in wheat, pea (Chapdelaine, 1992), rice, tobacco, *Arabidopsis* and soybean were examined (Figure 4.5). Primers were designed so that RT-PCR products were derived from partially-spliced RNA molecules in which at least one downstream intron had been excised (LB128 and LB19). This method was used because the methods for amplifying excised intron template (cf. Section 4.2) did not yield detectable amplification products. This may be due to the *trans*-splicing nature of this intron, with Y-branched excised intron molecules being more rapidly degraded than non-linear lariat forms. Alternatively, because D6 does not have a conventional helical bulged adenosine, splicing may not involve branch point formation, but may proceed through another mechanism such as hydrolysis (cf. Section 1.1.5.1).

In Figure 4.5B, the wheat, rice, *Arabidopsis*, soybean, tobacco and pea *nad1* intron 1 D5 and D6 have been aligned with DNA counterparts from *Oenothera* (Wissinger *et al.*, 1991) and the quillwort *Isoetes lacustris* (Malek and Knoop,1998). It can be seen that D6 is very poorly conserved relative to D5 where there is only one compensatory change within helical regions. For wheat, rice, pea and soybean mitochondria, a single C-to-U edit was seen 6 nt upstream of the 3' splice site and appeared relatively complete in the RNA population. In contrast, the tobacco and *Arabidopsis* counterparts showed no editing at this site, even though downstream exon editing positions were fully edited. Interestingly, neither of the two editing sites reported in *Oenothera* (Wissinger *et al.*, 1991) occur at the single position seen in the plants in this study, rather they are at positions -7 and -11 relative to the 3' splice site and were observed in 1 and 10 out of 10 cDNA clones, respectively (Figure 4.5B, red vs. purple boxes). At the position of one of the *Oenothera* editing sites there is an adenosine which is not an editing candidate in pea and *Arabidopsis*. Thus, although the DNA sequences of the terminal 8 bp of the intron are identical among these plants, editing decreases the RNA sequence similarity, rather than increasing it as within coding sequences.

Because D6 of the *nad1* intron 1 could not be folded into a classical helical structure which was conserved among plants, I used the program "mfold" (Zuker *et al.*, 1999; Mathews *et al.*, 1999) to assess potential compensatory base changes among plants and possible secondary structure models for this domain. Parameters were set so that the structures were constrained by imposing minimal classical group II intron features, namely the pairing of nucleotides -3, -4 from the 3' end of the intron with nucleotides +3, +4 downstream from domain 5. Using this strategy, no single structure which was supported by all plant sequences could be obtained. Similarly, when no constraints were imposed, I obtained no common structure. The most thermodynamically stable structures considering the above constraints are shown in Figure 4.5C. In the case of the dicots, D6 can be folded into a relatively long helix which lacks a bulging nucleotide. Editing appears to improve its stability in the case of pea and soybean (Figure 4.5C, red arrows) and it would have improved the structure in *Arabidopsis* (open blue arrow). In contrast, if editing had been observed in tobacco, it would decrease stability of the predicted structure by converting a GC pair to GU (Figure 4.5C, open blue arrow). However, the monocot (wheat, rice) sequences do not support this structure and the most stable one has a bulging dinucleotide on the other side of the helix and has a ΔG value of only -3.3. The maize sequence (Wolstenholme *et al.*, 1993) is not shown here but is identical to wheat and rice. It is notable that, relative to the dicot sequences, the D6 sequences of monocots have a poor structural stability and show a 5 bp deletion at the expected bulging adenosine position (Figure 4.5B, asterisk). Lack of common D6 features and poor conservation of this domain among plants is consistent with reduced constraint on the nucleotide sequence of this intron. In the lower plant, quillwort (*Isoetes lacustris;* Malek and Knoop, 1998), a conventional D6 structure is observed in the homologous cis-splicing *nad1* intron 1 (Figure 4.5C, asterisk) even though there is considerable sequence divergence. In contrast, D5 is very conservative in sequence with only a single deletion resulting in the conventional purine tetraloop compared to flowering plants (Figure 4.5B). The retention of group II intron features in quillwort in contrast to deviation from conventional structure in flowering plants, illustrates the different evolutionary pathways followed during the time since quillwort and flowering plants last shared a common ancestor.

4.5 Domain 6 sequence variation among plants within the cis-splicing *nad4* intron 2

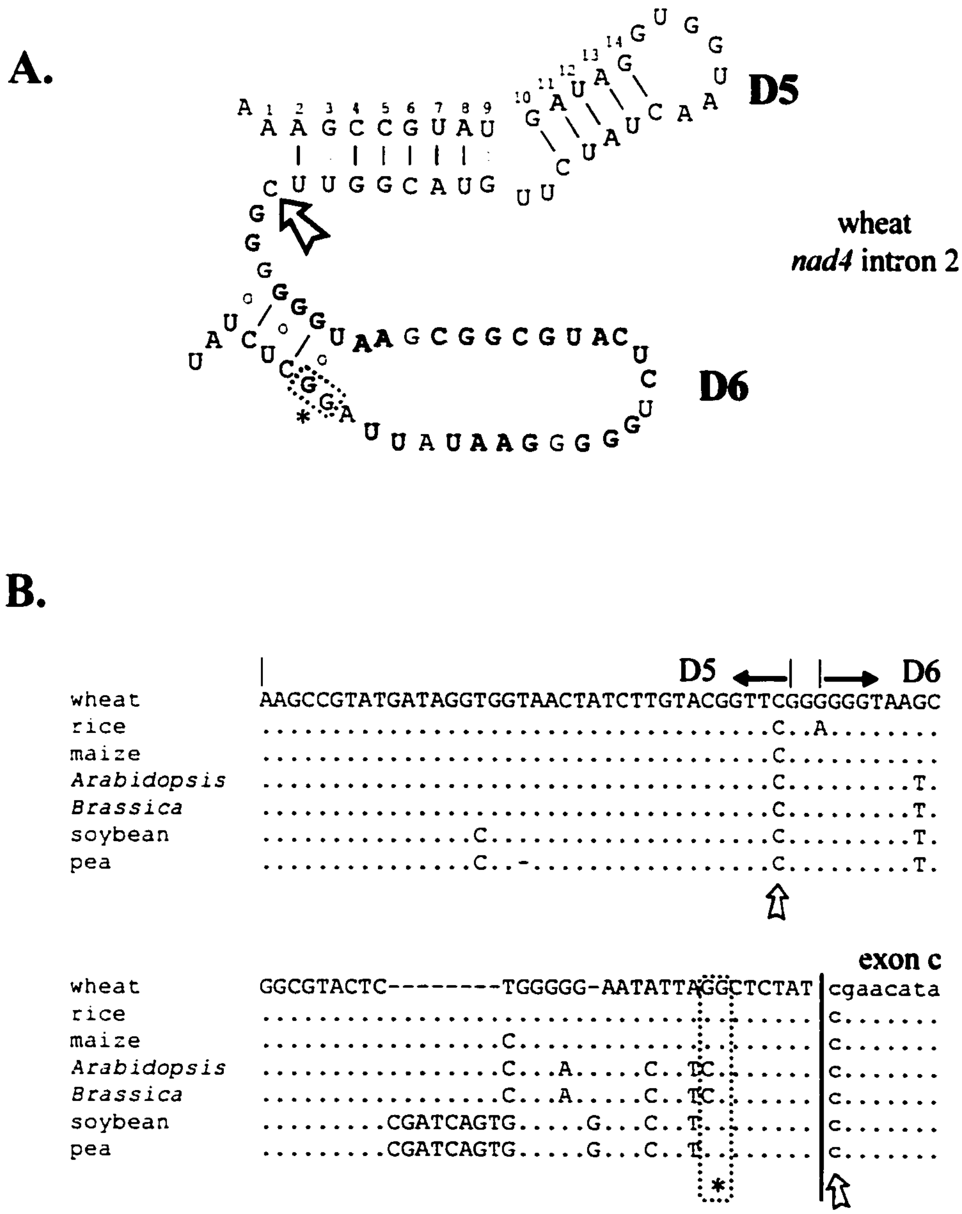
The second intron of *nad4* is similar to *nad1* intron 1 in that it also lacks a bulging adenosine and it has a long loop in D5 (Figure 4.6A) but it differs in that it is *cis*-splicing. I have examined the RNA sequences of *nad4* intron 2 from the mitochondria of wheat, rice pea, *Arabidopsis* and soybean (Figure 4.6B; cf. Figure 3.7C). D6 cannot be folded into a conventional structure in any of these plants and there is no adenosine at the expected branch point position 7 or 8 nucleotides upstream from the 3' splice site. Furthermore, position -8, which could potentially be used as a branch site, differs among plants. The closest adenosine is at position -9 in monocots and at -14 in dicots. As for *nad1* intron 1, attempts to obtain RT-PCR amplification products across a putative lariat structure were unsuccessful, and partially-spliced molecules were examined instead (Figure 3.7C). From this analysis it was seen that an AC mispair (A1-C39) at the base of D5 was not edited in any of the plants examined, nor was the first nucleotide of *nad4c* (Figure 4.6B, open blue arrows; cf. Chapter 3).

The sequence variation seen within D6 among plants is striking for *nad1* intron 1, and both this intron and *nad4* intron 2 cannot be folded into the conventional helix, nor into an alternative structure that is common to all of these flowering plants. This raises questions about the biochemical mechanism of splicing in the absence of a bulging adenosine. Mutational analysis of self-splicing introns has illustrated the importance of the presentation of the bulging adenosine for branch site formation in the first step of splicing (Chu *et al.*, 1998; Liu *et al.*, 1997; Podar *et al.*, 1998; Gaur *et al.*, 1997) and in its absence the reaction can proceed by hydrolysis where the nucleophile is a water molecule. It has also recently been demonstrated that RNA splicing of a branch site mutant group II intron in yeast mitochondria occurs by hydrolysis, releasing a linear form of the intron *in vivo* and *in vitro* (Podar *et al.*, 1998). The study of splicing of aberrant group II introns, such as those in plant mitochondria may provide evidence of such an alternate splicing pathway in naturally occurring introns.

Figure 4.6. Sequence variation within the D5/D6 region of *nad4* intron 2.

[A] RNA secondary structural model for wheat with domain 5 folded according to the 9 bp +5 bp model (Costa *et al.*, 1998). The unedited AC mispair is indicated by an open blue arrow and the position of the expected branchpoint (7-8 nt upstream of the 3' splice site) is boxed. Shading indicates conserved positions in domain 6.

[B] Alignment of D5/D6 sequences from wheat, rice, maize (L20488), *Arabidopsis* (Y08501), *Brassica* (X60794), soybean and pea. Positions identical to that of wheat are shown by dots and insertion/deletions are shown by dashes. Note the absence of adenosine at the expected branchpoint position in domain 6 (boxed, green asterisk). The open blue arrows indicate intron and exon positions at which editing was predicted but not observed in partially-spliced *nad4* transcripts in any of these plants (cf. Figure 3.7). Our analysis revealed several differences in the wheat DNA sequence compared to that reported by Lamattina and Grienenberger (1991; X57164, see Appendix II).


A.
D5
wheat
nad4 intron 2
D6
B.
D5 D6
wheat AAGCCGTATGATAGGTGGTAACTATCTTGTACGGTTCGGGGGGTAAGC
rice ....................................C..A........
maize ....C...........
Arabidopsis ....................................C.........T.
Brassica ....................................C.........T.
soybean ...............C....................C.........T.
pea ...............C..-.................C.........T.
exon c
wheat GGCGTACTC--------TGGGGG-AATATTAGGCTCTAT cgaacata
rice .... c.......
maize ....C..................... c.......
Arabidopsis .................C...A.....C..TC....... c.......
Brassica .................C...A.....C..TC....... c.......
soybean .........CGATCAGTG.....G...C..T........ c.......
pea .........CGATCAGTG.....G...C..T........ c.......

CHAPTER 5:
PHYSICAL FORM OF EXCISED PLANT MITOCHONDRIAL GROUP II INTRONS

5.1 Rationale

The focus of the work described in this chapter is on determining the physical form of excised introns in plant mitochondria to gain information on the splicing mechanism of these introns. Many plant mitochondrial introns lack features which are typical of the self-splicing group II introns. How do such unusual features affect splicing *in vivo*? While the lariat structure of excised group II introns has been well documented for yeast self-splicing introns, fungal mitochondrial and *Euglena* chloroplast introns (reviewed in Michel and Ferat, 1995) and has been recently described in chloroplasts of seed plants (Vogel *et al.* 1997B), the existence of these structures has only been observed in plant mitochondria by the present work using RT-PCR techniques (Chapter 4). In addition to lariat structures, there may also be other forms of excised intron which would not be detected by PCR methods.

I have concentrated on *nad7* introns 3 and 4 since preliminary characterization of these introns indicates some unusual features. Both introns have mispairings within D5 helical regions, with intron 3 helices being particularly weak relative to other plant mitochondrial introns (cf. Figures 3.1B and 6.3). The possible physical structures of excised introns would be lariat if splicing proceeds via normal mechanisms, linear if splicing takes place by first step hydrolysis (cf. Section 1.1.5.1) or circular (see below). The following sections describe experiments which are aimed at determining the relative levels of each of these forms for both *nad7* introns 3 and 4. The development of techniques for studying excised introns *in vivo* will be useful for studying more unusual introns, such as those lacking a branchpoint adenosine (i.e., *nad1* intron 1, *nad4* intron 2; Chapter 4).

5.2 Direct sequencing of *nad7* intron 4 RT-PCR products across branch junctions

In the course of RT-PCR experiments to study RNA editing events, where oligomers mapping to the 5' and 3' regions of the intron were used to amplify a region of the intron across the branch junction (Figure 5.1A and C, LB40 and LB159), lariat forms of excised introns, with the expected branch point adenosine (Figure 5.1B), were observed for wheat, rice, soybean, *Arabidopsis* and tobacco despite differences in editing status of D5 structures

Figure 5.1. Direct sequencing of RT-PCR products to detect lariat and circular forms of excised *nad7* intron 4.

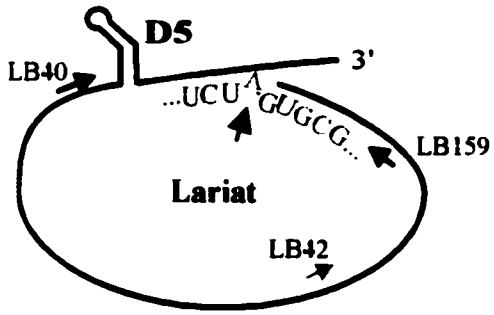
[A] Schematic of strategy for amplifying RT-PCR products from *nad7* intron 4. Grey arrows (LB42) indicate oligomers used for cDNA synthesis and black arrows represent primers used for RT-PCR (LB40 and LB159). The green arrow indicates the branch site. Domain 1 sequence is shown in pink and the domain 6 sequence is shown in blue.

[B] Direct sequencing of RT-PCR products from wheat (24 hr), rice (24 hr), soybean (24 hr), *Arabidopsis* and tobacco showing the branch junction region at the extreme 5' and 3' termini (domains 1 (pink) and 6 (blue), respectively) of the intron where the branchpoints are indicated with green arrow and asterisks. Note the misincorporation of thymidine rather than adenosine at the branch site during cDNA synthesis (see Vogel *et al.*, 1997).

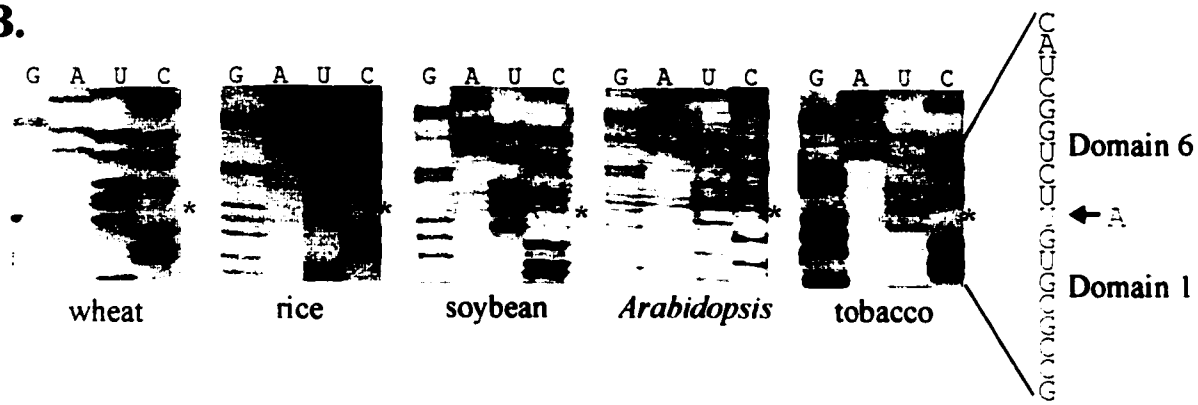
[C] Schematic circular RT-PCR product from wheat (24 hr) *nad7* intron 4. The green arrow indicates the branch site adenosine in [A&B]. Domain 1 sequence is shown in pink, domain 6 sequence is shown in blue and the sequence of the lariat tail is shown in black.

[D] Direct sequencing of RT-PCR products (primers LB40 and LB159) of *nad7* intron 4 generated using 24 hr wheat mtRNA demonstrating a circular form of the excised intron.

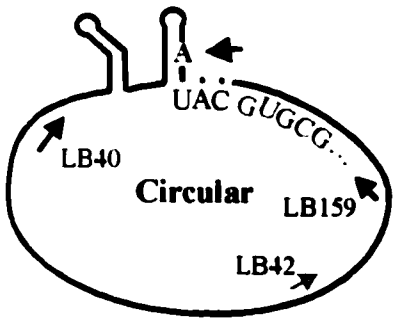
A.


D5
3'
LB40
...UCUAGUGCG...
LB159
Lariat
LB42

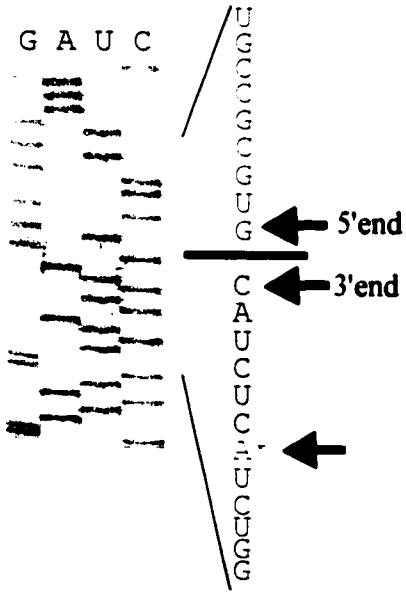
B.


G A U C
G A U C
G A U C
G A U C
G A U C
wheat
rice
soybean
Arabidopsis
tobacco
Domain 6
A
Domain 1

C.


A
UAC GUGCG...
LB40
Circular
LB159
LB42

D.


G A U C
5'end
3'end

within these introns (cf. Chapter 4). In wheat mtRNA from 24 hr embryos and 6 day seedlings, the prevalent species observed was a product derived from a lariat form of excised intron 4 (Figure 5.1B). However, when these RT-PCR products were cloned, a mixture of clones derived from both lariat and circular forms of the intron were obtained, with the majority of products representing the lariat form. For RT-PCR products derived from 6 day mtRNA 15 "lariat" and four "circular" excised intron clones were obtained and for products derived from 24hr mtRNA, 12 "lariat" and one "circular" intron derived clones were obtained. These results indicate, that for *nad7* intron 4 in wheat, the levels of the lariat form of the excised intron are high relative to the levels of the circular form, so that the circular form was observed in cloning experiments but was usually beyond the detection limit of direct sequencing. In one RT-PCR experiment which involved a reamplification step, the circular form of the excised intron was the predominant product observed by direct sequencing (Figure 5.1C and D); however, this observation likely does not represent the situation *in vivo*, where the circular form may be over-represented due to the re-amplification step used in this experiment.

The circular intron RT-PCR product may indicate the presence of circular forms of excised intron *in vivo*, may be the result of ligation of the ends of a linearly excised intron ,or may represent an artefact from the experimental methods used. The first possibility is that a circular form of excised intron is generated through an alternate splicing mechanism; however, a circular rather than lariat form of an excised group II intron has not yet been reported in the literature and would not be an expected product from the two transesterification reactions typical of group II intron splicing. Alternatively, RNA ligase activity in the wheat mitochondrion could be involved in the formation of a phosphodiester bond between the two ends of a linear intron. At least one type of RNA ligase is known to be present in the nucleus where it is involved in tRNA intron splicing (cf. Section 1.1.3); however, such ligases should require a 2' phosphate, 3' hydroxyl end, so that they should not have been able to catalyze this reaction (Weber *et al.*, 1996). An alternate RNA ligase, perhaps from a viral source, might carry out this ligation. The identification by RT-PCR of other circularized versions of RNA molecules which are known to be linear, such as mRNAs, would confirm such RNA ligase activity. This product may also have been generated as an

artefact of the experimental methods used. For example, if the ends of a linearized intron were close together due to the secondary structure of the intron, the reverse transcriptase may have "jumped" from the 5' end of the intron to the 3' end of the intron. The template for this could also have been excised lariat introns or introns which have not yet been excised; however, all of the 3' and 5' nucleotides of the excised intron are present and no additional nucleotides are observed. If the template for "jumping" was lariat intron, there would likely be some exonucleolytic digestion of the 3' end of the intron and if the template was precursor then some of the exon nucleotides would be expected to be incorporated into this product. Finally, it is possible that a circular form of excised linear intron is generated through a secondary reaction after splicing such as that of group I introns. However, for group I introns this reaction reduces the size of the intron whereas the sequence of the circular form of excised *nad7* intron 4 is complete. Interestingly, this observation also indicates that the ends of the intron were not subjected to any exonucleolytic degradation as might be expected for the unprotected RNA ends of a linear form of excised intron.

The presence of a circular form of *nad7* intron 4 is particularly interesting since putative 'excised intron' species with three different electrophoretic mobilities are observed in wheat (cf. Figure 3.2B). Comparison of profiles of *nad7* intron 3 and 4-containing transcripts between 6 day and 24 hr stages of development showed differences in both amounts (introns 3 and 4) and apparent sizes of excised introns (intron 4) indicating possible differences in physical forms of excised introns (cf. Figure 3.2B, asterisks). The signals representing excised introns could be composed of linear and circular forms of intron as well as the expected lariat form. Note that largest of these is consistent with the size of a splicing intermediate containing the downstream exon (discussed below). However, the difference in mobility of the remaining two excised intron species may be due to differences in physical form.

5.3 Differentiating between circular and lariat introns: oligomer hybridization and S1 nuclease protection

By direct sequencing of RT-PCR products, lariat forms of both *nad7* excised introns 3 and 4 are clearly present; however, since this method is highly sensitive, in that even low levels of excised lariat introns would result in an amplified product, it is not clear what

relative levels of excised lariat intron, compared to other physical forms, exist *in vivo*. Oligomer hybridization and S1 nuclease protection assays were used to assess the steady state levels of the lariat form (*nad7* introns 3 and 4) and the circular form (*nad7* intron 4) of the excised intron *in vivo*.

5.3.1 Oligomer hybridization: circular, lariat and splicing intermediates

5.3.1.1 *nad7* intron 3

Northern analyses were carried out to identify all excised *nad7* intron 3 species with a lariat conformation. An oligomer probe designed based on the sequence across the branch site of the excised lariat form of *nad7* introns 3 (Figure 5.2A , LB151) was used in conjunction with an internal intron oligomer (LB64) to differentiate between intron lariats and intron-containing precursors. Note that for the lariat oligomer (LB151) 10 nts are complementary to the 3' end of the intron, upstream of the branch site, and 7 nts are complementary to the 5' end of the intron. The lariat oligomer was designed with a thymidine residue complementary to the adenosine of the branch site, even though in sequencing across excised introns this site was paired with an adenosine (cf. Figure 4.1C, green arrow). An oligomer probe generated based on the circular intron sequence was not used for intron 3 since no circular forms of the intron were detected by RT-PCR (cf. Section 4.2). The optimal hybridization temperatures were determined by Southern hybridization at temperatures which ranged from 39°C to 53°C (C. Carrillo and S. Macdonald, unpublished observations). The DNA controls used for Southerns were pGEM clones containing wheat sequences generated from (1) an RT-PCR product (LB41 & LB64) of the excised lariat *nad7* intron 3 (Figure 5.2B, lane 1; cf. Section 4.2) or (2) a mtDNA PCR product (LB91 & LB92) encompassing *nad7* intron 3 and flanking exons (Figure 5.2B, lane 2). This latter control was used to demonstrate that the lariat probe could discriminate lariat introns from precursor, excised linear and/or circular forms of the intron. The optimal hybridization temperature was 51°C, a temperature at which the lariat probe bound to its complementary DNA sequence (Figure 5.2B, LB151 lane 1), but did not form short 7 or 10 bp hybrids (lane 2). As expected, the internal intron oligomer probe (Figure 5.2A, LB64) hybridized to both DNA controls under the conditions used (Figure 5.2B, LB 64 lanes 1 and 2). These conditions were used for northern analyses of wheat mitochondrial RNA from 24 hr embryos and 6 day seedlings.

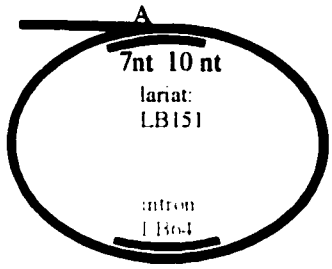
Figure 5.2. Northern blot hybridization analysis of wheat *nad7* intron 3 lariats using oligomer probes

[A] Schematic showing lariat 17-mer (red, LB151) and intron 20-mer (green, LB64) used for northern hybridization experiments.

[B] Southern blots of cloned lariat RT-PCR product generated using primers LB41 and LB64 (lane 1) and cloned *nad7* intron 3 mtDNA PCR product generated using primers LB91 and LB92 and consisting of wheat *nad7* exon 3, intron 3 and exon 4 digested with NarI (lane 2) were used to determine if the oligomer probes (LB64 and LB151) would form short hybrids (see text).

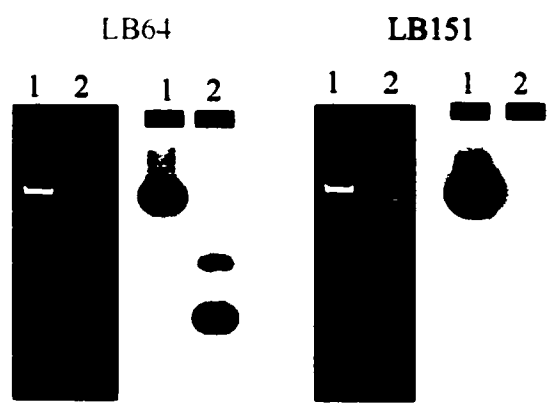
[C] Northern hybridization using intron (LB64) and lariat (LB151) probes with 24 hr and 6 day wheat mt RNA with RNA size marker indicated on the left and sizes of possible lariat-containing splicing intermediates and excised intron RNAs indicated on the right (boxes represent downstream exons, lines represent introns).

A.

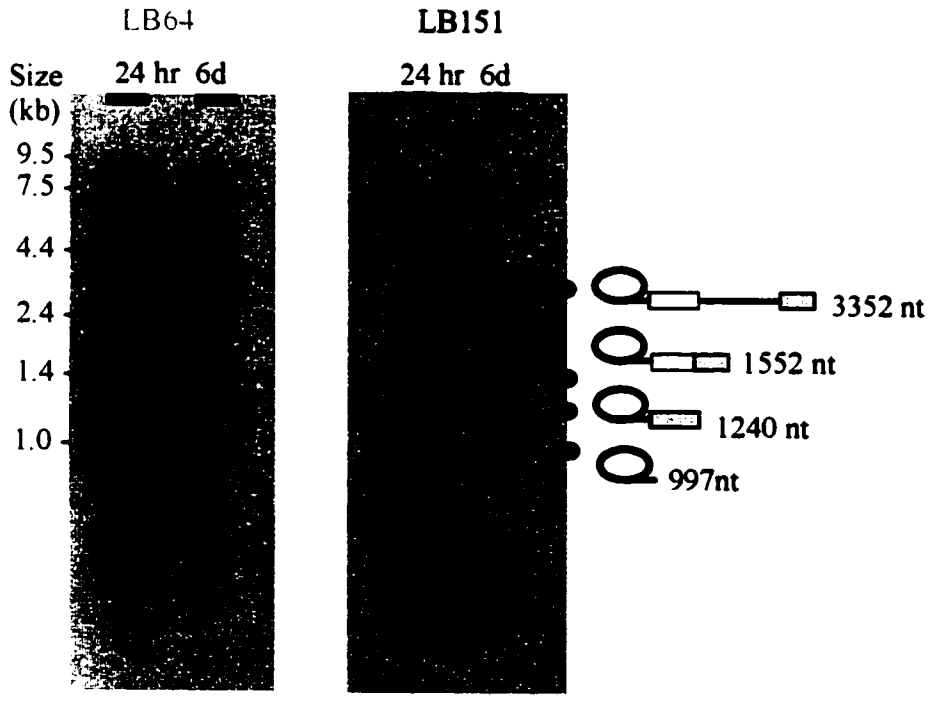


nad7 intron 3

B.



C.



Differences were observed between wheat mtRNA blots probed with an internal intron oligomer (Figure 5.2C, LB64) compared to those probed with the lariat oligomer (Figure 5.2C, LB151). The lariat LB151 and intron LB64 oligomers bound to species migrating at the expected excised intron position as well as to three higher molecular weight RNA species. However, the lariat probe did not bind to species greater than 3.4 kb representative of *nad7* precursors. This indicates that the lariat oligomer only bound to fully complementary sequences and not to all intron-containing sequences present on the blot. Thus, there are four signals which represent lariat-containing sequences. The smallest and most abundant RNA is the predicted excised intron size and the sizes of the larger lariat-containing species are consistent with sizes of the three possible *nad7* intron 3 splicing intermediates which include the intron 3 sequence and downstream *nad7* exon 4 and/or intron 4 and exon 5 (Figure 5.2C, black dots). Developmental differences were also observed. The 6 day mtRNA had relatively high levels of a 3.4 kb intron 4-containing splicing intermediate compared to excised introns and intron-containing precursors (Figure 5.2C, LB151). Levels of excised intron were higher in 24 hr mt RNA than in 6 day seedlings as was previously shown (cf. Figure 3.2B).

5.3.1.2 *nad7* intron 4

Similar northern analyses were performed for *nad7* intron 4. For this intron, two different oligomer probes were used to detect either lariat or circular forms of excised intron (Figure 5.3A, LB172 and LB165). For both oligomers, the last 7 nucleotides were complementary to the 5' end of the intron (Figure 5.3A, LB172 vs. LB165, yellow line) whereas the first 10 nucleotides differed between the probes (green vs red line) such that the circular probe (LB165) included the 6 nucleotides at the 3' end of the intron (green line) and the lariat probe (LB172) was complementary to sequences upstream and including the bulging adenosine (red line). The optimal hybridization temperatures were determined by Southern hybridization to "circular" and "lariat" forms of *nad7* intron 4 RT-PCR products (LB40 and LB159) which had been cloned into pGEM vectors (cf. Section 5.2). At 47°C in 5% formamide the circular LB165 probe bound to its full complement (Figure 5.3B, LB165, lane C) but not to cloned lariat RT-PCR product (LB172, lane L). Under these conditions, the lariat LB172 probe bound to its full complement (Figure 5.3B, lane L), but only a weak

Figure 5.3. Northern blot hybridization analysis of wheat *nad7* intron 4 excised introns using oligomer probes.

[A] Schematic showing oligomers used to detect lariat (LB172), circular (LB165), precursor (LB152) and all intron-containing transcripts (LB42). Exons are represented by boxes and introns are depicted as lines. Regions of the oligomer probes which are identical in sequence are shown in the same color. Green lines represents the lariat tail.

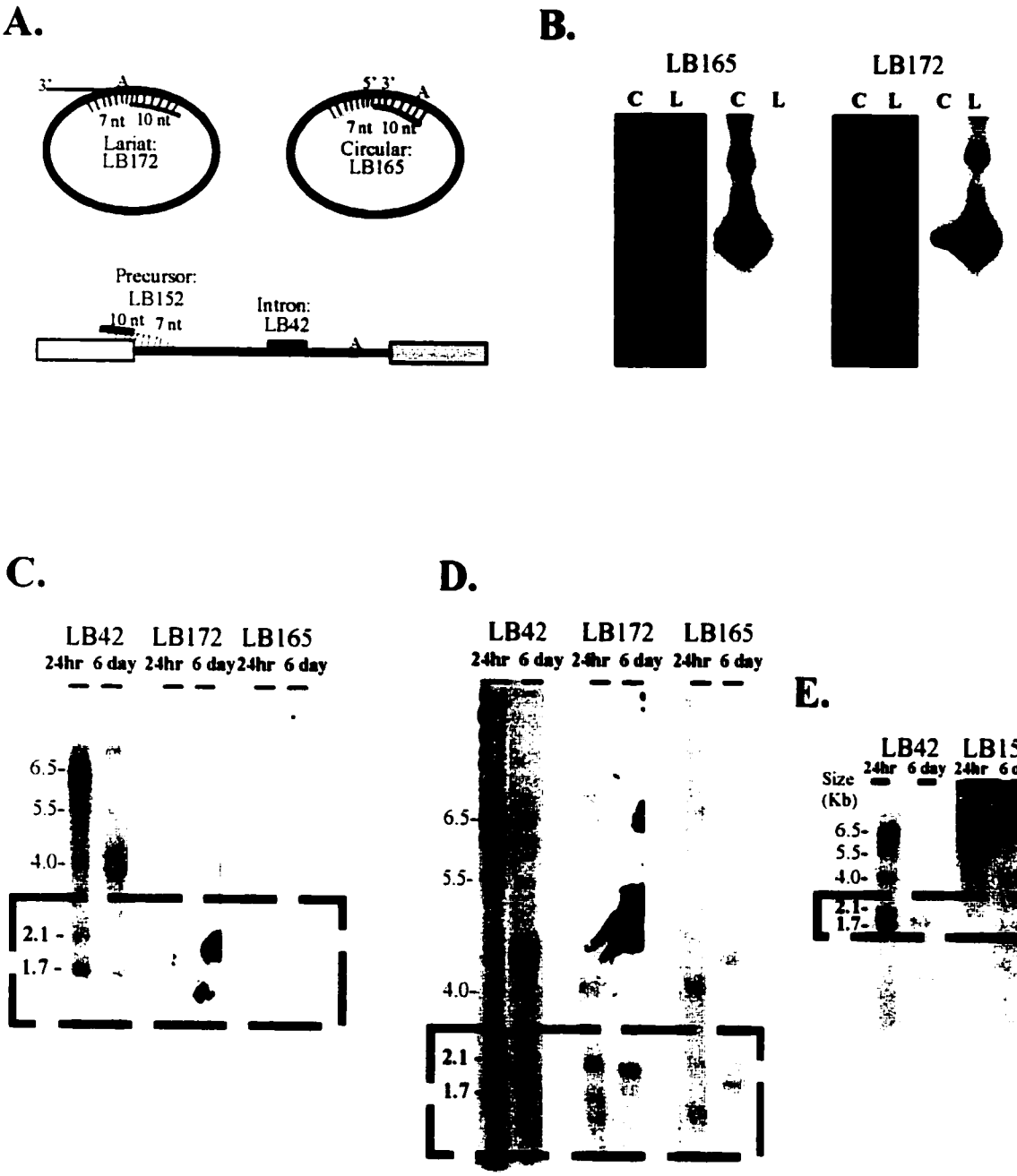
[B] Southern blot of cloned RT-PCR products generated with primers LB159 and LB40 which were derived from lariat excised introns (lane L) or circular excised introns (lane C) and were used to test for short hybrid formation (see text).

[C] Northern hybridization using circular (LB165), lariat (LB172) and intron (LB42) oligomers with wheat mtRNA from 24 hr embryos and 6 day seedling. The red box indicates the region on the blot where excised intron 4 (1.7 kb) or splicing intermediates consisting of intron 4 and exon e (2.0 kb) were expected to migrate.

[D] As for [C] but with longer electrophoresis and exposure times.

[E] Northern hybridization using precursor RNA oligomer probe (LB152) and intron oligomer probe (LB42).

[C,D&E] Sizes of transcripts are indicated on the left and are calculated based on RNA size markers (Gibco-BRL).


A.
3'
A
7 nt 10 nt
Lariat:
LB172
5' 3'
A
7 nt 10 nt
Circular:
LB165
Precursor:
LB152
10 nt 7 nt
Intron:
LB42
A
B.
LB165
LB172
C L C L
C L C L
C.
LB42 LB172 LB165
24hr 6 day 24hr 6 day 24hr 6 day
6.5-
5.5-
4.0-
2.1 -
1.7 -
D.
LB42 LB172 LB165
24hr 6 day 24hr 6 day 24hr 6 day
6.5-
5.5-
4.0-
2.1
1.7
E.
LB42 LB152
24hr 6 day 24hr 6 day
Size
(Kb)
6.5-
5.5-
4.0-
2.1-
1.7-

signal was observed in the lane containing the circular form of excised *nad7* intron 4 (lane C). When the temperature was decreased, the circular probe bound to all intron-containing products (data not shown).

When blots of wheat 24 hr embryo and 6 day seedling mtRNA were probed with these two oligomers, signals were absent (Figure 5.3C, LB165 and LB172) or very weak (Figure 5.3D, LB165 and LB172), relative to the internal intron oligomer probe (Figure 5.3C and D, LB42). The lariat probe was expected to produce stronger signals as an abundant signal from lariat forms of the excised intron have been detected by S1 nuclease protection assays (cf. Section 5.3.2). To increase these signal strengths, the temperature of the hybridization was lowered to 42°C; however, under these less stringent conditions short hybrids were able to form with the DNA controls so that the circular and lariat oligomer probes bound to both circular and lariat DNA sequence and the northern profiles were similar to ones observed using the internal intron oligomer probe (data not shown).

Even though weak, the signals in Figure 5.3D provided some clues about the possible identities of the excised *nad7* intron 4 species. In the 24hr wheat embryos, four signals which are consistent with the expected sizes of excised intron species and splicing intermediates were observed (Figure 5.3D, LB42, red box). The doublet migrating at the 1.7 kb position is consistent with the size of the excised intron 4. The two signals may represent introns with different electrophoretic mobilities due to differences in physical forms (cf. Section 5.4.1). This doublet is especially apparent in 6 day mtRNA. The slower migrating RNA species (2.1 kb) is the expected size of a splicing intermediate where the first step of splicing of intron 4 has taken place but where the intron is still associated with the downstream exon. In fact, a signal at this position is also observed for northerns probed with the downstream exon, *nad7*e (cf. Figure 3.2B, *nad7*e). In northern blots probed with an oligomer complementary to the upstream exon, *nad7*c, this signal is absent (LB48, see Figure 5.6C). However, results from the latter experiment were difficult to interpret since the species in the region of the 1.6 kb mRNA were not sufficiently separated so that the signal from the messenger RNA would overshadow the 2.1 kb region with longer exposures. The fourth, fastest migrating excised intron 4 species (1.45 kb) is not consistent with any expected excised intron species. This product may be derived from intron breakdown, where the intron is cleaved at positions

which are more accessible to endonucleases (i.e., looped out regions) so that breakdown products are of uniform size and migrate as a discrete band. A strong signal of approximately 4.5 kb was observed in the 6 day stage of development (Figure 5.3D, LB42). This species may represent an *nad7* transcript which includes intron 4 as well as one of the other three introns; however, this signal was not observed in all blots probed with LB42 (i.e., Figure 5.3E, LB42) and it may represent an experimental artefact.

For the *nad7* intron 4 lariat probe, three signals corresponding to excised intron species were observed; however, the relative intensities of these hybrids differed from the intron probe in that the signal strength of the 1.7 kb species was lower and there appeared to be only one band rather than a doublet in the 1.7 kb region. These results may be expected if the doublet were composed of both excised linear RNAs and lariats. If this were the case, the intron probe could bind to both species, whereas the lariat probe could only bind to one, giving it a reduced signal strength. Interestingly, the level of the putative splicing intermediate was higher than the signal from the excised intron in the 6 day stage of development as was observed for *nad7* intron 3. The splicing intermediate in 6 day wheat mtRNA migrated faster than its counterpart in 24 hr embryos. If this splicing intermediate is a stable, dead end product (see below) then the 3' end of this species is processed differently between 24 hr and 6 day stages of development. For blots probed with the circular probe (Figure 5.3D, LB165), weak signals corresponding to high molecular weight precursor species were observed in 24hr mt RNA indicating that the probe hybridized to all intron-containing species under the conditions used. It will be of interest to carry out additional experiments using longer exposure times and higher amounts of RNA to try to improve the signal strengths obtained in these experiments.

It is possible that the structure of the excised intron somehow prevented binding of the lariat probe due to a destabilizing bond in the center of the hybrid because of the internal branched adenosine. However, if this was the case, the *nad7* intron 3 lariats should not have been easily detected (cf. Figure 5.2, LB172). Perhaps the differences in sequence of the *nad7* intron 4 lariat probe resulted in a probe which formed a less stable hybrid, despite the equivalent lengths and GC content. To address this possibility, two 20-mers with an additional 3 nucleotides at the 5' end (lariat probe, LB183) or 2 nucleotides at the 5' end and

one at the 3' end (circular probe, LB182) were used with the expectation that the increased length of the region flanking the bulging adenosine would increase the stability of the hybrid (Carrillo and Niknejad, unpublished results). Since these oligomers were more prone to cross-hybridization due to the increased lengths of partial hybrids, hybridization at 47°C in 5% formamide (as for the LB165 and LB172 probes) resulted in binding of the probes to all intron 4 containing species. Therefore, to increase stringency, varying concentrations of formamide (from 5% to 25%) were assayed. At 45°C in 15% formamide, the circular probe did not bind to the lariat RT-PCR product and the lariat probe bound less efficiently to the circular RT-PCR product (data not shown). However, since the circular probe did not cross-hybridize to lariat clones, these experiments were deemed appropriate for detecting circular forms of excised introns. To date no detectable signals have been observed with RNA blots probed with these oligomers at 45°C in 15% formamide. It would be of interest to increase the levels of mtRNAs on northern blots to improve signal strengths.

5.3.1.3 *nad7* intron 4 precursors

A second set of experiments was carried out to further characterize *nad7* intron 4-containing transcripts by differentiating between intron-containing precursors and excised introns. An oligomer probe (Figure 5.3A, LB152) which included the 7 nt from the 5' end of intron 4 and 10 nt from the upstream exon was used along with the internal *nad7* intron 4 probe (LB42). Conditions under which the precursor probe could only bind to its complement and not to excised intron were determined by Southern hybridization with DNA controls (data not shown). Northern hybridizations with wheat 24hr embryo and 6 day seedling mtRNA were carried out at 47°C in 5% formamide. As expected, this oligomer bound to all higher molecular weight species and did not bind to lower molecular weight species (Figure 5.3E, LB152) compared to the intron probe (LB42) which bound to two additional excised intron species. This lends further support to the characterization of 2.1 kb *nad7* intron 4-containing species as a splicing intermediate and further confirms that the 1.7 kb signal represents excised intron.

Detection of splicing intermediates for both *nad7* intron 3 and intron 4 revealed an additional complexity to the observed RNA profiles. Not only are there precursor RNAs which contain different combinations of the four introns with no apparent polarity to the

splicing steps, but there are also products in which the first step, but not the second step of splicing has taken place. These intermediates may be dead end products, in which the 5' exon has become dissociated so that the second step of splicing can no longer take place. In fact, the 1240 nt band observed for *nad7* intron 3 (Figure 5.2) may represent such a dissociated upstream exon resulting from the first step of splicing of the downstream intron 4. The presence of splicing intermediates in 6 day wheat mtRNA was of particular interest since the levels of the other precursor RNAs were low at this stage of development, even though the level of *nad7* mRNA was rather equivalent to that of 24 hr embryos (cf. Figure 3.2B). The high level of these splicing intermediates may result from the stability of these transcripts. Regions at the 3' end of some mitochondrial mRNAs are capable of folding into stem-loop structures which are thought to associate with proteins which confer stability to the transcript (reviewed in Binder *et al.*, 1996). With the involvement of such stability enhancing elements, terminal exon-containing splicing intermediates would be expected to be stable due to the presence of the processed 3' end of the *nad7* mRNA. High levels of such group II intron-containing splicing intermediates have also been observed for *atpF* and *petD* introns in spinach chloroplasts (Kim and Hollingsworth,1993) and in branch site mutants of the aI5γ intron transformed into yeast mitochondria (Podar *et al.*, 1998).

In contrast, in self-splicing introns, 5' cleavage is the rate limiting step: the second step in splicing occurs rapidly so that splicing intermediates are rarely detected (Michel *et al.*,1989). However, mutants with second step splicing defects accumulate intron-3' exon splicing intermediates due to dissociation of the upstream exon. Disruption of the tertiary EBS2-IBS2 interaction (Michel and Jacquier, 1987), mutation of the first intron nucleotide (Peebles *et al.*, 1993) or length changes in the D5/D6 joining region (Boulanger *et al.*, 1996) have all been shown to affect the second splicing step. Similarly, when the γ-γ'tertiary interaction was mutated in the self splicing intron rI1, splicing was not impaired but was somewhat less efficient and resulted in the accumulation of splicing intermediates (Holländer and Kück, 1999B). Perhaps non-conventional structures in the plant mitochondrial group II introns result in a less efficient second splicing step.

5.3.2 S1 Nuclease Protection assays: circular, lariat and precursor RNAs.

To further characterize the physical forms of excised *nad7* introns 3 and 4, S1

nuclease protection assays were carried out. PCR products generated from clones of excised lariat *nad7* introns (cf. Sections 4.2 and 5.2) were uniformly labeled by incorporation of α^{32}P-dATP in the course of PCR reactions. Probes were subsequently hybridized to 24 hr wheat mtRNA then digested with S1 nuclease to generate protected fragments which were characterized by separation on denaturing polyacrylamide gels.

For *nad7* intron 3, lariat intron RNAs were expected to protect a fragment of 258 nt and precursor RNAs or linear excised introns would protect fragments of 84 and 174 nt (Figure 5.4A). Fragments of all three sizes were observed in this experiment and could be differentiated from the 306 nt input probe (Figure 5.4B and C). The fragment protecting precursor or linear forms of intron 3 is approximately 65% the size of the fragment protecting excised lariat forms of the intron and should therefore have a correspondingly lower signal intensity if the amounts of both species are roughly equivalent. The 84 nt fragment, which also represented precursor forms of the intron, had approximately half the signal strength of its 174 nt counterpart, so that longer exposure times were necessary to detect this fragment (Figure 5.4C). The signal from the 174 nt fragment representing precursor or linear forms of excised intron was roughly 35% of the strength of the signal from the fragment representing the lariat form of the intron (Figure 5.4B, 258 vs. 174 nt) indicating that the amount of lariat excised intron was approximately double the amount of precursor or linear forms of excised intron. This confirms observations in northern analyses, where levels of precursors were low relative to excised intron (cf. Figure 3.2B), and it suggests that the majority of the excised *nad7* intron 3 species observed are of a lariat rather than linear form.

Similar experiments were performed to characterize *nad7* intron 4 RNAs using a lariat RT-PCR product probe (cf. Section 5.2). In this case, *nad7* intron 4 RNA lariats would protect a 291 nt fragment and precursor RNAs and excised linear introns would protect both 189 nt and 102 nt fragments (Figure 5.5A). A circular form of the excised intron would also be expected to bind to the probe, resulting in a 291 nt protected fragment, since the entire length of the probe would be double stranded, except for an internal 6 nucleotides of the circular intron RNA which would be single stranded and susceptible to degradation by S1 nuclease (Figure 5.5A, green dotted line). Similarly, intron-containing precursors are capable of protecting a spliced mRNA probe (Sisodia *et al.*, 1987). Products of all three predicted

Figure 5.4. S1 nuclease protection analysis of wheat *nad7* intron 3-containing precursors and excised introns.

[A] Schematic of the ^{32}P uniformly labeled probe used for the S1 nuclease analysis. The probe was generated by PCR amplification of cloned *nad7* intron 3 lariat RT-PCR products (LB41 and LB64) using LB41 and forward sequencing primers. The red section of the probe complements the 5' end of the intron, the green section complements the 3' section of the intron and the dotted black line represents sequence from the pUC vector.

[B & C] Autoradiography of fragments protected from S1 nuclease and separated on a 7% denaturing polyacrylamide gel. Numbers correspond to the sizes of the protected fragments calculated based on the DNA sequencing ladder shown on the right [B] or left [C]. The lanes are labeled 0, 10, 50, 200 to indicate the number of units of S1 nuclease added to each reaction and untreated input probe run alongside digested products is indicated with an I when 1% of the probe was added and I2 when 2% of the probe was used.

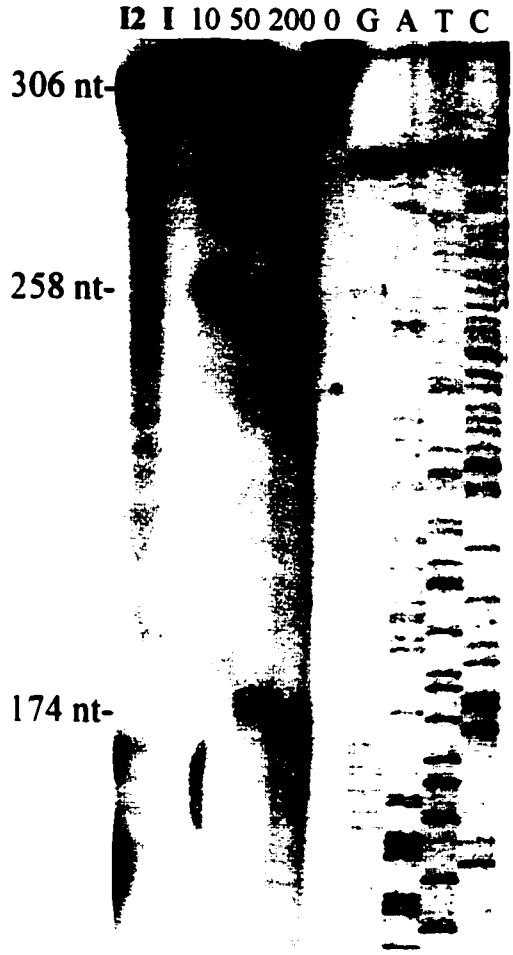
A.

lariat intron

linear intron

intron-containing precursor

B.



C.

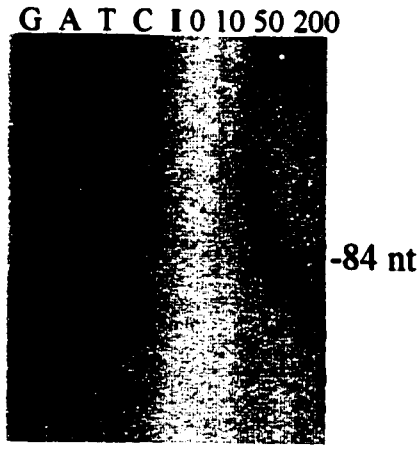


Figure 5.5. S1 nuclease protection analysis of wheat *nad7* intron 4 containing precursors and excised introns.

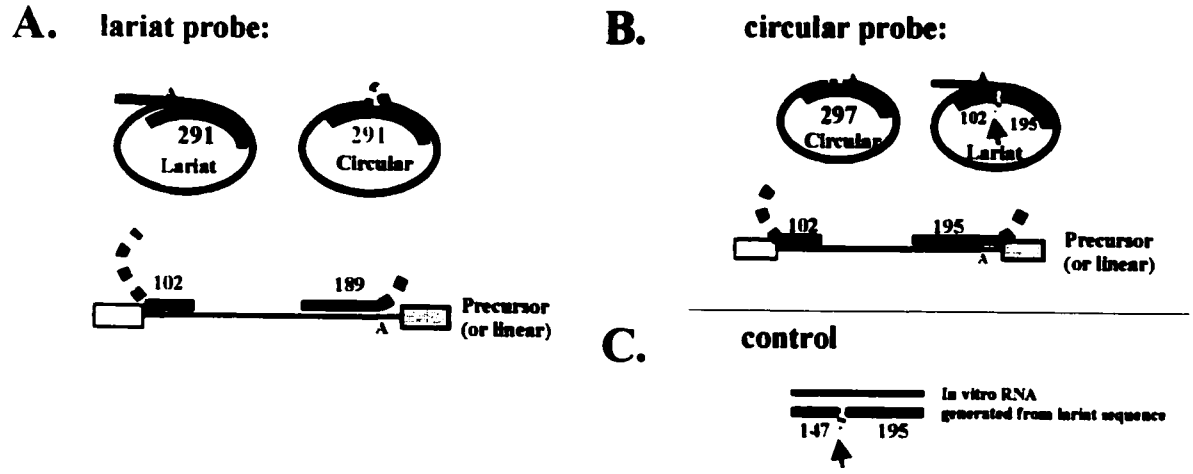
[A, B & C] Schematic showing the probes used for the S1 nuclease analysis. Probes were generated by PCR amplification (primers LB159 and reverse sequencing) of cloned (pGEM vector) RT-PCR products (primers LB40 and LB159) derived from lariat [A] or circular [B] wheat *nad7* intron 4 excised introns. Regions of identical sequence are shown as the same color [A] Predicted sizes of protected products from binding of the lariat probe to lariat (291 nt), circular (291) and precursor or linear forms of intron 4 (102 nt, 189 nt) are indicated.

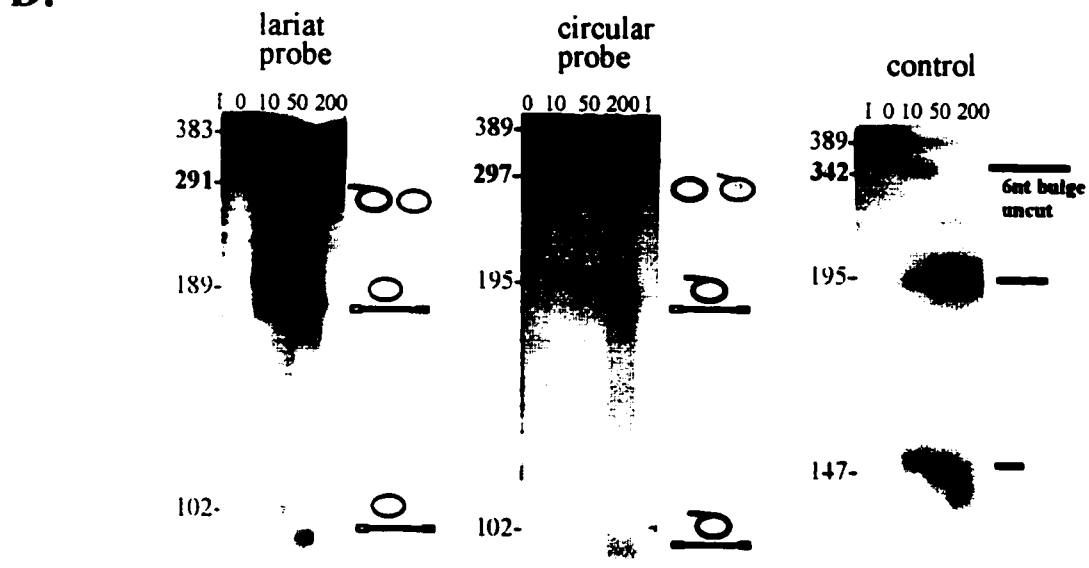
[B] Predicted sizes of protected products generated from binding of the circular probe to circular (297 nt) and precursor, lariat or linear forms of intron 4 (102 nt, 195 nt) are indicated.

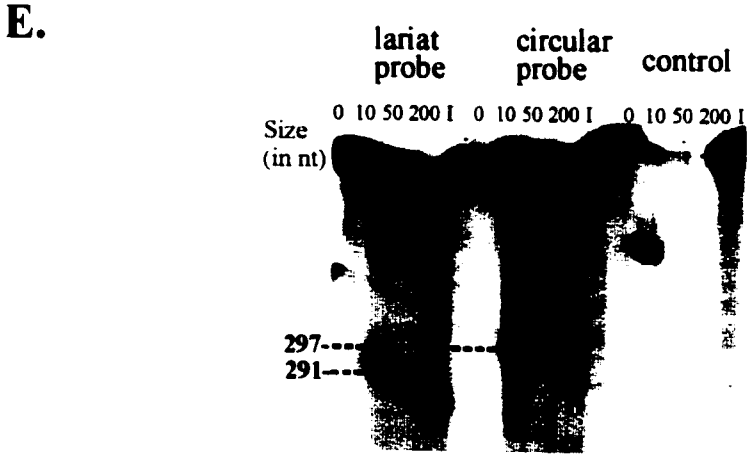
[C] Control RNA was generated from the cloned lariat intron 4 RT-PCR product (LB40 and LB159; cf. Section 5.2). Predicted sizes of the circular probe bound to the control RNA are 342 nt if the 6 nucleotide bulge is not digested and 195 and 147 nts after the digestion of this bulge.

[D] Autoradiography of protected fragments generated by S1 nuclease analysis. Fragment sizes indicated on the left are in nt and were determined according to a DNA sequencing ladder. Lanes are labeled according to the number of units of S1 nuclease used in each reaction (0, 10, 50 or 200) and according to the probe used. Untreated probe was run alongside digested products and is indicated by an 'I'. Possible identities of products (lariat, circular or precursor) are depicted on the right of each gel.

[E] Autoradiograph of [D] showing a greater separation of the protected fragments representing excised intron.


A.
lariat probe:
291
Lariat
291
Circular
102
189
A
Precursor
(or linear)
B.
circular probe:
297
Circular
102
195
Lariat
102
195
A
Precursor
(or linear)
C.
control
In vitro RNA
generated from lariat sequence
147
195


D.
lariat
probe
circular
probe
control
I 0 10 50 200
0 10 50 200 I
I 0 10 50 200
383
291
189
102
389
297
195
102
389
342
195
147
6nt bulge
uncut


E.
lariat
probe
circular
probe
control
Size
(in nt)
0 10 50 200 I
0 10 50 200 I
0 10 50 200 I
297
291

sizes were observed with the lariat probe (Figure 5.5D, lariat probe). The levels of excised intron were slightly lower than the levels of precursor/linear forms of the intron compared to *nad7* intron 3 where the level of excised intron were relatively higher. This corresponds to results from northern analyses of *nad7* intron 4-containing products where signals representing precursor RNAs were cumulatively stronger than those representing excised intron RNAs (cf. Figure 3.2B). The relative signal strength of the 102 nt fragment was low, but this was as expected since its length is 54% of that of the 189 nt fragment, which also represented precursor/linear RNAs, so signals would be expected to be about half as strong if the adenosine content of these fragments were roughly equivalent.

These results were extended by using a circular-intron derived PCR probe to detect circular forms of excised intron (Figure 5.5B). In this case, a circular intron would protect a 297 nt fragment and the lariat, precursor and linear forms of the excised intron would protect both a 102 nt and a 195 nt fragment (Figure 5.5B). Binding of the circular probe to the lariat excised intron generates a 6 nucleotide bulge due to the presence of the additional "tail" in the circular probe (Figure 5.5B, green arrow). This bulge was predicted to be digested by S1 nuclease since even two consecutive mispairs are susceptible to degradation (Chapdelaine, 1992). The results from this experiment were consistent with high levels of excised intron of a circular form *in vivo*, however the abundance of this product appeared to decrease as the amount of S1 nuclease was increased from 10 to 200 units (Figure 5.5D, circular probe). The differences in sizes of the lariat and circular excised intron protected fragments are shown with a better separation in Figure 5.5E. For the circular probe, the decrease in the level of the 297 nt fragment is accompanied by an increase in the level of the other protected fragments, indicating that perhaps hybrids of the probe with a lariat form of the excised intron are protected due to inefficient digestion of the 6 nt bulge within this hybrid. For the lariat probe, the levels of protected precursor fragments also increased at 200 units (Figure 5.5D, 189 and 102 nt) , however, this was not accompanied by a significant decrease in the level of the 291 nt lariat fragment (Figure 5.5 E, lariat vs. circular probes)

To test whether the 297 nt protected fragment observed with the circular probe was due to the lack of digestion of the 6 nt bulge, a control experiment where the circular probe was bound to RNA synthesized *in vitro* from the lariat RT-PCR product of wheat *nad7* intron

4 in a pGEM clone (cf. Section 5.2) was carried out (Figure 5.5C). The six nucleotide bulge was incompletely digested by the S1 nuclease at a concentration of 10 units of enzyme, although complete digestion was observed at concentrations of 50 and 200 units (Figure 5.5D and E, control). Thus, it is not possible by these methods to determine whether the protected "circular" product observed with the wheat mtRNA (at 10, 50 and 200 units) represents excised circular introns or whether the observed product represents the lariat form of the intron and reflects a less efficient enzyme activity, perhaps due to impurities in RNA extracts.

Taken together, these experiments indicate the presence of high levels of lariat forms of *nad7* introns 3 and 4 compared to levels of precursor or linear introns and the potential presence of a circular form of intron 4. This is the first time a lariat form of plant mitochondrial introns has been demonstrated experimentally. If the circular form of excised intron is generated by processes occurring *in vivo*, it may indicate novel splicing mechanisms of group II introns in plants.

5.4 Differentiating between lariat/circular and linear excised introns.

While a lariat form of excised intron is abundant for *nad7* intron 3 and most likely for intron 4 as well, this does not preclude the possibility that a second splicing pathway, where the intron is excised in a linear form, exists *in vivo*. Experiments performed with self-splicing introns *in vitro* have demonstrated the coexistence of both splicing mechanisms (Perlman and Podar, 1996). The experiments described above were designed to differentiate between circular or lariat forms of introns but could not be used to demonstrate the presence of linear forms of excised intron *in vivo* because they would give the same results for intron-containing precursors and linear introns. In the following section, several strategies which were used to detect linear excised intron and to determine the abundance of such a species relative to the excised lariat form of the intron are described.

5.4.1 Characterization of *nad7* introns 3 and 4 containing transcripts by random cleavage of RNA with Mg^{2+}.

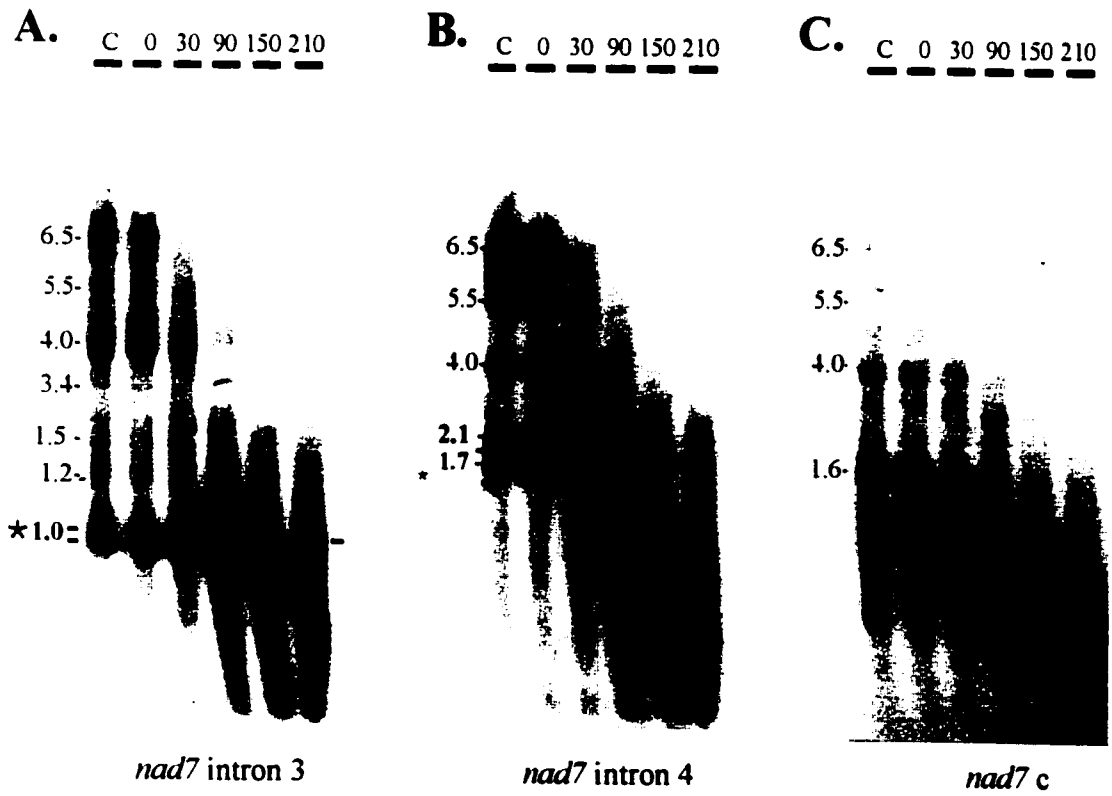
Lariat excised *nad7* introns were further characterized by subjecting them to random cleavage to alter their electrophoretic mobility. Bivalent and trivalent cations at low concentrations are capable of catalyzing the cleavage of phosphodiester bonds in RNA and

linearizing circular viroid RNA molecules (Sanger *et al.*, 1979). By this treatment, linear RNA molecules subjected to a single random internal cleavage would decrease in size and would not be detected as a discrete band, whereas lariat (or circular) molecules would be the same size but would become linear introns with a short 6 nt tail. Broken open lariats are known to have the same electrophoretic mobility as linear introns under these electrophoretic conditions (Perlman and Podar, 1996). I wanted to examine the effect of "linearizing" lariat introns on migration through a denaturing agarose gel. To this end, wheat mtRNA from 24 hr embryos was incubated in Mg^{2+}-containing glycine buffer at a slightly alkaline pH at 37°C. Aliquots of the mixture were removed at different time intervals and immediately ethanol precipitated. Samples were electrophoresed through denaturing agarose gels and analysed by northern hybridization.

A doublet corresponding to excised *nad7* intron 3 and a triplet corresponding to excised intron 4 (Figure 5.6A and B, asterisks) were observed by northern analysis. Such doublets were described by Podar *et al.* (1998), for the yeast mitochondrial a15γ group II intron where the more slowly migrating species represented linear introns and the other species represented lariat forms of the intron. For *nad7* intron 4, the four species observed in Figure 5.3D are not resolved and there appear to be three excised intron species. As the incubation time increased, the results of the magnesium treatment became less clear since the background in the region near the excised intron is high due to breakdown products of the precursor forms of the intron. Over time, random cleavage of the RNA resulted in a shift of the excised intron to a form which migrated more slowly on an agarose gel for both *nad7* intron 3 (Figure 5.6A) and intron 4 (Figure 5.6B). This is consistent with the lariat form of the intron migrating faster than the linear form and it indicates that the majority of the excised introns *in vivo* are not in a linear form, since linear introns should be rapidly degraded under these conditions. Interestingly, the signals representing broken open lariats for both introns are rather equivalent to those representing lariats. The *cis*-splicing *nad7* introns from wheat 24 hour embryos appeared to be exceptionally stable, since magnesium treatment of excised introns 3 and 4 lasting over 3 hours did not completely degrade these introns. Note that intron 3 appears to be more stable than intron 4 with this treatment. This may be due to the smaller size, and possibly a more compact structure of *nad7* intron 3 which

Figure 5.6. Northern hybridization analysis of wheat 24hr embryo mtRNA randomly cleaved with Mg^{2+}.

[A, B&C] Northern blots were hybridized to oligomer probes complementary to *nad7* intron 3 (LB64, panel [A],), *nad7* intron 4 (LB133, panel [B]) or *nad7* exon c (LB48, panel [C]). Lanes are numbered according to the time of incubation in 5 mM $MgCl_2$ (0 to 210 minutes). The control was untreated input RNA (I). Approximate sizes of signals are indicated on the left and were calculated based on RNA markers (Gibco-BRL).



nad7 intron 3

nad7 intron 4

nad7 c

may be less susceptible to cleavage by Mg^{2+}. The same blot probed for *nad7* exon c (Figure 5.6C) demonstrated that the linear mRNA (which is 1.6 kb and similar in size to *nad7* intron 4) was completely degraded by 210 minutes. Note that the RNA profile in the control which was not incubated with the Mg^{2+} buffer (Figure 5.6, lane C) is the same as for the sample taken at T = 0 minutes (lane 0) demonstrating that the ethanol precipitation and wash were efficient at removing salts from the treated samples so that there is no 'salt effect' causing a shift in migration in these experiments.

Although the results of these experiments are somewhat difficult to interpret due to the inevitable high levels of background resulting from RNA degradation, they are consistent with observations by Podar *et al.* (1998), who showed that excised introns in a lariat form migrated faster than linear introns (or broken open lariats) within a denaturing agarose gel system. In contrast, migration of lariat structures through polyacrylamide gels is extremely slow relative to linear species (Chu *et al.*, 1998). In both gel systems broken open lariats (linear molecules with a short tail) have about the same electrophoretic mobility as linear introns (Daniels *et al.*, 1996; Podar *et al.*, 1998). The migration properties of lariat introns in agarose gels suggest that they are compact structures, even under the denaturing conditions used. Perhaps denaturation of RNA helices within the intron is incomplete.

5.4.2 Specific cleavage of *nad7* intron 3 lariats using RNAse H.

In the experiment described above, a doublet was observed for excised *nad7* introns 3 and 4 and broken open lariats migrated more slowly than the lariat forms of the excised introns. It is not clear whether the slower migrating species in the doublet observed in untreated wheat mtRNA from 24 hr embryos consists of randomly broken open lariats, or if this species is also composed of introns specifically excised in a linear form. To differentiate between linear and circular/lariat (or broken open forms of these) excised introns, wheat mtRNA was cleaved by binding it to intron-specific oligomers (Figure 5.7A, red arrows, LB64 and LB94) and digesting the resulting complex with RNase H (Erster *et al.*, 1988) which specifically hydrolyzes RNA which is hybridized to DNA and not single stranded RNA. The digested products were characterized by performing northern hybridization analyses using an additional intron-specific oligomer as a probe (Figure 5.7A, black arrows, LB60). Using this method, linear forms of the excised intron would be cut into

Figure 5.7. Northern hybridization analysis of wheat *nad7* intron 3 RNA specifically cleaved with RNAse H.

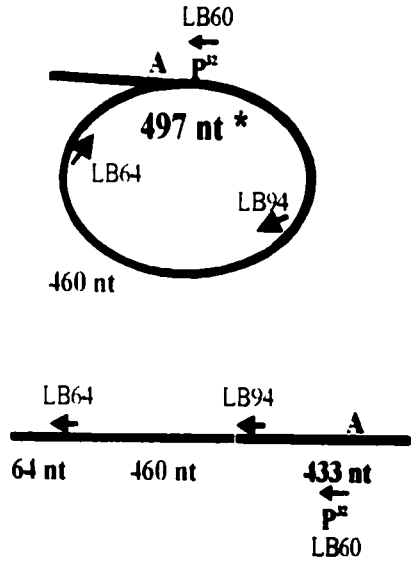
[A] Schematic showing the location of *nad7* intron 3 oligomers (red, LB64 and LB94) which were hybridized to *nad7* intron 3 RNA in order to direct specific digestion of the RNA at the location of the hybrid. The oligomer probe used in the northern hybridization is indicated in black (LB60).

[B] Autoradiograph of the northern blot of 24 hr wheat mt RNA. Lanes are labeled according to oligomers used in the RNase H digestion (LB64, LB94 or LB64+LB94) or control, C, where oligomers were not added. The table below the blot indicates the sizes of fragments expected to be generated in each lane from linear or lariat excised introns.

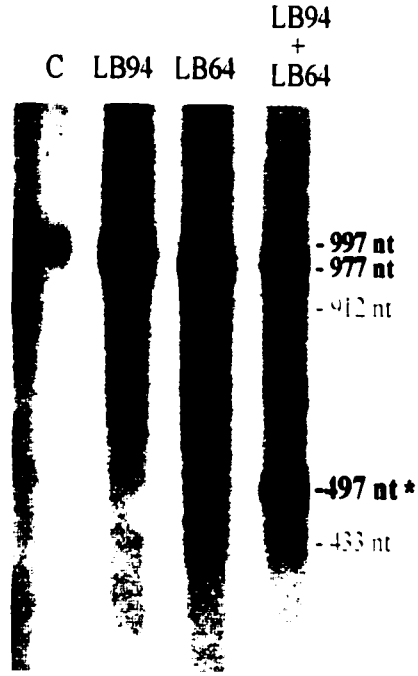
[C] Schematic of products expected from RNAse H cleavage of *nad7* intron 3 transcripts produced *in vitro* from a wheat mtDNA PCR (LB91 and LB92) cloned into a pGEM vector. Oligomers are the same as in [A].

[D] Autoradiograph of the northern blot of control RNAs generated in vitro. Lanes are labelled according to the oligomer(s) used for digestion of the RNA (LB94, LB94+LB64) or control, C, where no oligomers were added. [B&D] Approximate sizes of signals in nt are indicated on the right and were calculated based on RNA markers (Gibco-BRL).

A.

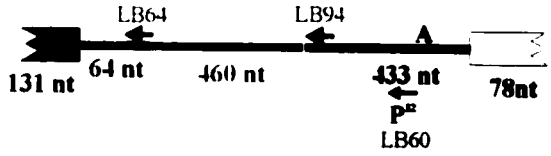


B.

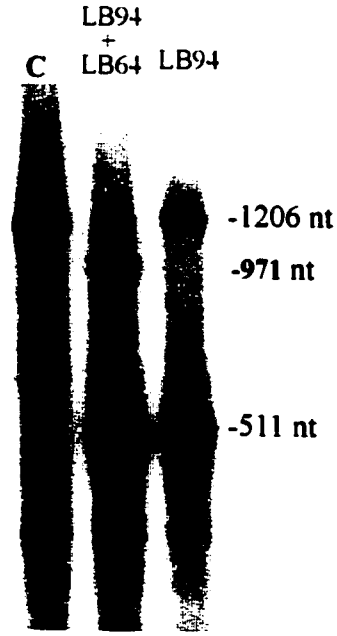


Expected size (in nt)				
Lariat	997	977	977	**497***
Linear	997	433	912	433

C.



D.



two pieces, so that only one piece would be detected by oligomer hybridization and this piece would be smaller than the full length intron (Figure 5.7A, 433 nt). In contrast, circular or lariat forms would retain their length and would only experience a small shift in migration in the denaturing agarose gel system due to their linearization (Figure 5.7B, 977 nt). Any lariats which had been broken open by cleavage at random sites within the RNA (such as those observed in section 5.4.1) would not be detected as they would not generate products of discrete size.

For *nad7* intron 3, these experiments demonstrated an abundance of the lariat form of the excised intron since no major shift in migration was observed using either oligomer LB64 or LB94 for digestion (Figure 5.7B, lanes LB94 and LB64, 977 nt). In this experiment, the broken open lariats are actually 20 nt smaller than the intact lariats (977 nt vs 997 nt) as the region bound to the 20-mer is digested in the RNase H step so that broken open lariats migrate slightly faster than in the untreated control (Figure 5.7B, lanes LB94 and LB64 vs. C) . To demonstrate that RNase H digestion was indeed taking place an additional control, where both oligomers were added to the same reaction mixture to cleave the intron RNA twice, was used to further decrease the size of the signal representing the lariat form of the excised intron to 497 nt (Figure 5.7A, asterisk). As expected, when both oligomers were used, a signal of about 500 nt was observed, indicating that the majority of the excised intron 3 species were of a lariat form (Figure 5.7B, lane LB64+LB94). If a linear form of the excised intron existed, it was not present at a level high enough to be detected by these methods (Figure 5.7B, 912 nt, 433 nt); however, this does not preclude the possibility that an alternate splicing pathway, where the first step occurs by hydrolysis, exists *in vivo* since linear forms of excised intron are degraded more rapidly *in vivo* (Podar *et al.*, 1998) and steady state levels of such introns may be difficult to detect.

RNA transcribed *in vitro* was used as a control for this experiment to demonstrate efficient hybrid formation and RNAse H cleavage. The RNA was transcribed from an *nad7* intron 3 PCR product (primers LB91 & LB92) cloned into a pGEM vector (constructed by J. Szeto, summer student). The uncut transcript would be 1206 nt, the transcript digested with LB94 or LB64 (Figure 5.7C, red arrows) would be detected as 511 or 971 nt signals using the LB60 oligomer probe (Figure 5.7C, black arrow). When primer LB94 was used to

direct RNA digestion, the majority of the *in vitro* RNA was digested (Figure 5.7D, LB94 vs. C) and when both LB94 and LB64 were used, bands representing digestion with LB94 (511 nt) and primer LB64 (971 nt) were observed (Figure 5.5D, LB94+LB64) indicating that digestion with both primers was successful. Furthermore, rather equivalent levels of transcripts were observed in all three lanes with the undigested product giving a slightly weaker signal, indicating that the conditions used in this experiment do not cause rapid degradation of linear RNAs.

Similar experiments were used to characterize *nad7* intron 4 in wheat mtRNA from 24 hr embryos. However, the results of these experiments were more difficult to interpret due to the presence of high levels of precursor forms of intron-containing RNAs relative to excised intron RNAs (data not shown). The digestion products from precursor RNAs were similar in size to the excised introns so that the two products could not be resolved. It should be possible to circumvent the problems caused by the precursors by using a different set of oligomers to carry out the experiment and by probing the blots with an oligomer mapping to the 5' region of the intron.

In summary, linear forms of excised *nad 7* intron 3 were not detected in 24 hr wheat mtRNA in contrast to abundant lariat and broken open lariat forms of the intron. It would be of interest to carry out these experiments with the other *nad7* introns and with other more aberrant introns, which are cis-splicing (i.e., *nad1* intron 2, *nad4* intron 2), to determine whether splicing resulting in linear intermediates occurs at any significant level *in vivo*.

5.4.3 Primer extension analysis of excised *nad7* introns 3 and 4.

To further characterize branch sites and differentiate between lariat and circular forms of excised introns, a set of primer extension experiments was performed. In the first experiment, an intron 3 oligomer (LB64) complementary to a region 84 nt downstream of the splice junction *nad7* exon c/intron 3 was bound to wheat mtRNA isolated from 24 hour embryos and was extended with AMV reverse transcriptase (Figure 5.8A). As expected, a single product of 84 nt was obtained, which was consistent with termination of reverse transcription at the 5' end of *nad7* intron 3 (Figure 5.8B). This product could have been produced either by stalling of the enzyme at the 5'–2' bond present in a lariat form of this intron or by the termination of extension at the end of a linear form of the excised intron

Figure 5.8. Primer extension analyses of excised *nad7* intron 3 and intron 4 RNAs from wheat mitochondria.

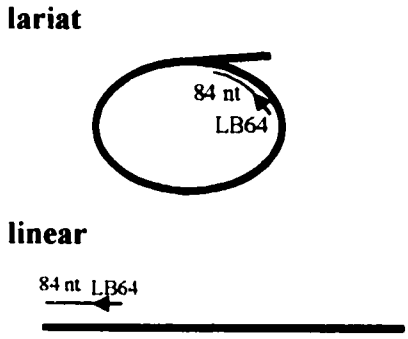
[A] Schematic of primer extension experiments for wheat *nad7* intron 3. The ^{32}P end-labeled oligomer used in this experiment (arrow, LB64) maps to the 5' end of the intron. The expected size for primer extension products (orange line) of excised linear or lariat introns is 84 nt.

[B] Autoradiograph of primer extension products separated on a 7% denaturing polyacrylamide gel. Both 30% (lane 1) and 15% (lane 2) of the extension reaction were loaded on the gel. The size of the product, shown on the left, is calculated based on the DNA sequencing ladder.

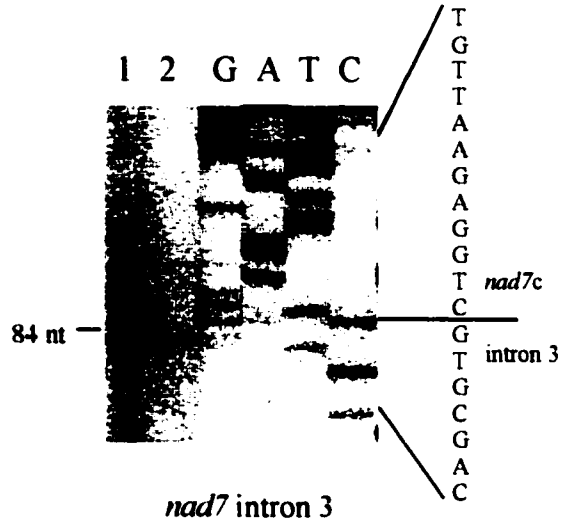
[C] Schematic of primer extension experiments for wheat *nad7* intron 4 showing the location of *nad7* intron 4 oligomer (green, LB178) which was hybridized to *nad7* intron 4 RNAs in order to direct specific cleavage of the RNA at the location of the hybrid. Black arrows indicate the end labeled oligomer (LB159) used in this experiment. Expected primer extension products are shown as orange, purple and black lines where the orange line represents the region at the 5' end of the intron, the purple line represents sequences complementary to the 3' end of the intron, and the black line represents the lariat tail. Expected sizes for extension products derived from linear (102 nt), lariat (117 nt) and circular (123 nt) excised introns are indicated.

[D] Autoradiograph of *nad7* intron 4 primer extension products generated using wheat mt RNA. Products are generated from mtRNA from 24 hr embryos (lanes 1 & 3) or 6 day etiolated seedlings (lanes 2 & 4) which were digested with RNAse H and LB178 (lanes 1 & 2) or were used as controls (lanes 3 & 4). Sizes of the expected products are indicated on the left and were calculated based on the DNA sequencing ladder. Arrows indicate high molecular weight bands which were observed but were not predicted.

A. ***nad7*** **intron 3**

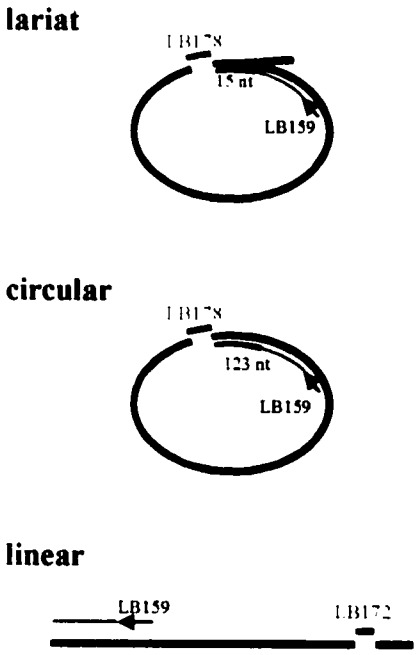


B.

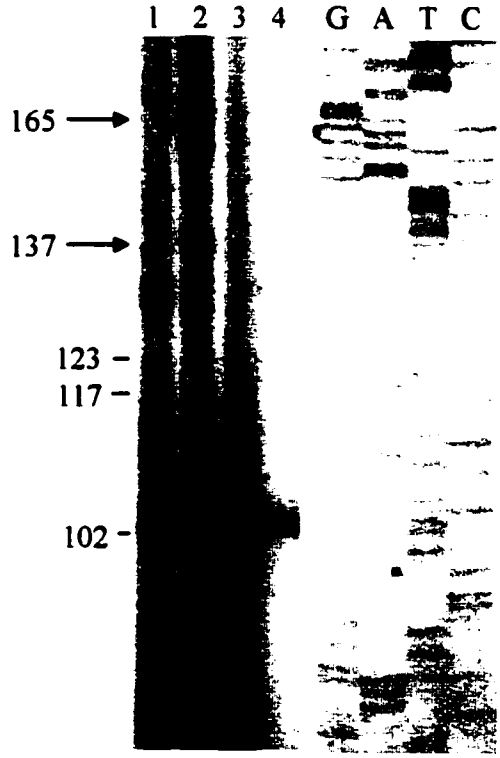


nad7 intron 3

C. ***nad7*** **intron 4**



D.



nad7 intron 4

(Figure 5.8A).

Primer extension experiments aimed at delineating the location of the branch site at the 3' end of the intron were also carried out. If intermediate splicing products in which the second transesterification step has not yet occurred are present in sufficient quantities, then it should be possible to demonstrate the location of the 2'–5' phosphodiester bond by both RNA sequencing and primer extension using a primer located within the downstream exon. Such experiments have been performed by Kim and Hollingsworth (1993) with *atpF* and *petD* introns in spinach chloroplasts where they observed high levels of splicing intermediates by northern analyses. However, none of the expected products were obtained for *nad7* introns 3 and 4 under the conditions used (data not shown) even though high levels of lariat-containing intermediates are observed in oligomer hybridization experiments with *nad7* intron 3 (section 5.3.1). These results may be improved by increasing the levels of RNA, or using RNA from 6 day seedlings.

For *nad7* intron 4, I performed a set of primer extension experiments aimed at differentiating between linear, lariat and circular forms of this intron and determining the relative amounts of each form in 24hr embryo and 6 day mtRNA preparations from wheat. A region close to the 3' end of *nad7* intron 4 was cleaved using an antisense oligomer (Figure 5.8C, LB178) and RNase H as in section 5.4.2. This cleavage was done to terminate primer extension at a specific position so that products where the reverse transcriptase had read through the branch site (see below) were of a discrete size rather than heterogeneous sizes due to stalling of the enzyme in regions of high secondary structure. A second ^{32}P end-labeled oligomer (Figure 5.8C, LB159) was bound to the 5' end of the intron and primer extension was performed using Superscript II reverse transcriptase which is able to read through 2'-5' bonds even though it stalls at such bonds (Vogel *et al.*, 1997A). This enzyme was used in RT-PCR experiments to generate cDNA across mitochondrial group II intron branch sites (cf. Section 5.2) whereas no RT-PCR products generated from across branch junctions could be amplified from cDNA generated with the AMV reverse transcriptase (data not shown). Three different extension products are expected (Figure 5.8C). The smallest product should stop at the 5' end of the intron, representing either linear intron, or a stop caused by a stalled enzyme at the unusual 2'-5' bond (102 nt). The next two possible products

would differ by six nucleotides, the smaller of which would represent the lariat form of the intron (117 nt), and the other a circular form of the intron (123 nt).

The results of preliminary experiments were unclear since numerous extension products were obtained (Figure 5.8D). Perhaps the RNA was somewhat degraded due to the RNase H treatment or the RT activity was inhibited due to impurities in the RNA sample or by extensive intron secondary structure. A 102 nt primer extension product was observed for both 24 hr and 6 day mt RNA which were cleaved with RNAse H (Figure 5.8D, lanes 1 and 2) and with 6 day mtRNA that was untreated (lane 4). Note that results observed for untreated 6 day mtRNA are similar to those obtained for *nad7* intron 3. There was a high background in the lane representing untreated 24 hr mtRNA (Figure 5.8D, lane 3), perhaps due to degradation of the RNA in this sample. In addition to multiple faint extension products, there were two major higher molecular weight species of unexpected sizes (Figure 5.8D, arrows) which were observed in mtRNA samples treated with RNAse H. These species may be the result of higher levels of specific degradation of the RNA under the conditions used for the RNAse H assay. Future experiments aimed at reducing levels of nonspecific products may require improving RNA purity, decreasing degradation of RNA during RNAse H digestion or decreasing RNA secondary structure by using higher temperatures for extension.

5.5 RT-PCR analysis of excised introns which have aberrant D6 structures.

Some plant mitochondrial introns have D6 structures which do not conform to the group II intron consensus structure in that they lack a bulging adenosine and/or cannot be folded into a helix (cf. Figure 6.4). RT-PCR experiments aimed at determining whether or not branch sites exist for such aberrant introns (*nad1* introns 1 and 2, *nad4* intron 2; cf. Chapter 4) were carried out in the same manner as for *nad7* introns 3 and 4 (cf. Sections 4.2 and 5.2). In these experiments, cDNA was synthesized using a pool of gene specific primers, so that *nad1*, *nad4* or *nad7* introns could all be amplified by PCR from the same cDNA preparation. No RT-PCR products were obtained for *nad1* introns 1 and 2 or for *nad4* intron 2, even though excised *nad7* introns 3 and 4 products were generated. A possible explanation for this is that, due to the absence of a branch site, the introns are excised by first step hydrolysis resulting in a linear, rather than lariat form of excised intron. If this was the case

then amplification of excised introns with these methods would not be possible.

To detect linear introns by RT-PCR, RNA ligase was used to circularize such introns. T4 RNA ligase, which catalyzes the ATP-dependent inter- and intramolecular ligation of single-stranded RNA, was used in dilute mtRNA preparations. RT-PCR was then performed on this population of RNA. RT-PCR products were generated in parallel with controls where no RNA ligase was added. In these experiments, several products were obtained for *nad1* introns 1 and 2. In both cases, a product of the expected size was observed with the ligation and was not present in the control where no RNA ligase was added. Such RT-PCR products were cloned and sequenced. For *nad1* intron 1 the RT-PCR product corresponded to the sequence of exon a and intron 1. Only one sequence was observed for *nad1* intron 2 and this did not correspond to any sequence in the EMBL data bank. These products may have been generated by mispriming of oligomers on related sequences, perhaps due to low stringency PCR conditions. Alternatively, the RNA ligase may not have functioned in the expected manner under the conditions used. A control for RNA ligase activity should be incorporated into future experiments.

5.6 Conclusions

Using a combination of experimental strategies, excised intron lariats and lariat containing splicing intermediates of *nad7* introns 3 and 4 were identified. No significant levels of linear forms of excised intron generated through an initial hydrolysis step in splicing were observed; however, this does not preclude the possibility that splicing via hydrolysis occurs to some degree *in vivo* since linear products may be rapidly degraded in the mitochondrion so that they are not detected. The unexpected identification of a circular form of excised *nad7* intron 4 by RT-PCR suggests the possibility that a novel splicing pathway exists *in vivo* in plant mitochondria. It will be of interest to carry out further experiments aimed at characterizing this product.

The experiments described in this chapter will be of use for determining the physical form of excised introns in plant mitochondrial introns lacking typical structures such as a bulging adenosine in D6. Preliminary characterization of such aberrant introns by RT-PCR indicates that these introns may not be excised via the typical transesterification mechanisms. Further examination of such introns will provide insight into possible alternate mechanisms of splicing in plant mitochondria.

CHAPTER 6:
GENERAL DISCUSSION

6.1 RNA editing in plant mitochondrial introns

Many plant mitochondrial group II introns appear to have a "degenerate" structure when compared to autocatalytic introns, particularly within the D5 and D6 helices. For example, most of the plant mitochondrial introns have mispairs within these helices. Part of this project involved the analysis of RNA editing of AC mispairs within group II intron helical regions or within regions involved in long range interactions to examine the role of editing of these regions in group II intron splicing in plant mitochondria. I have looked at editing within [1] D5, which is the most highly conserved helix of conventional group II introns and is part of the catalytic core of self splicing introns, [2] D6 which includes the bulging adenosine involved in the first step of splicing and [3] exon sequences close to splice junctions which are involved in long range pairing with the intron. It is notable that when excised introns were examined (*nad7* introns 3 and 4; cf. Chapter 4), lack of editing at certain positions indicated that this mispairing did not affect the splicing competency of these introns.

For D5, eight predicted AC mispairs were examined in intron-containing transcripts or in excised introns. The results of these analyses are summarized in the histogram in Figure 6.1B (red, pink and blue boxes). Six predicted editing sites occurred within *nad7* introns 1, 3 and 4 and these were each examined in at least five plants (*Arabidopsis*, pea, wheat, soybean, rice and/or tobacco). The remaining two positions occurred within *nad4* intron 2 and *nad5* intron 2 and these were each examined in at least three of the above plants including both monocots and dicots. Only 3 out of 8 positions showed editing (Figure 6.1B, red boxes), and these occurred within *nad7* intron 3 at position #11 and in *nad7* intron 4 at positions #6 and #11. Moreover, editing at these three sites was not conserved among plants: at position #11 of D5 from *nad7* intron 4, editing was not observed in tobacco and this position was only partially edited in the population of excised *Arabidopsis* introns. The comparable position in *nad7* intron 3 and at position #6 in *nad7* intron 4 was not conserved at the DNA level due to a genomically encoded T in some plants; however, in these cases, pairing at these positions was observed in all plants at the RNA level due to editing. Mispairings which could not be corrected at position #6, (e.g., UU, *nad4* intron 3) or were

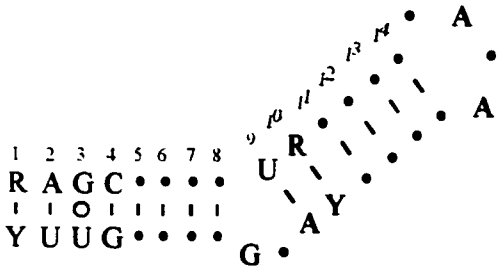
Figure 6.1. Summary of mispairing at helical positions in plant mitochondrial domain 5 structures.

[A] Classical group II intron domain 5 helical structure showing invariant and highly conserved positions (Y=pyrimidine, U=purine) (Michel and Ferat, 1995). The 8 bp and 6 bp helices are numbered as positions #1 to #14 extending from the central wheel to the loop.

[B] Alternative domain 5 model with 9 bp and 5 bp helices (Costa *et al.*, 1998).

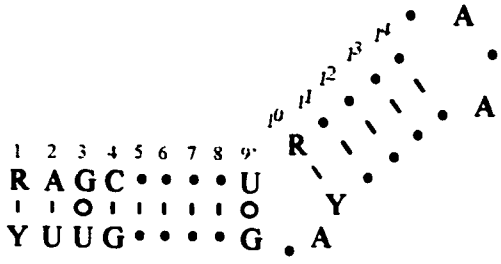
[C&D] Summary of locations of mispairs within domain 5 helices in [C] 24 flowering plant mitochondrial introns (cf. Figure 6.3; Appendix IV), excluding *rpl2* intron due to its aberrant nature, and [D] 25 group II introns from *Marchantia polymorpha* mitochondria (Oda *et al.*, 1992; Appendix V). An additional pairing position (9') represents the alternative pair #9 in the 9bp + 6bp domain 5 model. Red bars represent AC mispairs which are edited to conventional AU pairs in all plants examined, pink bars represents AC mispairs which are edited in only some species, blue bars represent unedited AC mispairs and open boxes indicate AC mispairs which have not yet been examined. Other mispairs are represented by yellow boxes (YY) and green boxes (RR).

A.



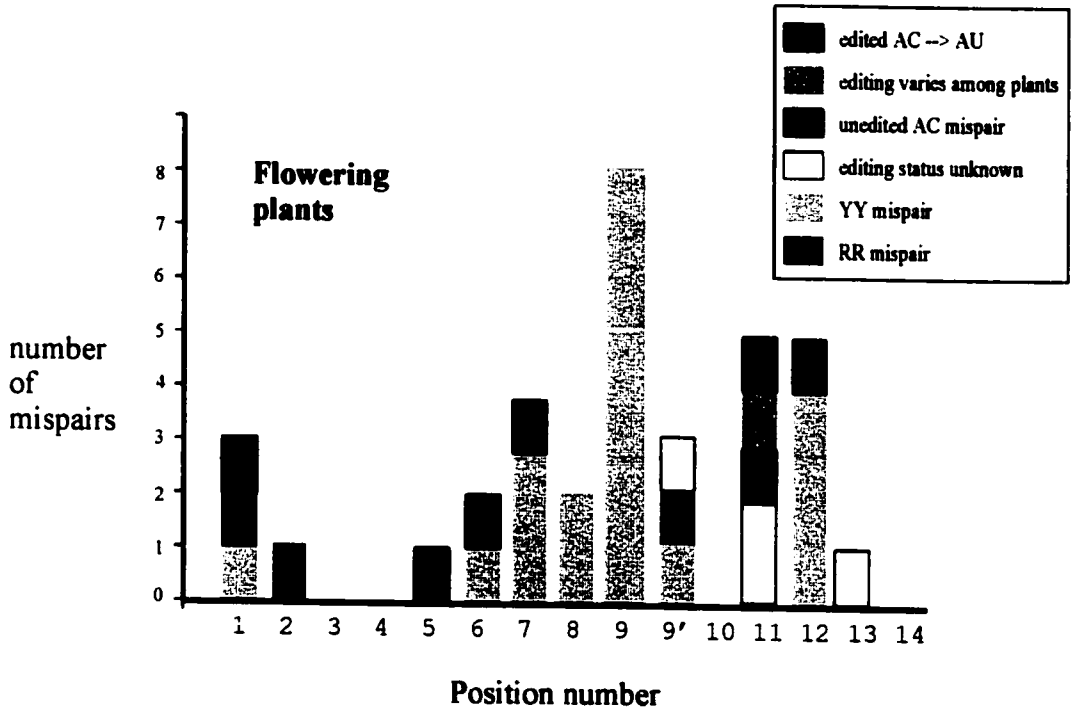
8+6 Structure

B.

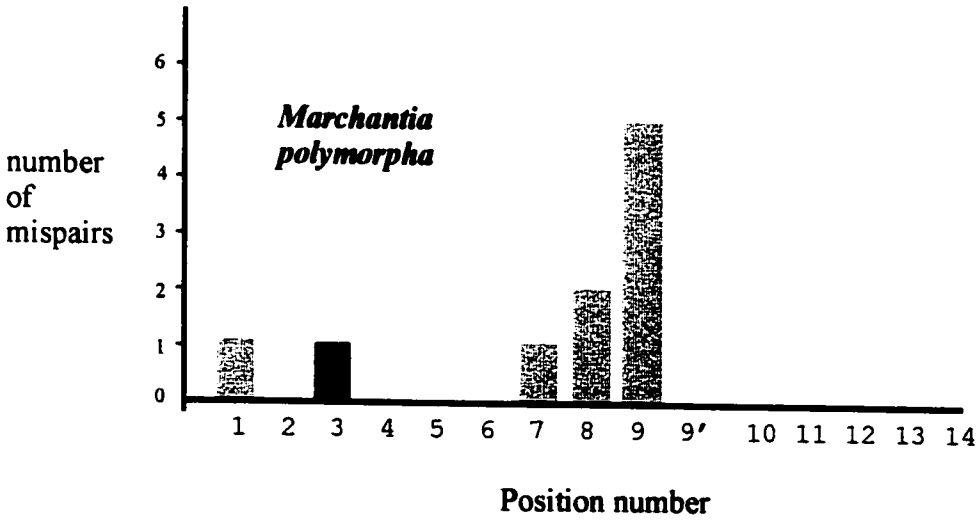


9+5 Structure

C.



D.



not corrected by editing at position #11, (AC, *nad5* intron 2) were observed in other introns (Figure 6.1B, yellow and blue boxes). This indicates that Watson-Crick pairing at positions where editing is observed is not absolutely required in all introns. An AC mispair within the *rpl2* intron was not included in the D5 histogram due to its unusually short structure lacking 2 base pairs (cf. Figure 6.3). The *rpl2* AC mispair was not edited in excised introns (Subramanian, 1999); however, if this structure was folded into a 9 bp + 3 bp structure rather than a 8 bp + 4 bp structure (cf. Appendix VIII), the AC mispair would no longer exist.

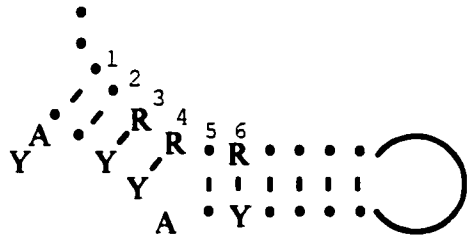
Within D6, editing in intron-containing transcripts or excised introns has also been examined at four predicted positions. Figure 6.2B summarizes my results along with other D6 editing positions which have been reported to date (Appendix III). Interestingly, 4 out of the 6 positions examined are edited, at either position #5 (*nad1* intron 3; Wissinger *et al.*, 1991) or position #6 in *nad1* intron 4 (Farré and Araya, 1999), *nad5* intron 2 (Knoop *et al*, 1991; this work) and *nad7* intron 2 (tobacco, this work; *Arabidopsis*, Giegé and Brennicke, 1999). Editing at position #6, the second pairing after the bulging adenosine, may be important for splicing since this position is either paired or edited in all of the seed plant introns examined; however, this position is mispaired in one out of the 25 introns in *Marchantia* (Figure 6.2C) and in *nad1* intron 4 this position is not edited in *Petunia* although an adjacent position is edited (Conklin *et al.*, 1991). Editing of an *nad1* intron 3 AC mispair at position #5, adjacent to the bulging adenosine, was observed *in vivo* and was shown to be essential for self-splicing in a chimeric *Oenothera* / yeast aI5γ ribozymic intron (Börner *et al.*, 1995). However, at comparable positions in other plant mt introns, mispairings that cannot be corrected by editing are observed in *nad4* intron 1 and *nad5* intron 3 of monocots (cf. Figure 6.4F). Furthermore, this position is mispaired in 6 out of 25 introns in Marchantia (Figure 6.2C), so that pairing at this position cannot be important for splicing in these introns. Interestingly, in conventional group II introns, the branch site adenosine is typically flanked by GU pairs and splicing of introns with GC pairs in these positions is less efficient *in vitro* (Chu *et al.*, 1998). Most plant mitochondrial intron branch sites are flanked by GC pairs (cf. Section 6.2). Perhaps mispairing in the region surrounding the branch site improves the efficiency of splicing for some introns. The requirement for editing at certain positions may vary among introns or among organisms. For example, *Marchantia* mitochondria may

Figure 6.2. Summary of mispairing at helical positions in plant mitochondrial domain 6 structures.

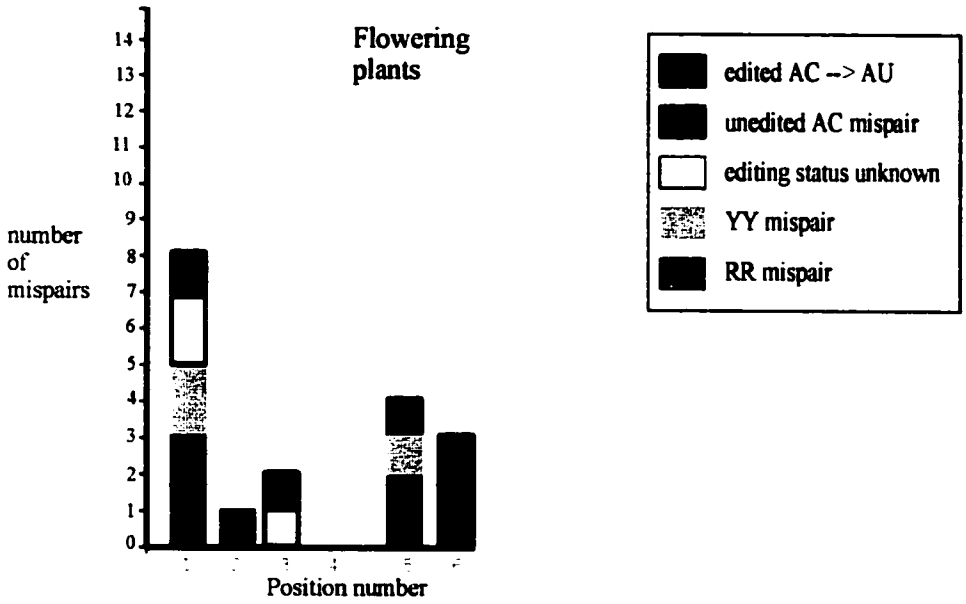
[A] Classical domain 6 helical structure showing invariant and highly conserved positions (Y=pyrimidine, U=purine)(Michel and Ferat, 1995). The 4 bp and 2 bp helices are numbered as positions #1-#6 extending from the central wheel to the loop. The conserved bulging adenosine is indicated with an asterisk.

[B&C] Summary of locations of mispairs within domain 6 helices in [B] 21 flowering plant mitochondrial introns excluding *nad1* intron 1 and 2, *nad 2* intron 1 and *nad4* intron 2 as these lack a bulging adenosine (Appendix VI) and [C] 25 group II introns from *Marchantia polymorpha* mitochondria (Oda *et al.*, 1992; Appendix VII). Red bars represent AC mispairs which are edited to conventional AU pairs in at least one plant examined, Blue bars represent unedited AC mispairs and open boxes indicate AC mispairs which have not yet been examined. Other mispairs are represented by yellow boxes (YY) and green boxes (RR).

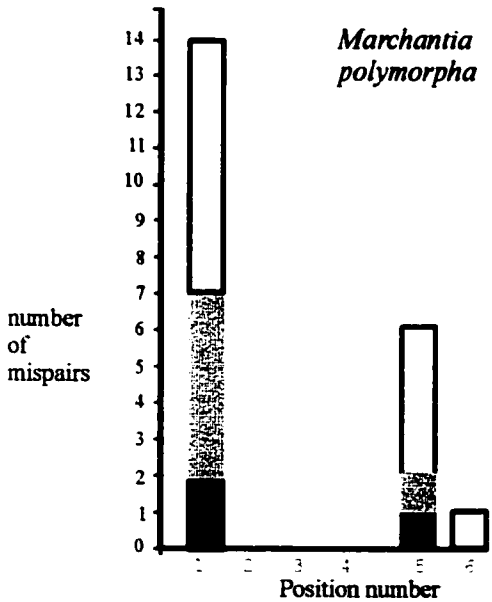
A.


1
2
3
4
5
6
A
Y
Y
R
R
Y
R
A
Y

B.


Flowering plants
number of mispairs
Position number
edited AC --> AU
unedited AC mispair
editing status unknown
YY mispair
RR mispair

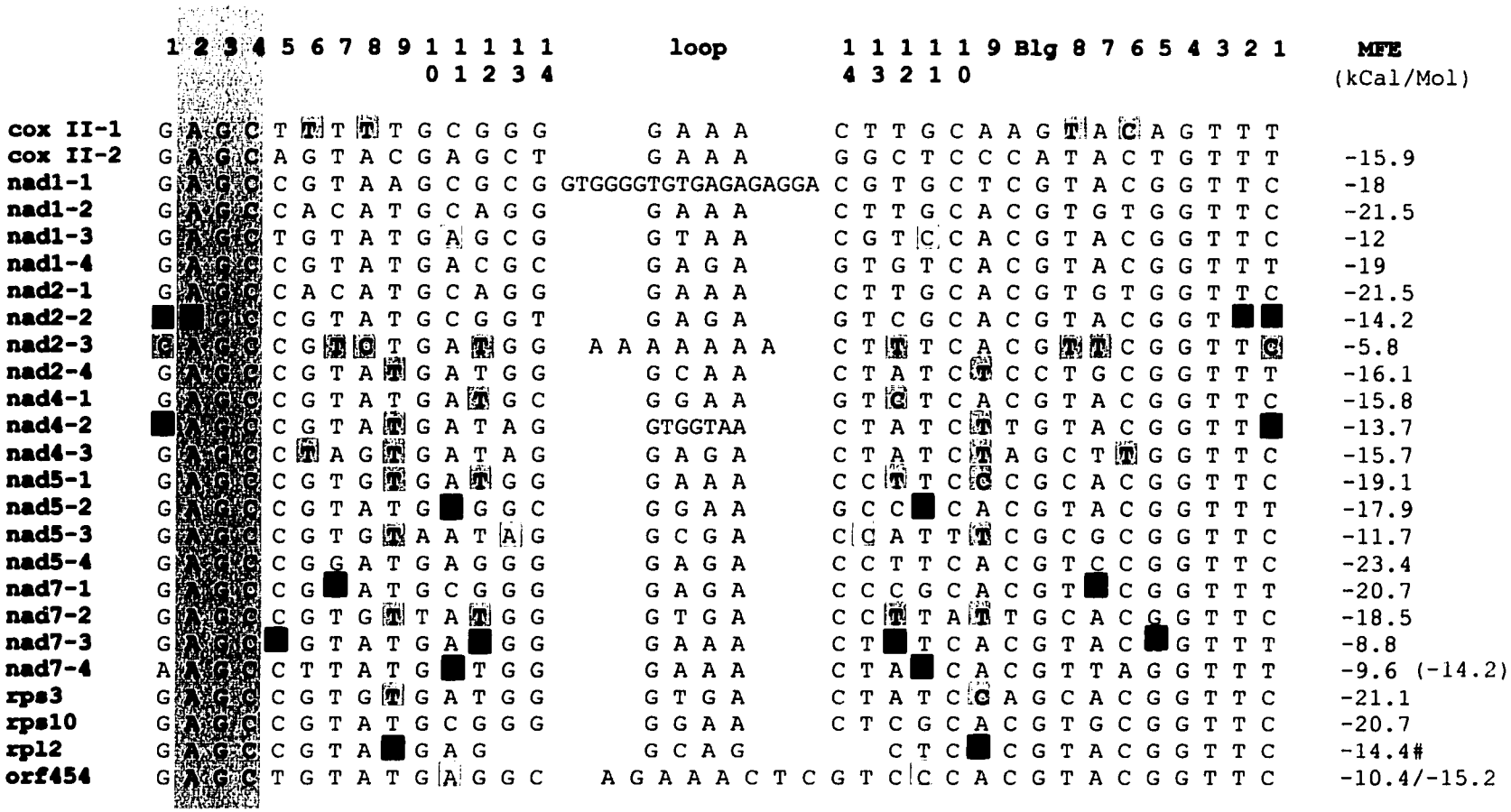
C.


Marchantia polymorpha
number of mispairs
Position number

contain factors which recognize D6 structures with a mispair past the bulging adenosine and which bind to such structures to assist in splicing. Of the positions which remained unedited, position #3 in *nad7* intron 2, located 2 nt upstream of the branch site is not edited in any of the plants examined (this work and in *Arabidopsis*, Giegé and Brennicke, 1999), and another at the base of the D6 helix (position #1) is unedited in *rpl2* of rice (Subramanian, 1999). Two other D6 editing positions occurring outside of conserved helical regions were also observed in this present work. Editing at these sites in *nad1* intron 1 and *nad7* intron 4 is not conserved among plants (cf. Chapter 4) and it is unclear whether editing at such positions is important for splicing.

In contrast to editing positions within coding sequences which can be easily predicted as they usually increase sequence similarity among plants, it is difficult to predict whether an AC mispair within helical structures of plant mt introns will be edited to an AU pair. Editing within introns does not appear to be dependent on the position within a helix: because editing is observed at a specific pairing in one intron does not mean that pairing at such a position is important for all introns. Do pairs which flank editing sites influence whether or not they are edited? In *nad7* intron 3, two AC mispairs occur side by side in wheat, compared with flanking Watson-Crick pairs at this position in *nad7* intron 4 and *nad5* intron 2. However, pairs which flank the unedited *nad5* intron 2 (GC/GU) are weaker than pairs which flank the *nad7* intron 4 position #11 (GC/AU) which is edited in some plants. The observed editing occurred in structures which had relatively low structural stability without editing. For example *nad7* introns 3 and 4 have minimum free energies of -6.3 and -9.6 kcal/mol respectively; whereas *nad5* intron 2 was more thermodynamically stable with a ΔG of -16.5 kcal/mol (Figure 6.3). It is possible that editing plays some role in improving the weakest structures within D5 for splicing. However, if thermodynamic stability was the only important factor, editing of *nad7* intron 4 would have been expected to be conserved among plants and the two other AC mispairs in *nad7* intron 3 should have been edited. Editing of core D6 sites was also confined to two positions, and these were distal to the bulging adenosine. It may be that in higher plants editing is important at such sites to properly position the bulging adenosine for the first step of splicing. Mutational analysis of self-splicing introns has illustrated the importance of the presentation of the bulging adenosine for branch site formation in the first step of splicing (Chu *et al.*,1998; Liu *et*

Figure 6.3. Alignment of domain 5 sequences from the 25 plant mitochondrial group II introns. Sequences are from the completely sequenced *Arabidopsis* mitochondrial genome (GenBank accession numbers Y08501 and Y08502). In this alignment DNA sequences from the carrot *coxII* intron 1 (X63625) and the pea *rps10* intron (X80854) were used since these introns are absent in *Arabidopsis*. Grey shading indicates nucleotide positions which are conserved among introns. The 8 bp and 6 bp helices are represented by positions 1-14 and 1'-14' (Figure 6.1A) and are separated by the dinucleotide bulge. Base-pairs other than AU, GC or GU within the helices are highlighted in yellow for pyrimidine mispairs and green for purine mispairs. AC mispairs corrected by RNA editing to AU are shaded in red, those not corrected are shaded in blue and if the editing status is unknown, mispairs are boxed in red. The minimum free energies (MFE) of domain 5 helices were determined (given in kcal/mol) using mfold version 3.0 by Zuker and Turner (Mathews *et al.*,1999; Zuker *et al.*,1999) constraining structures to either the 8+6 bp D5 structure (Figure 6.1A, Appendix IV) or the 9+5 bp structure (Figure 6.1B, Appendix IV). MFE values are given based on structures determined from DNA sequences (no editing), or from RNA sequences with observed editing (highlighted in red).


1 2 3 4 5 6 7 8 9 10 11 12 13 14
loop
14 13 12 11 10 9 Blg 8 7 6 5 4 3 2 1
MFE (kCal/Mol)
cox II-1 GACC TTTT TGCGGG GAAA CTTGCAAGTACAGTTT
cox II-2 GACC AGTACGAGCT GAAA GGCTCCCATACTGTTT -15.9
nad1-1 GACC CGTAAGCGCG GTGGGGTGTGAGAGAGGA CGTGCTCGTACGGTTC -18
nad1-2 GACC CACATGCAGG GAAA CTTGCACGTGTGGTTC -21.5
nad1-3 GACC TGTATGAGCG GTAA CGTCCACGTACGGTTC -12
nad1-4 GACC CGTATGACGC GAGA GTGTCACGTACGGTTT -19
nad2-1 GACC CACATGCAGG GAAA CTTGCACGTGTGGTTC -21.5
nad2-2 ■■CC CGTATGCGGT GAGA GTCGCACGTACGGT■■ -14.2
nad2-3 CACC CGTCTGAT GGAAAAAAA CTTTCACGTTCGGTTC -5.8
nad2-4 GACC CGTATTGA TGGA GCAA CTATCTCCTGCGGTTT -16.1
nad4-1 GACC CGTATGAT GGC GGAA GTCTCACGTACGGTTC -15.8
nad4-2 ■ACC CGTATTGA TAG GTGGTAA CTATCTTGTACGGTT■ -13.7
nad4-3 GACC CTAGTGAT AG GAGA CTATCTAGCTTGGTTC -15.7
nad5-1 GACC CGTGTGAT GG GAAA CCTTCCGCACGGTTC -19.1
nad5-2 GACC CGTATG■G GC GGAA GCC■CACGTACGGTTT -17.9
nad5-3 GACC CGTGTTAA TAG GCGA CCATTTCGCGCGGTTC -11.7
nad5-4 GACC CGGATGAG GG GAGA CCTTCACGTCCGGTTC -23.4
nad7-1 GACC CG■ATGCG GG GAGA CCCGCACGT■CGGTTT -20.7
nad7-2 GACC CGTGTTAT GG GTGA CCTTATTGCACGGTTC -18.5
nad7-3 G■CC GTATGA■G GG GAAA CT■TCACGTAC■GTTT -8.8
nad7-4 ACCC CTTATG■T GG GAAA CTA■CACGTTAGGTTT -9.6 (-14.2)
rps3 GACC CGTGTGAT GG GTGA CTATCGAGCACGGTTC -21.1
rps10 GACC CGTATGCG GG GGAA CTCGCACGTGCGGTTC -20.7
rpl2 GACC CGTA■GAG GCAG CTC■CGTACGGTTC -14.4#
orf454 GACC TGTATGAG GC AGAAACTCGTCCACGTACGGTTC -10.4/-15.2

al.,1997; Podar *et al.*,1998; Gaur *et al.*,1997). It would be interesting to examine positions of observed editing in a wider variety of plants to determine if these positions are unedited in some species.

Within introns, editing of certain AC mispairs may be essential for splicing if the improved structure is important for the catalytic activity of the intron RNA. However, such editing may simply improve the structure of the intron so that splicing is more efficient. Conversely, mispairs may be required at certain positions, perhaps for the recognition of the domain by RNA binding proteins. Editing at exon sites close to splice junctions may affect intron splicing if such positions are involved in long range pairings with the intron. For these positions editing status may be particularly important since at some positions (i.e., the δ-δ' interaction) only single nucleotides are involved in long range base pairing interactions. If such positions were edited before splicing, a putative GC pair could be converted to a less stable GU pair, which may reduce the efficiency of the splicing. In this case, editing before splicing may be selected against resulting in a temporal relationship where editing must occur after splicing.

Editing within group II introns is not as prevalent as it is within coding sequences and in many cases it is not required for splicing. While at certain positions intron editing improves structure within helical regions, in other cases it does not. Perhaps such positions are easily accessible to editing machinery or may be in the correct sequence context for editing (see below). The reduced or absence of editing in *Arabidopsis* and tobacco for both *nad7* intron 4 and *nad 1* intron 1 may indicate a reduced level of overall editing in these plants. In *Arabidopsis,* only 441 editing sites are observed (Giegé and Brennicke, 1999) compared to at least 1000 sites predicted for higher plant mitochondrial introns (Mulligan *et al.*, 1999). In some lower plants such as *Isoetes lacustris,* where exon editing is more prevalent, the case may be different. For example, D5 of *nad5* intron 2 of *Isoetes* (Malek and Knoop, 1998) has a low predicted stability based on DNA sequence alone, and thus cannot be folded into a typical structure without the assumption that certain AC mispairs are edited. The possibility that editing at certain positions in flowering plant mitochondria is necessary for splicing cannot be easily ruled out. Perhaps with increased knowledge about plant mtRNA editing mechanisms, the effects of inactivation of editing at certain sites within

group II introns could be characterized.

6.1.1 Recognition of RNA editing sites in plant mitochondria

The mechanism of RNA editing in plant organelles is still largely unknown. How does editing machinery within plant mitochondria specifically recognize hundreds of unique editing positions without editing at incorrect positions? Recent evidence from plastid transformation studies demonstrates that sequences flanking editing positions are important for delineating editing sites (reviewed in Smith *et al.*, 1997; Mulligan *et al.*, 1999). Generally sequences upstream of the editing site are important whereas nucleotides downstream of the editing site are required to a lesser degree (cf. Section 1.3.2). However, sequence requirements for editing vary depending on the site. In plant mitochondria, due to the lack of a transformation system, studies have been limited to looking at regions which are similar in sequence between closely related plants or the analysis of duplicated and transcribed segments of mitochondrial DNA (Mulligan *et al.*,1999). From these analyses it appears that sequence requirements for editing in plant mitochondria are similar to those required for RNA editing within plastids.

The study of editing within introns and exon sequences close to splice junctions has some interesting implications in regard to editing site recognition in plant mitochondria. I have looked at such positions in three different sequences where the exon editing site was within three nucleotides from the intron at positions both upstream and downstream from splice sites and editing was not observed in intron-containing transcripts (cf. Section 3.4). One possible explanation for this is that editing of these regions, which are involved in long range interactions with the intron, is required for splicing and is a "bottleneck" so that splicing occurs immediately after editing and intron-containing transcripts which are edited at such positions are not detected. An alternate explanation for the observed lack of editing near splice junction involves the mechanism of RNA editing. For example, the folding of group II introns may prevent the binding of RNA editing machinery at flanking exon sequences. However, if this was the case, editing should not have been reduced to a point where it was not detectable by direct sequencing or in any of the 22 clones examined for *nad7* exon e. The most likely explanation for the lack of editing in precursors is that RNA editing cannot occur until the intron is excised due to the requirement for a different sequence context. After intron splicing the sequence either upstream or downstream from the

editing site changes from intron sequence to the sequence of the adjacent exon. In all of the exon sequences near splice junctions that I have examined the adjacent exon sequences and intron sequences show no similarity. If the correct sequence context for editing occurred before splicing, editing of a non-silent position which did not take place before splicing could no longer occur in the mRNA resulting in errors in the translated protein.

As with plastid editing positions, sequence requirements for specifying mitochondrial editing positions vary depending on the site. For example, in a chimeric atp6 gene in rice sequence variation 1 nt downstream from the editing site did not affect editing (Kubo and Kadowaki, 1997); however, I found that editing of *nad7* exon c does not occur until the sequence 1 nt downstream of the editing site changes via excision of *nad7* intron 3. In *coxII* exon a in petunia, editing at the fourth nucleotide position upstream from the splice junction is less efficient than at other codon positions within the exon as it is observed in only 3 out of 9 clones examined (Sutton *et al.*, 1991). This also indicates a less pronounced effect of sequences downstream from an editing position. Similarly, such effects were observed in transformation studies in plastids where a change in the sequence 3 nucleotides downstream from an editing site reduced the efficiency of splicing to 45% (Bock *et al.*, 1996).

The importance of upstream sequences for delineating editing positions is supported by observations of editing in both *nad4* exon c and *nad7* exon e. However, as in plastids, upstream flanking sequence requirement is limited: I have found that editing at a position 12 nt downstream of the splice junction in *nad7* exon d occurred in 15 out of 20 clones examined (Figure 3.6A). This level of editing was comparable to that of other exon positions in the incompletely spliced transcripts examined so that editing at this position was not influenced by splicing.

The examination of editing at splice junctions is a particularly useful approach for determining sequence requirements for editing in plant mitochondria. The advantage of this approach is that there is no question as to the presence or absence of *trans* acting elements as for comparison of closely related genes between two plants. Furthermore, since functional copies of genes are being examined, rather than pseudogenes, editing is likely to be required at the sites examined. In this study, I have shown that sequences from 3 nucleotides upstream to one nucleotide downstream of the editing site are important for editing. The

study of editing at splice junctions in plants may provide valuable information as to sequence requirements for editing.

How are sequences flanking editing positions recognized? One possibility is that guide RNAs bind to transcripts to act as templates for editing. This mode of editing is found in trypanosomal mitochondria (reviewed in Smith *et al.*,1997) and it is the most likely scenario in plant mitochondria since many different editing sites must be identified and non-specific editing would be detrimental to the plant as it may cause unwanted changes to nonsilent codon positions within mRNA. If specificity of editing was conferred by short guide RNAs, then paired helical regions within RNA (such as those within D5 and D6) could potentially act in this capacity as internal guide RNAs. In this regard, editing has been observed within the acceptor stem of phenylalanine transfer RNA where it improves base-pairing and has been shown to be important for tRNA processing *in vitro* (Maréchal-Drouard *et al.*,1996; Marchfelder *et al.*,1996).

If D5 or D6 helices did serve as internal guide RNAs, then AC mispairs at homologous sites among different introns within a plant or at the same position among different plants would have all been expected to be edited. This was not the case as *nad5* intron 2 was not edited at position #11 (cf. Section 4.2; Bonen *et al.*, 1998) even though there was an AC mispair which was corrected at homologous sites in *nad7* introns 3 and 4 in wheat. Furthermore, editing at this position in *nad7* intron 4 was not conserved among plants. Perhaps the paired region within the D5 structure was too short to specify editing (only 1 to 2 base pairs before the editing site, Figure 3.1B) or the sequences flanking the editing position affected its editability. In this regard, exon positions are especially unlikely to be edited if preceded by a G (Smith *et al.*, 1997). The unedited AC mispair in D5 of *nad7* intron 1 fell into this category. In *Arabidopsis* 93.6% of the editing sites were preceded by a pyrimidine (63.1% U and 30.5% C; Giegé and Brennicke, 1999). Of the 17 predicted editing positions examined here, 8 out of 9 of the edited sites and 5 out of 6 unedited sites are preceded by pyrimidines. Thus, the presence of an upstream pyrimidine is not sufficient to confer editing of AC mispairs within helical regions. In the case of edited positions in *nad7* intron 3 (#11) and *nad7* intron 4 (#6), editing sites were flanked by a mispair. It is possible that editing machinery requires such a mispair flanking the editing site and that these sites were more easily recognized due to their relaxed helical structures. However, the

editing position #11 in *nad7* intron 4 was flanked by Watson-Crick pairs. Alternatively, stable helical structures at some sites may inhibit editing by reducing the efficiency of binding of the putative guide RNAs. This may help to explain why most of the predicted editing positions within D5 were not edited. The development of a plant organellar *in vitro* editing assay or the isolation and characterization of plant organellar editosomal complexes will help to extend our knowledge of the mechanism of editing in these systems.

6.2 Plant mitochondrial introns have unusual D5/D6 structures

In addition to non-canonical AC mispairs, the *nad7* D5/D6 helices also show three UU mismatches (cf. Figure 3.1B). Such pyrimidine-pyrimidine mispairs are expected to further destabilize helical structure (SantaLucia *et al.*, 1991). To assess the extent to which plant mitochondrial D5 sequences deviate from the conventional group II structure, sequences from the 25 different known introns were aligned (Figure 6.3, Appendix II) and sites which did not show the anticipated Watson-Crick or the GU wobble base-pairing (cf. Figure 6.1A) were scored (Appendix IV). Sequences from several different plants have been used because not all 25 introns are present or have been sequenced in any one plant. When sequence data from a number of plants were included in the analysis, similar results were obtained because the majority of nucleotide substitutions neither improve nor weaken base-pairing (Appendix II). Several length variations are seen among the D5 sequences of these introns: *nad1* intron1, *nad2* intron 3, *nad4* intron 2 and *orf454* introns have longer than normal terminal loops and the helix of *rpl2* is shorter by 2 bp (cf. Appendix VIII). The *rpl2* intron was consequently omitted from from the histogram in Figure 6.1C. The calculated thermodynamic stabilities of the D5 helices when constrained to the classical 14 bp helix (Figure 6.1A) are presented in Figure 6.3. It is evident that only some of the D5 helices with the least favourable free energies (cf. Section 6.1) are candidates for editing; others such as *nad4* intron 3, *nad2* intron 3 and *nad5* intron 1 have pyrimidine-pyrimidine mispairs.

In Figure 6.1, the profiles of mispairing at each of the 14 paired positions within the D5 helical region, based on the conventional structure (Figure 6.1A) or the 9bp + 5bp structure (Figure 6.1B), are shown for the mitochondrial introns of flowering plants (Figure 6.1C, Appendix IV) and those of the 25 mitochondrial group II introns of the bryophyte, *Marchantia polymorpha* (Figure 6.1D; Oda *et al.*, 1992, Appendix V). *Marchantia* was used in this comparison since RNA editing has not been observed nor is it predicted to be required

in this plant (Steinhauser *et al.*, 1999). It should be noted that only one of these introns, namely *nad2* intron 3, is located at homologous sites in flowering plants and *Marchantia polymorpha*. Approximately one-third of the plant D5 mispairs are AC mispairs which were editing candidates (cf. Section 6.1) and the vast majority of the rest are pyrimidine-pyrimidine mismatches (yellow boxes). It is clear that flowering plant mitochondrial introns have more mispairs within D5 helices than those of either *Marchantia polymorpha* or fungal mitochondrial or bacterial ribozymic group II introns, the latter having no such mispairs.

An alternate D5 structure consisting of 9 bp and 5 bp helical stretches (Figure 6.1B) has recently been proposed (Costa *et al.*, 1998). Using this model, the relaxed base pairing adjacent to the dinucleotide bulge is greatly reduced in both seed plants and *Marchantia* (Figure 6.1C & D, 9 vs 9'). In plants the number of mispairs is reduced by 5 when this model is applied to the D5 structure (Figure 6.1C, position 9 vs 9'). There are 3 mispairs at position 9' in the 9+5 model and 2 of these mispairs occur in both D5 foldings; however, position 9 in *nad1* intron 1 is a GA mispair in the 9+5 structure, where it is an AU pair in the 8+6 model. Interestingly, in the remaining plant mitochondrial introns this position is a GU wobble pair. The differences in the D5 folding models are more striking in *Marchantia*, where 5 mispairs at position 9 in the 8+6 model become a conserved GU wobble pair in all but one of the introns for the 9+5 model (Figure 6.1D, position 9 vs 9').

Within D5, positions #3 and #4, which are among the nucleotides which are known to be essential for self-splicing *in vitro* (Boulanger *et al.*,1995; Peebles *et al.*,1995), invariably show conventional pairing in higher plants (Figure 6.1C). The fact that higher plant mt introns have maintained these catalytically critical nucleotides and pairings within D5 indicates that, despite many unusual features, the D5 helix is still somewhat constrained. Perhaps proteins have become involved to strengthen the folding of this domain.

Plant mitochondrial D6 structures were also examined. For D6 conserved helical structures are shorter and the length and sequence of the entire domain is more variable than those of D5 (Appendix II). Only 4 pairings before and 2 pairings after the bulging adenosine are generally maintained (Figure 6.2A), even for catalytic introns. Several plant mitochondrial introns have D6 structures which do not conform to the group II intron consensus structure (Figure 6.4A). For example, the D6 structures of *nad1* introns 1 and 2, *nad2* intron 1 and *nad4* intron 2 all lack a bulging adenosine (Figure 6.4 B, C, D and E).

These introns were therefore not included in the histogram in Figure 6.2B since helical positions were scored relative to the bulging adenosine. In the structures examined, differences in levels of mispairing between flowering plant and *Marchantia polymorpha* introns were not as striking as those observed for D5 (Figure 6.2 B & C; Appendices VI and VII). In both cases, the position at the base of the helix was often mispaired (Figure 6.2B&C, position 1): there are 7 mispairs in plants and 14 in *Marchantia*. Pairing at position 4 and position 6 is conserved among all plants examined (Figure 6.2B). Interestingly, in all plant mitochondrial introns examined, position #4 preceding the bulging adenosine is a GC pair, except in *nad7* intron 3, where it is a GU wobble pair (Appendix VI). In *Marchantia*, pairing at positions #2, #3 and #4 is maintained in all introns; however, positions #5 and #6 were mispaired more often than in flowering plants.

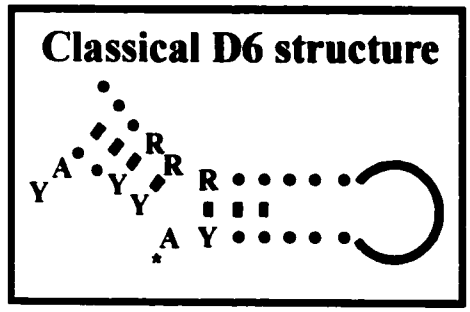
In addition to the conserved D6 structure, typical group II introns have several conserved features. For example, normally the bulging adenosine is flanked one or two GU wobble pairs (Michel *et al.*, 1989). In plants, only the *rps10* intron is flanked by two GU pairs and an additional four introns have at least one GU pair next to the bulging adenosine. Of the remaining introns, 11 branch sites are flanked by GC pairs. This observation is unexpected since the presence of flanking GC pairs reduces the rate of self-splicing of aI5γ introns three times and results in an increase in splicing by hydrolysis (Chu *et al.*, 1998). Other features of D6 which are common in group II introns, although not as highly conserved, are a terminal tetraloop which interacts with domain 2 (Chanfreau and Jacquier, 1996; Costa *et al.*, 1997) and an internal loop distal to the bulging adenosine (Chu *et al.*, 1998). While terminal loops were observed in predicted D6 structures of plant mitochondrial introns which I examined (i.e., *nad7* introns 3 and 4, *nad1* intron 1 and *nad4* intron 2), loops were not of the GNRA type normally involved in such long range interactions (Costa *et al.*, 1995). As to internal loops past the bulging adenosine, either loops or hairpin structures were present in most of the plant mitochondrial introns examined. Plant mitochondrial D6 structures are unusually long relative to those in self-splicing introns, with loops that range from 3 to 194 nt in length (cf. Appendix VI). In fact, a D6 structure from *nad1* intron 3 of *Oenothera* which was inserted into the yeast self-splicing aI5γ intron needed to be shortened to regenerate the self-splicing ability of this chimeric intron (Börner *et al.*, 1995).

Figure 6.4. Unusual domain 6 structures in plant mitochondrial introns.

[A] Structure of a typical domain 6 within group II introns showing invariant and highly conserved positions (Y=pyrimidine, U=purine, cf. Figure 6.3.)

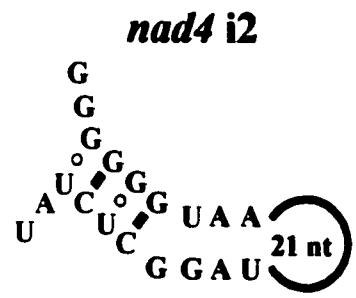
[B, C, D, E & F] Domain 6 structures with unconventional features in flowering plant mitochondria. Wheat sequences are shown for *nad1* intron 2 [B], *nad4* intron 2 [C], *nad1* intron 1 [D], *nad2* intron 1 and *nad5* intron 3 [F]. Regions which vary among plants are indicated.

A



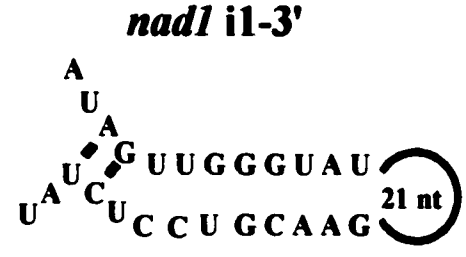
B

C



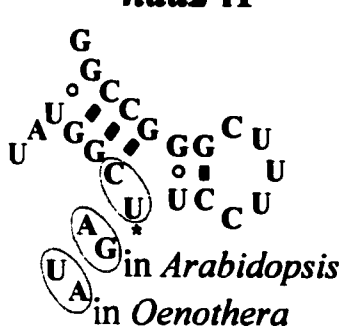
variable among plants

D

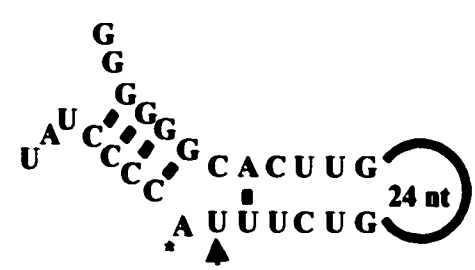


variable among plants

E



F



G in dicots

It is clear that seed plant mitochondrial group II introns have atypical features compared to ribozymic introns (Appendix VIII). In fact only five introns, namely *nad1* intron 2, *nad1* intron 4, *nad2* intron 1, *nad5* intron 4 and *rps*10 have typical D5 structures containing no mispairs (Appendix IV). Of these, *nad1* intron 2 and *nad2* intron 1 have atypical D6 structures (Figure 6.4 B and E), so that only three out of the 25 angiosperm mitochondrial introns identified to date have D5 and D6 helices which conform to the consensus structure. If we were to consider domain 1 and other core group II intron features, and the fact that *nad1* intron 4 is *trans*-splicing in some plants, this number may be decreased even further.

Regardless of unusual features, plant mt introns are readily identified as group II-type. For example, despite mispairings within helices, D5 structures are still usually conserved in length (34 nt) and, to some degree, in sequence with invariant residues A2,G3,C4 (Figure 6.3, yellow shading and Appendix II). In contrast, the plastids of the protists *Euglena gracilis* and *Astasia longa,* have group II-type introns with highly degenerate features (reviewed by Michel and Ferat, 1995). These introns are rather small (277-671 nt), and must be folded using relaxed pairing rules for stems and tertiary interactions. All *Euglena* group II introns have D5 and D6, however terminal loops of these domains can be variable and helical structures tolerate many mispairs. Similarly, in seed plant mt introns these terminal loops are also variable, however D5 has the typical terminal tetraloop in all but four introns. Interestingly, these protist introns are excised in a lariat form despite aberrant features (Copertino and Hallick, 1993). Plant mitochondrial introns may reflect an intermediate evolutionary phase, between self-splicing introns and degenerate ones such as those in *Euglena.*

Self-splicing yeast mitochondrial introns have been extensively studied *in vitro* and many of the features exhibited in plant mt introns have been shown to impede self-splicing *in vitro.* However, in a recent study where effects of certain mutations were examined *in vivo* and *in vitro* Hollander and Kuck (1999A) found that *in vitro* studies could not necessarily be applied to activities *in vivo.* Notably, when the entire D6 was removed, self-splicing occurred *in vitro* by first step and second step hydrolysis with inaccurate 3' splice site selection but *in vivo,* no excised intron was detected. In contrast, plant mt introns are capable of accurate splicing *in vivo,* despite aberrant features, but these naturally occurring introns

have likely co-evolved with factors which aid the intron in carrying out splicing.

6.3. Variation in nucleotide conservation of plant mitochondrial introns.

Some plant mt group II intron sequences are very highly conserved. For example, in the first intron of *nad4* the substitution rate is about half of that observed at silent positions within *nad4* coding sequences (Yang *et al.*,1999) and the first intron of the *coxII* is even more highly conserved, having a substitution rate roughly equivalent to that of the *coxII* exon sequences. Mitochondrial genes in flowering plants are known to accumulate nucleotide substitutions approximately three times slower than chloroplast genes and twelve times slower than nuclear genes due to differences in mutation rates among the three genomes (Wolfe *et al.*, 1987). There is some variation in rate of substitution among different mitochondrial genes. For example, the rate of substitution at silent sites in *nad4* was only 60% of that of the average silent rate in nine other mitochondrial genes examined (Laroche *et al.*, 1997). For *nad1* intron 2 and *nad2* intron 1 in *Oenothera* (and other higher plants), 48 nts which include the entire D5 strucure are identical due to a gene conversion event (Lippok *et al.*, 1994). In contrast some mitochondrial introns are not very highly conserved. For example, a survey of five introns and exons indicated that the rate of nucleotide substitution is generally similar to that at silent positions (Laroche *et al.*, 1997). However, in this study, variable loop regions were included in the calculations of nucleotide substitution rates. When intron sequences are compared by dividing them into individual domains, a clearer picture of their conservation emerges. In a study by Learn *et al.*(1992) the rate of substitution of individual group II intron domains was determined for a chloroplast intron within the gene encoding tRNA-Val^{UAC}. In this study four monocots (barley, maize, *Cenchrus setigerus* and *Pennisetum glaucum*) and two dicots (tobacco and pea) and *Marchantia polymorpha* introns were compared. Domain 2 was found to have the highest substitution rate which exceeded that of synonymous sites in coding regions, as did regions of domain 1. D5 and core regions of domain 1 were the most highly conserved with rates that were lower than nonsynonymous substitution rates in some protein coding genes, as would be expected if these regions of RNA are involved in catalyzing the splicing reaction. Interestingly, the loops of D5 and D6 were also relatively highly conserved. As expected, the conservation of intron regions correlates well to the importance of individual structures to self-splicing.

Comparative RNA analysis of the D5/D6 region of mitochondrial group II introns

from different seed plants suggests that some of these sequences and structures are not under the strong functional constraint observed for ribozymic group II introns (cf. Chapter 4, Appendix II). In some introns, sequence conservation among plants is unexpectedly poor and in several cases, differences in RNA editing status among plants further contributes to this divergence. Domain 5 structures exhibit fewer nucleotide substitutions (cf. Appendix II) compared to D6. The more unusual D6 structures appear to be more highly variable in core regions (i.e., *nad 4* intron 2, *nad2* intron 1, *nad1* intron 1; Figure 6.4C, D and E). A particularly good example of this is observed for *nad1* intron 1, where the quillwort (*Isoetes lacustris)* D5 structure is identical in sequence and has only one insertion/deletion event compared to the flowering plants. In contrast, the D6 sequence cannot be aligned and it has a conventional D6 structure in quillwort, but not in flowering plants (cf. Figure 4.5B and C). The lack of conservation of these unusual D6 structures in higher plants suggests that they are no longer under evolutionary constraint, perhaps due to their role being taken over by accessory proteins and/or RNAs. It would be interesting to determine if these D6 structures play a role in splicing in the introns in which they are found. Experiments performed *in vitro* with self-splicing introns have demonstrated that many D6 mutations do not inhibit splicing but reduce the efficiency of splicing and in some cases result in splicing without branch site formation (Podar *et al.*, 1998; Chu *et al.*, 1998). Such alternate splicing pathways have not yet been described for naturally occurring introns.

6.4 Evolution of group II introns

It has been proposed that group II introns are the evolutionary predecessors of pre-mRNA introns (Jacquier,1990; Sharp,1985; Sharp,1991; Cech, 1986; Weiner,1993). The model that the functions of the individual group II intron domains have been taken over by spliceosomal components as small generalized RNAs acting in *trans* may perhaps be paralleled by events occurring in the evolution of plant mitochondrial group II introns. The first step of such a model (cf. Stoltzfus, 1999) requires the duplication of a mitochondrial region containing an intron followed by the loss of features in the original *cis*-splicing intron. The functions of the degenerate domains would be supplied in *trans* from the duplicated intron. In the second step, other introns within the genome begin to lose their self splicing ability as they become dependent on RNAs supplied by the duplicated intron to carry out splicing. Finally, a number of other accessory factors, such as RNA binding proteins, become

involved in the splicing reaction thus forming the spliceosome.

The first step of the model is supported by observations in plant mitochondria. Six instances of fragmented introns are observed in the mitochondrial genome demonstrating that intron pieces are capable of re-associating to form splicing competent structures *in vivo* (see section 1.1.5.6). In these cases, intron pieces associate to regenerate the equivalent structure of a *cis*-splicing intron. Of particular interest is the *trans*-splicing *nad5* intron 3 of *Oenothera* which has been fragmented into three pieces (Knoop *et al.*, 1997). In this intron, an internal intron-only piece consisting of part of domain 1 through to part of domain 4 is thought to re-associate with the other intron pieces for splicing *in vivo*. This internal intron piece could potentially be involved in assisting with the splicing of other introns. Fragmented intron pieces may also be generated from duplicated intron-containing genes of which pieces are transcribed. However, even though the entire *Arabidopsis* mitochondrial genome sequence is available, no such pieces were identified using any of the plant mitochondrial intron sequences in BLASTN searches (data not shown). It may be difficult to identify pieces of a former group II intron based on nucleic acid sequence alone since intronic regions are variable in sequence despite conserved secondary structures. Group II intron pieces have been found in the nuclear genomes of plants (Knoop and Brennicke, 1994A). Notably, a region of approximately 270 kb from the *Arabidopsis* mitochondrial genome appears to have been transferred to the nuclear genome where it is incorporated into chromosome II (Lin *et al.*, 1999). Intron pieces from such transferred mitochondrial sequences could be transcribed in the nucleus and subsequently imported into the mitochondrion for splicing. RNA import into the mitochondrion is observed for several nuclear-encoded mitochondrial tRNAs (Dietrich *et al.*, 1996). If RNAs which are capable of participating in the splicing reaction became available, the function of the intron domain that they are replacing would no longer be necessary, so that any mutations to this domain would have no deleterious effects. Perhaps unusual structures which lack conservation, such as D6 of *nad1* intron 1, are examples of such domains.

As to the second and third steps in the model, they may be taking place concurrently. To date there are no reports of land plant organellar group II introns capable of self-splicing *in vitro* (Malek and Knoop, 1998) and even for fungal group II introns which are

autocatalytic, such as the yeast aI5γ intron, *trans*-acting factors are known to be required *in vivo* for efficient splicing (Grivell, 1995). Since plant mitochondria exhibit features which do not conform to conventional group II structure, proteins and/or small RNAs are expected to have become involved in their splicing. Such accessory factors could substitute for structural defects or could assist in folding of certain intronic regions. Identifying such factors in plants will be challenging since methods to isolate RNP splicing complexes involving group II introns have not yet been developed. Perhaps such binding proteins could be identified in mitochondrial extracts by their ability to bind certain intron domains.

6.5 Factors involved in plant mitochondrial intron splicing

It is likely that many of the proteins which are involved in splicing of plant mitochondrial group II introns were recruited from nuclear genes encoding proteins with RNA binding properties. Recently, a homologue of one of the factors known to be involved in group II intron splicing in yeast, the *MRS2* gene product (Wiesenberger *et al.*, 1992), has been identified in the *Arabidopsis* nuclear genome (Steinhauser *et al.*, 1998). Similarly, a nuclear-encoded mitochondrial RNA helicase has been identified in *Arabidopsis* (Gagliardi *et al.*, 1999A). RNA helicases are known to be involved in splicing in yeast mitochondria (Grivell, 1995). Other possible splicing factors include the *crs* genes required by group II introns in chloroplasts (cf. Section 1.1.5.4) which could potentially be targeted to the mitochondrion as well. It would be interesting to examine mitochondrial intron splicing in these mutants. Interestingly, within the *Arabidopsis* nuclear sequences which are now available, genes for potential splicing factors which are targeted to the plant organelles have been identified (http://www.mips.biochem.mpg.de/proj/thal/mitochloro.html). It would be of interest to examine the activity of such proteins *in vivo*.

A possible mitochondrial-encoded splicing factor is the *mat-r* gene product within *nad1* intron 4 (Farré and Araya, 1999). This intron-encoded ORF has a conserved domain X which is homologous that in yeast maturases where genetic studies have demonstrated a role for this domain in splicing of the group II intron in which it is encoded (Moran *et al.*, 1994). Similarly, the maturase (*mat-k*) encoded within the *trnK* intron retains a conserved domain X (Vogel *et al.*,1997B,1999). A plastid-encoded ORF is known to be required for splicing since ribosome deficient plastids are defective in splicing all group IIA and some

group IIB introns (Hess *et al.*,1994; Jenkins *et al.*,1997; Vogel *et al.*,1999). The *mat-k* ORF may encode a polypeptide which plays a role in splicing of a number of introns. The *mat-r* ORF may also perform a similar role for plant mitochondrial intron splicing.

Plant mt introns may be more dependent on *trans* factors for splicing than their chloroplast counterparts as they appear to be more degenerate in structure and there are six trans-splicing introns in mitochondria, including one in *Oenothera* which is tripartite, compared to only one *trans*-splicing intron in the higher plant chloroplast genome. In an algal plastid genome where two *trans*-splicing group II intron are observed in the *psaA* gene, at least 14 nuclear gene products are involved in splicing the two trans-splicing introns in *Chlamydomonas* chloroplasts (Goldschmidt-Clermont, 1990) including one which has recently been shown to be related to pseudouridine synthases (Perron *et al.*, 1999). Perhaps similar RNA-binding proteins are involved in plant mitochondrial intron splicing.

6.6 Concluding remarks

The complexity of post-transcriptional gene expression is a distinctive feature of plant mitochondria. Processes such as RNA splicing (including some cases of *trans*-splicing), RNA editing and processing of transcript termini are known to be important for the production of mature messenger RNAs and, along with factors conferring RNA stability, they can perhaps be implicated in regulation at the post-transcriptional level. Although there has been some study of these processes, information is still very incomplete in that very little is known about the mechanisms or the machinery involved in these processes. The asymmetric distribution of plant mitochondrial introns in genes encoding subunits of complex I (NADH dehydrogenase) invokes the possibility that introns are involved in regulation of these genes. If accessory factors necessary for splicing are encoded by the nucleus, then regulation of such nuclear genes could affect the expression of mitochondrial genes. The identification of factors involved in splicing in plant mt may help us understand how the nuclear spliceosomal introns evolved and may also provide insight into post-transcriptional regulation in plant mitochondria.

Differences in the steady state levels of intron-containing *nad7* precursors and excised introns have been observed for wheat in 24 hr embryos compared to 6 day seedlings and are consistent with developmental regulation of these transcripts (Chapter 3). The higher levels of precursors and excised introns in the 24 hr embryos may indicate higher levels of

splicing and perhaps different processing pathways. Such differences could be exploited to identify nuclear encoded factors which are involved in splicing which are more abundant in embryos. Studies aimed at examining post-transcriptional gene expression of plant mitochondria, with particular emphasis on splicing in this system, will serve to increase our knowledge of this unique genetic system.

REFERENCES

Abelson J, Trotta C R, Li H. 1998. tRNA splicing. *J Biol Chem 273:* 12685-12688.

Albertazzi, F J, Kudla J, Bock R. 1998. The *coxII* locus of the primitive angiosperm plant *Acorus calamus:* molecular structure, transcript processing and RNA editing. *Mol Gen Genet 259*: 591-600.

Araya A, Bégu D, Litvak S. 1994. RNA editing in plants. *Physiologia Plantarum 91*: 543-550.

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.

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

Binder S, Marchfelder A, Brennicke A. 1996. Regulation of gene expression in plant mitochondria. *Plant Mol Biol 32*: 303-314.

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

Blanc V, Jordana X, Litvak S, Araya A. 1996. Control of gene expression by base deamination: the case of RNA editing in wheat mitochondria. *Biochimie 78*: 511-517.

Bock R, Maliga P. 1995. *In vivo* testing of a tobacco plastid DNA segment for guide RNA function in psbL editing. *Mol Gen Genet 247*: 439-443.

Bock R, Hermann M, Kössel H. 1996. *In vivo* dissection of cis-acting determinants for plastid RNA editing. *EMBO J 15*: 5052-5059.

Bock R, Hermann M, Fuchs M. 1997. Identification of critical nucleotide positions for plastid RNA editing site recognition. *RNA 3*: 1194-1200.

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

Bonen L. 1993. Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J 7*: 40-46.

Bonen L, Williams K, Bird S, Wood C. 1994. The NADH dehydrogenase subunit 7 gene is interrupted by four group II introns in the wheat mitochondrial genome. *Mol Gen Genet 244*: 81-89.

Bonen L, Gray MW. 1997. The mitochondrial genome and its expression. *In* Plant Metabolism. *eds.* Dennis DT *et al.* Longman Press. pp 166-180.

Bonen L, Carrillo C, Subramanian S. 1998. Wheat mitochondrial gene expression: splicing, editing and gene transfer to the nucleus. *In* Plant Mitochondria: From Gene to Function. *Eds.* Moller IM, Gardestrom, P, Glimelius K, Glaser E. Backhuys Publishers, Leiden, The Netherlands, 153-158.

Boudvillain M, Pyle A M. 1998. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J 17*: 7091-7104.

Boulanger S C, Belcher S M, Schmidt U, Dib-Hajj S D, Schmidt T, Perlman P S. 1995. Studies of point mutants define three essential paired nucleotides in the domain 5 substructure of a group II intron. *Mol Cell Biol 15*: 4479-4488.

Boulanger S C, Faix P H, Yang H, Zhuo J, Franzen J S, Peebles C L, Perlman P S. 1996. Length changes in the joining segment between domains 5 and 6 of a group II intron inhibit self-splicing and alter 3' splice site selection. *Mol Cell Biol 16*: 5896-5904.

Börner G V, Mörl M, Wissinger B, Brennicke A, Schmelzer C. 1995. RNA editing of a group II intron in Oenothera as a prerequisite for splicing. *Mol Gen Genet 246*: 739-744.

Brown J W S, Simpson C G. 1998. Splice site selection in plant pre-mRNA splicing. *Annu Rev Plant Physiol Plant Mol Biol 49*: 77-95.

Caprara M G, Mohr G, Lambowitz A M. 1996. A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J Mol Biol 257*: 512-531.

Carola C, Eckstein F. 1999. Nucleic acid enzymes. *Curr Opin Chem Biol 3*: 274-283.

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

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

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

Chanfreau G, Jacquier A. 1996. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J 15*: 3466-3476.

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.

Chapdelaine, Y. 1992. Structure and expression of the trans-split gene for NADH dehydrogenase subunit 1 in wheat mitochondria. PhD thesis. University of Ottawa.

Chaudhuri S, Maliga P. 1996. Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site. *EMBO J 15*: 5958-5964.

Cho Y, Qiu Y L, Kuhlman P, Palmer J D. 1998. Explosive invasion of plant mitochondria by a group I intron [see comments]. *Proc Natl Acad Sci U S A 95*: 14244-14249.

Chu V T, Liu Q, Podar M, Perlman P S, Pyle A M. 1998. More than one way to splice an RNA: branching without a bulge and splicing without branching in group II introns. *RNA 4*: 1186-1202.

Combettes B, Grienenberger J M. 1999.Analysis of wheat mitochondrial complex I purified by a one-step immunoaffinity chromatography. *Biochimie 81:* 645-653.

Conklin, P.L., R.K. Wilson and M.R. Hanson. 1991. Multiple *trans*-splicing events are required to produce a mature *nad1* transcript in a plant mitochondrion. *Genes Develop 5:* 1407-1415.

Conley C A, Hanson M R. 1994. Tissue-specific protein expression in plant mitochondria. *The Plant Cell 6*: 85-91.

Copertino D W, Hallick R B. 1993. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem Sci 18*: 467-471.

Costa M, Michel F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J 14*: 1276-1285.

Costa M, Dème E, Jacquier A, Michel F. 1997. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J Mol Biol 267*: 520-536.

Costa M, Christian E L, Michel F. 1998. Differential chemical probing of a group II self-splicing intron identifies bases involved in tertiary interactions and supports an alternative secondary structure model of domain V. *RNA 4*: 1055-1068.

Costa M, Michel F. 1999. Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site. *EMBO J 18*: 1025-1037.

Covello P S, Gray M W. 1993. On the evolution of RNA editing. *Trends Genet 9*: 265-269.

Daniels D L, Michels W J, Pyle A M. 1996. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. *J Mol Biol 256*: 31-49.

Dean A D, Greenwald J E. 1995. Use of filtered pipet tips to elute DNA from agarose gels. *BioTechniques 18*: 980.

Dème E, Nolte A, Jacquier A. 1999. Unexpected metal ion requirements specific for catalysis of the branching reaction in a group II intron. *Biochemistry 38*: 3157-3167.

Dietrich A, Small I, Cosset A, Weil J H, Maréchal-Drouard L. 1996. Editing and import: strategies for providing plant mitochondria with a complete set of functional transfer RNAs. *Biochimie 78*: 518-529.

Doolittle W F. 1998. You are what you eat: a gene transfer ratchet could account for bacou are what you eat: a gene transfer ratchet could account for bacterial genes in eukaterial genes in eukaryotic nuclear genomes. *Trends Genet 14*: 307-311.

Dong F G, Wilson K G, Makaroff C A. 1998. The radish (Raphanus sativus L.) mitochondrial cox2 gene contains an ACG at the predicted translation initiation site. *Curr Genet 34*: 79-87.

Erster SH, Finn LA, Frendewey DA, Helfman DM. 1988. Use of RNase H and primer extension to analyze RNA splicing. *Nucleic Acids Res 16:* 5999-6014 .

Farré J C, Araya A. 1999. The mat-r open reading frame is transcribed from a non-canonical promoter and contains an internal promoter to co-transcribe exons nad1e and nad5III in wheat mitochondria. *Plant Mol Biol 40*: 959-967.

Gagliardi D, Kuhn J, Spadinger U, Brennicke A, Leaver C J, Binder S. 1999A. An RNA helicase (AtSUV3) is present in *Arabidopsis thaliana* mitochondria. *FEBS Lett. 458:* 337-42.

Gagliardi D, Leaver C J. 1999B. Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO J 18*: 3757-3766.

Gaur R K, McLaughlin L W, Green M R. 1997. Functional group substitutions of the branchpoint adenosine in a nuclear pre-mRNA and a group II intron. *RNA 3*: 861-869.

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

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

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

Goodall G J, Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J 10:* 2635-2644.

Grivell L A. 1995. Nucleo-mitochondrial interactions in mitochondrial gene expression. *Crit Rev Biochem Mol Biol 30*: 121-164.

Grohmann L, Thieck O, Herz U, Schröder W, Brennicke A. 1994. Translation of nad9 mRNAs in mitochondria from Solanum tuberosum is restricted to completely edited transcripts. *Nucleic Acids Res 22*: 3304-3311.

Grosskopf D, Mulligan R M. 1996. Developmental- and tissue-specificity of RNA editing in mitochondria of suspension-cultured maize cells and seedlings. *Curr Genet 29*: 556-563.

Gualberto J M, 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.

Guo H, Zimmerly S, Perlman P S, Lambowitz A M. 1997. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J 16*: 6835-6848.

Handa H, Mizobuchi-Fukuoka R, Pinyarat W. 1997. The rapeseed mitochondrial gene for subunit 2 of the NADH dehydrogenase complex: a trans-spliced structure is conserved in one of the smallest plant mitochondrial genomes. *Curr Genet 31*: 336-342.

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

Hanson M R, Sutton CA, Lu B. 1996A. Plant organelle gene expression: altered by RNA editing. *Trends Plant Sci 1:* 57-64

Hanson M R. 1996B. Protein products of incompletely edited transcripts are detected in plant mitochondria. *The Plant Cell 8*: 1-3.

Hess W R, Hoch B, Zeltz P, Hubschmann T, Kössel H, Börner T. 1994. Inefficient rpl2 splicing in barley mutants with ribosome-deficient plastids. *Plant Cell 6*: 1455-1465.

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

Hiesel R, Combettes B, Brennicke A. 1994. Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proc Natl Acad Sci U S A 91*: 629-633.

Holländer V, Kück U. 1999A. Group II intron splicing in Escherichia coli: phenotypes of cis-acting

mutations resemble splicing defects observed in organelle RNA processing. *Nucleic Acids Res 27*: 2339-2344.

Holländer V, Kück U. 1999B. Group II intron splicing in chloroplasts: identificationof mutations determining intron stability and fate of exon RNA. *Nucleic Acids Res 27*: 2345-2353.

Jacquier A, Rosbash M. 1986. Efficient trans-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5' exon-intron interaction. *Science 234*: 1099-1104.

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

Jenkins B D, Kulhanek D J, Barkan A. 1997. Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell 9*: 283-296.

Kämper U, Kück U, Cherniack A D, Lambowitz A M. 1992. The mitochondrial tyrosyl-tRNA synthetase of Podospora anserina is a bifunctional enzyme active in protein synthesis and RNA splicing. *Mol Cell Biol 12*: 499-511.

Karcher D, Bock R. 1998. Site-selective inhibition of plastid RNA editing by heat shock and antibiotics: a role for plastid translation in RNA editing. *Nucleic Acids Res 26*: 1185-1190.

Kim J K, Hollingsworth M J. 1993. Splicing of group II introns in spinach chloroplasts (in vivo): analysis of lariat formation. *Curr Genet 23*: 175-180.

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.

Knoop, V. and Brennicke, A. 1993. Group II introns in plant mitochondria-trans-splicing, RNA editing, evolution and promiscuity. In Brennicke, A. and Kück, U. (eds.), Plant Mitochondria. pp. 221-232.

Knoop V, Brennicke A. 1994A. Promiscuous mitochondrial group II intron sequences in plant nuclear genomes. *J Mol Evol 39*: 144-150.

Knoop V, Kloska S, Brennicke A. 1994B. On the identification of group II introns in nucleotide sequence data. *J Mol Biol 242*: 389-396.

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

Krug M S, Berger S L. 1987. First strand synthesis primed with oligo (dT). *Methods Enzymol 152:* 316-325.

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell 31:* 147-157.

Kubo N, Kadowaki K. 1997. Involvement of 5' flanking sequence for specifying RNA editing sites in plant mitochondria. *FEBS Lett 413*: 40-44.

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 A M, Belfort M. 1993. Introns as mobile genetic elements. *Annu Rev Biochem 62:* 587-622.

Laroche J, Li P, Maggia L, Bousquet J. 1997. Molecular evolution of angiosperm mitochondrial introns and exons. *Proc Natl Acad Sci U S A 94*: 5722-5727.

Learn G H, Shore J S, Furnier G R, Zurawski G, Clegg M T. 1992. Constraints on the evolution of plastid introns: the group II intron in the gene encoding tRNA-Val(UAC). *Mol Biol Evol 9*: 856-871.

Lewin B. 1994. Genes V. Oxford University Press Inc. New York. 911-939.

Liere K, Link G. 1995. RNA-binding activity of the matK protein encoded by the chloroplast trnK intron from mustard (Sinapis alba L.). *Nucleic Acids Res 23*: 917-921.

Lin X, Kaul S, Rounsley SD, Shea TP, Benito M-I, Town CD, Fujii CY, Mason TM, Bowman CL, Barnstead ME, Feldblyum TV, Buell CR, Ketchum KA, Lee JJ, Ronning CM, Koo H, Moffat KS, Cronin LA, Shen M, VanAken SE, Umayam L,Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Nierman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC. 1999. Sequence and analysis of chromosome II of *Arabidopsis thaliana. Nature 402:* 761-768.

Lippok B, Brennicke A, Wissinger B. 1994. Differential RNA editing in closely related introns in Oenothera mitochondria. *Mol Gen Genet 243*: 39-46.

Liu Q, Green J B, Khodadadi A, Haeberli P, Beigelman L, Pyle A M. 1997. Branch-site selection in a group II intron mediated by active recognition of the adenine amino group and steric exclusion of non-adenine functionalities. *J Mol Biol 267*: 163-171.

Logsdon JMJ. 1998. The recent origins of spliceosomal introns revisited. *Curr Opin Genet Dev 8:* 637-648.

Lu B, Hanson M R. 1994. A single homogeneous form of ATP6 protein accumulates in petunia mitochondria despite the presence of differentially edited atp6 transcripts. *Plant Cell 6*: 1955-1968.

Lu B, Hanson M R. 1996A. Fully edited and partially edited nad9 transcripts differ in size and both are associated with polysomes in potato mitochondria. *Nucleic Acids Res 24*: 1369-1374.

Lu B, Wilson R K, Phreaner C G, Mulligan R M, Hanson M R. 1996B. Protein polymorphism generated by differential RNA editing of a plant mitochondrial rps12 gene. *Mol Cell Biol 16*: 1543-1549.

Lupold D S, Caoile A G, Stern D B. 1999. Polyadenylation occurs at multiple sites in maize mitochondrial cox2 mRNA and is independent of editing status. *Plant Cell 11*: 1565-1578.

Mackenzie S. 1999. Higher plant mitochondria. *The Plant Cell 11*: 571-585.

Maier R M. 1997. Editing of plastid encoded transcripts. *Endocytobiosis & Cell Res. 12*: 41-48.

Malek O, Knoop V. 1998. Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA 4*: 1599-1609.

Marchfelder A, Brennicke A, Binder S. 1996. RNA editing is required for efficient excision of tRNA(Phe) from precursors in plant mitochondria. *J Biol Chem 271*: 1898-1903.

Maréchal-Drouard L, Ramamonjisoa D, Cosset A, Weil J H, Dietrich A. 1993. Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. *Nucleic Acids Res 21*: 4909-4914.

Maréchal-Drouard L, Cosset A, Remacle C, Ramamonjisoa D, Dietrich A. 1996. A single editing event is a prerequisite for efficient processing of potato mitochondrial phenylalanine tRNA. *Mol Cell Biol 16*: 3504-3510.

Mathews D H, Sabina J, Zuker M, Turner D H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol 288*: 911-940.

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

Michel F, 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 LII*: 201-212.

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

Michel F, Ferat J L. 1995. Structure and activities of group II introns. *Annual Review of Biochemistry 64*: 435-461.

Moenne A, Bégu D, Jordana X. 1996. A reverse transcriptase activity in potato mitochondria. *Plant Mol Biol 31*: 365-372.

Mohr G, Perlman P S, Lambowitz A M. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res 21*: 4991-4997.

Moran J V, Mecklenburg K L, Sass P, Belcher S M, Mahnke D, Lewin A, Perlman P. 1994. Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron al2. *Nucleic Acids Res 22*: 2057-2064.

Mulligan R M, Williams M A, Shanahan M T. 1999. RNA editing site recognition in higher plant mitochondria. *J Hered 90*: 338-344.

Niemer I, Schmelzer C, Börner G V. 1995. Overexpression of DEAD box protein pMSS116 promotes ATP-dependent splicing of a yeast group II intron in vitro. *Nucleic Acids Res 23*: 2966-2972.

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

Peebles C L, Perlman P S, Mecklenburg K L, Petrillo M L, Tabor J H, Jarrell K A, Cheng H L. 1987. A self-splicing RNA excises an intron lariat. *Cell 44*: 213-223.

Peebles C L, Belcher S M, Zhang M, Dietrich R C, Perlman P S. 1993. Mutation of the conserved first nucleotide of a group II intron from yeast reduces the rate allows accurate splicing. *J Biol. Chem. 268*: 11929-11938.

Peebles C L, Zhang M, Perlman P S, Franzen J S. 1995. Catalytically critical nucleotide in domain 5 of a

group II intron. *Proc Natl Acad Sci U S A 92*: 4422-4426.

Pereira d S, 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.

Perlman P S, Podar M. 1996. Reactions catalyzed by group II introns in vitro. *Methods Enzymol 264*: 66-86.

Perron K, Goldschmidt-Clermont M, Rochaix J D. 1999. A factor related to pseudouridine synthases is required for chloroplast group II intron trans-splicing in chlamydomonas reinhardtii. *EMBO J 18:* 6481-90.

Phreaner C G, Williams M A, Mulligan R M. 1996. Incomplete editing of rps12 transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. *Plant Cell 8*: 107-117.

Pla M, Mathieu C, De Paepe R, Chétrit P, Vedel F. 1995. Deletion of the last two exons of the mitochondrial nad7 gene results in lack of the NAD7 polypeptide in a Nicotiana sylvestris CMS mutant. *Mol Gen Genet 248*: 79-88.

Podar M, Chu V T, Pyle A M, Perlman P S. 1998. Group II intron splicing in vivo by first-step hydrolysis. *Nature 391*: 915-918.

Podar M, Perlman P S. 1999. Photocrosslinking of 4-thio uracil-containing RNAs supports a side-by-side arrangement of domains 5 and 6 of a group II intron. *RNA 5*: 318-329.

Puglisi J D, Wyatt J R, Tinoco I J. 1990. Solution conformation of an RNA hairpin loop. *Biochemistry 29*: 4215-4226.

Rasmusson A G, Svensson A S, Knoop V, Grohmann L, Brennicke A. 1999. Homologues of yeast and bacterial rotenone-insensitive NADH dehydrogenases in higher eukaryotes: two enzymes are present in potato mitochondria. *Plant J 20*: 79-87.

Saldanha R, Mohr G, Belfort M, Lambowitz A M. 1993. Group I and group II introns. *FASEB J 7*: 15-24.

Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz A M. 1999. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry 38*: 9069-9083.

Sambrook J, Fritsch E, Maniatis, T. 1989. Molecular Cloning: A laboratory manual, 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger F, Nicklen S, Coulson A R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A 74:* 5463-7.

Sanger H I, Ramm K, Domdey H, Gross H J, Henco K, Riesner D. 1979. Conversion of circular viroid molecules to linear strands. *FEBS letts 99*: 117-119.

SantaLucia J J, Kierzek R, Turner D H. 1991. Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs. *Biochemistry 30* : 8242-8251.

Schmelzer C, 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.

Schmidt U, Maue I, Lehmann K, Belcher S M, Stahl U, Perlman P S. 1998. Mutant alleles of the MRS2 gene of yeast nuclear DNA suppress mutations in the catalytic core of a mitochondrial group II intron. *J Mol Biol 282*: 525-541.

Schuster W, Brennicke A. 1994. The plant mitochondrial genome: Physical structure, information content, RNA editing, and gene migration to the nucleus. *Annu Rev Plant Physiol Plant Mol Biol 45*: 61-78.

Scott W G. 1999. RNA structure, metal ions, and catalysis. *Curr Opin Chem Biol 3:* 705-709.

Sharp P A. 1985. On the origin of RNA splicing and introns. *Cell 42*: 397-400.

Sharp P A. 1991. Five easy pieces. *Science 254* : 663-663.

Shields D C, Wolfe K H. 1997. Accelerated evolution of sites undergoing mRNA editing in plant mitochondria and chloroplasts. *Mol Biol Evol 14*: 344-349.

Simpson G G, Filipowicz W. 1996. Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. *Plant Mol Biol 32*: 1-41.

Sisodia S S, Cleveland D W, Sollner-Webb B. 1987. A combination of RNase H and S1 nuclease circumvents an artefact inherent to conventional S1 analysis of RNA splicing. *Nucleic Acids Res 15:* 1995-2011.

Smart C J, Monqger F, Leaver C J. 1994. Cell-specific regulation of gene expression in mitochondria during anther development in sunflower. *Plant Cell 6*: 811-825.

Smith H C, Gott J M, Hanson M R. 1997. A guide to RNA editing. *RNA 3*: 1105-1123.

Sontheimer E J, Gordon P M, Piccirilli J A. 1999. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev 13*: 1729-1741.

Soole K L, Menz R I. 1995. Functional molecular aspects of the NADH dehydrogenases of plant mitochondria. *J Bioenerg Biomembr 27*: 397-406.

Spinelli S L, Malik H S, Consaul S A, Phizicky E M. 1998. A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proc Natl Acad Sci U S A 95*: 14136-14141.

Steinhauser S, Malek O, Schock I, Beckert S and Knoop V. 1998. Group II intron in plant mitochondria: Trans-splicing and transgenic approaches. *In* Plant mitochondria: From Gene to Function. *Eds.* Moller IM, Gardestrom P, Glimelius K and Glaser E. Backhuys Publishers, Leiden, The Netherlands, 201-206.

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

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. *J Mol Evol 49:* 169-181.

Subramanian S. 1999. Analysis of Plant Mitochondrial Ribosomal Protein L2 Genes and their Expression. PhD thesis. University of Ottawa.

Sutton C A, Conklin P L, Pruitt K D, Hanson M R. 1991. Editing of pre-mRNAs can occur before cis- and trans-splicing in Petunia mitochondria. *Mol Cell Biol 11*: 4274-4277.

Sutton C A, Zoubenko O V, Hanson M R, Maliga P. 1995. A plant mitochondrial sequence transcribed in transgenic tobacco chloroplasts is not edited. *Mol Cell Biol 15*: 1377-1381.

Thomson M C, Macfarlane J L, Beagley C T, Wolstenholme D R. 1994. RNA editing of mat-r transcripts in maize and soybean increases similarity of the encoded protein to fungal and bryophyte group II intron maturases: evidence that mat-r encodes a functional protein. *Nucleic Acids Res 22*: 5745-5752.

Topping J F, Leaver C J. 1990. Mitochondrial gene expression during wheat leaf development. *Planta 182*: 399-407.

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

van der Veen R, Arnberg A C, van der Horst G, Bonen L, Tabak H F, 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.

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

van der Veen R, de Haan M, Grivell L A. 1988. RNA splicing in yeast mitochondria: DNA sequence analysis of mit- mutants deficient in the excision of introns a11 and a12 of the gene for subunit I of cytochrome c oxidase. *Curr Genet 13*: 219-226.

Vogel J, Hess W R, Börner T. 1997A. Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res 25*: 2030-2031.

Vogel J, Hubschmann T, Börner T, Hess W R. 1997B. Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. *J Mol Biol 270*: 179-187.

Vogel J, Börner T, Hess W R. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Res 27*: 3866-3874.

Wank H, SanFilippo J, Singh R N, Matsuura M, Lambowitz A M. 1999. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell 4*: 239-250.

Weber U, Beier H, Gross HJ. 1996. Another heritage from the RNA world: self-excision of intron sequences from nuclear pre-tRNAs. *Nucl Acids Res 24:* 2212-2219.

Weiner A M. 1993. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell 72*: 161-164.

Whitehouse D G, Moore A L. 1995. Regulation of oxidative phosphorylation in plant mitochondria. *In* Levings III Ch S, Vasil I K (eds). The Molecular Biology of Plant Mitochondria. Kluwer Academic Publishers. The Netherlands. 313-344.

Wiesenberger G, Waldherr M, Schweyen R J. 1992. The nuclear gene MRS2 is essential for the excision of group II introns from yeast mitochondrial transcripts in vivo. *J Biol Chem 267*: 6963-6969.

Wilson A J, 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, 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, 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, Brennicke A, Schuster W. 1992. Regenerating good sense: RNA editing and trans splicing in plant mitochondria. *Trends Genet 8*: 322-328.

Wolfe K H, Li W H, Sharp P M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A 84*: 9054-9058.

Wolstenholme D R, Macfarlane J L, Beagley C T, Thomson M C, Okada N A, Fauron C M-R. 1993. Maize mitochondrial DNA: the *nad1*gene-mat-r gene complex, a maturase-related pseudogene linked to a *nad2* exon and *nad* gene intron interrelationships. In Brennicke, A. and Kück, U. (eds.), Plant Mitochondria. pp. 221-232.

Yang A J, Mulligan R M. 1991. RNA editing intermediates of *cox2* transcripts in maize mitochondria. *Mol Cell Biol 11*: 4278-4281.

Yang A J, Mulligan R M. 1993. Distribution of maize mitochondrial transcripts in polysomal RNA: evidence for non-selectivity in recruitment of mRNAs. *Curr Genet 23*: 532-536.

Yang Y W, Lai K N, Tai P Y, Li W H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *J Mol Evol 48*: 597-604.

Yoshinaga K, Iinuma H, Masuzawa T, Uedal K. 1996. Extensive RNA editing of U to C in addition to C to U substitution in the rbcL transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res 24*: 1008-1014.

Yu W, Sánchez H, Schuster W. 1998. Assays for investigating RNA editing in plant mitochondria. *Methods 15*: 63-74.

Zanlungo S, Quiñones V, Moenne A, Holuigue L, Jordana X. 1995. Splicing and editing of rps10 transcripts in potato mitochondria. *Curr Genet 27*: 565-571.

Zeltz P, Kadowaki K, Kubo N, Maier R M, Hirai A, Kössel H. 1996. A promiscuous chloroplast DNA fragment is transcribed in plant mitochondria but the encoded RNA is not edited. *Plant Mol Biol 31*: 647-656.

Zuker, M., Mathews, D.H. and Turner, D.H. 1999. In Barciszewski, J. and Clark, B.F.C. (eds.), RNA Biochemistry and Biotechnology. Kluwer Academic Publishers.

APPENDIX I: Databank accession numbers of plant sequences used in analyses.

	Plant	Accession number
CoxII	*Triticum aestivum*	X01108
	Zea mays	V00712
	Zea perennis	U16993
	Oryza sativa	X01088
	Nymphaea alba	X79411
	Beta vulgaris	X57694, X57695, X55297, X57696
	Petunia hybrida	X17394, X17395
	Helianthus annuus	X62341
	Avena sativa	X79410
	Daucus carota	X63625
	Magnolia grandiflora	X78418
	Brassica rapa	AF036383, AF036384, AF036385, AF037264
	Arabidopsis thaliana	Y08501
	Raphanus sativus	AF036387
nad1	*Triticum aestivum*	X57967, X57966, X57965
	Oryza sativa	this work
	Zea mays	M18339, U09986, U09987
	Citrullus lanatus	X04130
	Arabidopsis thaliana	X98301, X98300, Y08501, Y08502, this work
	Oenothera berteriana	M63033, M63034
	Pisum sativum	this work
	Glycine max	this work, U09988
	Petunia hybrida	X60401, X60402, X60403
	Nicotiana tabacum	X04019, this work

	Isoetes lacustris	Y17812
	Pinus sylvestris	AJ223312
	Picea abies	AF142642
	Vicia faba	M30176
	Osmunda regalis	Y17815
	Equisetum	Y17811
	Solanum tuberosum	AF095277
nad2	*Triticum aestivum*	Y14433, Y14434
	Zea mays	Wolstenholme *et al.*, 1993
	Arabidopsis thaliana	Y08501, Y08502
	Oenothera berteriana	M81725, M81726
	Panax ginseng	AF034121, AF034119
	Solanum tuberosum	X93575
	Asplenium nidus	Y07910
	Marsilea drummondii	Y07911
	Isoetes lacustris	Y17813
	Beta vulgaris	X16828
	Brassica napus	Handa *et al.*, 1997
nad4	*Triticum aestivum*	X57164
	Zea mays	L20487, L20488, L20489, this work
	Oryza sativa	D32052
	Arabidopsis thaliana	Y08501
	Brassica campestris	X60794
	Lactuca sativa	L12246
	Glycine max	this work
	Pisum sativum	this work

nad5	*Triticum aestivum*	M74157, M74158, M74159, this work
	Arabidopsis thaliana	X60045, Y08501, X60047, X60048
	Oenothera berteriana	X07566, X60046, X60049
	Vicia faba	Y12731, L36945, L36946
	Dolichos biflorus	X72287
	Malus domestica	D37958
	Glycine max	P. Clark
	Beta vulgaris	X55786
	Zea mays	X57445, X14709, M74160, X07378
	Brassica napus	U10423
	Asplenium nidus	Y07912
	Pisum sativum	P. Clark
	Oryza sativa	D21251, D50565, D50566
	Isoetes lacustris	Y17814
	Anthoceros crispulus	Y17809
nad7	*Triticum aestivum*	X75036
	Nicotiana sylvestris	X86706
	Arabidopsis thaliana	Y08501,this work
	Oryza sativa	this work
	Pisum sativum	this work
	Glycine max	this work
	Nicotiana tabacum	this work
rpl2	*Arabidopsis thaliana*	X72616, Y08501
	Oenothera berteriana	X80170
	Oryza sativa	D78336
rps3	*Zea mays*	X57445

	Oryza sativa	D21251
	Arabidopsis thaliana	Y08501
	Oenothera berteriana	X69140
	Brassica napus	X63088, X68726
	Petunia hybrida	X67028
rps10	*Pisum sativum*	X80854, X14409
	Solanum tuberosum	X74826
orf454	*Oenothera berteriana*	X78036
	Arabidopsis thaliana	Y08501

APPENDIX II. Alignments of D5 and D6 sequences of plant mitochondrial introns.

Sequences highlighted in green were determined or reconfirmed in this study. Regions of nucleotide substitution relative to the first sequence shown for each intron are highlighted in pink (flowering plants) or yellow (vascular plants). Cytosines predicted to be edited are bold, those found to be edited are red and those which remained unedited are blue. Accession numbers for these sequences are listed in Appendix I. The full names of the plants are listed below.

Cox2-1

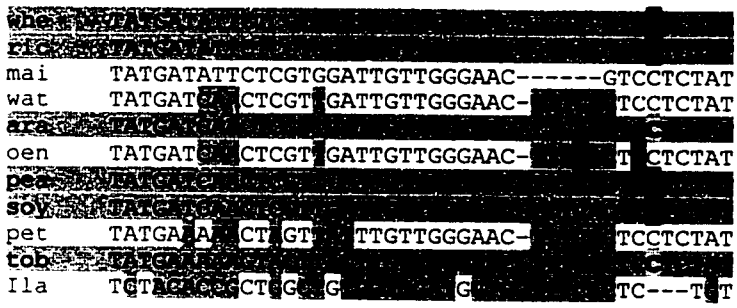
```
wh     taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
ric    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
mai    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
oat    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
car    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
pet    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
Bvul   taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
sun    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
mag    taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
Nalb   taggcgtgga-GAGCTTTTTGCGGGGAAAC-TTGCAAGTCAAGTTTGGGGGGAGGCGGGCGTCGACCCAAC
Acal   tagatgtgga-GAGCTTGTTGCGGGGAAACGTGGCAAGTCAAGTTTGGGGGAGGCGGGCGTCGACCCAAC

wh     cttatgagtattcggactat
ric    cttatgagtattcggactat
mai    cttatgagtattcggactat
oat    ct
car    cttatgagtattcggactat
pet    cttatgagtattcggactat
Bvul   cttatgagtattcggactat
Acal   cttatgagtattcggactat
```

cox2-2

```
car    ctaggagtgtGAGCAGTACGAGCTGAAAGGCTCCCATACTGTTTGGAGGGCAGGGGGCATAGATGCCAAA
Brap   ctaggagtgtGAGCAGTACGAGCTGAAAGGCTCCCATACTGTTTGGAGGGCAGGGGCATAGATGCCAAA
Rsat   ctaggagtgtGAGCAGTACGAGCTGAAAGGCTCCCATACTGTTTGGAGGGCAGGGGCATAGATGCCAAA
ara    ctaggagtgtGAGCAGTACGAGCTGAAAGGCTCCCATACTGTTTGGAGGGCAGGGGGCATAGATGCCAAA

car    CAAACCTGACCCCTCTctatcgtcgtagaagctgtt
Brap   GAAAGCTGACCCCTATctatcgtcgtagaagctgtt
Rsat   GAAAGCTGACCCCTATctatcgtcgtagaagctgtt
ara    GAAAGCTGACCCCTATctatcgtcgtagaagctgtt
```

nad1-1

```
whe    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGATGTGCTCGTACGGTTCATAGTTGGG
ric    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGATGTGCTCGTACGGTTCATAGTTGGG
mai    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGATGTGCTCGTACGGTTCATAGTTGGG
wat    ggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
ara    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
oen    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
pea    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
soy    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
pet    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
tob    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACAGAGGACGTGCTCGTACGGTTCATAGTTGGG
Ila    ttggggtataGAGCCGTAAGCGCGGTGGGG-GTGACGTGCTCGTACGGTTCTAAGTTGG
```



nad1-2

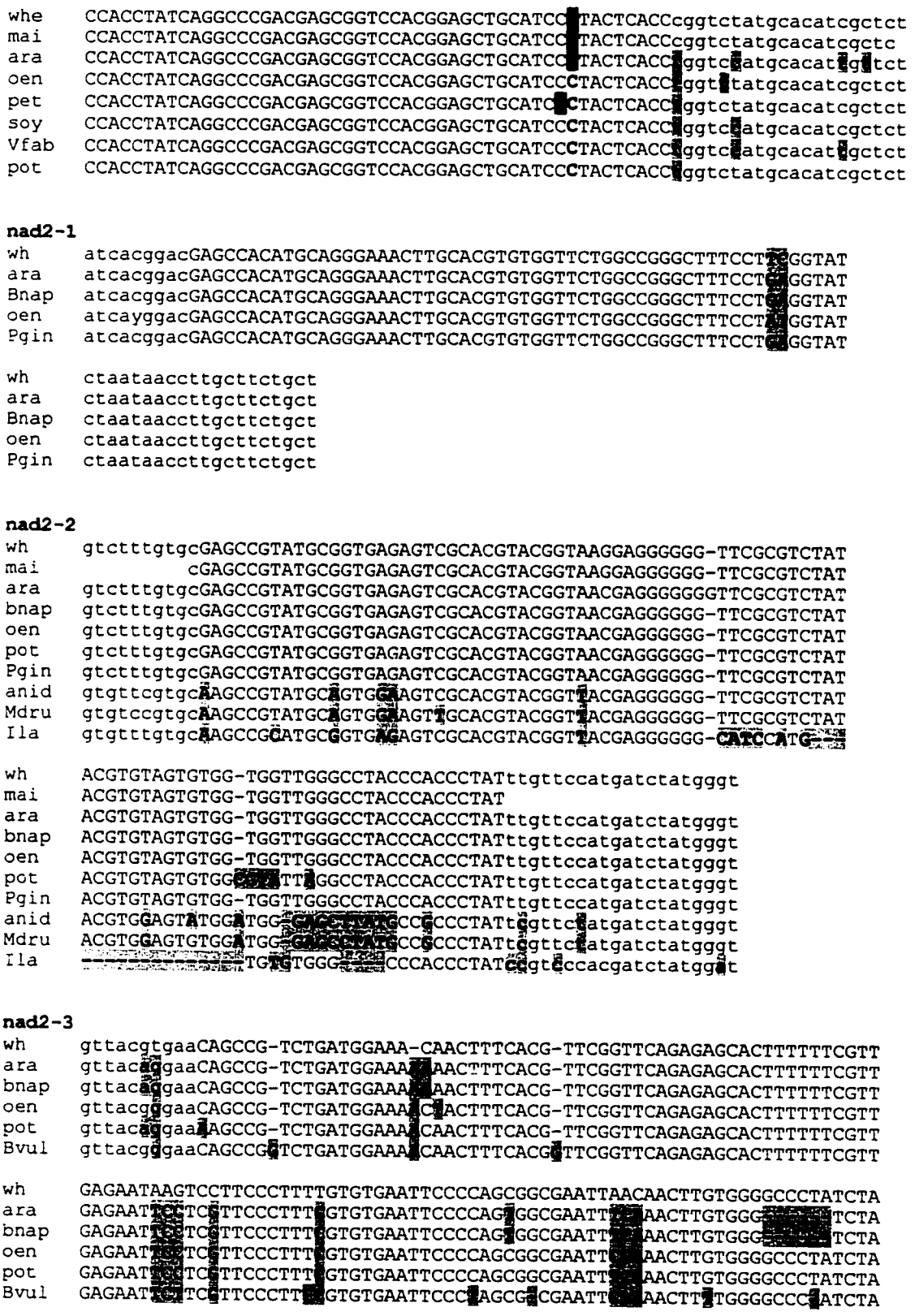
```
wh      aagcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGGA---CCCCGGTAT
ara     aagcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGGACGGCCCCGGTAT
oen     aagca ggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGAGA---CCCCGGTAT
pet     aagcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGGA---CCCCGGTAT
wat     aagcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGGA---CCCCGGTAT
pine    aagcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGCCTG CCGGGGA---CCCCGGTAT
spruc   aagcgcggacGAGCCACATNCAGGGAAACTTGCACGTGTGGTCTG CCGGGGA---CCCCGGTAT


wh      actgtactaatatgtgtagg
ara     actgtactaatatgtgtagg
oen     actgtactaatatgtgtagg
pet     actgtactaatatgtgtagg
wat     actgtactaatatgtgtagg
pine    actgtacCaatatgtgtagg
```

nad1-3

```
Wh      tgctggctggGAGCTGTATGAGCGGTAACGTCCACGTACGGCTCCGTGAGAA----GGTGGACGGAAA
mai     tgctggctggGAGCTGTATGAGCGGTAACGTCCACGTACGGCTCCGTGAGAA----GGTGGACGGAAA
ara     tgctggctggGAGCTGTATGAGCGGTAACGTCCACGTACGGTTCCGTGAGAACGGGGGTGGACGGAAA
oen     tgctggctggGAGCTGTATGAGCGGTAACGTCCACGTACGGCTCCGTGAGAACGGGGGTGGACGGAAA
soy     tgcggctggGAGCTGTATGAGCGGTAACGTCCACGTACGGCTCCGTGAGAACGGGGGTGGACGGAAA
pet     tgctggctggGAGCTGTATGAGCGGTAACGTCCACGTACGCTCCGTGAGAACGGGGGTGGACGGAAA
Vfa     tgcggctggTAGCTGTATGAGCGGTAACGTCCACGTACGGCTCCGTGAGAACGGGGGTGGGCGGAAA
Ore     taaggctggGAGCTGTATGAGCGGTAACGTTCACGTACGGTTCTGTGAGAAGCGGGGTGGACGGAAA
Ete     gttggctggGAGCCGTATGGGCGATAACGTCCACGTACGGCTCTGTGAGAAGCGGGGTAGACGGAAA

Wh      TGGCCTTGTTGTACCTCACTCCCGtcttcaatggggtctgctct
mai     TTTGTTGTACCTCACTCCCGtcttcaatggggtctgctct
ara     TGGCCTTGTTGTACCTTACTCTCGtcttcaatggggtctgctct
oen     TGGCCTTGTTGTACCT ACTCTCGtcttcaatggggtctgctct
soy     TGGCCTTGTTGTACCTCACTCTCGtcttcaatggggtctgctct
pet     TGGCCTTGTTGTACCTCACTCTCGtcttcaatggggtctgctct
Vfa     TGGCCTTGTTGTACCTCACTCTCGtcttcaatggggtct
Ore     TGGCC--ATCGTACCTTACTCTCGtcgtcaatggggtttgctct
Ete     TGGCC--ATCGTACCTTACTCTCGtcgtcaatgggattttgctgt
```

nad1-4

```
whe     ttccaggacgGAGCCGTATGAGGCGAGAGTCTCACGTACGGTTCCTTTGAGAAGGGTGTGATACCA
mai     ttccaggacgGAGCCGTATGAGGCGAGAGTCTCACGTACGGTTCCTTTGAGAAGGGTGTGATACCA
ara     ttccggggcgGAGCCGTATGAGGCGAGAGTGTCACGTACGGTTTCTTTGAGAAGGGTGTGATACCA
oen     ttccgggacgGAGCCGTATGAGGCGAGAGTGTCACGTACGGTTCCTTTGAGAAGGGTGTGATACCA
pet     ttccgggacgGAGCCGTATGAGGCGAGAGTGTCACGTACGGTTCCTTTGAGAAGGGTGTGATACCA
soy     ttccgggacgGAGCCGTATGAGGCGAGAGTCTCACGTACGGTTCCTTTGAGAAGGGTGTGATACCA
Vfab    ttccgggacgGAGCCGTATGAGGCGAGAGTCTCACGTACGGTTCCTTTGAGAAGGGTGTGAGACCA
pot                                                                   CCA
```


whe CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCTACTCACCcggtctatgcacatcgctct
mai CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCTACTCACCcggtctatgcacatcgctc
ara CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCTACTCACCggtctatgcacatcgtct
oen CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCCTACTCACCggtctatgcacatcgctct
pet CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCCTACTCACCggtctatgcacatcgctct
soy CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCCTACTCACCggtctatgcacatcgctct
Vfab CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCCTACTCACCggtctatgcacatcgctct
pot CCACCTATCAGGCCCGACGAGCGGTCCACGGAGCTGCATCCCTACTCACCggtctatgcacatcgctct
nad2-1
wh atcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGCTTTCCTCGGTAT
ara atcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGCTTTCCTCGGTAT
Bnap atcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGCTTTCCTCGGTAT
oen atcayggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGCTTTCCTCGGTAT
Pgin atcacggacGAGCCACATGCAGGGAAACTTGCACGTGTGGTTCTGGCCGGGCTTTCCTCGGTAT
wh ctaataaccttgcttctgct
ara ctaataaccttgcttctgct
Bnap ctaataaccttgcttctgct
oen ctaataaccttgcttctgct
Pgin ctaataaccttgcttctgct
nad2-2
wh gtctttgtgcGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAAGGAGGGGGG-TTCGCGTCTAT
mai cGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAAGGAGGGGGG-TTCGCGTCTAT
ara gtctttgtgcGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAACGAGGGGGGGTTCGCGTCTAT
bnap gtctttgtgcGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAACGAGGGGGG-TTCGCGTCTAT
oen gtctttgtgcGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAACGAGGGGGG-TTCGCGTCTAT
pot gtctttgtgcGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAACGAGGGGGG-TTCGCGTCTAT
Pgin gtctttgtgcGAGCCGTATGCGGTGAGAGTCGCACGTACGGTAACGAGGGGGG-TTCGCGTCTAT
anid gtgttcgtgcAAGCCGTATGCAGTGGAAGTCGCACGTACGGTTACGAGGGGGG-TTCGCGTCTAT
Mdru gtgtccgtgcAAGCCGTATGCAGTGGAAGTTGCACGTACGGTTACGAGGGGGG-TTCGCGTCTAT
Ila gtgtttgtgcAAGCCGCATGCGGTGAGAGTCGCACGTACGGTTACGAGGGGGG-CATCCATG---
wh ACGTGTAGTGTGG-TGGTTGGGCCTACCCACCCTATttgttccatgatctatgggt
mai ACGTGTAGTGTGG-TGGTTGGGCCTACCCACCCTAT
ara ACGTGTAGTGTGG-TGGTTGGGCCTACCCACCCTATttgttccatgatctatgggt
bnap ACGTGTAGTGTGG-TGGTTGGGCCTACCCACCCTATttgttccatgatctatgggt
oen ACGTGTAGTGTGG-TGGTTGGGCCTACCCACCCTATttgttccatgatctatgggt
pot ACGTGTAGTGTGGCGGATTTGGCCTACCCACCCTATttgttccatgatctatgggt
Pgin ACGTGTAGTGTGG-TGGTTGGGCCTACCCACCCTATttgttccatgatctatgggt
anid ACGTGGAGTATGGATGG-GAGCTTATGCCGCCCTATtcgttcgatgatctatgggt
Mdru ACGTGGAGTGTGGATGG-GAGCTTATGCCGCCCTATtcgttcgatgatctatgggt
Ila ----------TGTGTGGG----CCCACCCTATCCgtCccacgatctatggat
nad2-3
wh gttacgtgaaCAGCCG-TCTGATGGAAA-CAACTTTCACG-TTCGGTTCAGAGAGCACTTTTTTCGTT
ara gttacaggaaCAGCCG-TCTGATGGAAAACAACTTTCACG-TTCGGTTCAGAGAGCACTTTTTTCGTT
bnap gttacaggaaCAGCCG-TCTGATGGAAAACAACTTTCACG-TTCGGTTCAGAGAGCACTTTTTTCGTT
oen gttacggaaCAGCCG-TCTGATGGAAAACCAACTTTCACG-TTCGGTTCAGAGAGCACTTTTTTCGTT
pot gttacaggaaAAGCCG-TCTGATGGAAAACAACTTTCACG-TTCGGTTCAGAGAGCACTTTTTTCGTT
Bvul gttacggaaCAGCCGGTCTGATGGAAAACAACTTTCACGGTTCGGTTCAGAGAGCACTTTTTTCGTT
wh GAGAATAAGTCCTTCCCTTTTGTGTGAATTCCCCAGCGGCGAATTAACAACTTGTGGGGCCCTATCTA
ara GAGAATTCCTCGTTCCCTTTTGTGTGAATTCCCCAGTGGCGAATTTCAACTTGTGGGTTTTTTCTA
bnap GAGAATTCCTCGTTCCCTTTTGTGTGAATTCCCCAGTGGCGAATTTCAACTTGTGGGTTTTTTCTA
oen GAGAATTCCTCGTTCCCTTTTGTGTGAATTCCCCAGCGGCGAATTTCAACTTGTGGGGCCCTATCTA
pot GAGAATTCCTCGTTCCCTTTTGTGTGAATTCCCCAGCGGCGAATTTCAACTTGTGGGGCCCTATCTA
Bvul GAGAATTCCTCGTTCCCTTTTGTGTGAATTCCCCAGCGGCGAATTTCAACTTGTGGGGCCCTATCTA

```
wh      TTCAATCTCTCGAGCCCC-AAGAAAACCCCCCCTCTCCCTCGGACTCAATCTCTCTTTTGACTCTATat
ara     TTCAATCTCTCGAGCCCCAAGAAAACCCCCCTCCCCTCGGACTCATTCTCTTTTGACTCTATat
bnap    TTCAATCTTTCGAGCCCCAAGAAAACCCCCCTCCCCTCGGACTCATTCTCTTTTGACTCTATat
oen     TTCGATCTCTCGAGCCCCAAGAAAACCCCCCCTCCCTCGGACTCATTCTCTTTTGACTCTATat
pot     TTCGATCTCTCGAGCCCCAAGAAAACCCCCCCTCCCTCGGACTCATTCTCTTTTGACTCTATat
Bvul    TTCGATCTTTCGAGCCCC  GAAAACCCCCCCTCCCTCGGACTCAATTCTCTTTTGACTCTATat

wh      atgtgggcacctgatatc
ara     atgtgggcaccagatatc
bnap    atgtgggcaccagatatc
oen     atgtgggcacctgatatc
pot     atgtgggcacctgatatc
Bvul    atgtgggcacctgacatc
```

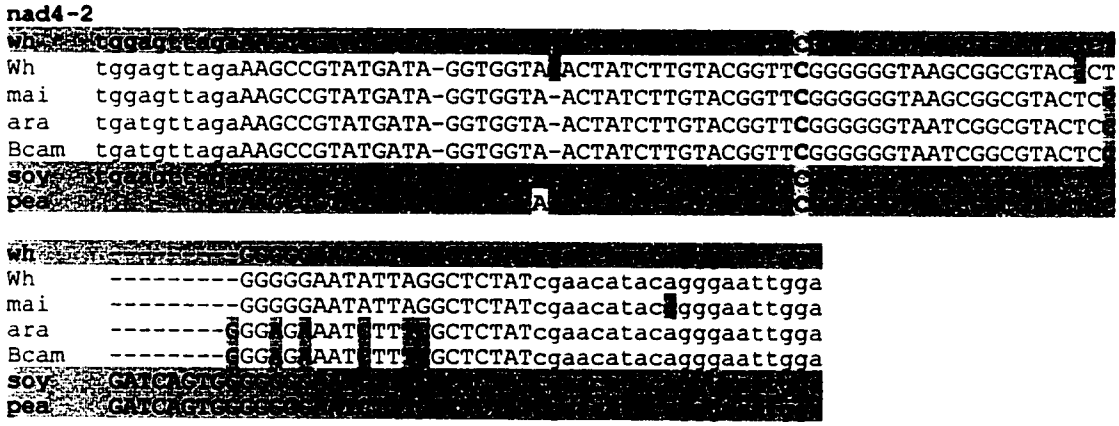
nad2-4

```
wh      tgaagttggtGAGCCGTATGATGGGCAACTATCTTCCTGCGGTTCGGAGAGGACTCAGCTGTTAGTTA
ara     ccaagttggtGAGCCGTATGATGGGCAACTATCTTCCTGCGGTTTGGAGAGGACTCAGCTGTTAGTTA
bnap    gcaagttggtGAGCCGTATGATGGGCAACTATCTTCCTGCGGTTTGGAGAGGACTCAGCTGTTAGTTA
pot     tgaagttggtGAGCCGTATGATGGGCAACTATCTTCCTGCGGTTCGGAGAGGACTCAGCTGTTAGTTA
oen     tgaagttggtGAGCCGTATGATGGGCAACTATCTTCCTGCGGTTCGGAGAGGACTCAGCTGTTAGTTA
Bvul    tgaagttggtGAGCCGTATGATGGGAAACTATCTTCCTGCGGTTCGGAGAGGACTCAGCTGTTAGTTA

wh      GTACCCCC-TTGGTTTCGGGGTGGACCCTTT--CACTCTATtttattatatacgcttagcg
ara     GTATCCCC-TTGGTTTCGGGGTGGACCCTTTTCACTCTATtttattatatacgcttagtg
bnap    GTATCCCC-TTGGTTTCGGGGTGGACCCTTT--CACTCTATtttattatatacgcttagtg
pot     GTACCCCC-TTGGTTTCGGGGTGGACCTTTT--CACTCTATtttattatatacgcttagtg
oen     GTACCCCCGTGGTTTCGGGGTGGACCCTTT--TACTCTATtttattatatacgcttagcg
Bvul    GTACCCTC-TTGGTTTCGGTGTGGACCCTTT--CACTCTATtttattatatacgcttagtg
```

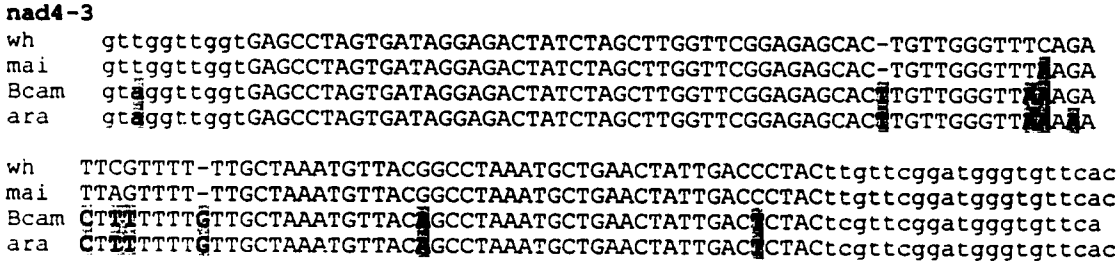
nad4-1

```
wh      gctcgggacgGAGCCGTATGATGCGGAAGTCTCACGTACGGTTCCCTGAGAAGGGAGTGGCTACCCAC
mai     gctagggacgGAGCCGTATGATGCGGAAGTCTCACGTACGGTTCCCTGAGAAGGGAGTGGCTACCCAC
ric     gctcgggacgGAGCCGTATGATGCGGAAGTCTCACGTACGGTTCCCTGAGAAGGGAGTGGCTACCCAC
ara     gctcgggacgGAGCCGTATGATGCGGAAGTCTCACGTACGGTTCCCTGAGAAGGGAGTGGCTACCTAC
Bcam    gctcgggacgGAGCCGTATGATGCGGAAGTCTCACGTACGGTTCCCTGAGAAGGGAGTGGCTACCTAC
let     gctcgggacgGAGCCGTATGATGCGGAAGTCTCACGTACGGTTCCCTGAGAAGGGAGTGGCTACCTAC

wh      TGGAGCTTCGACCAACTACCACCGGTCAATTCCGCTTTGGGGCCACCCCTGACTCTACcatcattata
mai     TGGAGCTTCGACCAACTACCACCGGTCAATTCCGCTTTGGGGCCACCCCTGACTCTACcatcattata
ric     TGGAGCTTCGACCAACTACCACCGGTCAATTCCGCTTTGGGGCCACCCCTGACTCTACcatcattata
ara     TGGAGCTTCGACCAAGTACCTCCGGTCAATTCCGCTTTGGGGCCACCCCTTACTCTACcattattata
Bcam    TGGAGCTTCGACCAAGTACCTCCGGTCAATTCCGCTTTGGGGCCACCCCTTACTCTACcattattata
let     TGGAGCTTCGACCAATTACCACCGGTCAATTCCGCTTTGGGTCCACCCCTTACTCTACcattattata

wh      ggggtatggg
mai     ggggtatggg
ric     ggggtatggg
ara     ggagtatggg
Bcam    ggagtatggg
let     ggggtatggg
```

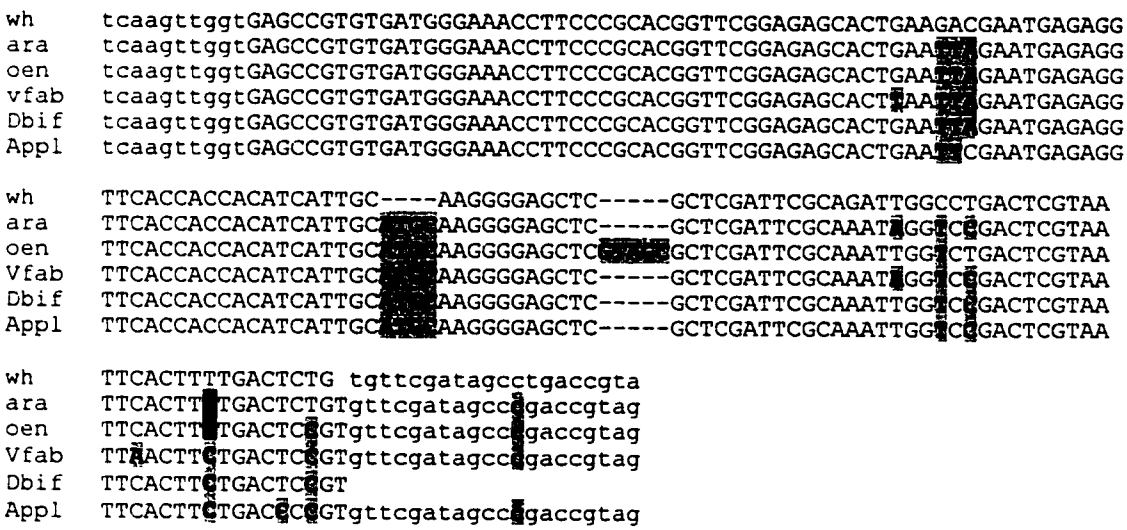
nad4-2



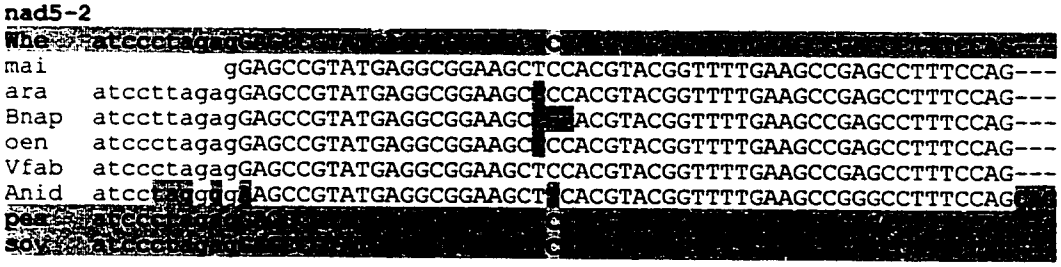
nad4-3

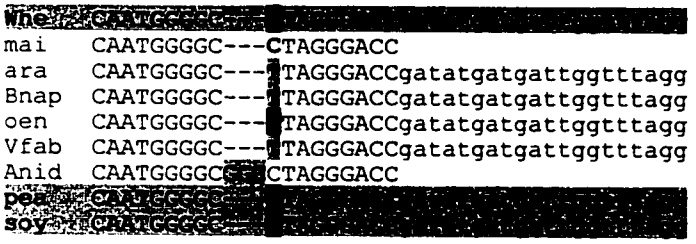


nad5-1

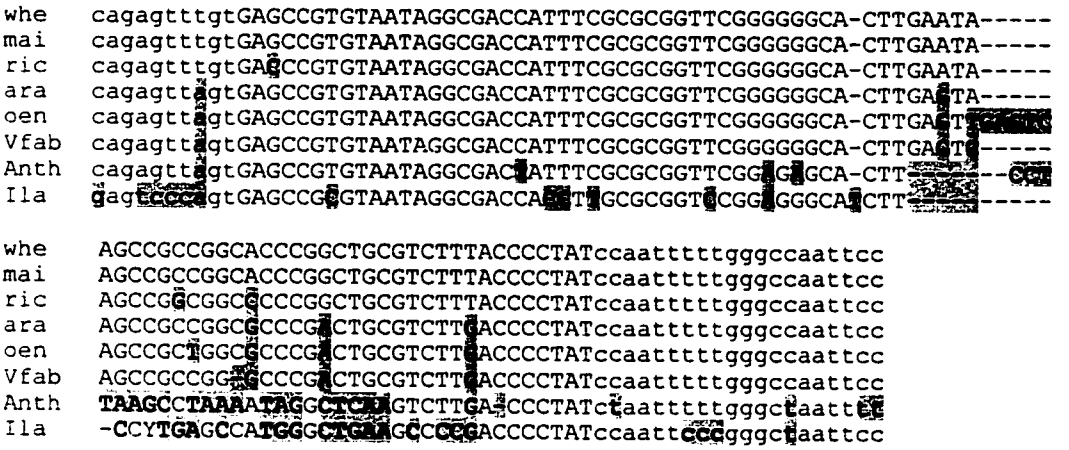


nad5-2

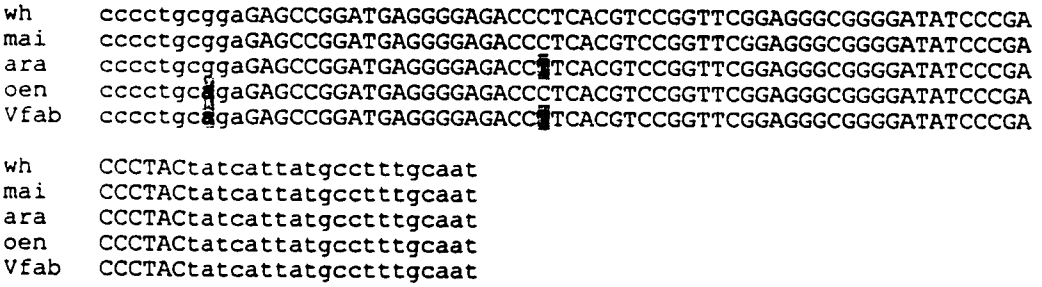




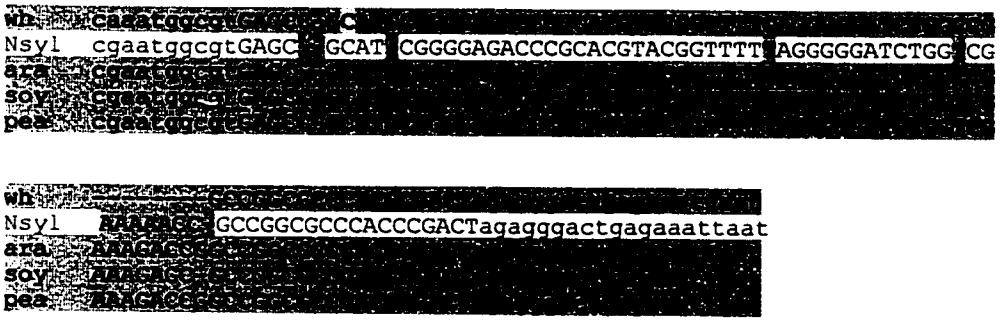
nad5-3



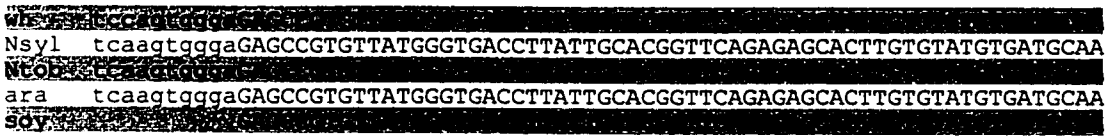
nad5-4



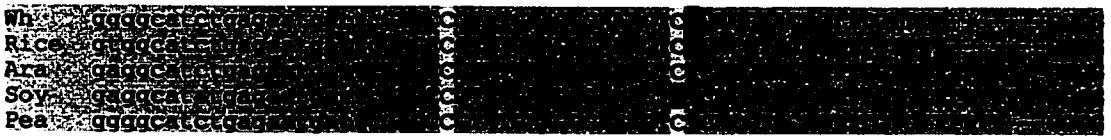
nad7-1

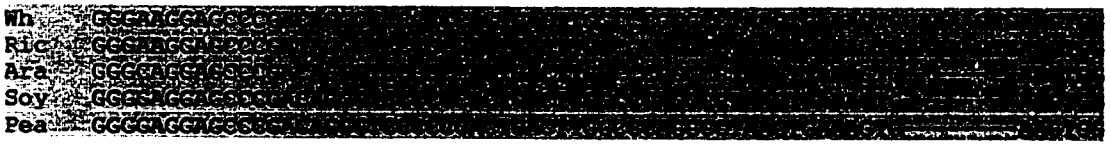


nad7-2

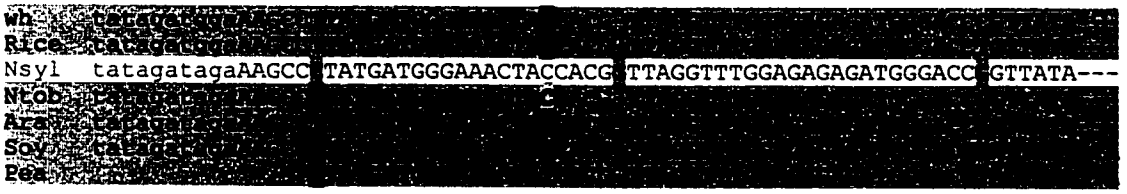


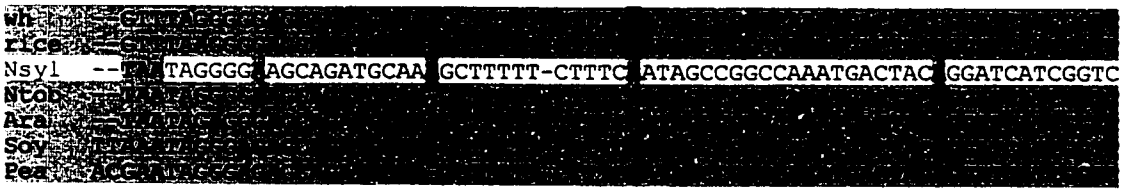
nad7-3

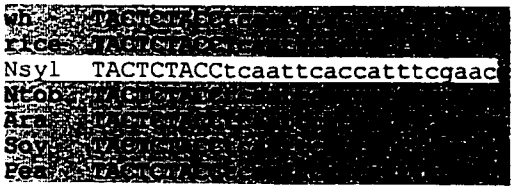




nad7-4







rpl2

```
ara    ttcttaggaaGAGCCGTACGAGGCAGCTCACGTACGGTTCGGGAGCCGAGCCCCTGCACAGGGGCTTAGG
oen    ttcttaggaaGAGCCGTACGAGGCAGCTCACGTACGGTTCGGGAGCCGAGCCCCTGCACAGGGGCTTAGG
ric    ttcttaggaaGAGCCGTACGAGGCAGCTCACGTACGGTTCGGGAGCCGAGCCCTGCACAGGGGCTTAGG

ara    TCAAcacttatataatagccagtc
oen    TCAACacttatataatagccagtc
ric    TCAACacttatatatagccagtc
```

rps3

```
mai    ttgggtttgaGAGCCGTGTGATGGGTGACTATCTAGCACGGTTCAGAGAGCACGTGTATGTAGTCTGCGC
ric    ttgggtttgaGAGCCGTGTGATGGGTGACTATCTAGCACGGTTCAGAGAGCACGTGTATGTAGTCTGCGC
ara    ttgggtttgaGAGCCGTGTGATGGGTGACTATCTAGCACGGTTCGGAGAGCACTTTTAGTCTTGG
oen    ttgggtttgaGAGCCGTGTGATGGGTGACTATCTAGCACGGTTCGGAGAGCACTTTTAGTCTGCGC
pet    ttgggtttgaGAGCCGTGTGATAGGTGACTATCTAGCACGGTTCGGAGAGCACTTTTAGTCTGCGC
bnap   ttgggtttgaGAGCCGTGTGATGGGTGACTATCTAGCACGGTTCGGAGAGCACTTTTAGTCTTGG

mai    TGGTGAATGGAAGCCCCCGCCG----GCAAAAAAGAAGCGGCTTTTCCCACGGCTCTGTCACCATTGACT
ric    TGGTGAATGGAAGCCCCCGCC-----GCAAAAAAGAAGCGGCTTTCCCACGGCTCTGTCACCATTGACT
ara    TGGTGAATGGAAGCCCCC[illegible]GCAAAAAAAAAGCGGCTTTTCTACGGCTTGTCACTATTGACT
oen    TGCTGAATGGAAGCCCCC[illegible]GCAAAAAAGAAGCGGCTTTCCCACGGC[illegible]ACCATTGACT
pet    TGGTGAATGGAAGCCCCC[illegible]GCAAAAAAGAAGCGGCTTTCCCACGGC[illegible]GTCACCATTGACT
bnap   TGGTGAATGGAAGCCCCC[illegible]GCAAAAAAGAAGCGGCTTTTCTACGG[illegible]GTCACTATTGACT

mai    CTATttattattatggtaaatca
ric    CTATttattattatggtaaatca
ara    CTATttattattatggtaaat[illegible]
oen    CTATttattattatggtaaatca
pet    CTATttattattatggtaaatca
bnap   CTATttattattatggtaaat[illegible]
```

rps10

```
pot    ttcgccacttGAGCCGTATGCGGGGGAACTCGCACGTGCGGTTCTTAGGGGGGGAGAGCTAGTAGGAGCC
pea    ttcgccacttGAGCCGTATGCGGGGGAACTCGCACGTGCGGTTCTTAGGGGGG-AGAGCTAGTAGGAGCC

pot    ATCCCATCCCAATagcgtatatttggagctcaa
pea    ATCCTATCCCAATagcgtatatttggagctcaa
```

orf454 = ccb6c = ccb452

```
ara    tcttcggcccGAGCTGTATGAGGCAGAAACTCGTCCCACGTACGGTTCGGAGGCCGAGCCCCACCCC
oen    tcttcggtccGAGCTGTATGAGGCAGAAACTCGTCTCACGTACGGTTCGGAGGCCGAGCCCCACCCC

ara    AGCAGTAATGGTGCGGCTTAGGTCAACtaacacaaagaagatacagtt
oen    TTAAGGTTGGTGCGGCTTAGGTCAACtaacacaaagaagatacagct
```

Abbreviations used in alignment for seed plants:

oat *Avena sativa* (oat)
Acam *Acorus calamus* (sweet flag)
ara *Arabidopsis thaliana*(thale cress)
Bcam *Brassica campestris* (field mustard)
Bnap *Brassica napus* (rape)
Bvul *Beta vulgaris* (beet)
Brap *Brassica rapa* (field mustard)
car *Daucus carota* (carrot)
let *Lactuca sativa*
mai *Zea mays* (corn)
mag *Magnolia grandiflora*
Ntob *Nicotiana tobaccum* (tobacco)
Nsyl *Nicotiana sylvestris*
oen *Oenothera berteriana* (Evening primrose)
pea *Pisum sativum*
pet *Petunia hybrida*
Pgin *Panax ginseng* (Korean ginseng)
pot *Solanum tuberosum* (potato)
ric *Oryza sativa* (rice)
Rsat *Raphanus sativus* (radish)
soy *Glycine max* (soybean)
sun *Helianthus* (sunflower)
Vfab *Vicia faba* (faba bean)
wat *Citrullus lanatus* (watermelon.)
wh *Triticum aestivum* (wheat)
Nalba *Nymphaea alba* (water lily)
apple *Malus domestica*
pine *Pinus sylvestris* (Scotch pine)
spruce *Picea abies* (Norway spruce)

Abbreviations used in alignment for vascular plants:

Anid *Asplenium nidus* (bird's nest fern)
anth *Anthoceros crispulus* (hornwort)
Cpu *Ceratodon purpureus* (moss)
Dbif *Dolichos biflorus* (horse gram)
Ete *Equisetum telmateia* (giant horsetail)
Ila *Isoetes lacustris* (quillwort)
Mdru *Marsilea drummondii* (fern)
Ore *Osmunda regalis* (fern)
Ppat *Physcomitrella patens* (moss)

APPENDIX III:

Summary of editing sites identified in plant mitochondrial introns

Gene	Location of edit	Improves base pairing?	Number of clones	Reference
nad1	stem D6, intron 1 *Oenothera*	Yes?	1/10	Wissinger *et al.*, 1991
nad1	stem D6, intron 1 *Oenothera*	No	10/10	Wissinger *et al.*, 1991
nad1	stem D6, intron 1 pea and wheat	No	4/4	Bonen and Chapdelaine, unpublished
nad1	stem D6, intron 3 *Oenothera*	Yes	-	Wissinger *et al.*, 1991
nad1	stem D6, intron 4 maize, wheat, *Arabidopsis*	Yes	-	Wolstenholme *et al.*, 1993; Farré and Araya, 1999; Giegé and Brennicke, 1999
nad1	stem D6, intron 4, *Petunia*	No	-	Conklin *et al.*, 1991
nad2	stem D1, intron 2 *Oenothera*	Yes	6/6	Binder *et al.*, 1992
nad2	stem D4; intron 2 *Oenothera*	Yes	4/4	Binder *et al.*, 1992
nad2	upstream D5; intron 1 *Oenothera*	No	-	Binder *et al.*, 1992; Knoop and Brennicke, 1993.
nad2	loop D4; intron 2 *Oenothera*	No	3/4	Knoop and Brennicke, 1993.
nad5	stem D6, intron 1 *Oenothera, Arabidopsis*	No	-	Knoop *et al.*, 1991; Giegé and Brennicke, 1999
nad5	loop D6, intron 1 *Arabidopsis*	Yes	-	Giegé and Brennicke, 1999
nad5	stem D6, intron 2 *Oenothera*, wheat, soybean, pea	Yes	-	Knoop and Brennicke, 1993; Bonen *et al.*, 1998; Clark, 1994
nad5	2 consecutive sites upstream D5, intron 4, *Oenothera*	?	-	Knoop *et al.*, 1991

Gene	Location of edit	Improves base pairing?	Number of clones	Reference
nad7	stem D6, intron 2 *Arabidopsis*, tobacco	Yes	-	Giegé and Brennicke, 1999; this work
nad7	stem D5, intron 3 wheat, rice	Yes	14/20 wheat	this work
nad7	stem D5, intron 4 soybean, tobacco, pea	Yes	-	this work
nad7	stem D5, intron 4 wheat, rice, *Arabidopsis*	Yes	19/22 wheat	this work
nad7	loop D6 wheat, rice	Yes?	17/22 wheat	this work
rps10	stem D1, potato	Yes	8/10	Zanlungo *et al.*, 1995

APPENDIX IV

Domain 5 structures in wheat and *Arabidopsis*.

Pairings were scored in wheat (W) and Arabidopsis (A) at each of the 14 D5 positions (cf. Figure 6.1). Pairs are indicated by a "O" for a GU pair, a "/" for a GC pair, a "-" for an AU pair and an E for an AC mispair which is predicted to be edited. Blue boxes indicate editing sites which are not edited and red shading indicates positions which are edited. Yellow boxes are used to indicate YY mispairs and green boxes indicate RR mispairs. Blue text is used to indicate introns with non-conventional features (cf. Section 6.2). Trans splicing introns are indicated by asterisks

Intron	1	2	3	4	5	6	7	8	9 9+5	9 8+6	10	11	12	13	14	loop	bulge 9+5	bulge 8+6	MFE 9+5	MF E 8+6
1. cox2-1W	O	-	O	/	-	-	CU	UU	O	-	/	/	O	O	/	GAAA	AA	GA	-14.4	-14.4
2. cox2-2A	O	-	O	/	-	/	-	-	E	CC	/	-	/	/	O	GAAA	CC	AC	-15.9	-15.9
3. nad1-1W* A*	/	-	O	/	/	/	-	-		-	/	/	O	/	O	17nt 18 nt	CU	GC		-15.5 -18
4. nad1-2W A	/	-	O	/	/	-	/	-	O	-	/	/	-	O	/	GAAA	CA	GC	-21.3	-21.5
5. nad1-3W* A*	/	-	/ O	/	O	/	-	/ -	O	-	/	E	O	/	/	GUAA	CA	GC	-16.0* -11.8	-14.4 -12
6. nad1-4W* A	/ O	-	O	/	/	/	-	-	O	-	/	-	/	O	/	GAGA	CA	GC	-19.9 -18.8	-20.1 -19
7. nad2-1W* A*	/	-	O	/	/	-	/	-	O	-	/	/	-	O	/	GAAA	CA	GC	-21.3	-21.5
8. nad2-2W* A*			O	/	/	/	-	-	O	-	/	/	/	O	O	GAGA	CA	GC	-14.0	-14.2

9. nad2-3W A	CC	-	O	/	/	/	UU	UC	O	-	/	-	UU	-	/	AACAAA AAAAAAA	CA	GC	-4.1 -2.6	-6.3 -5.8
10. nad2-4W A	/ O	-	O	/	/	/	O	-	CU	UU	/	-	-	O	/	GCAA	UU	CU CC	-16.6 -16.1	-16.6 -16.1
11. nad4-1W A	/	-	O	/	/	/	-	-	O	-	/	-	UC	O	/	GGAA	CA	GC	-15.6	-15.8
12. nad4-2W A		-	O	/	/	/	-	-	O	UU	/	-	-	-	/	8 NT 7 NT	UU	GU	-14.0 -13.7	
13. nad4-3W A	/	-	O	/	/	UU	-	/	O	UU	/	-	-	-	/	GAGA	AU	GA	-15.7	
14. nad5-1W A	/	-	O	/	/	/	-	/	O	CU	/	-	UU	/	/	GAAA	CC	GC	-19.1	
15. nad5-2W* A*	O	-	O	/	/	/	-	-	O	-	/		O /	/	/	GGAA	CA	GC	-16.3 -17.7	-16.5 -17.9
16. nad5-3W* A*	/	-	O	/	/	/	O	/	O	UU	-	-	-	E	/	GCGA	CU	GC	-11.7	
17. nad5-4W A	/	-	O	/	/	/	/	-	O	-	/	-	/ O	/	/	GAGA	CA	GC	-25.9 -23.2	-26.1 -23.4
18. nad7-1W A	O	-	O	/	/	/		-	O	-	-	/	/	/	/	GAGA	CA	GC	-19.3	-20.7
19. nad7-2W A	/	-	O	/	/	/	-	/	O	UU	-	-	UU	/	/	GUGA	UU	GU	-18.5	
20. nad7-3W A	O	-	O	/		/	-	-	O	-	/			O	/	GAAA	CA	GC	-6.1 (-8.6ed) -8.6	-6.3 (-8.8ed) -8.8

21. nad7-4WA	-	-	O	/	/	-	UU	-	O	-	/		-	O	/	GAAA	CA	GC	-7.5 (-12.1ed)	-9.6 (-14.2 ed)
22. Rpl2 RA	/	-	O	/	/	/	-	-	/		/	-	/			GCUG	CA	GC	-14.4	
23. rps3 RA	/	-	O	/	/	/	-	/	O	UU	/	-	-	O	/	GUGA	AU	GA	-21.1	
24. rps10 P	/	-	O	/	/	/	O	-	O	-	/	/	/	O	/	GGAA	CA	GC	-20.5	-20.7
25. orf454A	/	-	O	/	O	/	-	-	O	-	/	E	/	O	/	8 NT	CA	GC	-10.2 (-15 ed)	-10.4 (ed - 15.2)

Summary of mispairs in D5. Grey shading indicates most commonly observed pair

Pair #	G-C	A-U	GoU	A C	YY	GA
1	14	1	4	1 (U)	1	1
2						1
3	1					
4						
5		1	1	1 (U)		
6	16	5			1	
7	4	13	1	1 (U)	3	
8	6	14			2	
9 (9+5)	1					1
9 (8+6)		15		1 (U)	6	
10		3				
11	7	11		4(1 U,2 E, 1 ?)		
12	5	8	4	1 (U)	3	
13	7	3	10	1 (?)		
14			2			

Summary of nucleotide composition in D5 terminal loop and internal bulge.
Grey shading indicates most commonly observed nucleotide.

		G	A	U	C
LOOP	1				
	2	2	13	3	2
	3	8	11	1 rpl2	
	4	1 rpl2			
BULGE 9+5	1		3	2	
	2			3	2
BULGE 8+6	1				
	2		3	2	

Loops in 20 introns (long loops excluded from analysis)

APPENDIX V

Domain 5 structures in *Marchantia polymorpha*

Structure of Domain 5 in 25 mitochondrial group II introns of *Marchantia polymorpha*. Pairings were scored at each of the 14 D5 positions (cf. Figure 6.1 A & B). Pairs are indicated by a "O" for a GU wobble pair, a "/" for a GC pair, a "-" for an AU pair. Pyrimidine mispairs are shaded in yellow (i..e., UU) and purine mispairs are shaded in green (i.e., AG). The second intron in atpA-2 has a D5 structure which has an additional base pair compared to conventional D5 structure (pink box).

	1	2	3	4	5	6	7	8	9 9+5	9 8+6	10	11	12	13	14	loop	bulge 8+6	bulge 9+5
cox1-1	/	-	O	/	/	/	-	/	O	-	/	-	-	O	/	AAGA	GA	AA
cox1-2	/	-	O	/	/	/	-	UU	O	-	/	-	/	/	O	GAAA	GC	CA
cox1-5	/	-	O	/	O	O	-	-	O	-	/	-	/	O	-	GAAA	GC	CA
cox2-1	/	-	O	/	-	/	-	/	O	-	/	-	-	O	/	UAAA	GC	CA
cox2-2	/	-	O	/	/	/	/	-	-	-	/	/	-	/	/	GUGA	UA	AU
cox3-1	/	-	O	/	/	/	-	UU	O	UU	/	-	/	/	O	GAAA	GC	CU
cox3-2	/	-	O	/	/	/	/	-	O	-	/	/	/	O	/	GAAA	GC	CA
cob-1	/	-	O	/	/	-	/	-	O	-	/	/	-	O	/	GAAA	GC	CA
cob-2	/	-	O	/	/	/	-	O	O	UU	/	-	-	O	/	GCGA	GC	CU
cob-3	/	-	O	/	/	/	-	-	O	-	/	-	O	/	/	GUAA	GC	CA
atpA-1	/	-	O	/	/	/	-	-	O	-	/	-	-	/	/	GAAA	GC	CA
atpA-2	/	-		/	/	/	/	-	O	-	/	-	/	O	O -	GAAA	GC	CA

atp9	/	-	O	/	/	-	/	-	O	-	/	/	-	O	/	GAAA	GC	CA
nad2	/	-	O	/	/	/	-	-	O	UU	/	-	-	O	/	GUCA	GA	AU
nad3	/	-	O	/	/	/	-	/	O	UU	-	-	-	O	/	GCAA	GC	CA
nad4	/	-	O	/	/	/	-	-	O	-	/	/	O	-	/	GAGA	GC	CA
nad4L-1	/	-	O	/	/	/	-	-	O	-	/	/	O	-	/	GAAAU	GC	CA
nad4L-2	/	-	O	/	/	/	-	-	O	UU	/	-	-	/	/	GCGA	GC	CU
nad7-1	E	-	O	/	/	/	UC	-	O	-	/	/	O	-	/	GAAAU	GC	CA
nad7-2	/	-	O	/	/	/	-	-	O	-	/	-	/	O	/	GAAA	GC	CA
rpl14	O	-	O	/	/	/	-	-	O	-	/	-	-	O	/	GAGA	GC	CA
rpl2	/	-	O	/	/	/	-	-	O	-	/	/	/	O	/	GAGA	GC	CA
18srRNA	/	-	O	/	/	-	-	-	O	-	/	/	/	O	/	GAAA	GC	CA
26sRNA	/	-	O	/	/	/	O	-	O	-	/	/	O	O	/	GAGA	GC	CA
tRNAser	/	-	O	/	/	/	-	-	O	-	/	/	/	O	/	AAAA	GC	CA

Summary of D5 pairing in *Marchantia polymorpha*
Grey shading indicates pairings most commonly observed.

Pair #	G-C	A-U	GoU	A C	YY	GA
1	23		1	1		
2						
3						1
4						
5		1	1			
6		3	1			
7	5	18	1		1UC	
8	3	19	1		2UU	
9 (9+5)		1				
9 (8+6)		20			5UU	
10		1				
11	11	14				
12	9	11	5			
13	6	3	16			
14		1	3			

Nucleotide composition of the terminal tetraloop and the internal dinucleotide bulge in *Marchantia polymorpha*
Grey shading indicates nucleotides most commonly observed.

		G	A	U	C
LOOP	1		2	1	
	2		19	3	3
	3	8	16		1
	4				
BULGE 9+5	1		3		
	2			5	
BULGE 8+6	1			1	
	2		3		

APPENDIX VI

Domain 6 structure in wheat and Arabidopsis.
Pairs at each of the 6 helical positions in D6 (cf. Figure 6.2) are shown as a "O" for a GU pair, a "/" for a GC pair, a "-" for an AU pair and an E for an AC mispair. Blue boxes indicate editing sites which are not edited and red shading indicates positions which are edited. Yellow boxes are used to indicate YY mispairs and Green boxes indicate RR mispairs. Blue text is used for introns with non-conventional features (cf. Section 6.2).

		J(56)	1	2	3	4	5	6	Tail	loop
1. Cis	cox2-1 wh	GGG		/	/	/		/	AC	8
2. Cis	cox2-2 Ara	GG	E	/	/	/	/	-	UAU	24
3. Trans	nad1-1									
4. Cis	nad1-2 wh **Ara**	UG	O	/	/	/	/	/	AU	36
5. Trans	nad1-3 wh **Ara**	CG	**UC** **UU**	/	-	/		-	CG	2731
6. Cis/trans	nad1-4 wh **Ara**	CUU	-	/	-	/	-		CC	54
7. Cis /A	nad2-1 wh	UG	O	/	/	/	O	O	AU	6
8. Trans /A	nad2-2 wh **Ara**	GG **CG**	-	/	/	/	/	/	AU	3739
9. Cis / A	nad2-3 wh **Ara**	AG	-	/	-	/	/	-	AU	1e+05
10. Cis /A	nad2-4 wh Ara	GG	-	/	-	/	/	-	AU	4456
11. Cis /A	nad4-1 wh **Ara**	CC	**UU**	/	-	/		-	AC	6465
12. Cis /A	nad 4-2 wh **Ara**	GG	O	/	O	/	O		AU	2526
13. Cis /A	nad4-3 wh **Ara**	GG	-	/	E -	/	/	-	AC	5558
14. Cis /A	nad5-1 wh **Ara**	GG	-	/	-	/	/	-	GU	8690

15. Trans /A	nad5-2 wh **Ara**	UGA			/	/	O		CC	19
16. Trans /B	nad5-3 wh* **Ara**	GG	/	/	/	/	**UC /**	-	UAU	32
17. Cis /A	nad5-4 wh **Ara**	GG	-	/	/	/	/	/	AC	9
18. Cis /B	nad7-1 wh **Ara**	UC **UU**	AG	/	/	/	/	/	ACU	1725
19. Cis /A	nad7-2 wh **Ara**	AG	-	/		/	/		AU	69
20. Cis /B	nad7-3 wh **Ara**	GUUU U **G + 8U**	/	/	/	O	/	/	AAC	7769
21. Cis /B	nad7-4 wh **Ara**	GG	-	/	-	/	-	/	ACC	8082
22. Cis /B	rpl2 rice **Ara**	GGG		O	/	/	O	-	AAC	16
23. Cis /A	rps3 rice	AG	-	/	-	/	/	-	AU	78
24. Cis /B	rps10pea	UU	E	/	/	O	O	/	AAU	17
25. Cis /B	orf454 Ara	GGA	/	O	/	/	O	-	AAC	27

Summary of pairing at each D6 position. Note that D6 structures with non-conventional features were not included in the histogram in Figure 6.2 (blue highlighted numbers).

	GC	AU	GU	AC	YY	RR
-4	3	10	3	3	2	2 GA, 1AA
-3	17, 3		2			GG
-2	10, 2	8, 1(A)	1	2		
-1	19, 3		1			
1	11, 1(A), 1	2, 2(A)	3, 2	1	1UC(W)	1AG(W), 1AG
2	6, 1	12, 1(A)	1	3, 1 (A)		AG(W)

APPENDIX VII

Domain 6 structure in *Marchantia polymorpha*.
Pairs at each of the 6 helical positions in D6 (cf. Figure 6.2) are shown as a "O" for a GU pair, a "/" for a GC pair, a "-" for an AU pair and an E for an AC mispair. Blue boxes indicate editing sites which are not edited and red shading indicates positions which are edited. Yellow boxes are used to indicate YY mispairs and Green boxes indicate RR mispairs. Blue text is used for introns with non-conventional features (cf. Section 6.2).

	J (56)	1	2	3	4	5	6	Tail	loop
cox1-1	CA	/	/	-	/	-	-	AC	10
cox1-2	UG	E	/	/	/	/	-	UAC	20
cox1-5	GG	-	/	/	/	/	-	AU	8
cox2-1	UU		/	/	O	O	/	AAC	20
cox2-2	GG	-	/	-	/	CU	-	AU	30
cox3-1	GG	E	/	/	/	/	-	UAC	20
cox3-2	UUA	/	O	/	O	O	/	AAU	20
cob-1	UGA	/	/	/	O	O	/	UAU	9
cob-2	GG	-	/	-	/	/	-	AU	56
cob-3	UG	UU	/	-	/	-	-	CG	29
atpA-1	GG	UC	O	-	/	O	/	UAC	8
atpA-2	UU	UU	/	-	/	-	/	CU	8
atp9	UGA	/	/	/	O	O	/	UAU	9
nad2	GG	-	/	-	/	/	-	AU	485
nad3	GG	E	/	/	/	/	-	UAU	30
nad4	GG	E	O	/	-	E	/	UUA	98
nad4L-1	GG	E	/	/	/	E	/	CUAC	325
nad4L-2	GG	-	/	-	/	/	/	AU	70
nad7-1	GG	E	O	/	-	E	/	UUAC	51
nad7-2	UGAG	UU	-	/	/	/	E	AAC	21
rps14	UGAA	UU	/	-	/	O	/	AAC	21
rpl2	GG	E	O	/	-	E	/	CUAC	34
18s rRNA	GG	-	/	/	/		/	AC	30

26s rRNA	UU		/	/	O	O	/	AC	21
tRNA-Ser	UUA	/	/	/	O	O	/	GAC	21

Summary of pairing in D6 of Marchantia mitochondrial group II introns.
The most commonly observed pairing is highlighted in grey.

Pair #	G-C	A-U	GoU	A C	YY	RR
1	5	6		7	4UU 1UC	2AA
2	19	1	5			
3	16	9				
4	16	3	6			
5	8	3	8	4	1CU	1AG
6	15	9		1		

APPENDIX VIII:

Aberrant D5/D6 structures of plant mitochondrial introns.

Classical D5/D6 structure is shown in panel A, and features of plant mitochondrial introns which deviate from this structure are shown in panels B, C, D, E and F. AC mispairs are shown with coloured arrows where solid red indicates edited, open blue indicates unedited and open red indicates that position has not been examined.