uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

u Ottawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Sophie Calixte

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biology)

GRADE / DEGREE

Department of Biology

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

RNA Processing of the CCMFN-RPS1 and RPL5-ΨRPS14-COX3 Loci in Wheat Mitochondria during Seeding Development

TITRE DE LA THÈSE / TITLE OF THESIS

Dr. L. Bonen

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Dr. S. Aris-Brosou

Dr. D. Johnson

Dr. J. Vierula

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

RNA PROCESSING OF THE *CCMFN-RPS1* AND *RPL5-ΨRPS14-COX3* LOCI IN WHEAT MITOCHONDRIA DURING SEEDLING DEVELOPMENT

Sophie Calixte

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the

Ottawa-Carleton Institute of Biology

Thèse soumise à la
Faculté des études supérieures et postdoctorales
Université d'Ottawa
en vue de l'obtention de la maîtrise ès sciences

L'Institut de biologie d'Ottawa-Carleton

Your file Votre référence
ISBN: 978-0-494-41655-6
Our file Notre référence
ISBN: 978-0-494-41655-6

Abstract

Plant mitochondria possess a gene expression system in which post-transcriptional events, such as transcript end maturation and turnover mechanisms play a key role in regulating the transcriptome. In addition, during early developmental stages of embryo germination, differing transcript profiles have been seen. This research focuses on two loci in wheat mitochondria, *ccmFn-rps1* and *rpl5-Ψrps14-cox3*, to elucidate the transcription and post-transcriptional events involved in their expression. Northern analysis of the *ccmFN-rps1* genes during early seed-to-seedling development reveals a 3.2 kb primary transcript and a 2.7 kb bicistronic mRNA. A 0.7 kb monocistronic *rps1* mRNA is detectable up to 2d but there is no detectable monocistronic *ccmFN* transcript during the stages examined. Transcript ends were mapped using circular-RT-PCR and phosphatase treatment at three different developmental stages and revealed two processing sites as well as a single 3' end common to all three transcripts. The 5' ends of the processed *rps1* transcripts are heterogeneous and do not always include the start codon, questioning the *rps1* transcript functionality.

Gene order varies between plant species due to the high recombination rate in mitochondrial genomes, as is seen for *rpl5–Ψrps14* in wheat and rice. In both plants, the functional *rps14* gene is encoded in the nucleus and the mitochondrial *rps14* copy is a pseudogene. In wheat, *rpl5–Ψrps14* are co-transcribed with *cox3* as two RNA species of 3.5 kb and 2.7 kb at 24hr post-imbibition and exhibit developmentally-specific differences in abundance in seedlings. Two promoter regions were mapped in wheat upstream of *rpl5* and both transcripts have the same 3' end. In rice 24hr and 6d however, *rpl5–Ψrps14* are co-transcribed as a 1.4 kb bicistronic mRNA. This presumably reflects the different regulatory signals used in different species. In addition, *rpl5* has been subject to several independent gene transfers to the nucleus in the cereal lineages. For example, there is a functional copy of *rpl5* in the mitochondria and the nucleus in wheat but it is absent from the mitochondria in rye and maize. In oat mitochondria, *rpl5* appears to be a pseudogene and in barley, rearrangements at the 3' end and low transcript levels question its functionality.

The characterization of transcription initiation sites, processing sites and 3' ends for these two loci reflect the relaxed nature and flexibility of signals exploited by plant mitochondria. This research supports the significant role of post-transcriptional events in the regulation of gene expression in plant mitochondria.

Régulation de l'ARN des gènes mitochondriaux *ccmFN-rps1* et *rpl5-Ψrsp14-cox3* pendant le développement des graines de blé

Résumé

Les génomes mitochondriaux des plantes possèdent un système d'expression des gènes où les processus post-transcriptionnels, tels que la maturation des extrémités 5' et 3' ainsi que les mécanismes de dégradation, jouent un rôle majeur dans la régulation du transcriptome. Les profiles de transcription sont souvent plus complexes lors des premiers stades de développement de la germination des graines. Cette recherche se concentre sur deux régions du génome du blé, *ccmFN-rps1* et *rpl5-Ψrps14-cox3*, dans le but d'éclaircir les événements transcriptionnels et post-transcriptionnels impliqués dans l'expression de ces deux loci. Les analyses de Northern pendant le développement de la graine en pousse révèlent la présence d'un précurseur d'ARN de 3,2 kb et d'un ARN messager bicistronique de 2,7 kb. Durant les stades de développement précoces, un ARN messager de *rps1* monocistronique de 0,7 kb est détectable jusqu'à 24h mais aucun ARN messager monocistronique de *ccmFN* n'a été détecté. Les extrémités 5' et 3' des transcrits ont été établies en faisant usage de la technique de la transcription inverse suivie de la réplication en chaine de la polymérase (TR-RCP) circulaire et du traitement avec la phosphatase pendant les trois stades de développement et révèlent la présence de deux sites de clivage 5' ainsi qu'une extrémité 3' commune aux trois transcrits. Comme l'extrémité 5' des transcrits secondaires monocistroniques de *rps1* est hétérogène et ne contient pas toujours le codon d'initiation, ceci remet en question leur fonctionnalité.

La séquence génomique peut varier entre espèces à cause du grand taux de recombinaison, tel qu'observé pour les gènes *rpl5–Ψrps14* dans le blé et le riz. Le gène fonctionnel de *rps14* se trouve dans le noyau et la copie mitochondriale de *rps14* est un pseudogène. Dans le blé, ils sont co-transcrits avec *cox3* en deux formes d'ARN de 3,2 kb et 2,7 kb après 24h après d'imbibition et démontre une transcription différente en abondance à 6 jours. Deux promoteurs ont été établis en amont de *rpl5* dans le blé et les deux transcrits ont la même extrémité 3'. Par ailleurs, dans le riz, *rpl5–Ψrps14* sont co-transcrits en un ARN messager bicistronique de 1,4 kb à 24h et à 6 jours. De plus, *rpl5* a été sujet à plusieurs transferts vers le noyau dans les lignées de céréales. Une copie fonctionnelle de *rpl5* est présente dans le génome mitochondrial et le génome nucléaire chez le blé mais *rpl5* est absent dans les mitochondries du seigle et du maïs. *rpl5* parait être un pseudogène dans

l'avoine et dans l'orge où des réarrangements à l'extrémité 3' et les bas niveaux de transcrits remettent en question sa fonctionnalité.

La caractérisation de sites d'initiation de transcription, de sites de clivages et des extrémités 3' pour ces deux loci reflète la nature relâchée et la flexibilité des signaux utilisés dans les mitochondries des plantes. Cette recherche supporte le rôle crucial des événements post-transcriptionnels dans la régulation de l'expression des gènes dans les mitochondries des plantes.

Acknowledgments

I would like to sincerely acknowledge my supervisor Dr. Linda Bonen, for her continual support and advice throughout my graduate studies. Her encouragement and assistance is greatly appreciated and I am always grateful for her decision to take a chance on me as a newly graduated undergrad student.

My appreciate thanks extend to my committee members Dr Doug Johnson and Dr John Vierula for their helpful feedback. Special thanks to Dr. Johnson for friendly neighbourly advice.

A special notice is given to my colleagues, Dr Jennifer Li-Pook-Than and Dr. Thomas Hazle, for their friendship, encouragement and support in and out of the laboratory. This work couldn't have been done without them, before they even held Ph.Ds. I greatly appreciate the contribution of former lab members (JC, NN, MF, CC, SS, PG) and thanks is also given to the former honours students, summer students and technicians in the lab, especially K. Mellet (oat *rpl5*), O. Antoniuk, J. Hardy (barley *rpl5* and wheat *rpl5* editing status) and B. Allen for technical assistance and other valuable contributions. I acknowledge Alain Toutloff for his work as a 4th year honours student on the *rpl5-Ψrps14* loci in cereals. Thanks is also given to Dr. Candace Webb, Neda Massourian and Jaimie Moise for their friendship and help, as well as other biology graduate students over the years.

I would like to thank the support staff and the members of the Biology Department, notably Dr. J. Basso, D. Smith and the support staff.

Finally this work would not have been accomplished without the support and love of my parents, Nicole and Jean-Marie Calixte and my sisters Laurence and Brigitte. I will be forever indebted to them for their patience and encouragement, this work is dedicated to them.

Table of contents

List of figures

List of tables

List of appendices

List of Abbreviations

ATP	adenosine triphosphate
atp	adenosine triphosphate synthase subunit genes
bp	base pairs
BSA	bovine serum albumin
ccm	cytochrome c biogenesis subunit genes
cDNA	complementary DNA
CDS	coding sequence
CMS	cytoplasmic male sterility
cob	cytochrome bc1 oxidoreductase subunits genes
cox	cytochrome oxidase subunits genes
ddH_2O	double distilled water
ddNTP	dideoxyribonucleotide triphosphate
dNTPs	deoxyribonucleotide triphosphates (G, A, T, C)
dsRNA	double-stranded RNA
DTT	dithiolthreitol
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
kb	kilobases
mat-r	plant mitochondrial intron-encoded maturase (reverse transcriptase)
mRNA	messenger RNA
mt	mitochondria, mitochondrial
mtt	membrane targeting and translocating genes
MYA	million years ago
µL	microliter
µg	microgram
mL	milliliter
mg	milligram
$MgCl_2$	magnesium chloride
MMLV	Moloney murine leukemia virus
NAD	nicotinamide adenine dinucleotide
nad	NADH dehydrogenase subunit

	genes
nt(s)	nucleotide(s)
	R = Purine
	A = adenosine
	G = guanosine
	Y = Pyrimidine
	T = thymidine
	U = uridine
	C = cytidine
ORF	open reading frame
PNPase	polynucleotide phosphorylase
PCR	polymerase chain reaction
PPR	pentatricopeptide repeat
rpl	protein of the large ribosomal subunit
rps	protein of the small ribosomal subunit
RNase	ribonuclease
rRNA	ribosomal RNA
RT, cr-RT	reverse transcriptase, circularized reverse transcriptase (with RNA ligase)
sdh	succinate dehydrogenase subunit genes
SDS	sodium dodecyl sulphate
snRNA	small nuclear RNA
ssRNA	single stranded RNA
SSII	Superscript II RT
t-element	tRNA-like element
TAP	tobacco acid phosphatase
TBE	Tris-borate-EDTA buffer
Tris	tris(hydroxymethyl) aminomethane
tRNA	transfer RNA
U	units
UTP	uricile triphosphate
UTR	untranslated region
V	volts
Ψ	pseudogene

Chapter 1: Introduction

1.1 Plant mitochondrial genomes

1.1.1 Origin of mitochondria, function and gene content

Mitochondria arose after the engulfment of an α-proteobacteria by a pre-eukaryotic cell or archaebacterial host, and this is known as the endosymbiont origin of mitochondria (reviewed in Gray *et al.* 1999; Martin and Koonin, 2006). Following this initial engulfment, the mitochondrial genome lost many genes required for free-living and many more genes were transferred over evolutionary time to the nucleus where they are now encoded. The products resulting from these genes are targeted back to the mitochondria in cases where their function still lies in the organelle. Mitochondria provide the energy of the cell through the production of ATP via the electron transport pathway: the oxidative phosphorylation pathway is composed of a series of five complexes (Table 1.1) which oxidize the products of the tricarboxylic acid cycle (TCA) and glycolysis.

Present day plant mitochondrial genomes contain only about 30 protein-coding genes as well as ribosomal RNAs (18S, 5S and 26S) and some of the 20 tRNAs necessary for translation, while the rest are imported from the cytosol (reviewed in Knoop *et al.* 2004). Some ribosomal proteins of the large and small subunits are encoded in the mitochondria as well as components of the respiratory chain complexes (Table 1.1). Notably, respiration in plants can bypass some of these complexes via alternative oxidoreductase respiration pathways, changing the constraint on the bypassed complexes (Eubel *et al.* 2004). The set of genes encoded in the mitochondria in different plant species varies slightly due to ongoing gene transfer to the nucleus, which can also affect the functional status of the mitochondrial genes. Transferred gene remnants can be seen as pseudogenes (Ψ) in the mitochondria (see section 1.2.10)

Table 1.1: Gene content in the wheat and rice mitochondrial genomes

Gene content based on the complete mitochondrial genome sequences of wheat (AP008982, Ogihara *et al*, 2005) and rice (BA000029, Notsu *et al*, 2002). Genes are categorized by their function: members of the respiratory chain complexes, ribosomal proteins, structural RNAs (rRNA and tRNA) and others (*mat-r*, *mttB*). Members of the complex II (succinate dehydrogenase: CoQ), *sdh* genes are nuclear encoded in these two species. + means mitochondrially encoded. Numbers refer to the number of copies. mt: mitochondrial origin, Ψ: pseudogene, cp: chloroplast origin, exons numbers (a-e) shown as cis spliced (-) or trans spliced (,). Excluded from the rice genome are duplicate gene copies in repeated regions.

		comments	
Function	**Gene**	Wheat	Rice
Complex I NADH dehydrogenase	***nad1***	**a, b-c, d, e**	**a, b-c, d, e**
	nad2	**a-b-c-d-e**	**a-b-c-d-e**
	nad3	+	+
	nad4	**a-b-c-d**	**a-b-c-d**
	nad4L	+	+
	nad5	**a-b, c, d-e**	**a-b, c, d-e**
	nad6	+	+
	nad7	**a-b-c-d-e**	**a-b-c-d-e**
	nad9	+	+
Complex II succinate dehydrogenase: CoQ	***sdh***	**nuclear encoded components**	
Complex III cytochrome bc1 oxidoreductase	***cob***	+	+
Complex IV cytochrome oxidase	***cox1***	+	+
	cox2	**a-b**	**a-b**
	cox3	+	+
Complex V ATP synthase	***atp1***	+	+
	atp4 (orf25)	+	+
	atp6	**2**	**1**
	atp8	**2**	**1**
	atp9	+	+
Cytochrome c biogenesis	***ccmB***	+	+
	ccmC	+	+
	ccmFC	**a, b**	**1 exon**
	ccmFN	+	+
Ribosomal proteins	***rpl2***	ψ	**functional**
	rpl5	+	+
	rpl16	+	+
	rps1	+	+
	rps2	+	+

	rps3	a-b	a-b
	rps4	+	+
	rps7	+	+
	rps11	absent	ψ
	rps12	+	+
	rps13	+	+
	rps14	ψ	ψ
	rps19	ψ	functional
ribosomal RNAs	*rrn5*	3	2
	rrn18	3	2
	rrn26	2, 1 cp	2
transfer RNAs	*trnA*	1 cp	absent
	trnC	cp	1 cp +1Ψ mt
	trnD	2	1
	trnE	+	+
	trnF	cp	cp
	trnfM	3	1
	trnN	absent	1
	trnI	+	1 mt, 1 Ψcp
	trnK	3	1
	trnM	+	1 mt, 1 cp
	trnN	cp	cp
	trnP	2	1 mt, 1 Ψcp
	trnQ	3	1
	trnS	2 mt + 1 cp	2 mt + 1 cp
	trnW	cp	cp
	trnY	+	+
Other	*mat-r*	+	+
	mttB (orfX)	+	+

1.1.2 Mitochondrial genome characteristics: size variation and gene organization

Plant mitochondrial genome sizes range from about 180 kb to 2400 kb (reviewed in Kubo and Newton, 2008). Those genome sizes are big compared to animal mitochondrial genomes that are only ~16 kb and where the genes are tightly packed. This variation in size among mitochondrial genomes is mostly due to the presence of lineage specific intergenic sequences that can account for 90% up to 98% of the genome size (Kubo and Newton, 2008). Genes are spread around the genome and can be separated by long stretches of non-coding sequence (up to several hundreds of kilobases). Gene linkages vary between closely related species but in a few cases genes have conserved the linkage found in bacteria like *Escherichia coli* (*E. coli*), such as the ribosomal protein linkages *rpl5-rps14* and *rps3-rpl16*, reflecting their ancestral bacterial origin (reviewed in Gagliardi and Binder 2007, Kubo and Mikami, 2007).

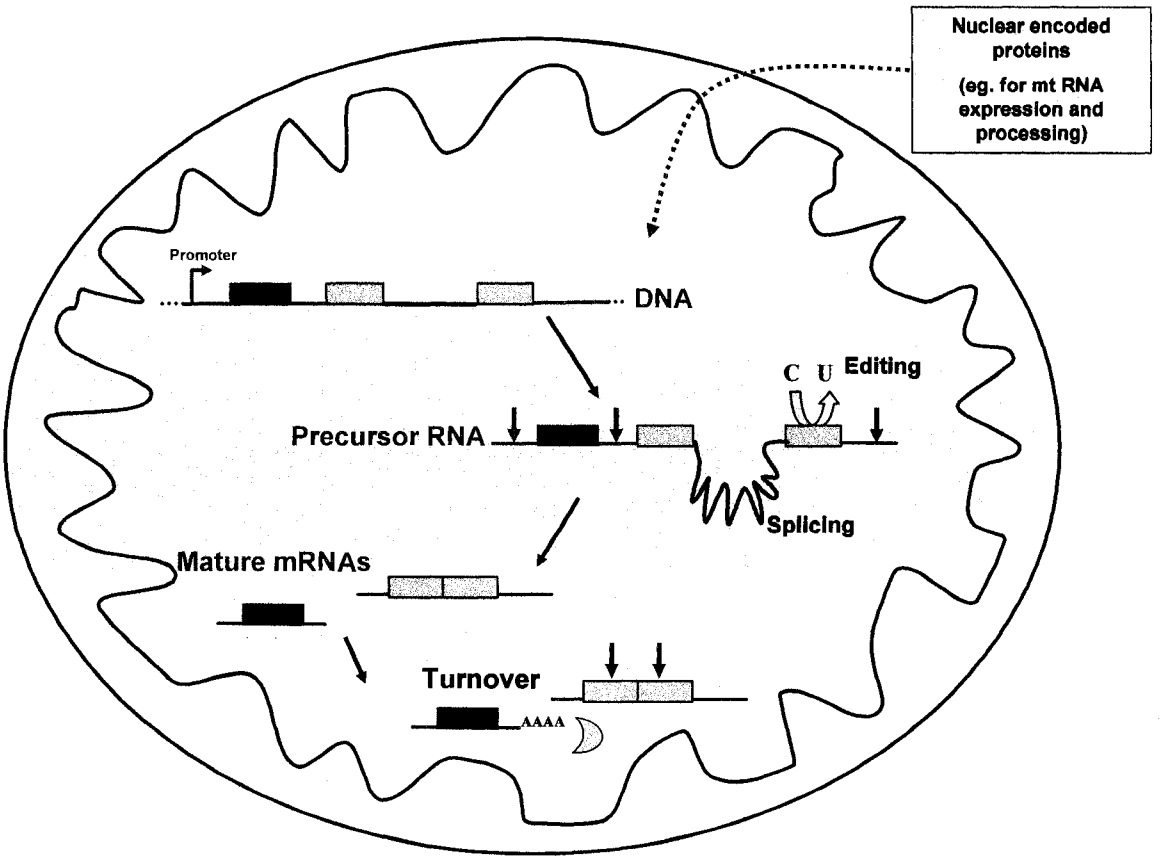
Plant mitochondrial genomes have a low nucleotide substitution rate compared to mammalian mitochondria (Wolfe *et al.* 1987) and exhibit a high degree of rearrangement. For example, in maize and wheat mitochondrial genomes there are 6 and 16 recombinational repeats respectively, which lead to the formation of many different subgenomic forms of the genome, present at different levels (Bendich, 1993), such that the master chromosome genome may not be seen as such *in vivo*.

Rearrangements in the genome can shuffle sequences around genes that contain regulatory signals for transcription and/or translation, creating differing genomic contexts for the same gene in different species: for example *rpl5-Ψrps14* in wheat is linked to *cox3-tRNA*^{E} whereas *rpl5-Ψrps14* are located far away from other genes and *cox3* is linked with *atp4* in rice (Ogihara *et al.* 2005, Notsu *et al.* 2002).

Faced with these various constraints at the DNA level, plant mitochondrial genomes have evolved new and fascinating ways to express their genomes. Following transcription, the precursor RNAs undergo several post-transcriptional events leading to the mature messengers (Figure 1.1). These events are not fully understood as of yet but our knowledge and characterization of the steps and machineries involved are expanding. The goal of this study is to investigate the transcriptional and post-transcriptional events leading to the

Figure 1.1: Transcription and RNA processing events in plant mitochondria

Processing events shown are 5' and 3' end maturation, polycistronic transcript cleavage, C-to-U editing, group II intron splicing and RNA degradation. Filled boxes represent genes, grey lines denotes UTR sequences and black lines show group II type intron in the DNA and in its secondary structure in the RNA. The bent arrow represents the promoter and grey lines with arrow heads are endonucleolytic processing sites. AAAA shows the presence of poly(A) tail and half-moon represents presumed degradation machinery. Import of nuclear-encoded proteins is denoted by dashed arrow. Figure adapted from a lab schematic.


Nuclear encoded proteins
(eg. for mt RNA expression and processing)
Promoter
DNA
C U Editing
Precursor RNA
Splicing
Mature mRNAs
Turnover
AAAA

transcript profiles of two mitochondrial loci during early stages of wheat seedling development.

1.2. Mitochondrial gene expression in plants

1.2.1 Transcriptional machinery

Despite its bacterial origin, plant mitochondrial transcription is achieved by a single subunit bacteriophage-type T3-T7 like enzyme, the RNA polymerase RpoTm, which is nuclear encoded (reviewed in Ikeda and Gray 1999a). A chloroplast specific RNA polymerase, RpoTp, is also present and originated from the duplication of RpoTm and plastid transcription is done by this polymerase as well as a eubacterial type RNA polymerase (Hedtke *et al.* 1997). In wheat two RNA polymerase genes have been identified (corresponding to RpoTm and RpoTp) (Ikeda and Gray 1999a). In *Arabidopsis* mitochondria, an additional RNA polymerase is also present, RpoTmp, which is dually targeted to the mitochondria and the chloroplast, and has diverged more recently from RpoTm than RpoTp (Hedtke *et al.* 1997, 2000).

A study in *Arabidopsis thaliana* (Kuhn *et al.* 2007) demonstrated the *in vitro* activity of the three RNA polymerases of this species. Each purified recombinant enzyme was able to initiate transcription *in vitro* from mitochondrial and plastid promoter regions, at varying levels of efficiency. Although the enzymes seem to have intrinsic recognition properties *in vitro*, co-factors might be needed *in vivo* for proper and efficient transcription initiation. Auxiliary factors involved in transcription initiation are yet to be identified in plant mitochondria, although in wheat a potential co-factor of 63 kDa enhances *in vitro* specific transcription and is possibly involved in mtDNA binding (Ikeda and Gray 1999b).

1.2.2 Promoter architecture

In plant mitochondria, genes are spread around the genome and several promoters exist to transcribe them. Genes in plant mitochondrial genomes can also be closely linked and the presence of polycistronic transcripts is not unusual, reminiscent of bacterial operon-like structure (Binder *et al.* 1996).

In plant mitochondria, transcription initiation sites do not seem to follow a strictly conserved consensus sequence although the motif C/YRTA appears to be part of some higher

plant mitochondrial promoters. This motif is apparent among mapped transcription sites for some genes in the eudicot soybean (Brown *et al.* 1991) and in the monocot wheat (Covello and Gray, 1991), although other promoters in the same plants do not correspond to it (Fey and Maréchal-Drouard, 1999). There is slightly more similarity among eudicot plants, extending the consensus to the nonanucleotide sequence CRTAAGAGA (reviewed in Gagliardi and Binder, 2007). Identification of transcript ends in several plants revealed the presence of several 5' ends and promoters per gene (Zhang and Liu, 2006 and reviewed in Gagliardi and Binder, 2007). Recent studies characterized the use of multiple promoters for a single gene in *Arabidopsis thaliana* (Kuhn *et al.* 2005, Forner *et al.* 2007), the former study identifying 9 out of 12 genes transcribed from multiple promoters. Certain motifs of promoter architecture were observed but more than half of the mapped promoters did not correspond to any consensus and no tissue-specific difference in promoter selection was observed between flower and leaf (Kuhn *et al.* 2005). It is therefore hard to predict promoter regions and transcription initiation sites based solely on sequence. The abundance of promoters and variation in promoter sequences suggest a relaxed specificity in promoter recognition. However caution should be exercised when looking at studies carried out with tissue culture material as they might not reflect activities *in planta* (Forner *et al.* 2007). Different promoter structures may reflect different transcription strength and hence serve a regulatory function, possibly in different developmental stages due to the import of co-factors which might be delayed during early development (Mulligan *et al.* 1991) (see section 1.2.8). In maize, two copies of *cox2* are found downstream of a direct repeat and different promoters differentially regulate the expression of each copy (Lupold 1999a).

Thus plant mitochondrial transcription initiation differs from that in bacteria, where the RNA polymerase recognises a promoter region, a fairly conserved sequence (-35 TTGACA and -10 TATAAT). Bacterial RNA does not typically undergo post-transcriptional processing. Genes are ordered in operon linkages which are transcribed together (reviewed in Browning and Busby, 2004). In human mitochondria on the other hand, the tightly packed genes are separated by tRNAs. Transcription from two promoters, one on each strand, creates long polycistronic transcripts, from which the tRNAs are processed, releasing the protein coding mRNAs (reviewed in Boore, 1999).

1.2.3 Transcript end maturation

Plant mitochondrial primary transcripts can undergo several processing steps, resulting in the presence of intermediate precursor RNAs, before achieving mature ends essential for translation and/or stability (reviewed in Gagliardi and Binder 2007; figure 1.2). In *Arabidopsis*, 5' UTR length varies from ~20 nucleotides up to 1898 nt (*atp1*) with an average of ~250 nts and 3' UTR length is also variable (Kuhn *et al.* 2005, Forner *et al.* 2007), from 10 nt to 498 nt with an average of ~130 nts.

The machinery involved in the processing of transcripts is still under investigation. In addition to the RNaseP and Z enzymes involved in tRNA maturation (see section 1.2.3.1), two exonucleases in *Arabidopsis,* PNPase (polynucleotide phosphorylase, targeted to the mitochondria) and RNaseII (ribonuclease, targeted to the mitochondria and the chloroplast), have been identified (Perrin *et al.* 2004a, b). The involvement of PNPase in *Arabidopsis* transcript degradation and turnover is discussed below in regards to rRNAs end processing and non-coding transcripts but there is no direct evidence implicating PNPase in other RNA post-transcriptional events. Proteins involved in the processing steps are nuclear encoded and imported from the cytosol.

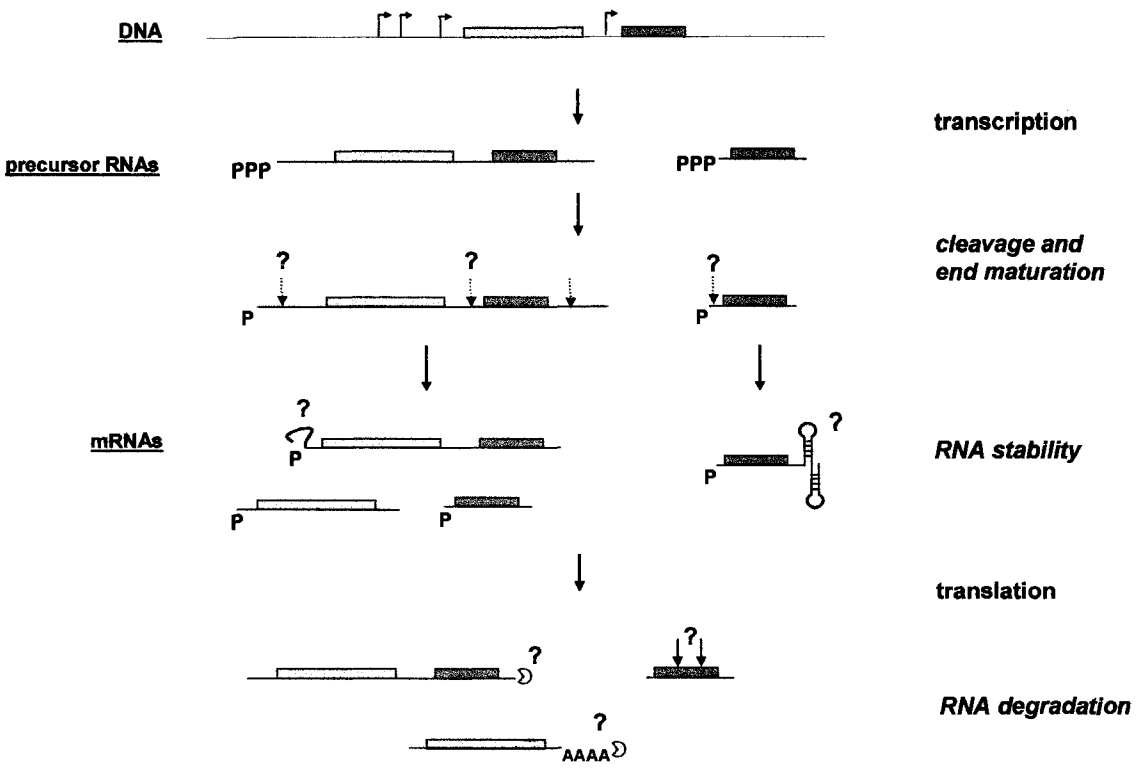
1.2.3.1 tRNA and rRNA processing

The steps involved in the processing of plant mitochondrial tRNAs are best understood to date. Following transcription, the mature tRNA is generated by two successive endonucleolytic cuts occuring at the 5' and 3' ends and the addition of the triplet CCA at the 3' end. An RNase-P-like enzyme cuts at the 5' end of the mature tRNA when it is folded in the typical clover-leaf structure. An RNase-Z-like enzyme cuts at the position of the mature 3' end or one nucleotide downstream. A nucleotidyl transferase then adds the CCA at the 3' terminal sequence (reviewed in Gagliardi and Binder 2007). These enzyme activities have been characterized for their involvement in the tRNA maturation process although the enzyme themselves have not yet been isolated.

Details about the maturation of the rRNA genes have been clarified in recent years, with work on mutant Arabidopsis plants where PNPase was downregulated (Perrin et al. 2004b). In Arabidopsis, the 18S rRNA and 5S rRNA genes are co-transcribed. Two successive endonucleolytic cuts are made at the 5' end of both coding sequences and result in

Figure 1.2: RNA post-transcriptional events in plant mitochondria

Filled boxes are genes, bent arrows denote promoter regions, tri- (PPP) and mono-(P) phosphate groups are shown at the 5' ends of transcripts and black lines with arrow heads represent processing sites. AAAA shows the presence of poly(A) tail and half-moon represents a putative polynucleotide phosphorylase. Question marks indicate processing events or secondary structures that are not seen in all transcripts or that are speculated.


DNA
transcription
precursor RNAs
PPP
PPP
cleavage and
end maturation
?
?
?
P
P
mRNAs
?
P
?
P
P
RNA stability
translation
?
?
RNA degradation
?
AAAA

two transcripts: one containing the 18S-rRNA precursor linked to the intergenic region transcript and one containing the 5S-rRNA transcript, as well as an RNA species containing the 18S 5' leader sequence. The latter and the intergenic region are polyadenylated at the 3' end and this acts as a tag for degradation by the PNPase enzyme (Perrin et al. 2004b, and reviewed in Gagliardi and Binder 2007). In PNP mutants, the 18S-rRNA precursor linked to the intergenic region accumulates as the intergenic region is no longer degraded (Perrin et al. 2004b).

1.2.3.2 5' end processing of protein-coding transcripts

Plant mitochondrial genes often exhibit transcripts with several 5' ends, created by transcription initiation (or initiations) and 5' end processing event (or events) (Kuhn *et al.* 2005, Forner *et al.* 2007). The comprehensive study in *Arabidopsis* by Forner *et al.*(2007) proposes mechanisms for the formation of processed 5' ends by endonucleolytic cleavage downstream of secondary structures. Mitochondrial endonucleases directly involved in these processes have not been isolated but RNaseZ and P are likely candidates. Two exonuclease genes in *Arabidopsis* (PNPase targeted to the mitochondria and RNaseII to the mitochondria and the chloroplast) have been characterized (Perrin *et al.* 2004a, b), but no direct evidence for their role in 5' end formation has been seen as of yet. Searches for signals used for recognition of processing sites have so far failed to characterize a consensus sequence (Kuhn *et al.* 2005, Forner *et al.* 2007). Developmentally specific differences have not been assayed for processing sites and variation at different developmental time points might be seen since processing events are dependent upon imported proteins. Other accessory factors involved in 5' end processing are still to be identified, as well as its temporal relationship with other RNA maturation events.

1.2.3.3 3' end processing of protein-coding transcripts

In contrast to the several 5' ends that can be found for a single gene, plant mitochondrial transcripts mostly exhibit a single 3' end. Transcription termination most likely occurs further downstream and processing at the 3' end takes place in the primary transcribed RNA (reviewed in Gagliardi and Binder, 2007) (Figure 1.2). Inverted repeats allowing the formation of stem loop structures are found at the 3' end of some transcripts

(e.g. *Arabidopsis atp9*, *nad1*, *cox2*; maize *atp9*) (Hanic-Joyce *et al.* 1990, Dombrowski *et al.* 1997, Kuhn J *et al.* 2001, Forner *et al.* 2007) and similarly to the formation of tRNAs 3' end, could be signals for an endonucleolytic cut. Although this model seems attractive, many genes in mitochondria are not followed by a sequence capable of folding in such recognizable structures as has been demonstrated by other mapped 3' ends (reviewed in Gagliardi and Binder, 2007).

In *Arabidopsis,* the 3' ends of some transcripts are located upstream of the genomically encoded stop codon (*nad6*, *ccmC*): t-element structures were found to be located at the 3' terminal of their coding sequences (Forner *et al.* 2007). Surprisingly in *Arabidopsis* and cauliflower, *nad6* and *ccmC* mRNAs, even though lacking a stop codon, are still functional and translated (Raczynska *et al.* 2006).

1.2.4 C-to-U-editing

C-to-U editing is a process in plant mitochondria and chloroplasts where specific C nucleotides are converted into U nucleotides in RNA molecules through a deamination process (Shikanai 2006). C-to-U editing is considered an early event in transcript processing (Gualberto *et al.*1991). Approximately 10% of editing sites are silent (e.g. 3rd codon position) while ~90% of sites are non-silent, usually restoring the codon to the conservative one, in regards to the amino acid sequence similarity with other organisms. In some cases it restores the initiating ATG start codon (*nad1* in wheat, Chapdelaine and Bonen, 1991) or forms a stop codon (*atp9* in petunia mitochondria, Wintz and Hanson, 1991). The number of editing sites in plant mitochondria varies between species: 362 sites in *Arabidopsis* (Bentolila *et al.* 2007), 427 in *Brassica* (Handa, 2003), 357 in *Beta vulgaris* (Mower and Palmer, 2006) and 491 in rice (Notsu *et al.* 2002), although those numbers reflect editing sites mostly found in coding sequences as few have been looked at in intergenic, non-coding regions. Notable examples include the middle nucleotide of the 3 nts intergenic sequence between *rpl2-rps19* in cereals (Kubo *et al.* 1996) and two editing sites in the 5'UTR of *rps1* of legumes (Hazle and Bonen, 2007a).

No enzymes have been characterized in relation to this process but co-factors are speculated to be important, especially for specificity in site recognition (Mower, 2007). A phylogenetic approach proposes that a domain in the PPR protein family may act as the

catalytic domain in the editing machinery (Salone *et al.* 2007). Sequence context is important but no obvious shared motifs have been found (Choury *et al.* 2004, Neuwirt, 2005, Mulligan *et al.* 2007, Farré *et al.* 2007).

1.2.5 RNA splicing

In the mitochondria of flowering plants (angiosperms), introns are of the group II type. They are ribozymic mobile genetic elements and are speculated to have given rise to spliceosomal introns. Group II introns contain 6 domains and have a fairly conserved secondary structure (reviewed in Bonen and Vogel, 2001). Genes can be found in cis- or trans- configuration where in the latter the intron is split in two (usually within domain IV) and the exons are spread around the genomes. The correct coding sequence is restored through splicing at the RNA level where the two halves of the intron find each other to correctly splice and join the surrounding exons.

Splicing occurs through two trans-esterification steps, where a bulged A nucleotide in domain VI attacks the upstream exon-intron junction. Subsequently, the 3' end of the upstream exon, now free, attacks the downstream exon-intron junction to join the two exons. The spliced intron is released in a lariat form (reviewed in Bonen and Vogel, 2001). New evidence for factors involved in the splicing machinery are emerging: a mutant nuclear encoded maturase-like protein shows impaired *nad4* splicing (Nakagawa and Sakurai, 2007) and a PPR protein has been implicated in the trans-splicing of *nad1in1* in *Arabidopsis thaliana* (de Longevialle *et al.* 2007). A maturase type coding sequence (*mat-r*) is also found in flowering plant mitochondria. Its sequence is homologous with the reverse transcriptase and maturase of retroelements containing a RNA binding domain (Chapdelaine and Bonen 1991). Other co-factors imported from the cytosol are thought to play a role (Bonen, 2008).

Interestingly in plant mitochondria, aberrant forms of group II introns are found where the sequence in domain V and VI cannot fold into the usual secondary structure. Spliced introns are found in circular or linear forms (Li-Pook-Than and Bonen 2006) and other splicing mechanisms have to be invoked such as a hydrolytic pathway or the use of an external attacking nucleophile. Notably, editing sites are found in exons at or very close to the exon-intron junction. The "late" editing of these sites is consistent with the access to the

editing site being blocked by the presence of the intron or recognition site being created by the exon-exon junction (Li-Pook-Than *et al.* 2007).

1.2.6 mRNA stability and turnover

The post-transcriptional processing and turnover of transcripts are key in regulating the transcriptome of plant mitochondria (Giegé *et al.* 2000, Holec *et al.* 2006 and Kuhn J *et al.* 2001). A study in *Arabidopsis* revealed that the different levels of transcriptional activity are balanced into similar steady state levels of RNA as determined by transcript degradation (Giegé *et al.* 2000). Those steady state levels of RNA reflect the stoichiometry of the proteins in their functional subunits better than the transcription rates, demonstrating the important role of RNA turnover in regulating the transcriptome.

The balance between stability and degradation of mRNAs is managed by various factors. Evidence for the involvement of the PNPase enzyme in the degradation of processing by-products and transcripts was seen in PNP down regulated *Arabidopsis* mutants (Perrin *et al.* 2004a, b and Holec *et al.* 2006) where there was a marked increase in the accumulation of rRNA processing by-products and in RNA originating from regions lacking known genes. Inverted repeats at the 3' end of transcripts might also form structures that would allow for the binding of proteins, conferring stability to the mRNA molecule as is seen in higher plant chloroplast mRNAs (reviewed in Herrin and Nickelsen 2004).

The presence of non-encoded nucleotides has been reported at the 3' end of mRNAs. Polyadenylation of transcripts is thought to promote degradation of mRNAs (Holec *et al.* 2006, Kuhn J *et al.* 2001, Lupold *et al.* 1999b, Gagliardi and Leaver 1999). Although signals for polyadenylation of transcripts are still unknown, polynucleotides tails of diverse lengths can occur (from 1 nucleotide up to 25 have been observed, and in some cases up to 50-100 nt; Forner *et al.* 2007), mostly comprised of A nts, a few nucleotides upstream of the 3' end (Kuhn J *et al.* 2001, Gagliardi *et al.* 2004). The presence of non-encoded nucleotides can be detected in experiments using the circular-RT-PCR method (see section 1.3.3) where A, C and sometimes T and G are present as short stretches at the 3' end of mRNAs.

Organellar RNAs differ from their bacterial ancestor in that they exhibit relatively stable messengers and precursors despite the lack of the stabilizing effect of poly(A) tails, for example seen in eukaryotic mRNAs (reviewed in Gagliardi *et al.* 2004). Stability of mRNA

can be conferred through different means and it has been suggested that 5' UTRs play in role through secondary structure (Gagliardi *et al.* 2004) or binding of proteins at the 5' or 3 end (as in chloroplasts, reviewed in Herrin and Nickelsen 2004), although there is no direct evidence for this so far.

Insight into chloroplast post-transcriptional events has been gained through work on higher plant plastids and the model system, *Chlamydomonas reinhardtii*, a green alga. In chloroplasts, stabilization of the mRNAs can be mediated through the interaction of proteins with cis-elements in the 5' and 3' UTRs (reviewed in Herrin and Nickelsen 2004). Stem loop structures at the 3' end of transcripts also block the progress of exoribonucleases. Polyadenylated mRNAs in chloroplasts are tagged for degradation, most likely with the help of a plastid PNPase enzyme and other components of a plastid degradosome. Polyadenylation sites can occur within the coding sequences of genes as well, signifying that endonucleolytic cuts in the mRNA happen first and are followed by poly(A) addition (the non-encoded stretch of nucleotides is not exclusively comprised of adenosines and can include up to 25% of guanosines) (reviewed in Schuster *et al.* 1999; Herrin and Nickelsen 2004).

In *E. coli* RNA degradation is carried out by endonucleolytic cleavage in the RNA molecules followed by the addition of short poly(A) tails, which are degraded by exonuclease activity. The degradosome contains PNPase and RNase E enzymes as well as other enzymes and proteins that work as a complex: for example the RhlB enzyme helps unwind the secondary structures of the RNA that impedes the 3' to 5' exonuclease activity by PNPase (Sarkar and Fisher, 2006).

RNA processing and maturation is also of importance in *Saccharomyces cerevisae* mitochondria. Detailed mechanisms of 5' and 3' mRNA ends processing are not known (Gagliardi *et al.* 2004) but mRNAs are not capped or polyadenylated. Yeast mitochondria lack PNPase activity and RNA degradation is dependent on the mitochondrial degradosome as the main exoribonuclease: it is composed of two protein subunits, an RNaseII-like exoribonuclease and a NTP-dependent RNA helicase, which unwinds dsRNA regions and degrades ssRNA in a 3' to 5' direction (Rogowska *et al.* 2006, Malecki *et al.* 2007).

1.2.7 Plant mitochondrial translation

The translational machinery in plant mitochondria is encoded in both the mitochondria and the nucleus. All three ribosomal RNAs (26S, 18S and 5S) are encoded in the mitochondria. Plant mitochondria use the same genetic code as the cytoplasm translation system. There are three categories of tRNAs: mitochondrial-encoded, cytosol-imported of nuclear origin and mitochondrial-encoded of plastid origin. Ribosomal proteins are encoded in both compartments and their distribution in different plants varies due to gene transfer to the nucleus (see section 1.2.10).

In bacteria, the recognition of the correct translation start codon is done through the binding of the 3' end of the SSU rRNA to an antisense sequence 6-7 nts upstream of the start codon, known as the Shine-Dalgarno motif (reviewed in Martinchev and Wagner, 2005). In eukaryotes, the ribosome recognizes capped mRNAs and scans the RNA until the initiation codon. In plant mitochondria however, no evident motif or consensus sequences for proper translation initiation recognition can be found upstream of genes. The 3' end sequence of the mt SSU rRNA has diverged from that in bacteria like *E. coli* and sequences corresponding to the antisense of the mt SSU rRNA are rarely found upstream of start codons in RNA transcripts (<25%, Hazle and Bonen, 2007b). Furthermore indels very close to the start codon are often seen upstream of protein-coding genes, showing a relaxed constraint on this region (Hazle and Bonen, 2007b). The machinery and signals used in translation hence seem to be varied and do not conform to the expected bacterial system.

1.2.8 Early embryo-to-seedling stages of development: mitochondrial transcription/post-transcriptional events coupling

Seed germination is a period of high mitochondrial biogenesis (Logan *et al.* 2001): embryos from dry seeds in maize contain two types of mitochondria (heavy/type 1 and light/type 2). At 0 hr the majority of mitochondria is of the light/type 2 kind and represents the remains of active mitochondria in the developing seeds (Logan *et al.* 2001). This mitochondrial fraction degrades during desiccation and contains a number of proteins for import but does not have a functional tricarboxylic acid (TCA) cycle. The heavy/type 1 mitochondria (pro-mitochondria) are present in dry embryos but are not yet fully differentiated. Import of cytosolic components and synthesis of mitochondrial-encoded

proteins starts at the onset of imbibition and by 6hr they possess all the components for mitochondrial biogenesis. By 24hr these mitochondria are fully functional (functional TCA cycle and election transport chain) (Logan *et al.* 2001). A burst in oxygen uptake from 0 hr to 24hr indicates the time of high mitochondrial biogenesis (Logan *et al.* 2001), also seen in wheat embryos (Khanam *et al.* 2007).

Work done during early germination in rice suggests that the import apparatus is present in pro-mitochondria, before germination is initiated and that at later times the composition of the proteome shifts from components for import and assembly to components for the metabolic function (Howell *et al.* 2006). In the early stages of development, while awaiting the assembly of components of the respiratory chain (encoded in the nucleus and the mitochondria) respiration might be relying on the alternative pathways present (Howell *et al.* 2006).

A tight coupling and close communication with cytosolic protein import are necessary to achieve proper expression since many post-transcriptional processing events undergone by mitochondrial transcripts are carried out by imported proteins. Although the nature of most proteins involved in those processes is still unknown (see above and reviewed in Gagliardi and Binder 2007), many candidates are found in the pentatricopeptide repeat (PPR) family of proteins. These nuclear encoded proteins are part of a very large family in higher plants [442 in *Arabidopsis* (Lurin *et al.* 2004) and 655 predicted in rice, (Andres *et al.* 2007, Geddy and Brown 2007)] and PPR proteins have been found to be involved as specific trans-factors in many processes in the mitochondrion and chloroplast such as RNA degradation, cleavage and stabilisation (reviewed in Andres *et al.* 2007) and editing in chloroplast (Okuda *et al.* 2006, 2007). Notably, PPR proteins have been identified as restorers of cytoplasmic male sterility (CMS) in species such as Petunia, *Brassica*, radish and rice where aberrant transcripts from fused ORFs are degraded or cleaved (reviewed in Chase 2007).

Studying transcriptional and post-transcriptional events during seed germination is of particular interest since transcript profiles often appear more complex when observed during these developmental stages, with high relative abundances of precursors at the onset of germination (Li-Pook-Than *et al.* 2004). There is a shift between precursors and mRNA relative abundances over development, consistent with a lag in the coupling of these events.

Observations are compatible with a bottleneck due to a deficit in imported machinery: transcription and processing might not be properly coordinated in the early stages of development, while the machinery is being imported and made available. The efficiency of this coupling appears to increase over development. Transcripts present in dry embryos and during very early time of seed imbibition / rehydration represent stored messengers and not *de novo* transcribed RNA. Macroarray analysis of the wheat mitochondrial transcriptome reveals variability in transcript abundances during embryo germination (Khanam *et al.* 2007). Until the start of *de novo* transcription during seed germination, the mitochondrion relies on stored messenger RNAs and the machinery in place that maintains respiration in dry embryos (Bewley, 1997).

1.2.9 Effect of mitochondrial DNA rearrangements on gene expression

The highly recombinogenic nature of plant mitochondrial DNA allows genes to be shuffled around the genomes and acquire new surrounding sequences. For a gene to be successfully relocated new regulatory signals must be present, whether gained or brought along during rearrangement (Adams and Palmer, 2003). Gene linkages are often disrupted and previously co-transcribed genes must acquire new expression signals. Plant mitochondrial genomes are comprised of a high amount of intergenic sequences and recruitment of expression signals can be done from these non-coding sequences. In some cases, "expression cassettes" are present upstream of several genes across species and within a genome, maybe representing a "good" set of expression signals (Hazle and Bonen 2007b). Indeed, copies of sequences upstream of functional genes (or part of sequences upstream of genes) are present in spacer regions (Hazle and Bonen 2007b). A rearrangement event could bring those previously unused signals in proximity of a gene, which would in turn be able to use this new context for expression. The presence of duplicated copies of upstream sequences also implies the potential expression of non-coding sequences throughout the genome; indeed in *Arabidopsis* a duplicated copy of the *rrn26* promoter is functional and expresses the non-coding region it precedes but the resulting transcript is rapidly turned-over by PNPase activity in *Arabidopsis*, as inferred from PNP mutants (Holec *et al.* 2006). It should be noted that previously uncharacterized sequences can also play the role of regulatory signals after rearrangements occur in their proximity. The plasticity of the

regulatory signals allows these processes to be more prevalent among plant mitochondrial genomes.

1.2.10 Gene transfer to the nucleus

Plant mitochondrial genomes undergo ongoing gene transfer to the nucleus over evolutionary time, which results in slightly different gene content among species. Gene transfer can happen for most genes but is most rampant for the ribosomal protein genes (Adams *et al.* 2002a). A survey of 280 angiosperms (flowering plants) for the presence of ribosomal protein coding genes in the mitochondria reveals that some ribosomal protein genes appear to have been transferred to the nucleus several times independently (up to 33 times in the case of *rps1*) (Adams *et al.* 2002a).

Transfer is hypothesized to occur when a mitochondrion (or several) bursts and genes are incorporated in the nuclear genome (Adams and Palmer, 2003). This is thought to occur via cDNA since mitochondrial genes in the nucleus are found without mitochondrial group II introns and in their edited form. Once the transferred gene is integrated in the nuclear genome, it acquires proper expression signalling, including a promoter for nuclear transcription, termination signals and translation signals as well as protein targeting signals to the mitochondria (Adams and Palmer, 2003). These might arise by the duplication of signals associated with another gene (e.g. upstream of wheat *Rpl5*, duplication of *Rpl4* preceding sequences, Sandoval *et al.* 2004) or the integration in the intron of a pre-existing gene and the use of alternative splicing during expression (e.g. *Rps14* is located in the *sdh2* intron in cereal species, Figueroa *et al.* 1999 and Kubo *et al.* 1999). In some cases the transferred genes acquire regulatory signals from a sequence of unknown origin (maize *Rpl5*, Sandoval *et al.* 2004). When the nuclear copy is activated, there is a functional copy in the mitochondria and in the nucleus, a period called the transition stage. In wheat *rpl5* is in a transition-state (Sandoval *et al.* 2004) and *cox2* is an example of such a state persisting in legumes (Adams *et al.* 1999). At this time, either copy could take over and become the sole copy of the gene. The gene transfer is successful once the mitochondrial copy is deactivated and the functional copy used is encoded in the nucleus. It is possible to detect pseudogenes or gene remnants of transferred genes in the mitochondrial genome because of its low rate of nucleotide substitution (Wolfe *et al.* 1987).

In some cases, the mitochondrial copy is lost but the protein is functionally replaced by a chloroplast copy of the gene encoded in the nucleus or by a nuclear gene. In *Arabidopsis*, the mitochondrial genes *rps8* and *rps13* have been replaced by the cytosolic gene *Rps15A* and the chloroplast gene *rps13* respectively (Adams *et al.* 2002b).

Once a mitochondrial copy is a pseudogene and no longer under constraint, rearrangements can occur that may cause the loss of this gene from the mitochondrial genome. Interestingly since mitochondrial genes can be closely linked, if one of the genes is a pseudogene it is possibly being kept because of the close association with the functional upstream or downstream gene: such speculation is made for *Ψrps14* in the grass lineages (Ong and Palmer 2006).

1.3 Analyzing RNA in plant mitochondria

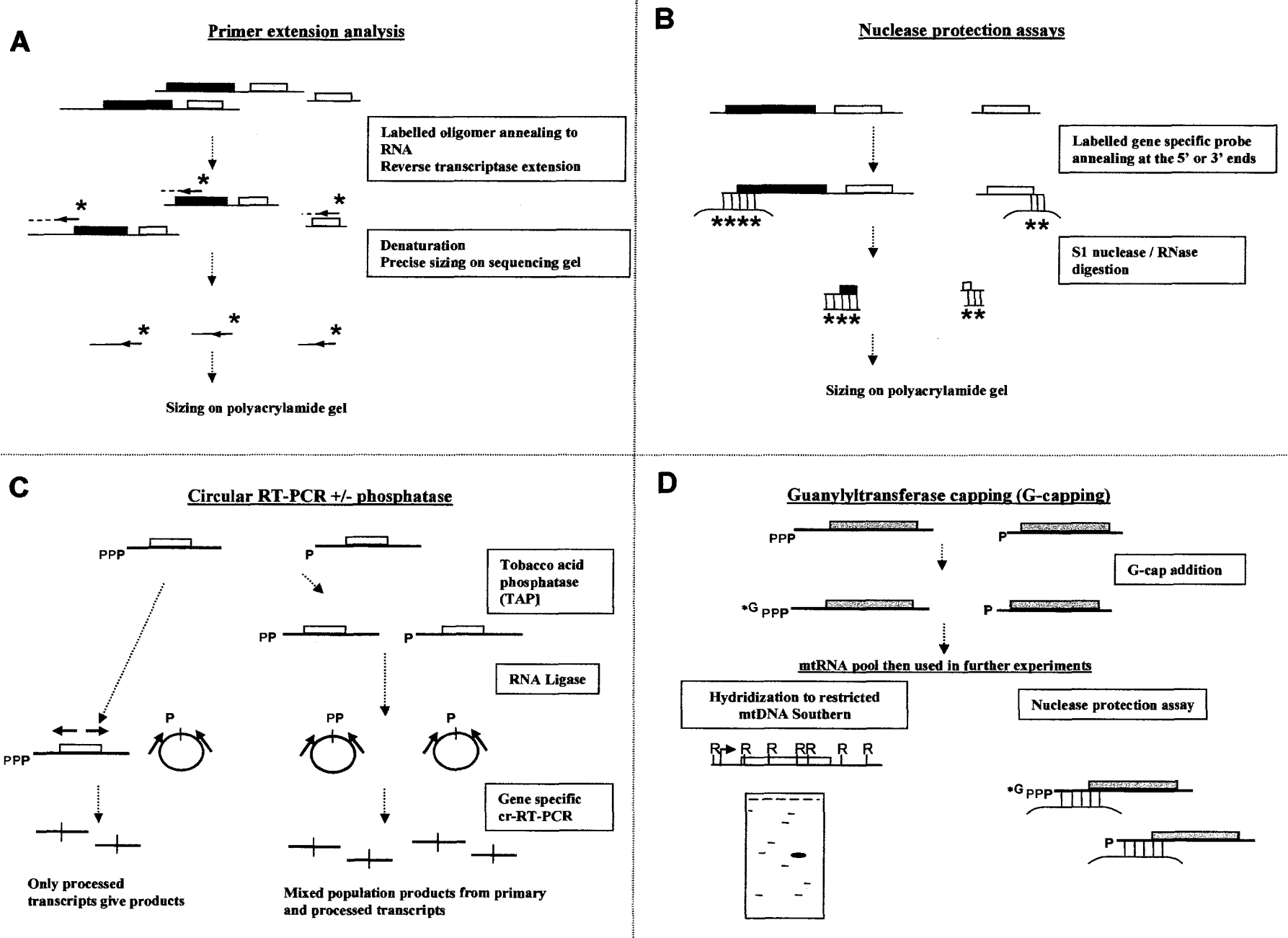
In order to look at the transcriptional and post-transcriptional events in plant mitochondria, several experimental approaches can be used. Key aspects of the RNA analysis done in this study include mapping ends of transcripts, differentiating between primary and processed transcripts and identifying sites of transcription initiation. Old and new methodologies have been exploited and central techniques used in this study will be further described here (Figure 1.3).

1.3.1 Primer extension

The primer extension experiment is a common technique to determine the 5' end of a transcript (Figure 1.3A). A radioactive end-labelled oligomer is used in a reverse-transcriptase reaction on mitochondrial RNA. The product extends to the 5' end of the RNA. The resulting cDNA is denatured and electrophoresed on a polyacrylamide denaturing gel. This technique does not differentiate between primary and processed transcripts but precisely maps the 5' end to the nucleotide. However cDNA extension can be affected by RNA secondary structures in regions resulting in premature extension termination (Sambrook *et al.* 1989). This technique was used in mapping the ends of the *ccmFN-rps1* transcript in 6d wheat seedlings (Gonzalez *et al.* 1993).

Figure 1.3: Schematic of the experimental techniques used to study RNA in plant mitochondria

A) Primer extension analysis B) Nuclease protection assays C) Phosphatase circular RT-PCR and D) Guanylyltransferase capping experiment. Filled boxes represent genes, tri- (PPP) and mono-(P) phosphate groups are shown at the 5' end of transcripts, horizontal arrow heads are oligomers. Hybrid formation and probe annealing are shown with hydrogen bond lines. In C and D, blue lines are primary transcripts, red lines are processed transcripts. Red stars represent radioactive labelling. Experimental steps are boxed on the right. R represents restriction sites in the DNA sequence


A
Primer extension analysis
Labelled oligomer annealing to RNA
Reverse transcriptase extension
Denaturation
Precise sizing on sequencing gel
Sizing on polyacrylamide gel
B
Nuclease protection assays
Labelled gene specific probe annealing at the 5' or 3' ends
S1 nuclease / RNase digestion
Sizing on polyacrylamide gel
C
Circular RT-PCR +/- phosphatase
PPP
P
Tobacco acid phosphatase (TAP)
PP
RNA Ligase
Gene specific cr-RT-PCR
Only processed transcripts give products
Mixed population products from primary and processed transcripts
D
Guanylyltransferase capping (G-capping)
G-cap addition
mtRNA pool then used in further experiments
Hydridization to restricted mtDNA Southern
Nuclease protection assay

1.3.2 Nuclease protection assays

Determining the 5' and 3' ends of RNA transcripts is also possible using nuclease protection assays, where a radioactive labelled DNA fragment (or antisense RNA) is placed in hybrid with RNA and the single stranded molecules not in hybrid are digested by the nuclease enzyme. The protected product is then denatured and electrophoresed on a polyacrylamide denaturing gel (Figure 1.3B).

A common experiment involves S1 nuclease, which can digest single stranded RNA or DNA molecules. A gene specific PCR product or restriction fragment is generated to be longer than where the ends of the RNA are expected to be. These products are uniformly labelled or 5' end labelled respectively denatured and placed in hybrid with mitochondrial RNA. Once the RNA ends are protected by being annealed in hybrid, the single stranded RNA and DNA not in hybrid are digested by S1 nuclease leaving a labelled protected fragment. This product is electrophoresed next to a size marker. This maps the 5' or 3' ends of transcripts precisely (Sambrook *et al.* 1989). This technique was used in mapping the ends of the *cox3* transcript in 6d wheat seedlings (Gualberto *et al.* 1990).

1.3.3 Circular-RT-PCR

This method allows the identification of both 5' and 3' ends at once (Figure 1.3C). RNA samples are treated with T4 RNA ligase to circularize the transcripts, in a mono-molecular fashion. RNA ligation requires a di- or mono-phosphate 5' end, excluding the ligation of primary transcripts which have a tri-phosphate at their 5' end. After ligation, a gene specific RT-PCR is performed to amplify products over the newly formed 5'-3' junction (cr-RT-PCR), which are then cloned and sequenced (Kuhn and Binder, 2002). This technique was used in mapping the ends of *rps12, cox2* and *atp9* transcripts in maize mitochondria (Williams *et al.* 2000).

1.3.4 Differentiating between primary and processed transcripts: Tobacco-acid-phosphatase-cr-RT-PCR and guanylyltransferase-capping

The methods described previously do not allow the identification of transcription initiation sites specifically. The guanylyltransferase enzyme is used in order to add a labelled GTP to the 5' end of tri-phosphate primary transcripts (similar to the G-cap of eukaryotic

cytosolic mRNAs) and this labelled RNA is used in a nuclease protection experiment, in hybrid with PCR products or restriction fragments of the region of interest. The signal seen corresponds to the 5' end of primary transcripts only (Figure 1.3D) (Sambrook *et al.* 1989). This technique was used in the definitive mapping of transcription initiation sites in wheat (Covello and Gray 1991).

Primary transcripts have a tri-phosphate 5' end whereas processed transcripts have a mono-phosphate. A new method exploits this difference in 5' end chemistry to allow to experimentally differentiate between primary and processed transcripts when using circular-RT-PCR (Figure 1.3C). Tri-phosphate transcripts are not able to undergo RNA ligation. In order to ligate such primary transcripts, RNA samples are first treated with tobacco acid phosphatase (TAP) which removes the outer most phosphate, hence allowing circularisation. Followed by gene specific cr-RT-PCR, the products are a mixed population of primary and processed 5' ends. The phosphatase treatment was in a slightly different ligation reaction by Kuhn *et al.* (2005) to map the 5' ends of *Arabidopsis thaliana* transcripts. Since T4 RNA ligase is able to ligate di-phosphate 5' ends, a small amount of damage at the 5' end would allow ligation (Forner *et al.* 2007) Differences in TAP treated and non-TAP treated samples must hence be interpreted carefully.

1.4 Objectives

The objectives of my thesis are to learn more about RNA processing events in plant mitochondria during early seed-to-seedling development. Sequences of whole mitochondrial genomes are becoming available and sequence comparisons between species have revealed gene linkages that have been conserved over long evolutionary time, notably *ccmFN-rps1-mat-rnad1e-nad5c* in rice, wheat and maize (Clifton *et al.* 2004). Previous work on *ccmFN-rps1* in wheat led us to investigate this region further, particularly in regards to its transcript profile during early embryo development. Studies of mitochondrial gene transfer to the nucleus have identified genes in transition state in mitochondrial genomes. This is the case of *rpl5* in wheat (Sandoval *et al.* 2004) which prompted our interest in characterizing its transcript profile in mitochondria. My organism of interest is primarily the cereal species wheat (*Tricitum aestivum*), an important crop plant worldwide and widely used in studying mitochondrial type events. Studies with other cereal species, notably rice (*Oryza sativa*) have

also been included to provide comparative data and assess the conservation of signals and cis-elements.

How do the transcript profiles of the *ccmFN-rps1* and *rpl5-Ψrps14-cox3* loci in wheat mitochondria differ during early embryo-to-seedling development? In wheat 6d seedlings, *ccmFN-rps1* are co-transcribed as described by Gonzalez *et al.* (1993) and no monocistronic transcripts for either gene are detectable. A northern hybridization of 2d wheat mtRNA (T. Hazle, a former Ph.D. graduate in our laboratory, unpublished results) revealed a more complex transcript profile however, prompting us to investigate this locus in wheat mitochondrial RNA during early seed-to-seedling developmental stages. In wheat, *rpl5* is in a transition state and its co-transcription with *Ψrps14* has been characterized (Sandoval *et al.* 2004). They are linked downstream to *cox3*, for which a promoter has been identified (Covello and Gray, 1992). The expression profiles of the *rpl5-Ψrps14-cox3* loci during early times of development will be examined.

What are the regulatory signals and processing events leading to the formation of discrete transcripts for these loci? By mapping the ends of the transcripts for *ccmFN-rps1* and *rpl5-Ψrps14-cox3* loci, we hope to discover the regulatory signals for these regions and how the processing events are linked to form their respective transcript profiles. The identification of cis-elements for these two gene loci and their processing over development will help us understand how mitochondrial RNA species are processed post-transcriptionally.

How are *rpl5* and *Ψrps14* transcribed in two closely related plants, wheat and rice, with different genomic context? In wheat, the *rpl5-Ψrps14* locus is found ~600 bp upstream of *cox3*, whereas in rice these genes are located far from other genes. We are interested in comparing the RNA processing pathways in wheat and rice who have diverged ~50 MYA, since they share a certain amount of surrounding sequences. In addition, *rpl5* is in a transition-state in wheat (Sandoval *et al.* 2004). What is the status of these genes in other closely related plants like oats and barley? Furthermore, in wheat and rice *rps14* was transferred to the nucleus in a common gene transfer event and the *rps14* pseudogene has persisted in the mitochondrial genome of most cereals. Does it hold a specific role in the transcription of *rpl5?*

Chapter 2 Materials and methods

2.1 Plant material

Wheat (*Tricitum aestivum* var. Frederick), oat (*Avena sativa* AC Goslin) and barley (*Hordeum vulgare* OAC Kippen) seeds were kindly provided by Dr R. Pandeya (Agriculture and Agri-food Canada). Rice (*Oryza sativa* var. Drew), rye (*Secale cereale* var. Gazelle), orchard grass (*Dactylis glomerata*), brome (brome grass, *Bromus*) and *lolium* (perennial rye grass, *Lolium perenne*) seeds were commercially purchased from Ritchie Feed and Seeds Inc. (Ottawa, ON, Canada). Maize seeds (*Zea mays* var. D39) were kindly provided by Direct Seeds, Inc. (Chatham, ON, Canada)

2.2 Mitochondrial nucleic acid isolation

50g of seeds were surface-sterilized in 1:6 dilution of Javex in distilled water and in 10mN HCl then rinsed in distilled water, and embryos were dissected from the endosperm prior to imbibition on 3mm filter paper in petri dishes with sterile water (6hr, 12hr, 18hr and 24hr) or germination of whole seeds in vermiculite (2d and 6d seedlings) in the dark at room temperature.

Mitochondrial DNA and RNA were isolated using procedures as previously described (Subramanian *et al.* 2001). Briefly, after homogenization in cold buffer I (0.44M sucrose, 50 mM Tris pH 8.0, 3mM EDTA, 1mM β-mercaptoethanol, 0.1% BSA) using cold mortar and pestle and filtration through cheesecloth and Miracloth (Calbiochem), crude mitochondria were isolated using differential centrifugation (twice five minutes at 500g and once 25min at 12000g) to remove cytosolic contaminants.

For DNA isolation, the crude mitochondrial pellet was resuspended in Buffer II (50mM Tris pH 8.0, 20 mM EDTA) and mitochondria were lysed using Buffer III (20mM Tris pH 8.0, 100 μM EDTA, 200mM NaCl, 2% SDS, 200 mM β-mercatoethanol) at 65°C for 20 minutes. DNA was subsequently precipitated with KOAc, isopropanol with NH_4OAc, ethanol and isopropanol with NaOAc. After collection of the pellet by centrifugation and vaccum drying, mtDNA was resuspended in TE buffer (10 mM Tris pH 7.5, 1mM EDTA) and stored at -20°C. Yield from this extraction was typically between 0.5 and 1 μg per gram

of wet weight tissue, with varying amount of nuclear DNA contamination depending on the species used.

RNA isolation of crude mitochondrial pellets obtained from differential centrifugation (as previously described in Subramanian *et al.* 2001) was done as follows. Pellets were resuspended in Buffer IV (10mM Tris pH 7.5, 50mM KCl, 10mM $MgCl_2$) and Buffer IV with 8% Triton-X 100. Mitochondria were lysed using detergent mix (2% tri-isopropylnapthalene sulfanate, 12% sodium p-aminosalicytate, 0.1M NaCl, 20mM Tris pH 7.4) and RNA was extracted twice with phenol (1.5 volume of phenol saturated in TE). Precipitation of nucleic acids was done using 0.1 volume 5M NaCl and 2 volumes 95% EtOH. Yields were typically between 5-10μg of RNA per gram of wet weight tissue.

RNA samples were treated with FPLC pure, RNAse-free DNAseI (Pharmacia) with 1X DNAse buffer (40mM Tris, pH 7.5), 6 mM $MgCl_2$) and 20 units of RNase Inhibitor (Promega) for 30minutes at 37°C. Sampled were subsequently extracted once with equal volume of TE-saturated phenol and once with phenol-chloroform-IAA (25:24:1) before ethanol precipitation overnight in 2.5X volume EtOH and 0.1 volume 5M NaCl.

2.3 Southern and northern hybridization

2.3.1 Southern and northern blots

Procedures were adapted from standard protocols (Sambrook *et al.* 1989).

For Southern blots, ~1 to 2μg of mtDNA was cut with restriction enzymes (as per standard protocol, Sambrook *et al.* 1989) and electrophoresed on 0.8% agarose gels in 1X TBE. Gels were ethidium bromide stained and photographed. Treatment with 0.25N HCl for 15 minutes was followed by denaturation 20min in 1.5M NaCl and 0.5M NaOH and neutralization 30 min in 3M NaOAc (pH 5.5). Transfer to a nylon membrane was done overnight by capillary action to a stack of paper in 20X SSC buffer (6M NaCl, 0.6M sodium citrate, pH 7.0), air dried the next day and UV cross-linked 1minute.

For northern blots approximately 5μg mtRNA was denatured in loading buffer containing 1X MOPS buffer pH 7.0 (40mM 3-N-morpholino-propane-sulphonic acid, 10mM sodium acetate, 1mM EDTA pH 8.0), 6.6% formaldehyde, 50% formamide, RNA loading buffer (50% glycerol, 1mM EDTA, bromophenol blue) and 40ng Et-Br by heating at 70°C 10min with frequent vortexing. The RNA was electrophoresed in a submarine gel system on

a denaturing 1.25% agarose gel with 7.2% (v/v) formamide in 1X MOPS buffer which was pre-run for 1 hour at 60V (~24mA). Electrophoresis was performed for the first hour at the same voltage/amperage then increased to 80-100 V for up to 3 hours. The RNA gel was then transferred to a nylon membrane overnight by capillary action to a stack of paper in 20X SSC buffer (6M NaCl, 0.6M sodium citrate, pH 7.0), air dried the next day and UV cross-linked 1 minute.

2.3.2 Southern and northern hybridization

Southern and northern membranes were pre-hybridized 2 hours in 10-15 ml of the appropriate hybridization buffer in rotating hybridization ovens at 40-42°C, before hybridization overnight, with 1-3 ml fresh hybridization buffer and 5-15 µl of radiolabelled probe. Hybridizations with end labelled oligomers were done overnight in oligo-mix (5% deionized formamide, 5X SSC, 0.1% SDS, 50µg/ml yeast tRNA) and random labelled probes were hybridized overnight using alpha-hybridization mix (5X Denhardt's solution, 5X SSC, 1% sodium dodecyl sulphate (SDS), 50mM $NaPO_4$, pH 7.0 and 250 µg of denatured herring sperm DNA).

Membranes were then washed twice in 20ml 20X SSC, 0.1% SDS for 20 minutes at the same hybridization temperature in order to remove non-hybridized probe. Exposure for the appropriate amount of time was done either on phosphoimaging screens (Kodak) at room temperature or on X-ray films (Kodak) at -80°C. Phosphoimaging screens were scanned in the phosphoimager scanner (Bio-rad Molecular imager FX) and X-ray films were developed using Kodak developer and fixer reagents.

2.4 Radiolabeled probes for hybridization

Two types of ^{32}P probes were used for hybridization: 5' end-labelled oligomers and random-labelled PCR products.

Oligomer probes were prepared using 100 ng of 20 nt oligomers incubated with 40 µCi γ-^{32}P-ATP (3000 Ci/mmol, Amersham), 5 units of T4 polynucleotide kinase (Invitrogen) and 1X kinase buffer (50mM Tris pH 9.5, 10 mM $MgCl_2$, 5mM DTT) for 45 minutes at 37°C. 37.5 µl of TE was added to stop the reaction prior to spinning through a Sephadex G-50 column (previously equilibrated with TE pH 8.0) in order to remove unincorporated radioactive label. 50 µl TE was passed twice more through the column in order to get three

eluants. The eluant with the highest specificity as measured by Geiger counter (1st or 2nd) was used for hybridization.

Random-labelled PCR product probes were prepared using the Megaprime labelling system (Amersham Biosciences). Approximately 25ng of PCR product was boiled 5 mins with the recommended amount of random nonamer primers from the kit and placed on ice 5 mins. The labelling reaction was done at 37°C for 45 minutes with 1unit of DNA polymerase I (Klenow) in the kit's buffer, dCTP, dTTP, dGTP and 25μCi α-^{32}P dATP (3000Ci/mmol, Amersham), according the kit's protocol. The reaction was stopped by the addition of 75.5μl TE and spun through a Sephadex G-50 column in order to remove unincorporated radioactive label. Activity of the probe and column was assayed with a Geiger counter and typically showed a 40-50% incorporation. Before use in hybridization experiments, probes were boiled 5mins and placed on ice 5 mins.

2.5 Phosphatase and RNA ligase treatment, reverse-transcriptase and polymerase chain reaction of mtRNA (cr-RT-PCR)

Wheat mtRNA (~5μg) was incubated in 1X supplied tobacco acid phosphatase reaction buffer, BSA and 0.16U/μl tobacco acid phosphatase (TAP) (Epicentre, Madison, WI), as per supplier's protocol, supplemented with 1.3U/μl RNAsin (Promega) at 37°C for 1h, subsequently extracted with phenol:chloroform:IAA (1:1:25), and precipitated with 2.5 vol 95% EtOH and 0.5 vol 5M NaCl. After recovery, wheat mtRNA was circularized and RT-PCR was performed as previously described by Li-Pook-Than *et al.* (2006). Briefly, mtRNA was incubated overnight at 14°C with T4 RNA ligase (New England Biolabs), 0.1 vol 10X supplied T4 RNA ligase buffer, supplemented with BSA (50μg/ml) and 0.5U/μl RNAsin (Promega) and once again precipitated with 2.5 vol of 95% EtOH and 0.5 vol of 5mM NaCl precipated.

After denaturation of the RNA 10min at 70°C, reverse transcriptase reactions were performed as recommended by supplier using 10U/μl Superscript II (Invitrogen), 0.2vol 5X 1st strand buffer, 10mM DTT and 0.5mM dNTPs for 1h30 at 42-45°C, using appropriate primers and a control. PCR reaction were performed as recommended by the supplier with DNA *Taq* polymerase (Invitrogen) for 30 cycles (94°C 30s, 50-60°C 30s-1min and 72°C 30s-1 min) with selected primers and in some instances reamplified with nested primers, for verification of products. Products were analyzed on 1.2%-1.5% agarose gels with appropriate

DNA size markers (MBI Fermentas). The 3 RNA treatments are: with tobacco acid phosphatase (TAP) and with RNA ligase (+T+L), without TAP and with RNA ligase (- +L) and without TAP and without RNA ligase (-T-L, negative control). Oligomers (Table 2.1) were designed based on sequences from the NCBI databank (Table 2.2) and ordered from Invitrogen, ON, Canada.

2.6 Cloning of PCR products and sequencing

PCR products were gel extracted using UltraClean 15 kit (MoBio Laboratories Inc.) and ligated overnight at 4°C into pGem-T-easy plasmid vectors with the pGem-T-easy vector system (Promega). The ligated products were cloned in competent TB1 strain *E.coli* cells, prepared according to Sambrook *et al.* 1989. Plasmids were isolated from bacterial colonies with the QIAprep spin Miniprep kit (Quiagen).

2.7 S1 nuclease protection assay

PCR products for S1 nuclease protection assays were either 5' end labeled or uniformly labeled. The uniformly labeled products were done as follows: PCR reactions were set up with 1 ng mtDNA, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.01mM dNTPs, 125ng of each primer, 0.5U *Taq* polymerase (Invitrogen) and 15μCi α-^{32}P dATP (3000Ci/mmol, Amersham) and run for 30 cycles (94°C 30s, 52°C 30s, 72°C 1min). The reaction was filtered through a Sephadex G-50 column in order to remove unincorporated radioactive label.

Approximately 5 μg of mtRNA dissolved in 5μl water and 3μl 5X PIPES buffer (0.2 M PIPES pH 6.5, 5 mM EDTA, 2 M NaCl) was dried and resuspended in 12 μl deionized formamide and mixed with 3 μl (~10 ng) of the denatured probe. Incubation at 85°C for 15min was followed by a slow cooling to 47°C and hybrids were allowed to form overnight. The RNA/DNA hybrids were then cooled down further to 43°C over ~45 min and placed on ice before adding 300 μl cold S1 buffer (0.25 M NaCl, 30 mM NaOAc pH 5.5, 1 mM zinc acetate, 20 μg/ml denatured herring sperm DNA). The reaction was divided in 4 tubes with 0, 5, 10 and 25 Units of S1 nuclease (Invitrogen) respectively and incubated 30 min at 30°C. 25μl of stop buffer (4M NH_4OAc, 50 mM EDTA, 0.05 μg/μl tRNA) was added and precipitation of nucleic acids was done with 240 μl 95% EtOH at -20°C overnight. RNA/DNA hybrids were pelleted by centrifugation 20 minutes and washed once in 95% EtOH before drying 10 min at 60°C and resuspension in 4μl TE and 8μl loading buffer. The

Table 2.1: Sequence and position of oligomer used in this study.

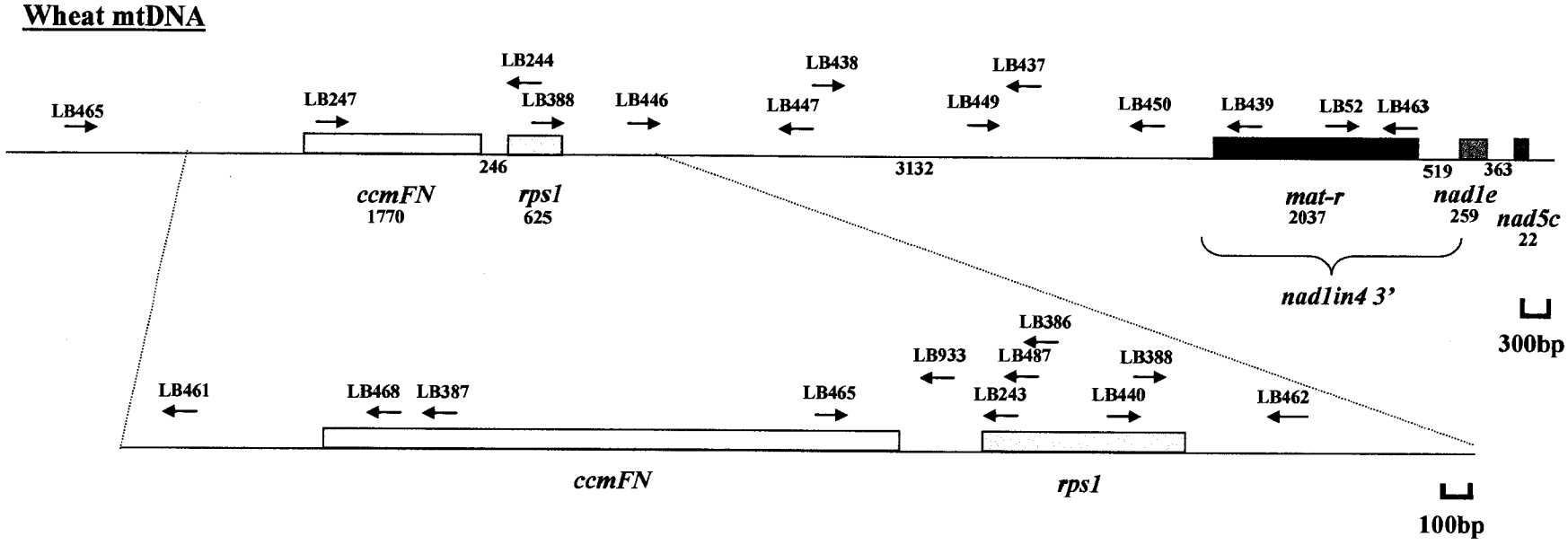
Oligomers name, position and direction referred to in figure 2.1, location based on accession number given. Oligomers were ordered from Invitrogen (CA, USA)

	Location	Position	Sequence (5'-3')	Orientation	Accession no.
A) *ccmFN*					
LB469	51865-52243	upstream *ccmFN*	TTTCTACAACAGGTGCTTGG	sense	AP008982
LB461	52262-52243	upstream *ccmFN*	TTTCTTCAGTCTCACAAGCG	antisense	AP008982
LB247	215-234	5' end *ccmFN*	GTGCAGCACTTGCATTTTGG	sense	X69205
LB468	447-428	5' end *ccmFN*	TGAGACATTATGGCTTTGGG	antisense	X69205
LB387	526-507	5' end *ccmFN*	GTTGGAGAGATAGAATGGAG	antisense	X69205
LB465	1539-1558	3' end *ccmFN*	TCTTCATTCTTGGACCTCGC	sense	X69205
B) *rps1*					
LB933	1961-1942	intergenic *ccmFN-rps1*	ATCAGAAGTATCCAGCAGCG	antisense	X69205
LB243	2010-1991	5' end *rps1*	TCGACTCCAATAGATGCTCA	antisense	X69205
LB487	2099-2080	5' end *rps1*	TCTTCCCTTAAGCGTAAGCG	antisense	X69205
LB386	2186-2167	5' end *rps1*	CCTGGTGGCTCGGTTGATTG	antisense	X69205
LB440	2380-2399	3' end *rps1*	TCAGACAAAACCGAGCTTGG	sense	X69205
LB388	2407-2426	3' end *rps1*	TGAAGAAGATTTGGCGAACG	sense	X69205
LB244	2501-2482	3' end *rps1*	AGTAGTAAGACCCGCGATGG	antisense	X69205
LB446	42928-49209	*rps1-mat-r* intergenic	ATCATGCGTGACTGGCGGAG	sense	AP008982
LB462	48966-48985	*rps1-mat-r* intergenic	TTCAGTAAACTGACCTTGGC	antisense	AP008982
LB438	48385-48376	*rps1-mat-r* intergenic	TTGATTCCGTTCGTTCCTGC	sense	AP008982
LB447	48299-48318	*rps1-mat-r* intergenic	TTACCCACGAGAGTCAGAGG	antisense	AP008982
LB437	47382-47401	*rps1-mat-r* intergenic	ACACAAGCGGCAATCATCCG	antisense	AP008982
LB449	47380-47361	*rps1-mat-r* intergenic	AATTTCATTGCTGTGGGTCC	sense	AP008982
LB450	137-114	*rps1-mat-r* intergenic	AAGGCTCTTTAAGGAAGTAGTCCC	antisense	X57965
C) *mat-r*					
LB439	558-539	*mat-r*	TTCTCACCCCCTACGAGTCG	antisense	X57965
LB52	1190-1215	*mat-r*	CGATATGCCGACGACTTACTACTG GG	sense	X57965
LB463	1746-1727	*mat-r*	TTCTTCAGGACACTCATGCC	antisense	X57965
C) *rpl5*					
LB509	30759-30778	upstream *rpl5*	AGGACTAGAGTGCTAAGCAG	antisense	AP008982
LB406	30658-30677	upstream *rpl5*	TTCGATATTTCACTCCGTGG	antisense	AP008982
LB510	277-258	5' end *rpl5*	TGGAGTGGAAACATCATGGC	antisense	AJ535507
LB352	319-300	5' end *rpl5*	TTGAGCAACAGATCCTGACG	antisense	AJ535507
D) *rps14*					
LB283	862-881	5' end *rps14*	CACAAACGTAGATTGCTCGC	sense	AJ535507
LB489	992-1111	3' end *rps14*	ATGGCACGAGTAAGAAACCG	sense	AJ535507
LB313	1233-1214	intergenic *rps14-cox3*	ATGACTCCGTTAGTTTCGGT	antisense	AJ535507
E) *cox3-tRNA^E*					
LB345	652-633	5' end *cox3*	GTACATCACACCTCCTACGG	antisense	X15944
LB314	999-980	middle of *cox3*	ATGAGCCCAAGTTACGGCAG	antisense	X15944
LB350	739-758	3' end *cox3*	TCTACGTGAATGCACGTTGG	sense	X15944
LB346	1267-1286	3' end *cox3*	AGCTGCATGGTACTGGCATT	sense	X15944
LB511	1347-1366	3' end *cox3*	AGAAGGGAACGAATAAGTGG	sense	X15944
LB407	27325-27344	intergenic *cox3-tRNA^E*	TGAATGGCTCGTGTTGCGCT	antisense	AP008982
LB374	27060-27079	*tRNA^E*	AATCGAACCCGTGTCTTCGA	antisense	AP008982
D) 18S rRNA					
LB211	56895-56914	*18rRNA*	GTGATCATTGGTCCGATGCT	antisense	AP008982

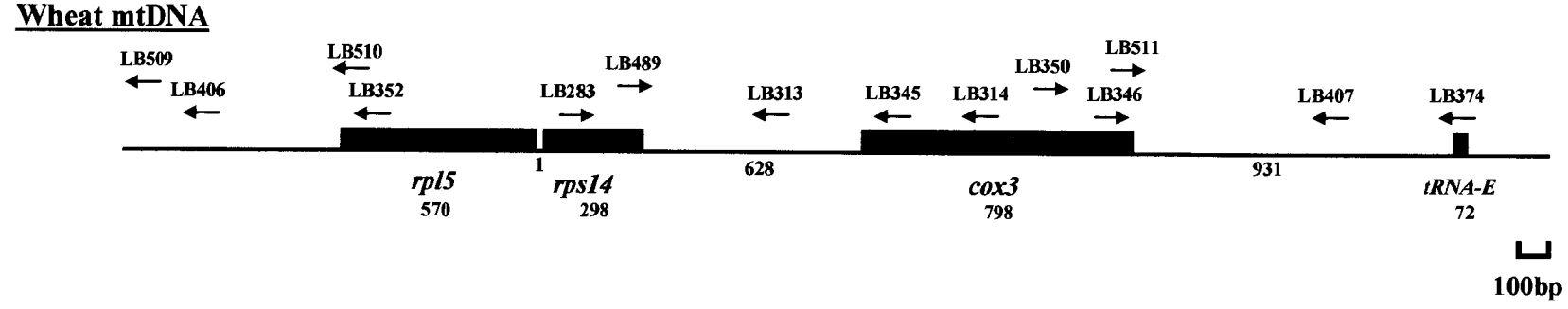
Figure 2.1: Schematic showing the positions of the oligomers used in this study

Based on the wheat mtDNA (AP008982) for the genomic regions A) *ccmFN-rps1-mat-r-nad1e-nad5c* and B) *rpl5-Ψrps14-cox3*. Filled boxes represent genes, oligomer names referred to in Table 2.1. Arrow heads represent oligomers with orientation. Names of genes and sizes of the coding sequences and intergenic spaces are written below

A


Wheat mtDNA
LB465
LB247
LB244
LB388
LB446
LB447
LB438
LB449
LB437
LB450
LB439
LB52
LB463
246
3132
519
363
ccmFN
1770
rps1
625
mat-r
2037
nad1e
259
nad5c
22
nad1in4 3'
300bp
LB461
LB468
LB387
LB465
LB386
LB933
LB487
LB388
LB243
LB440
LB462
ccmFN
rps1
100bp

B


Wheat mtDNA
LB509
LB406
LB510
LB352
LB489
LB283
LB313
LB345
LB314
LB350
LB511
LB346
LB407
LB374
1
628
931
rpl5
570
rps14
298
cox3
798
tRNA-E
72
100bp

protected products were electrophoresed on a 7% urea denaturing polyacrylamide gels (7 M urea, 5X TBE (25mM Tris pH 8.0, 1mM EDTA, 10mM boric acid, 7% acrylamide) as per standard procedures (Sambrook *et al.* 1989).

2.8 Primer extension analysis

Primer extension analysis was done using ~5μg of mtRNA dissolved in 6μl low TE with 5μl (~10 ng) of ^{32}P end-labelled oligomers. The mix was incubated at 70°C for 10min before adding 4 μl 5X first strand buffer (Invitrogen), 2 μl 100 mM DTT, 1 μl 10 mM dNTPs and 1 μl (40 U) RNasin® (Promega) and incubation at 37°C for 2min. The reaction was incubated 1.5 hr at 37°C with 1 μl (200 U) M-MLV Reverse Transcriptase (Invitrogen) and stopped by adding 10 μl loading buffer. Products were denatured by boiling 5 mins before a quick cool on ice and were then electrophoresed on a 7% urea denaturing polyacrylamide gels as previously described.

2.9 DNA sequencing

Automated sequencing of isolated clones was done by the Ontario Health Research Institute DNA sequencing facility (OHRI, StemCore laboratories) or the McGill University/ Génome Québec Innovation Center (Montréal).

For assessing editing status, RT-PCR products were sequenced directly by dideoxy chain termination using the Sequenase version 2.0 DNA sequencing kit (US Biochemicals) and α-^{35}S-dATP as per supplier's instruction. Sequencing products were electrophoresed on 7% urea denaturing polyacrylamide gels as previously described.

2.10 Bioinformatics analysis

Sequences were obtained from the NCBI (National Institute of Biotechnology Information) databank for primer design (http://www.ncbi.nlm.nih.gov) and sequence analysis. Accession numbers used are given in table 2.2. BLAST searches were done using the NCBI website online (Altschul *et al.* 1990, http://www.ncbi.nlm.nih.gov/blast). ClustalW alignements were done online at the European Bioinformatics Institute website (Chenna *et al.* 2003; http://www.ebi.ac.uk/Tools/clustalw/index.html). MultiPipmaker analyses were done online using the Multipipmaker software (Schwartz *et al.* 2000): this program aligns

regions of similar identity between two or more nucleotide sequences and results are shown as a percent identity plot (*pip*).

RNA secondary structure predictions were done online with RNAfold (Hofacker *et al.* 1994; http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi), mfold (Zuker, 2003; http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html) and GeneBee (http://www.genebee.msu.su/genebee.html).

50 and 100 nts around 5' and 3' ends were used in secondary structure prediction programs and settings for the RNA folding programs were used as follows:

RNAfold: Fold algorithm partition and pair probabilities using preset RNA parameters, at 37°C, avoid isolated base pairs.

mfold: linear RNA, at 37°C, 5 percent free energy increment delta(G), 1 M Na^{+}, 0 M Mg^{2+}, 50 maximum number of foldings to be computed, no structure annotation, 0 structure rotation angle and 5' base number 1.

Genebee: Greedy method, energy threshold for helices -4.0, conserved factor (Coefficient of increasing free energy for conservativeness of a pair of complementary position): 2 Kkal/mol, compensated factor (Coefficient of increasing free energy for a pair of complementary positions, having compensatory changes in the alignment): 4 Kkal/mol, cluster factor 2, greedy factor (Number of variants, tried out for inclusion into the structure on every step of the greedy algorithm): 2 Kkal/mol, using all the sequence.

Table 2.2 Accession numbers of sequences used in this study.

Accession numbers correspond to those in the NCBI databank (National Center of Biotechnology Information, National Institute of Health, USA). Species, locus and literature references are given. Gene names with capital letter refer to nuclear encoded genes as opposed to gene names without capital letter that are mitochondrially encoded.

Species		Accession number (NCBI databank)	References
wheat	mitochondrial genome	AP008982	Ogihara *et al,*, 2005
rice	"	BA000029	Notsu *et al,*, 2002
maize	"	AY506529	Clifton *et al,*, 2004
Arabidopsis	"	NC_001284	Unseld *et al,*, 1997
Brassica	"	AP006444	Handa, 2003
tobacco	"	NC_006581	Sugiyama *et al,*, 2005
sugarbeet	"	BA000009	Kubo *et al,*, 2000
	Locus		
wheat	*ccmFN-rps1*	X69205	Gonzalez *et al,*, 1993
"	*mat-r-nad1e*	X57965	Chapdelaine and Bonen, 1991
"	*nad5c*	M74158	Pereira de Souza *et al,, 1991*
wheat	*rpl5-Ψrps14*	AJ535507	Sandoval *et al,*, 2004
wheat	*cox3*	X15944	Gualberto *et al,*, 1990
wheat	*rps2*	AB158222	Kubo *et al,*, 2005
wheat	*Rpl5*	BJ269357	EST_others databank
"	*Rpl5*	BQ806344	"
"	*Rpl5*	BJ269331	"
"	*Rpl5*	CD919000	"
"	*Rpl5*	BJ269357	"
barley	*Rpl5*	BM368977	"
"	*Rpl5*	BU973461	"
"	*Rpl5*	CV053501	"
"	*Rpl5*	Bi779154	"
"	*Rpl5*	BE602472	"
oats	*Rpl5*	CN821212	"

Chapter 3 Characterization of transcript profiles for the *ccmFN-rps1* locus in wheat mitochondria

3.1 Rationale

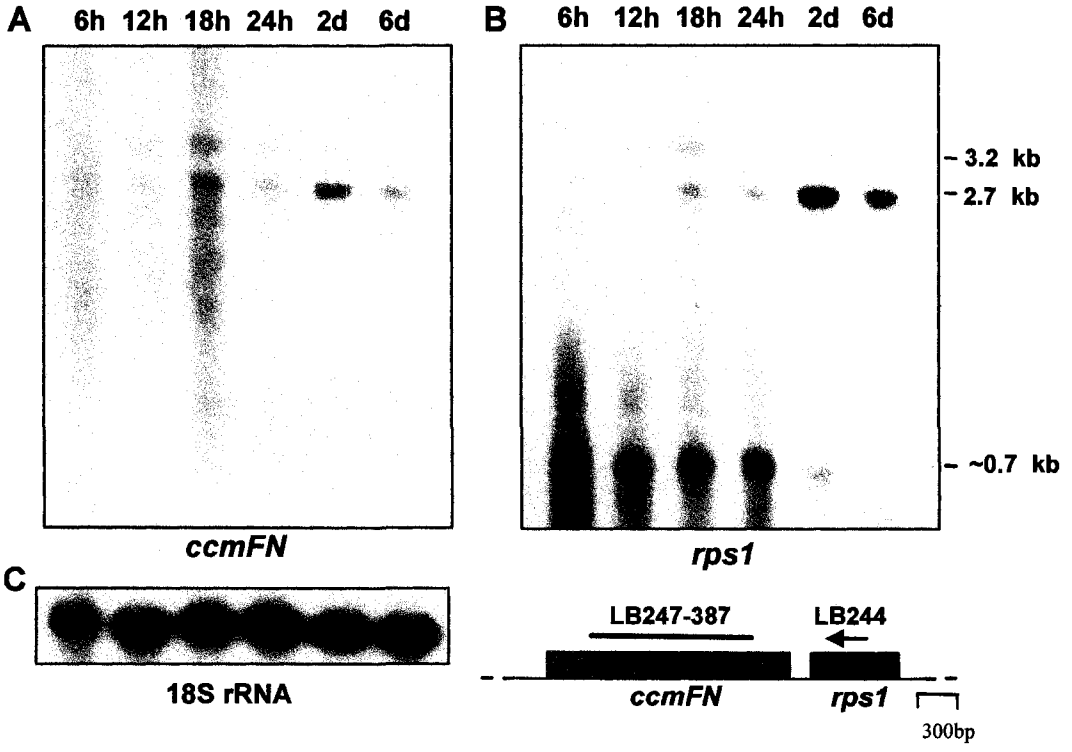
The mitochondrial *ccmFN-rps1* genes encode a cytochrome c maturation protein and the ribosomal protein S1, and are part of a conserved linkage in wheat, rice and maize. In wheat 6d seedlings, the ends of a 2.7 kb bicistronic transcript were characterized (Gonzalez *et al.* 1993). However preliminary northern analysis from our group at an earlier developmental stage revealed a more complex transcript profile, including a precursor RNA and a monocistronic *rps1* species in 2d wheat (T. Hazle unpublished data), prompting our interest in studying this region for post-transcriptional events during seed-to-seedling development. Furthermore, this locus is part of a longer linkage of ~10 kb (*ccmFN-rps1-mat-r-nad1e-nad5c*, including a group II intronic maturase and exons 4 and 3 of *nad1* and *nad5* genes of the NADH complex) conserved between wheat, rice and maize which raises interesting questions about the conservation of potential processing and regulatory signals in the intergenic region between *rps1* and *mat-r*. This chapter is currently being prepared as a manuscript.

3.2 Transcript profiles of the *ccmFN-rps1* locus in wheat during early seed germination

Wheat mitochondrial RNAs from different developmental stages during embryo-to-seedling development (6hr, 12hr, 18hr, 24hr, 2d and 6d after imbibition) were extracted and northern analysis of the *ccmFN-rps1* region was performed using random-labelled PCR fragment and end-labelled oligomers (Figure 3.1 A and B). Transcript profiles of *ccmFN* and *rps1* between 6hr and 6d show two high molecular weight bicistronic transcripts of ~ 3.2 kb and ~2.7 kb and a *rps1* monocistronic transcript of ~0.7 kb (Figures 3.1A and 1B). A random-labeled PCR probe was used for the *ccmFN* gene, which could potentially hybridize to antisense transcripts coming from the same region but no difference was seen between such a probe and an oligomer-probe (data not shown).The coding sequence of *ccmFN* is 1770 bp and the *rps1* coding sequence is 638 bp, when using the genomically encoded stop codon (cf. an edit site at nucleotide 619 creates an early stop codon in wheat, Gonzalez *et al.*

Figure 3.1: Northern blot analysis of the ccmFN-rps1 locus for wheat mtRNA during embryo-to-seedling development (6h, 12h

Sizes in kilobases (kb) of the transcripts are given on the right. The filled bars below represent the genes in the wheat mitochondrial genome, with arrow heads showing the position of the random-labelled (LB247-387 PCR product) and end-labelled oligomer (LB244) probes used, *ccmFN* in panel A, *rps1* in panel B. Hybridization with 18S rRNA was used as standardization for RNA loading (panel C).


A
6h 12h 18h 24h 2d 6d
ccmFN
B
6h 12h 18h 24h 2d 6d
~3.2 kb
~2.7 kb
~0.7 kb
rps1
C
18S rRNA
LB247-387
LB244
ccmFN
rps1
300bp

1993). As proposed by Hazle and Bonen (2007a) using comparative analysis between monocot and eudicot species, the *rps1* start codon is actually believed to be 111bp upstream of the one reported in the literature, in which a sequencing error caused a frameshift (Gonzalez *et al.* 1993). The wheat *ccmFN-rps1* intergenic distance is 135 bp (Figure 3.1 A). These sizes allow about 650 nt and 200 nt for total UTR length for the 3.2 kb and 2.7 kb transcripts respectively.

The abundances of the three transcripts shift over development (Figure 3.1 A and B): in early development (6hr) only the *rps1* monocistronic transcript is detected and it is abundant until about 24hr and is no longer detected at 6d. At 18 hr after imbibition, the two high molecular weight transcripts are detectable. In 6d seedlings, only the 2.7 kb mRNA is seen and the 3.2 kb precursor transcript is no longer detectable; it might no longer be *de novo* transcribed or is processed quickly into the 2.7 kb mRNA.

In order to determine whether the 500 nt difference between the two high molecular weight transcripts is due to a difference in length of the 5' end or the 3' end, end labelled oligomer hybridizations were done with probes specific to the 500nt upstream and downstream of the 5' and 3' ends of the 2.7 kb transcript, based on the published data (Gonzalez *et al.* 1993): the upstream one hybridized to the 3.2 kb transcript in 24hr wheat mtRNA (Figure 3.2A) and in earlier stages of development (data not shown), indicating that the difference between the two transcripts includes at least part of the 5' UTR.

3.3 Determining the precise 5' ends of the two *ccmFN-rps1* bicistronic transcripts in wheat

To map the transcript ends of the two *ccmFN-rps1* transcripts in wheat mtRNA at 12hr and 24hr post-imbibition, where both high molecular weight transcripts and the monocistronic *rps1* transcript are detected, phosphatase-circular-RT-PCR experiments were done (Figure 3.3). The RNA is first treated with tobacco acid phosphatase (TAP) to remove the 5' α phosphate of primary transcripts enabling their circularization in the subsequent step. The reaction with T4 RNA ligase circularizes TAP-treated primary transcripts and as well as processed transcripts that have a 5' monophosphate (Figure 1.2 C for details). RT-PCR over the newly formed 5'-3' junction of the self-ligated RNA is performed with transcript specific primers, providing simultaneous information about the 5' and 3' ends.

Figure 3.2: Northern blot analysis of the ccmFN-rps1 locus and the region upstream for wheat mtRNA at 24hr embryos and 6d seedlings

Sizes in kilobases (kb) of the transcripts are given on the right. The filled bars below represent the genes in the wheat mitochondrial genome, with arrowheads showing the position of the end-labelled oligomer probes used, upstream of *ccmFN* in panel A, *rps1* in panel B (same blot A and B), *ccmFN-rps1* intergenic in panel C (different blot).

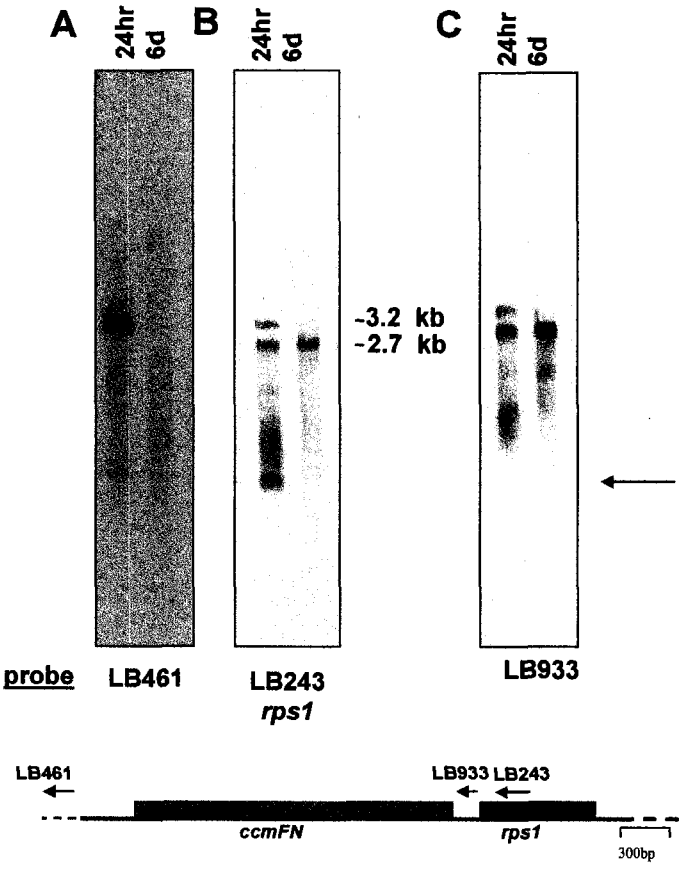
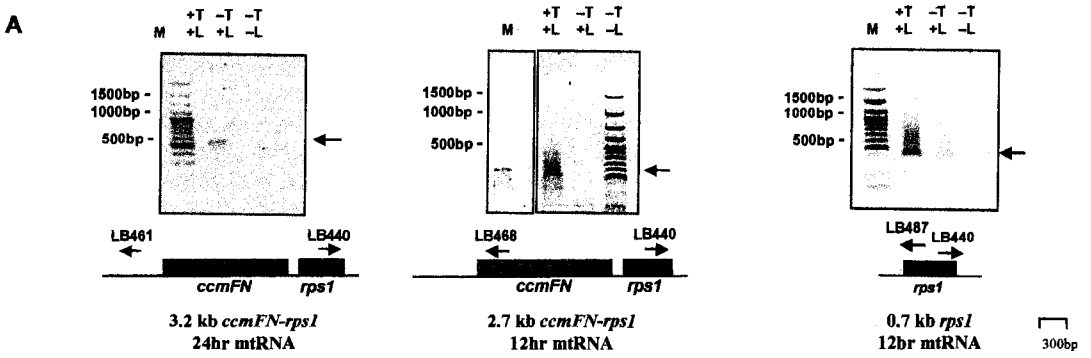
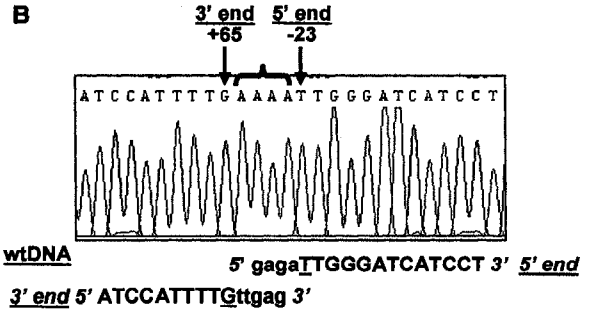

A
24hr
6d
B
24hr
6d
C
24hr
6d
-3.2 kb
-2.7 kb
probe
LB461
LB243
rps1
LB933
LB461
LB933
LB243
ccmFN
rps1
300bp

Figure 3.3: cr-RT-PCR analysis for the ccmFN-rps1 locus for the two major ccmFN-rps1 transcripts and the rps1 monocistronic transcript and sequence chromatogram of an exemplar cDNA sequence (B)

TAP-circular-RT-PCR analysis of the *ccmFN-rps1* transcripts in wheat mtRNA on agarose gels with sizes given on the left. The three RNA treatments are given above: with tobacco acid phosphatase (TAP) and with RNA ligase (+T +L), without TAP and with RNA ligase (-T +L) and without TAP and without RNA ligase (-T -L, negative control). M signifies the DNA ladder lane. Arrows point to the products extracted from the gel and cloned for sequencing.
(B) Chromatogram sequence of a cr-RT-PCR clone from TAP treated 24hr wheat for the 2.7 kb *ccmFN-rps1* transcript. Arrow heads demark the 5' and 3' ends of the transcript and a bracket denoted the non-encoded nucleotides found at the junction. The wheat mtDNA sequence is shown underneath with small caps letter representing nucleotides not in the cr-RT-PCR clone.


A
M
+T +L
−T +L
−T −L
1500bp -
1000bp -
500bp -
LB461
LB440
ccmFN
rps1
3.2 kb ccmFN-rps1
24hr mtRNA
LB468
2.7 kb ccmFN-rps1
12hr mtRNA
LB487
0.7 kb rps1
12br mtRNA
300bp


B
3' end
+65
5' end
-23
ATCCATTTTGAAAATTGGGATCATCCT
wtDNA
5' gagaTTGGGATCATCCT 3'
5' end
3' end 5' ATCCATTTTGttgag 3'

A cr-RT-PCR product for the 3.2 kb transcript of ~500 bp was expected and was yielded in the +TAP/+ligase samples. Sequencing of clones of + TAP cr-RT-PCR products of 12hr and 24hr post-imbibition wheat mtRNA template reveal that a major 5' end of the 3.2 kb transcript is located 589 nt upstream of the *ccmFN* start codon (Table 3.1 and Figure 3.5). Faint products were sometimes seen in -TAP/+ligase samples (Figure 3.3A) and clones from -TAP treated samples had 5' ends that mapped to the A nucleotides between -588 and -585 and are likely representatives of transcripts that have undergone 5' exonuclease degradation (Table 3.1). Four clones had longer 5' ends: one at -596, two at -614 and one at -615 nt upstream of the ccmFN start codon (Table 3.1). This might represent processing products from a longer primary transcript. Although there was no sign of a longer transcript in the northern analysis, its under-representation in clones might indicate its existence at low levels. It is also possible that the 3.2 kb transcript seen in the northern analysis represents two different transcripts whose sizes would be slightly different and undifferentiable with northern-level resolution and whose cr-RT-PCR product sizes would be similar.

S1 nuclease assays for the 3.2 kb dicistronic transcript from 24hr wheat mtRNA confirmed the two 5' ends at -588 and -615, the latter represented by a weaker band (Figure 3.4). It is a possibility that there is a longer transcript with a processing site at -615. The probe used in this nuclease protection assay experiment extended 759 nt upstream of the *ccmFN* start codon and would have detected 5' ends in this region but it is possible that the 5' end of a longer transcript is further upstream.

The 2.7 kb transcript was expected to give rise to cr-RT-PCR products of ~600 bp (Figure 3.3B) and this was reproducibly found in +TAP/+ligase and -TAP/+ligase treatments, indicating processed ends for these transcripts or a mixture of processed and primary transcripts. The 5' end of the 2.7 kb transcript at -23 upstream of the *ccmFN* coding sequence in all three developmental stages examined (12hr, 24hr and 6d) agrees with the data from 6d seedlings (Gonzalez *et al.* 1993) (Table 3.1 and Figure 3.5) and its presence in +TAP and -TAP treated RNA confirms it as a processing site (Figure 3.3B and Table 3.1). Two clones out of 10 showing shorter 5' ends at -10 and +9 most likely represent degradation products (Table 3.1).

Table 3.1: Table showing the sequences from cr-RT-PCR clones for ccmFN-rps1 major transcripts, aligned with the wheat mtDNA sequence.

Clones are categorized by developmental stage (12hr, 24hr and 6d), and mtRNA treatment with or without tobacco acid phosphatase TAP (+ / -).
Based on the wheat mtDNA sequence, AP008982, start codon highlighted in grey, and 5' and 3' ends in italics (Gonzalez *et al.* 1993), depicting the 5' and 3' ends determined in this study (highlighted) of the 3.2 kb, 2.7 kb and 0.7 kb *ccmFN-rps1* transcripts. Non-encoded nucleotides (nts) found at the 5'-3' junction were removed from the sequence for alignment purposes and are shown on the right. Small caps t nucleotide in a 0.7 kb clone sequence show where an extra T nucleotide was found.

A

Transcript		-615 -589	5′ end	3′ end	+68	non encoded nts
3.2 kb		TACCTTTAAGTCTGAGAGAAGGCTTTTTCTTTCTATTAAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGCACTTTATTCC		
12hr	+	TTTCTTTCTATTAAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		
	+	AAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCC		A
	+	AAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		A
	+	AAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCAT		
24hr	+	TCTGAGAGAAGGCTCTTTCTTTCTATTAAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		AA
	-	CTGAGAGAAGGCTTTTTCTTTCTATTAAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		
	-	TCTGAGAGAAGGCTTTTTCTTTCTATTAAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		26rRNA
	+	TAAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AAAAAA
	+	AAAAGATCACAGATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		AAAAA
	+	AAAAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGCACTTATTC		
	+	AAGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGC		
	+	AGATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		
	+	ATCACAAATCAGAGGGGTGATCAGTCCTCTTTCAATCCTT		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		A
	+	AGTCCTCTTTCAATCCTT		TTAT		
	-	CCTTC		TTATGAAAAATCCGAAGCCCACCTTAATC		

B

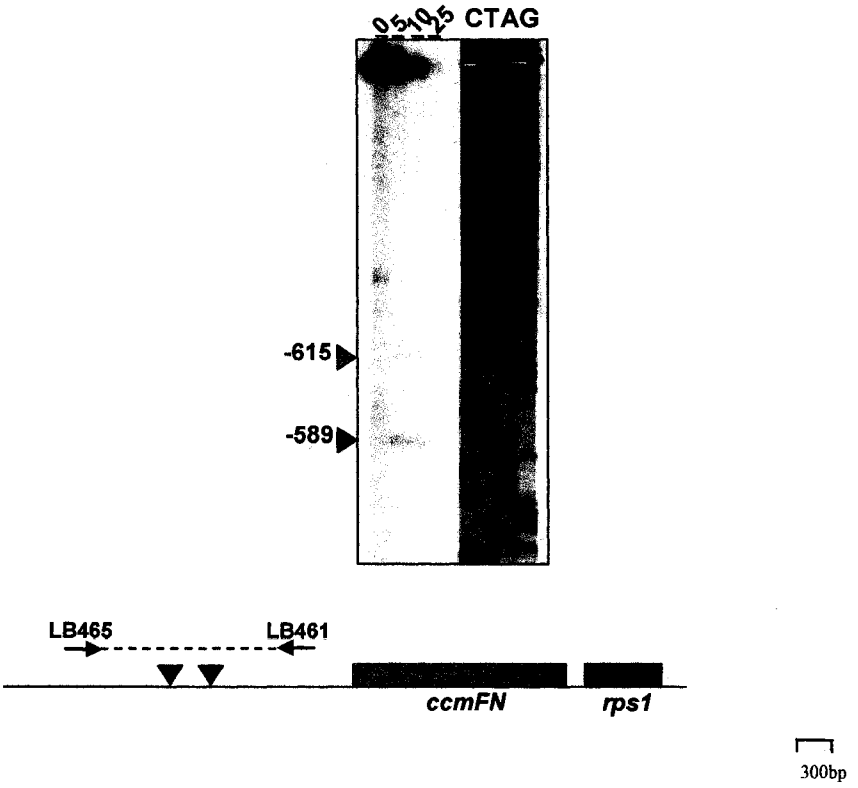
Transcript		-25	5′ end	3′ end	+68	non encoded nts
2.7 kb		TATCAAAAAGATCTATTACTAAAAGGAGATTGGGATCATCCTGTGGTTACCGG[redacted]ATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGCACTTTATTCC		
12hr	+	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AAAAACA
	+	TGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAG		18rRNA
	-	TGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AAA
	-	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AA
	-	TGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		
24hr	+	AAAAAGATCTATTACTAAAAGGAGATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AA
	+	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		A
	-	ANTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		AA
	+	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		AAA
	+	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCANTTTGT		AA
	-	ATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTT		
6d	+	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		AA
	+	ATTGGGATCATCCTGTGGTTACCGGATGATGGGAATAAC		TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		

C

Transcript		5′ end	3′ end	+68	non encoded nts
0.7 kb		AGGAGGAAAAAAAAGCTT[redacted]ATGAGCATCTATTGGAGTCGATCATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGCACT		
12hr	+	GCTTATGATGAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		
	-	ATGATGAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AAAC
	-	GCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		AA
	-	TCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAA		
	+	CAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATT		AAA
	+	ATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		
24hr	+	GGAGGAAAAAAAAGCTTATGATGAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		
	-	AAGCTTATGATGAGCATCTATGGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		
	-	GCTTATGATGAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTT		
	-	ATGATGAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGtTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTT		
	+	TGAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		
	+	GAGCATCTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGCA		
	+	TTTATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		
	-	ATTGGAGTCGATTATTTCCAAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AA
	-	AAGATCTAATTCTAGTTTCTTATTATGTAGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTG		AAAATGGT
	-	AGTGGAAAC	TTATGAAAAATCCGAAGCCCACCTTAATCCATTTTGT		AACAA

Figure 3.4: S1 nuclease protection analysis with 24hr wheat mtRNA for the 3.2kb ccmFN-rps1 co-transcript

Sizes in nucleotides are given on the left. The filled bars below represent the genes in the wheat mitochondrial genome, with arrow heads showing the position of the random-labelled PCR product used in the hybrid annealing. Light gray boxes represent the protected products. Block arrow heads show the two 5' ends at -589 and -615 relative to the *ccmFN* start codon. Units of S1 nuclease enzyme used in each lane are written on top and sequencing reaction of the LB465-461 PCR product was used as a size marker (C, T, A, G refer to the dideoxyribonucleotide used in the sequencing reaction)


0.5
1
2.5
CTAG
-615
-589
LB465
LB461
ccmFN
rps1
300bp

3.4 Analysis of the two *ccmFN-rps1* bicistronic transcripts

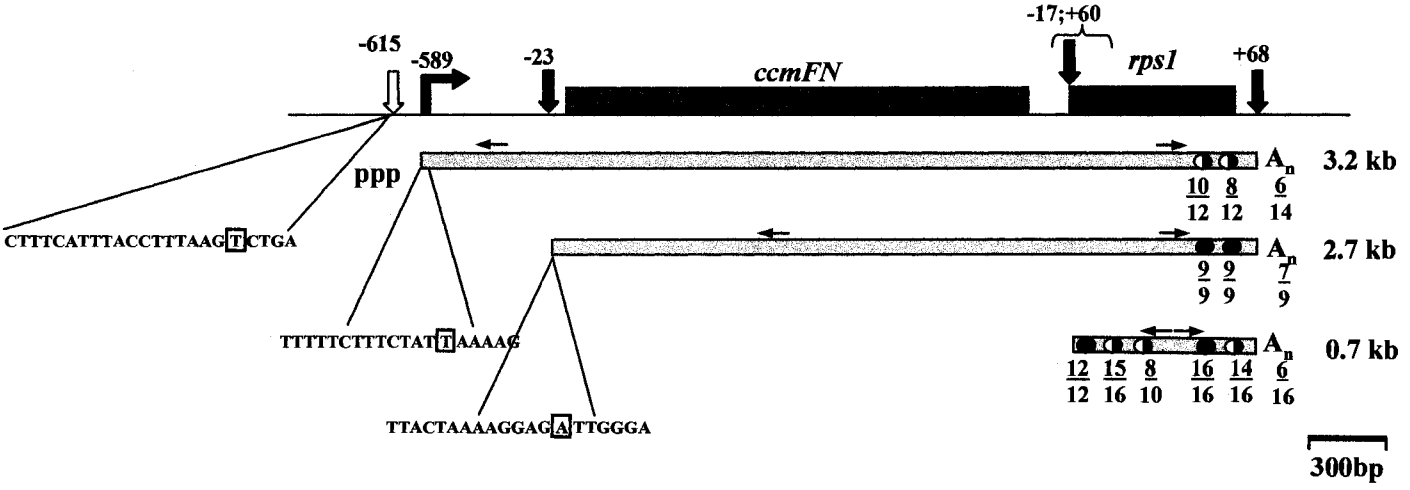
The noticeably different transcript profiles between 6hr up to 6d for the *ccmFN-rps1* locus is consistent with previous observations that during early development there is a lag between transcription and RNA processing (Li-Pook-Than *et al.* 2004), in that precursors of intron-containing genes in wheat were abundant between 0 and 24hr but no longer detectable by northern analysis in 6d seedlings. During early germination and embryo development, the machinery used for transcription and RNA processing is hypothetically being assembled and not yet fully efficient. While the components for *de novo* transcription are being put into place, mitochondria rely on stored mRNAs in the seeds. The *rps1* monocistronic transcripts are particularly abundant in early developmental stages (6hr), and like the monocistronic *rps7* transcripts present in 0 hr (Li-Pook-Than *et al.* 2004), might represent stored messenger RNAs. At later stages primary transcripts appear to be quickly processed into messenger RNAs and are no longer detectable by northern analysis.

The characterization of two tentative promoter regions upstream of *ccmFN* reveal that neither the -615 nt nor the -589 nt *ccmFN-rps1* 5' ends exhibit the monocot C/YRTA promoter consensus sometimes seen at transcription initiation sites in plant mitochondria (Figure 3.5 and Table 3.1; reviewed in Gagliardi and Binder 2007). The -589 promoter is preceded by a slightly modified version of this motif, namely TCTA (one nucleotide upstream). The -589 promoter is in a relatively AT rich region (62% in the 50nt downstream of the -589 promoter of *ccmFN)*, a DNA sequence that can easily denature, consistent with the accessibility and positioning of the transcription machinery. In *Arabidopsis*, promoter core motifs were often preceded by an A/T rich sequence (Kuhn *et al.* 2005). Furthermore, the two transcription initiation sites for *ccmFN-rps1* did not share sequence similarity (except for an adenosine at position -2) and were not found elsewhere in the wheat mitochondrial genome through BLAST searches. The presence of gene specific promoters is not unusual and different promoter structures may reflect different transcription strength and hence serve a regulatory function (Mulligan 1991, Gray 1992).

In *Arabidopsis thaliana*, the 5' ends of some transcripts are speculated to be generated by cleavage at the 3' ends of a t-element or stem-loop structures by the endonucleolytic enzymes RNaseZ or P (Forner *et al.* 2007). No evident secondary structures or t-element structures were found close to the -23 position or at the -615 position if this is a

Figure 3.5: Schematic of the ccmFN-rps1 genes in the wheat genome and the three major transcripts of 3.2 kb, 2.7 kb and 0.7 kb seen during embryo-to-seedling development, showing C-to-U editing sites and transcript ends

The bent arrow represents the promoter upstream of *ccmFN* and filled block arrows show processing sites at the 5' and 3' ends with position in nucleotide relative to start and stop codons respectively, empty block arrow denotes a putative processing site. Primer positions used in cr-RT-PCR are shown by horizontal arrow heads. Sequence at the promoter and the 5' end processing sites are shown underneath with 5' end nucleotide boxed. Editing sites based on Gonzalez *et al,* 1993. Editing sites examined in the cr-RT-PCR products are shown in the transcripts, where half filled circles represent partial editing and filled circles represent complete editing, with the number of clones shown underneath. A_n indicates the presence of non-encoded nucleotides with the number of clones shown underneath.


-615
-589
-23
ccmFN
-17;+60
rps1
+68
ppp
CTTTCATTTACCTTTAAGTCTGA
TTTTTCTTTCTATTAAAAG
TTACTAAAAGGAGATTGGGA
An
10 8 6
12 12 14
3.2 kb
An
9 9 7
9 9 9
2.7 kb
An
12 15 8 16 14 6
12 16 10 16 16 16
0.7 kb
300bp

processing site as well. Processing signals could also be conferred by long-range RNA interactions or proteins binding to cis-elements.

3.5 Determining the 5' end of the monocistronic *rps1* transcripts in wheat

The 0.7 kb transcript was expected to give rise to cr-RT-PCR products of ~400 bp (Figure 3.3C) and this was reproducibly found in +TAP/+ligase and -TAP/+ligase treatments, indicating that processed ends are found for these transcripts or arise from a mixture of processed and primary transcripts. Results from cr-RT-PCR revealed that the 5' ends of the monocistronic *rps1* transcripts at 12 hr and 24hr are heterogeneous (from -17 upstream of the *rps1* start codon to +60 downstream of the *rps1* start codon) (Table 3.1 and Figure 3.5). cr-RT-PCR products for the *rps1* monocistronic transcript were always seen as smears on the agarose gels, using different primer pairs, RT-PCR conditions and mtRNA preparations (see material and methods section). The presence of 0.7 kb *rps1* monocistronic transcripts in 6d wheat mtRNA was also investigated using cr-RT-PCR, in case of low steady-state levels not seen in northern analysis but it consistently yielded no products (data not shown). Unexpectedly, sequencing revealed that 10/16 cr-RT-PCR clones of the 0.7 kb *rps1* transcript did not include the start codon and are thus unlikely to represent functional transcripts. If a downstream in-frame ATG is considered, although all the clones contain this second in frame start codon, the 5' ends of the transcripts are heterogeneous.

Preliminary data on the 0.7 kb *rps1* transcript on 24hr wheat mtRNA using primer extension experiments show several 5' ends at +2, -15 and -42. Preliminary data with S1 nuclease assay on the 0.7 kb *rps1* transcript on 24hr wheat show a 5' end at the start codon. This agrees with the heterogeneous nature of the *rps1* transcript 5' end (Appendix 2 A and B). In northern hybridization, the 0.7 kb *rps1* transcript is a broad band possibly representing a heterogeneous population of *rps1* monocistronic transcripts (Figure 3.1B). To verify that the 0.7 kb *rps1* RNA species seen in the northern analysis did not include transcripts with 5' ends upstream of the most upstream 5' end from the clones sequenced, an end labelled oligomer specific to the *ccmFN-rps1* intergenic region (31bp upstream of the *rps1* start codon) was used in a northern blot hybridization and did not hybridize to the 0.7 kb *rps1* transcript (Figure 3.2 C)

3.6 Analysis of the *rps1* monocistronic transcript

Taken together the data about the 0.7 kb *rps1* transcript raises questions about its origin: it may be a stored mRNA in the seed and is being degraded during early development. Later developmental times might then be expected to contain *rps1* transcripts with more 5' end degradation. However there was no difference at the 5' ends of *rps1* transcripts between 12h and 24hr developmental stages (Table 3.1). Northern analyses revealed a slight shift in the size of the monocistronic transcript at 2d (Figure 3.1B). It would be of interest to characterize the ends of the transcript at this stage of development.

During embryo-to-seedling development, either the 3.2 kb precursor or the 2.7 kb *ccmFN-rps1* mRNA could be processed by endonucleolytic cleavage into a stable 0.7 kb *rps1* transcript and an unstable, quickly turned-over monocistronic *ccmFN* transcript. We noticed the absence of a *ccmFN* monocistronic transcript (Figure 3.1A), which indicates that the 2.7 kb transcript can be considered the bicistronic mature messenger RNA for *ccmFN*. The presence of a *ccmFN* monocistronic transcript was examined using *ccmFN* specific primers in the cr-RT-PCR reaction: no product was detected under different conditions, with different mtRNA preparations and with different primer pairs (data not shown, see material and methods section). Dicistronic transcripts in *Arabidopsis* (*rpl5-cob, rpl2-mttB, nad3-rps12, nad4L-atp4 and rps3-rpl16*) were investigated for monocistronic products by cr-RT-PCR and failed to yield products in 9/10 cases, only *rpl2* seem to give rise to potential monocistronic transcripts (Forner *et al.* 2007). Another technique to detect if an endonucleolytic cleavage takes place upstream of *rps1* would involve the ligation of an anchor to the 3' end of the mtRNA population followed by RT-PCR using an antisense oligomer specific to the anchor and a sense oligomer specific to the sequence upstream of the cleavage site, in this case specific to the *ccmFN* coding sequence.

The 0.7 kb *rps1* transcript could also be a primary transcript driven by a promoter in the *ccmFN-rps1* intergenic sequence, possibly used only during early developmental stages or during seed biogenesis. Although cr-RT-PCR products were yielded reproducibly with and without TAP treatment (Figure 3.3C), the 0.7 kb *rps1* transcript might be a primary transcript that is susceptible to 5'-to-3' exonuclease degradation. At later times it might no longer be *de novo* transcribed, which would explain the absence of the 0.7 kb transcript at 6d post-imbibition.

If a processing event does take place in the *ccmFN-rps1* intergenic sequence, there are several possibilities that can explain the absence of a stable monocistronic *ccmFN*, the counterpart of the monocistronic *rps1* transcript. The latter might have inherent characteristics that render it stable such as secondary structures at its 3' end. It might be due to association of proteins at the 3' end, as seen for chloroplast transcripts (Herrin and Nickelsen 2004). A hypothetical *ccmFN* monocistronic transcript would be quickly-turned over, maybe due to a lack of stabilising structures. The cis-element conferring the signal for processing is unknown due to the heterogeneous nature of the *rps1* 5' end. The 135bp *ccmFN-rps1* intergenic sequence in wheat is unique in the genome. In maize and rice, the two genes are also linked but the *ccmFN* stop codon is further downstream due to indels of 4 and 5 bp and although the sequence is homologous, the length of the intergenic sequence is different (135bp wheat, 73bp rice and 29bp maize, figure 3.6). It would be of interest to look at the transcript profiles for these cereal species to see whether a monocistronic *rps1* transcript exists. Preliminary northern analysis on rice 24hr and 6d mtRNA show putative low level *ccmFN-rps1* transcripts of similar sizes to those present in wheat (Appendix 3) but no abundant monocistronic *rps1* transcript.

Transcriptional analysis of the *rps1* gene in other plants such as legumes reveals the presence of a monocistronic transcript in rest harrow only but not for soybean, pea or bean (Hazle and Bonen 2007a). Its 5' end is upstream of the breakpoint between eudicots and monocots (60-70bp), meaning that in more distantly related plants there is no evidence for a processing signal or a promoter in the common sequence upstream of *rps1*. Interestingly in the eudicot *Oenothera*, *rps1* is co-transcribed with *atp9* but lacks stable monocistronic transcripts and is not edited (Mundel and Schuster 1996).

3.7 Determining the 3' ends of the three transcripts from the *ccmFN-rps1* locus in wheat

Based on cr-RT-PCR data for the two *ccmFN-rps1* transcripts and the *rps1* transcripts a common 3' end was found for the three major transcripts (as well as the products with a longer 5' end at -615). It is located 65 nt downstream of the *rps1* stop codon, with a slight scattering of 3' ends around this region, and agrees with the 3' end mapped by Gonzalez *et al.* 1993 (Table 3.1 and Figure 3.5). 5/39 and 3/39 clones (~20%) had a shorter and a longer 3' end respectively, which may represent degradation products and improper 3' end

Figure 3.6: Alignment of the ccmFN-rps1 intergenic sequence from wheat, rice and maize mitochondrial genomes (A) and ccmFN C-terminus amino acid alignment between wheat, rice, maize and Aradidopsis thaliana (B)

ClustalW alignment. The *ccmFN* stop codon is highlighted in red and the *rps1* start codon is shown in green.

A

```
wheat    CCGTTATGTTATGTTATGGAAGAATT[redacted]CTTCTTGCTGGGCTGGTTATTCAGAGCCAGC 60
rice     CC-----GTTATGTTATGGAAGAATTTAGCTTATTGCTGGGCTGGTTATTCAGAGCCAGC 55
maize    CC-----GTTATGTTATGGAAGAATTTAGCTTCTTGCTGGGCTGGTTATTCCGAGCCAGC 55
         **     ************************ ****************** ********

wheat    AACTGGC----TGCCGGATTTCGTCCCGTCCGTAGCGCTTCGGAAGTCACTCTGGCGTGG 116
rice     AACTGGC----TGCCGGATTTCGTCCCGTCCG[redacted]CGCTTCGGAAGTCACTCTGGCGTGG 111
maize    AACTGGCTGGCTGCCGGATTTCGTCCCGTCCGTAGCGCTTCGGAAGTCACTCTGGCGTGG 115
         *******    ************************************************

wheat    CGCTGCTGGATACTTCTGATCAATAGGAATAGGAGGAAAAAAAAAGCTT[redacted]ATGAGCATC 176
rice     CGCTGCTGGATACTTCTGATCAATAGGAATAGGGGGAAAAAAAAAGCTT[redacted]ATGAGCATC 171
maize    CGCTGCTGGATACTTC[redacted]TCAATAGGAATAGGAGGAAAAAAAAGCTT[redacted]ATGAGCATC 175
         ********************************* *************************
```

B

```
wheat          VMLCYGRI---------------------------------- 8
rice           VMLWKNLAYCWAGYSEPATGCRISSRP---------------- 27
maize          VMLWKNLASCWAGYSEPATGWLPDFVPSVALRKSLWRGAAGYF 43
Arabidopsis    VMLWKN------------------------------------ 6
               ***
```

processing respectively (Table 3.1). The 3' end at +65 is most likely the site of a processing event, transcription termination having occurred further downstream.

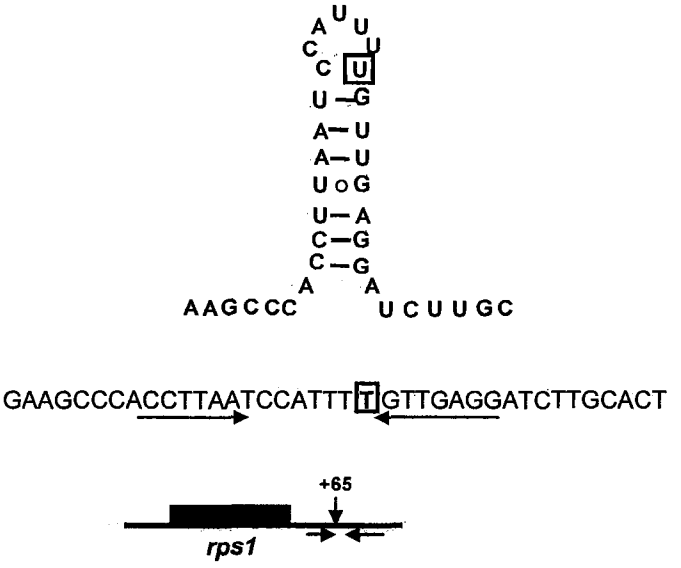
Interestingly, non-encoded nucleotides were found at the junction of circularized transcripts in about 50% of cr-RT-PCR clones (23/44 clones) (Table 3.1 and Figure 3.5). Of these clones, ~80% contained a series of A nucleotides as non-encoded nucleotides (1 to 6) but in 3 cases there was a C nucleotide and in one case the stretch AAAATGGT was seen. Similar results were seen for genes in *Arabidopsis thaliana*, where a number of circular-RT-PCR clones had non-encoded As and sometimes Cs at the 5'-3' junction: 58% of clones of mRNA and rRNA transcripts, between 1 and 19 non-encoded nucleotides (Former *et al.* 2007). This study also included clones from flanking regions and intergenic spaces, where 53% and 45% of the clones respectively had non-encoded nucleotides, mostly less than 10 but up to 56 nt and 101 nt in length (Forner *et al.* 2007). These non-encoded stretches could be polynucleotide tails added as degradation signals, as has been hypothesized previously (Kuhn *et al.* 2001, Holec *et al.* 2006, Binder and Brennicke 2002, Marchenfeld 2004). However, in our study most of the clones with non-encoded nucleotides were for the mature 2.7 kb *ccmFN-rps1* messenger RNA which is functional and would not be expected to be tagged for degradation, whereas the 0.7 kb *rps1* transcript where at least some of the clones did not appear to be functional transcripts had the least number of clones with non-encoded nucleotides. The significance of these non-encoded nucleotides is unclear to date.

3.8 Analysis of the 3' end of the *ccmFN-rps1* transcripts and the presence of non-encoded nucleotides

The signals conferring 3' end maturation are unknown for this locus and a secondary structure is shown as a possible processing recognition site (Figure 3.7). The sequence downstream of the 3' end of *rps1* in wheat has a 7 nt inverted repeat, creating a single stem loop. The 3' end of the transcripts is located at the end of the loop region (Figure 3.6). Other studies have suggested that stem loop structures and t-element-like signals provide processing signals for the 3' end processing at the 3' end of those structures ((Forner *et al.* 2007), reviewed in Gagliardi and Binder, 2007), reminiscient of tRNA processing. This is unlike what is shown here, where the endonucleolytic cut would be made in the loop region (Figure 3.6). No other probable secondary structures were evident in the 3' UTR using RNA folding programs (RNAfold, Genebee, mfold). The 3' UTR of ccmFN-rps1 transcripts is

Figure 3.7: Putative secondary structure at the 3' end of the three ccmFn-rps1 major transcripts

Structure based on Genebee RNA folding program. Sequence of the wheat mtDNA with the 3' end of the transcripts boxed, and the regions participating in the stem loop base pairing underlined. Hydrogen bond is shown as a line and o represents a wobble G-U pair in the stem loop structure. Position shown relative to stop codon.


AAGCCC
UCUUGC
GAAGCCCACCTTAATCCATTTTGTTGAGGATCTTGCACT
+65
rps1

very A rich (50% in 66 nucleotides), reducing the chances to have stable secondary structures. RNA stability signals are not well understood and might be due to association of proteins at the 3' end as is seen in chloroplast mRNAs (Gagliardi and Binder, 2007).

The 3' UTR sequence is unique in the wheat mitochondrial genome (BLAST search, data not shown) (short stretches of simple sequence were found elsewhere but they were not located close to genes). The sequence downstream of *rps1* is conserved between cereals but the breakpoint with more distantly related species such as soybean and tobacco is only a few nucleotides from the genomically encoded stop codon. In wheat, rice and maize, the *ccmFN-rps1* is part of a longer conserved gene linkage with *matr-nad1e-nad5c* (~3 kb downstream of *rps1*). The conservation of a gene linkage, including the ~3 kb *rps1-mat-r* intergenic region, for ~50 million years is quite unusual. This points towards this region being under constraint, perhaps due to the conservation of other regulatory signals or small non-coding RNAs encoded in this region (see section 5.4).

The presence of non-encoded nucleotides at the 5'-3' junction of circularized transcripts has been seen for other mitochondrial transcripts. Using a similar method with RNA ligase (where an anchor was ligated to the 3' end of transcripts, using the same T4 RNA ligase enzyme), Williams *et al.* (2000) found non-encoded nucleotides at the 3' end of mature *atp9* mRNA as well as *cox2* and *rps12* RNAs involved in the degradation process in maize (Williams *et al.* 2000). They pointed towards the role of the modifying enzymes in the 3' modification: the sequence and structure of the RNA prior to the ligation might affect the reaction, and the RNA ligase requires a single unpaired nucleotide at the 3' terminus for the reaction (Williams *et al.* 2000). The T4 RNA ligase buffer contains ATP for the enzyme reactivity. Since there are cases where only 1 or 2 As were seen in our data set, it is not excluded that these are experimental artefacts. It should be noted that only 2/23 cr-RT-PCR clones from the *rpl5-rps14* transcripts obtained in this study (Chapter 4, section 4.6) contained non-encoded nucleotides, which speaks to this phenomenon being more than just experimental artefact due to RNA ligase. In addition two cr-RT-PCR clones from the *ccmFN-rps1* transcripts included fragments of rRNA at the junction (Table 3.1), which were most likely added *in vitro*. RNA fragments coming from the processing of rRNA transcripts are present in high abundance in mitochondria and would be free to be ligated to the transcripts before circularization occurred *in vitro* (Perrin *et al.* 2004b). The nature of

mitochondrial RNA extraction used in this study should exclude the precipitation of small fragments and free nucleotides, however rRNAs represent up to 90% of the total transcriptome and their high concentrations increases the chance of their presence in mtRNA preparations. Potentially a RNA ligation reaction done in a higher volume would dilute the free nucleotides and DNA/RNA fragment and limit the number of *in vitro* ligations of these prior to transcript circularization.

3.9 Editing status of the three *ccmFN-rps1* transcripts at 24hr and 12hr post-imbibition

The editing levels of certain editing sites in the *ccmFN-rps1* and *rps1* transcripts were also examined since predicted editing sites were included in cr-RT-PCR products sequenced (Figure 3.5). Based on the sequence published by Gonzalez *et al.* (1992), there are 32 editing sites in *ccmFN* (sites #1 to 8 were available in my cr-RT-PCR products) and 6 in *rps1* (sites #1, 2, 3, 5 and 6 available). Gonzalez *et al.*(1992) refer to the *rps1* editing sites #1 and 2 as being in the intergenic sequence, however the start codon being upstream of the one they refer to, these sites are in the coding sequence presented here. The site #6 in *rps1* creates a stop codon 12nt upstream of the genomically encoded stop codon.

The 2.7 kb transcript was fully edited at all time points and at all editing sites examined in both *ccmFN* and *rps1* coding sequences (not shown and Figure 3.5) as expected for a mature messenger RNA, since C-to-U editing is regarded as an early event in transcript processing (Gualberto *et al.* 1991). In the 3.2 kb *ccmFN-rps1* transcript, two *rps1* sites at the 3' end [#5 and 6 (stop)] were not fully edited which is consistent with this transcript being an immature RNA.

The cr-RT-PCR clones for the monocistronic 0.7 kb *rps1* transcript had 4 or 5 editing sites (depending on the oligomer used for RT) available for observation. 4/15 clones were not fully edited (3 at 24hr and 1 at 12hr), two of which were not edited at the stop codon site. The monocistronic *rps1* transcript would be expected to be fully edited if this transcript is processed from the mature 2.7 kb messenger, which is completely edited. Editing levels might vary if it came from the processing of the 3.2 kb *ccmFN-rps1* precursor transcript, which is not fully edited. Editing of the 0.7 kb transcript would therefore happen after the processing event. If it arose from its own promoter, partial editing could represent non-

mature transcripts. These clones might represent monocistronic *rps1* transcripts that were wrongly processed at the 3' or 5' end and were not fully edited or tagged for degradation.

3.10 Future directions

Information gained from this study of wheat mtRNA for the *ccmFN-rps1* locus can be complemented by further study. Characterizing the 5' end of the *rps1* monocistronic transcript at earlier developmental times (0hr, 6hr) would help define its nature as a processed or a primary transcript. The presence of a longer precursor *ccmFn-rps1* transcript is also of interest: a promoter further upstream of the one identified here might be used in other developmental times or under different conditions. Mutational analysis of sequences surrounding the processing site at -23 would help determine the cis-signals essential in the recognition of the processing site, although this is limited by the lack of the mitochondrial transformation system.

Additional cr-RT-PCR experiments on earlier (0hr) and later (6d) developmental times would be of interest in regards to the presence of non-encoded nucleotides. To look at *in vivo* levels of transcripts with non-encoded stretches, a modified nuclease protection assay can be performed: using a labelled probe encoding the transcript 3'end with an additional poly(A) tail in hybrid with the transcript population to protect the RNA species containing non-encoded adenosine stretches. The *in vivo* representation of the poly(A) tail containing transcripts will give us clues about whether they hold a biological function in the mitochondrial transcript regulation.

The observations made in this study point towards a relaxed transcriptional system where post-transcriptional events like processing and RNA stability regulation are control mechanisms used in monitoring the levels of transcripts in the mitochondria (Giege *et al.* 2000, Holec *et al.* 2006, Kuhn *et al.* 2005). The machinery involved in post-transcriptional RNA processing events is not well known but is imported from the cytosol, requiring proper mitochondrial-nuclear cross-talk. Proteins such as members of the PPR (pentatricopeptide repeat) gene family are good candidates as proteins involved in post-transcriptional RNA processing (reviewed in Andres *et al.* 2007). Future work on plant mitochondrial transcripts will help us determine the role of post-transcriptional events in the regulation and control of gene expression as well as the coupling of cytosolic machinery with mitochondrial expression events over development in plant mitochondria.

Chapter 4 Characterization of transcript profiles of the *rpl5-Ψrps14* locus in wheat and rice

4.1 Rationale

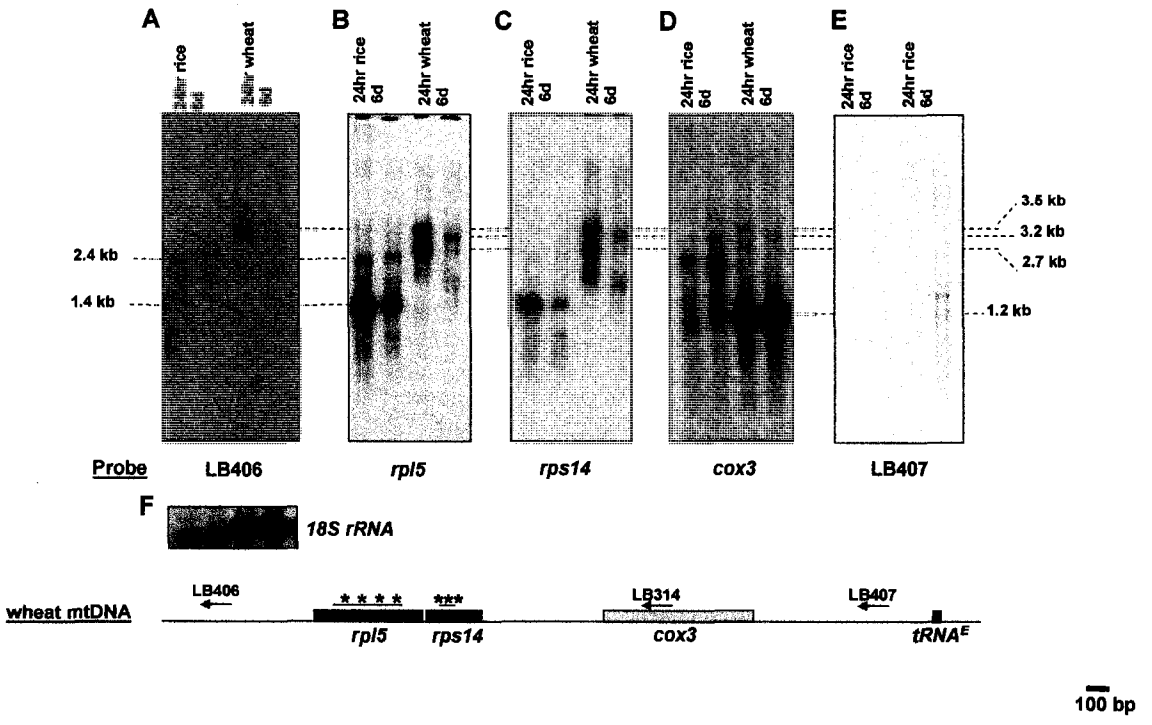
The organization of genes in plant mitochondrial genomes can vary between species due to the highly recombinogenic nature of the genomes, however bacterial-ancestral-type linkages are sometimes seen in present day species. Such an example is the *rpl5-rps14* transcriptional unit, which is spaced by one nucleotide only and encodes or is a pseudogene copy of ribosomal proteins of the large and small subunits respectively. This linkage is conserved among flowering plants. The genomic context surrounding the two genes is different in wheat and rice: in wheat *rpl5-ψrps14* are co-transcribed with downstream *cox3* while in rice *rpl5-ψrps14* stand alone. We are interested in the transcript profiles and regulatory signals used in these two species to transcribe *rpl5-Ψrps14* and subsequent RNA processing of the transcript (transcripts). This ancestral linkage is of interest in cereals where *rps14* is a pseudogene (due to frameshift mutations) after a common gene transfer to the nucleus (Sandoval *et al.* 2004). Co-transcription of mitochondrial *rpl5* and *rps14* in wheat was first noted by Sandoval *et al.* (2004) using RT-PCR experiments. Furthermore, *rpl5* was functionally transferred to the nucleus in wheat and thus this gene seems to be in a transition state where both nuclear and mitochondrial copies are functional (Sandoval *et al.*, 2004), whereas in rice *rpl5* is functional in the mitochondria.

4.2 Transcript profiles of *rpl5-Ψrps14-cox3* in wheat mitochondria

Using northern hybridization, the transcript profiles of the *rpl5-Ψrps14* region in the wheat and rice mitochondrial genomes were investigated (Figure 4.1). For wheat mt RNA of 24hr embryos, probes for *rpl5* and *rps14* genes hybridized to RNA species of ~3.5 kb and 2.7 kb. Hybridizations with *rpl5* and *rps14* probes did not show either a *rpl5-rps14* bi-cistronic mRNA or monocistronic transcripts for either gene (Figure 4.1 B and C). In 6d wheat seedlings, a co-transcript of ~ 3.2 kb was seen (Figure 4.1 B and C) and its abundance is lower compared to the transcripts seen in 24hr. This aspect is of interest regarding the transition-state of *rpl5* (see section 4.7). A random-labeled PCR probe was used for the *rpl5* gene, which could potentially hybridize to antisense transcripts coming from the same region but no difference was seen between such a probe and an oligomer-probe (data not shown).

Figure 4.1: Northern blot analysis of the rpl5-Ψrps14 locus in wheat and rice for 24hr and 6d mtRNA

A) upstream of *rpl5*, B) *rpl5*, C) *rps14*, D) *cox3*, E) downstream of *cox3*. Hybridization with 18S rRNA was used as standardization for RNA loading (F). Sizes in kilobases (kb) are shown on the right side (wheat) and left side (rice). Filled boxes below show the gene organization in the wheat mitochondrial genome, with sizes in nucleotides, arrow heads representing the position of the end-labelled oligomers used and lines with asterisks showing position of the random-labelled probes used.


A
B
C
D
E
24hr rice
24hr wheat
6d
2.4 kb
1.4 kb
3.5 kb
3.2 kb
2.7 kb
1.2 kb
Probe
LB406
rpl5
rps14
cox3
LB407
F
18S rRNA
wheat mtDNA
LB406
LB314
LB407
rpl5
rps14
cox3
tRNA^E
100 bp

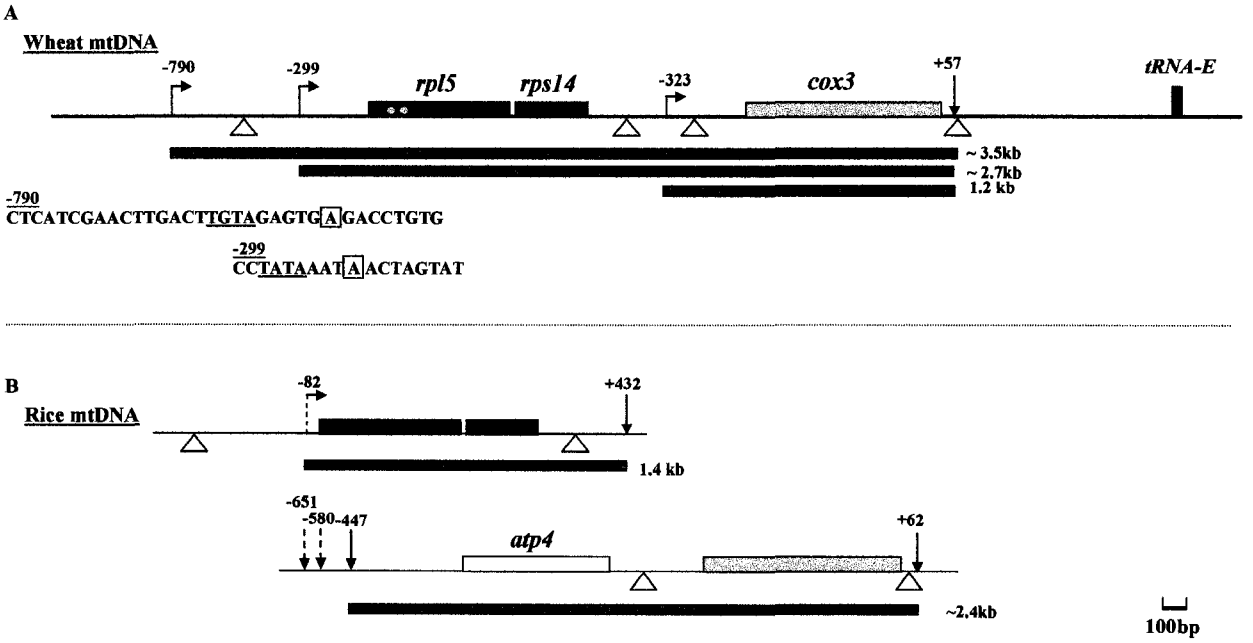
The *cox3* gene is found 628 bp downstream of *rpl5-rps14* (as first noted by Sandoval *et al.* 2004). Northern hybridizations with *cox3* oligomer probes revealed that it is co-transcribed with *rpl5-rps14* in the two high molecular weight RNA species (Figure 4.1D). Hybridizations also revealed an abundant 1.2 kb monocistronic *cox3* transcript (Figure 4.1 D). Previous studies have demonstrated that *cox3* has a promoter at -323 nt upstream of *cox3* and the 3' end of the monocistronic transcripts is located at +55 (Gualberto *et al.*, 1990 and Covello and Gray, 1991; figure 4.2). The *rpl5-Ψrps14-cox3* co-transcription is unexpected since *cox3* expression is driven by its own promoter but it may be due to a loss of transcription termination and 3' end processing signals downstream of *rps14* in wheat.

In order to determine whether the length difference between the two *rpl5-Ψrps14-cox3* transcripts in wheat is due to a difference in length of the 5' or 3' UTR, hybridizations with probes 300 bp upstream of *rpl5* and 700 bp downstream of *cox3* were done (Figure 4.1 A and E): the upstream probe hybridized to the 3.5 kb RNA species, indicating that the two transcripts have a different 5' end. This oligomer also hybridized to the 3.2 kb RNA species found in 6d wheat (Figure 4.1 A). The downstream probe did not hybridize to any stable, steady-state level RNA species (Figure 4.1E).

In the wheat mitochondrial genome, the *tRNA-Glu* gene is found 913 bp downstream of *cox3*. The 3.5 kb RNA species could represent a mixed population of transcripts: *rpl5-rps14-cox3* with a longer 5'UTR and *rpl5-rps14-cox3-trnaE*. However northern hybridizations with a *trnaE* specific oligomer did not hybridize to the 3.5 kb high molecular weight transcript (data not shown). Although there is no sign of a longer RNA species containing all four genes and previous work on *cox3* transcription did not show co-transcription with this tRNA gene, it is not excluded that there exists a large RNA precursor that is quickly processed to release the *trnaE* RNA (see section 1.2.3.1 for details on tRNA processing events). Northern blot analysis with a probe specific to the *cox3-trnaE* intergenic sequence did not show hybridization with a precursor RNA or a processing by-product containing the intergenic sequence at 24hr and 6d wheat mtRNA (Figure 4.1E), although it cannot be excluded that it is present at earlier times, at very low steady-state levels or quickly turned-over.

Figure 4.2: Schematic of the rpl5–Ψrps14 genes in the wheat (A) and rice (B) mitochondrial genomes and their major transcripts during embryo-to-seedling development (light gray boxes)

Sizes are given in kilobases (kb) on the right. The bent arrow represents the promoters upstream of *rpl5* and arrow heads show the processing sites at the 3' ends or mapped 5' ends with position in nucleotide relative to start and stop codons respectively. Transcript ends of *cox3* based on Gualberto *et al*, 1990a and *atp4-cox3* in rice based on Liu *et al*, 1992. Sequence at the promoter and the 5' end processing sites are shown underneath with 5' end nucleotide boxed and putative YRTA promoter sequence underlined. Empty block arrow heads underneath demonstrate the breakpoint in homology between rice and wheat. Editing sites in rpl5 and rp14 are shown by filled dots based on Sandoval *et al*, 2004.


A
Wheat mtDNA
-790
-299
rpl5
rps14
-323
cox3
+57
tRNA-E
~3.5kb
~2.7kb
1.2 kb
-790
CTCATCGAACTTGACTTGTAGAGTGAGACCTGTG
-299
CCTATAAATAACTAGTAT
B
Rice mtDNA
-82
+432
1.4 kb
-651
-580
-447
atp4
+62
~2.4kb
100bp

4.3 Transcript profiles of *rpl5-Ψrps14* in rice mitochondria

Northern analysis with *rpl5* and *rps14* probes using 24hr and 6d rice mtRNA show a messenger RNA of ~ 1.4 kb and no monocistronic transcript for either gene (Figure 4.1 B and C). Total UTR length for the mRNA would be ~530 nt based on the *rpl5* and *rps14* coding sequence. The messenger RNA is more abundant in 24hr than in 6d. The LB406 oligomer upstream of *rpl5* detects a faint hybridization signal to a larger RNA species of ~ 2.4 kb, also seen with *rpl5* probes but not with *rps14* probes. This RNA species could correspond to a RNA transcript with ~1.6kb of total UTR length.

In rice, *cox3* is co-transcribed with *atp4* (formerly called *orf25*) (Liu *et al.* 1992). The wheat *cox3* promoter region in wheat is upstream of the breakpoint with rice and in rice rearrangements have place *cox3* downstream of *atp4* creating a new co-transcriptional unit (Liu *et al.*, 1992) (Figure 4.3).

4.4 Mapping the 5' and 3' ends of the *rpl5-Ψrps14-cox3* transcripts in wheat

Circular-RT-PCR experiments were used to map the 5' and 3' ends of the 3.5 kb and 2.7 kb *rpl5-Ψrps14-cox3* polycistronic transcripts in 24hr wheat embryos. Circular-RT-PCR products corresponding to both transcripts were obtained in TAP-treated mtRNA samples (Figure 4.3A). Faint cr-RT-PCR products were sometimes seen in non-TAP treated samples although these products were not seen in all experiments. This leads us to believe that the *rpl5-Ψrps14-cox3* transcripts are primary transcripts. Sequencing of these products revealed two major 5' ends at -743 (P1) and -299 (P2) upstream from the *rpl5* start codon (Table 4.1A). 1/6 clone of 24hr TAP treated wheat mtRNA had a 5' end at -790. Sequencing of the non-TAP treated cr-RT-PCR revealed that 2/7 clones had shorter 5' ends (clone no. 10 and 13), suggestive of 5' to 3' exonuclease, and two had the same 5' end as clones from TAP treated samples (clone no. 11 and 12) (Table 4.1A). Since T4 RNA ligase is able to ligate di-phosphate 5' ends and a small amount of damage at the 5' end would allow ligation, the latter two clones might represent primary transcripts that were damaged in such a way. Notably, 3/7 clones of 24hr wheat non-TAP treated had 5' ends around -86. These clones could represent a processing site although more clones are needed to verify this, as products from non-TAP treated wheat RNA were not reproducible. Circular-RT-PCR data from 6d wheat mtRNA is needed to determine the end of the ~3.2 kb transcript. Preliminary attempts

Table 4.1: Table showing the sequences from cr-RT-PCR clones of rpl5-Ψrps14-cox3 co-transcripts in 24hr wheat mtRNA and rpl5-Ψrps14 transcripts for 24hr and 7d rice mtRNA, aligned with the wheat and rice mtDNA sequences.

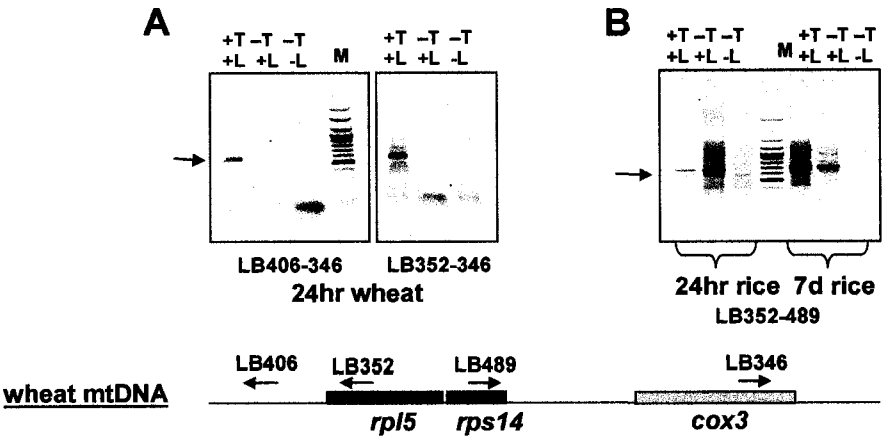
Based on the wheat and rice mtDNA sequence (AP008982 and BA000029, bold) depicting the 5' (A) and 3' (B) ends determined in this study (grey highlight) of the *rpl5-rps14-cox3* transcripts in wheat and *rpl5-rps14* transcript in rice, with precise end position written in the right. Non-encoded nucleotides found at the 5'-3' junction were removed from the sequence for better alignment purposes and are shown on the right. Clones are categorized by developmental stage (24hr and 7d), and mtRNA treatment with tobacco acid phosphatase TAP (+ / -). Clone number (clone no.) shown on the right denote corresponding clones between A and B. Small caps a nucleotide in a rice *rpl5-rps14* clone sequence show where an extra A nucleotide was found.

A		**-790**	**-743**	**5′ end**		**clone no.**
wheat mtDNA	***rpl15***	**AGATAGTAGGCACAGGTCCGCTAAGAGCTCATCGAACTTGACTTGTAGAGTGAGACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG**				
24hr wheat +TAP		AGTAGGCACAGGTCCGCTAAGAGCTCATCGAACTTGACTTGTAGAGTGAGACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG			-790	4
	P1	GAGACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG			-743	5
		GAGACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG			-743	6
24hr wheat -TAP		TCTCGG			-703	10
		GAGACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG			-743	11
		GAGACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG			-743	12
		ACCTGTGCTCGATATGGTTCATTTCAGTGCTCTTTCTCTCGG			-738	13
		-299				
wheat mtDNA	***rpl15***	**CCTATAAATAACTAGTATCTCAAAAATCCAAATAA**				
24hr wheat +TAP	**P2**	AACTAGTATCTCAAAAA-CCAAATAA			-299	1
		AACTAGTATCTCAAAAA-CCAAATAA			-299	2
		AAAAA-CCAAATAA			-287	3
wheat mtDNA	***rpl15***	**CTCGTACTTGACGTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC--AAAAAAAGCGGA**				
24hr wheat -TAP		GTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC--AAAAAAAGCGGA			-79	7
		TCACAGTGCTCATTCTC--AAAAAAAGCGGA			-35	8
		TTGACGTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC--AAAAAAAGCGGA			-84	9
		-82				
rice mtDNA	***rpl15***	**CTCGTAGTTGACGTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA**				
24hr rice +TAP		GACGTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTCaAAAAAAAAGCGGA			-82	14
		GACGTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-82	15
24hr rice -TAP		AAAAAAAGCGGA			-13	16
		ATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-41	17
		TTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-53	18
		CGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-71	19
7D rice +TAP		TTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-74	20
		GTTGGTTTCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-73	21
7D rice -TAP		TCGGTTCGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-72	22
		CGATCAACTATCCTTTTTGAAGCGGATAGTTCACAGTGCTCATTCTC-AAAAAAAAGCGGA			-66	23

B		**3′ end** **+58**	**non-encoded nts**		**clone no.**
wheat mtDNA	***cox3***	**AAGCTCGAAGACAAAGAGAACTTCTCCGGGT**			
24hr wheat +T		AAGCTCGAAGACAAAGAGAACTTCT		+58	1
		AAGCTCGAAGACAAAGAGAACTTCT		+58	2
		AAGCTCGAAGACAAAGAGAACTTCT	A	+58	3
		AAGCTCGAAGACAAAGAGAACT		+56	4
		AAGCTCGAAGACAAAGAGAACTTCT		+58	5
		AAGCTCGAAGACAAAGAGAACTTCTC		+59	6
24hr wheat -T		AAGCTCGAAGACAAAGAGAACTTCT		+58	7
		AAGCTCGGAGACAAAGAGAACTTCT		+58	8
		AAGCTCGAAGACAAAGAGAACTTCTC		+59	9
		AAGCTCGAAGACAAAGAGAACT		+56	10
		AAGCTCGAAGACAAAGAGAACTTCT		+58	11
		AAGCTCGAAGACAAAGAGAACTTCT		+58	12
		AAGCTCGAAGACAAAGAGAACTTC		+57	13
		+433			
rice mtDNA	***rps14***	**GAGAAGAGTGGGTATGAGGGCTTCGCTCGCTTGTT**			
24hr rice +T		GAGAAGAGTGGGTATGAGGGCTTCGCTCGCT		+433	14
		GAGAAGAGTGGGTATGAGGGCTTCGCTCGCT		+433	15
24hr rice -T		GAGAAGAGTGGGTATGAG		+420	16
		GAGAAGATCGGGTATGAGGGCTTCGCTCG		+431	17
		GAGAAGAGTGGGTATGAGGGCTTCGCTCGCT		+433	18
		GAGAAGAGTGGGTATGAGGGCTTCGCTCGC		+432	19
7D rice +T		GAGAAGAGTGGGTATGAGGCCTTCGCTCG	AA	+431	20
		----------CTCG		+431	21
7D rice -T		GAGAAGAGTGGGTATGAGGGCTT		+425	22
		GAGAAGAGTGGGTATGAGGGCTTC		+426	23

Figure 4.3: cr-RT-PCR analysis of the rpl5-Ψrps14-cox3 co-transcripts in 24hr wheat mtRNA (A) and rpl5-Ψrps14 transcripts for 24hr and 7d rice mtRNA

The three RNA treatments are given above: with tobacco acid phosphatase (TAP) and with RNA ligase (+T +L), without TAP and with RNA ligase (-T +L) and without TAP and without RNA ligase (-T -L, negative control). M signifies the DNA ladder lane and oligomer used in PCR are written below. Arrows point to the products extracted from the gel and cloned for sequencing. Filled bars below represent gene organization in the wheat mitochondrial genome and arrow heads show oligomers position.


A
+T –T –T
+L +L -L
M
+T –T –T
+L +L -L
LB406-346
LB352-346
24hr wheat
B
+T –T –T
+L +L -L
M
+T –T –T
+L +L -L
24hr rice 7d rice
LB352-489
wheat mtDNA
LB406
LB352
LB489
LB346
rpl5
rps14
cox3

to do this experiment failed to give products as of yet, and adjustments should be done to account for the low level of *rpl5-Ψrps14-cox3* transcript at 6d (Figure 3.1 A and B). At later developmental times only one promoter might be used and maybe this is due to the proteins and machinery available, which could be used by one promoter only. Analysis of the sequences around these two promoters upstream of *rpl5* in wheat reveal that the -743 P1 transcription initiation site is preceded by TGTA 5 nucleotides upstream and the -299 P2 transcription initiation site is preceded by TATA, sequences corresponding to the YRTA promoter consensus (Figure 4.2 A). The second promoter is located in a A/T rich region (74% in the downstream 50nt) while the first promoter is followed by a sequence of typical plant mitochondrial genome G/C composition (average 44% G/C) (46% in the downstream 50nt) (Chaw *et al.* 2008).

The 3' ends of all the 24hr mtRNA wheat cr-RT-PCR clones were the same at +58 downstream from the *cox3* stop codon (with a slight scattering around this position +55 to +59, table 4.1B). Furthermore this 3' end is close to the 3' ends from +54 to +57 of the *cox3* mRNA previously mapped by Gualberto *et al.* 1990. A putative stem loop structure was found downstream of the *cox3* 3' end by Gualberto *et al.* which puts the 3' end endonucleolytic cut at the 6th nucleotide from the base of the hairpin (Gualberto *et al.* 1990)

4.5 Mapping the 5' and 3' ends of the *rpl5-Ψrps14* transcripts in rice

The same technique of cr-RT-PCR was used on 24hr and 7d rice mtRNAs to map the end of the major 1.4 kb *rpl5-Ψrps14* transcript. Products were obtained with and without TAP treatment (Figure 4.3B). A major end was seen at – 82 upstream of *rpl5* in 24hr rice (Table 4.1A and Figure 4.2B). Clones from experiments without TAP generally exhibited shorter 5' ends (Table 4.1). Clones from 7d rice mtRNA also exhibited shorter 5' ends, possibly showing a degree of 5' to 3' exonuclease activity. More clones are needed to definitively characterize the 5' end in rice.

Northern analysis indicated that the probe located upstream of *rpl5,* hybridizes to a faint ~2.4 kb high molecular weight *rpl5-Ψrps14* transcript in rice (Figure 4.1A). This low level higher molecular weight transcript could represent a primary transcript. Preliminary attempts at cr-RT-PCR using the oligomer upstream of *rpl5* primer failed to give products. If the 3' end is the same as for the 1.4 kb transcript (see below), it would map the 5' end of this

transcript roughly ~1.2 kb upstream of the *rpl5* start codon (based on rice mt genome sequence, BA000029) and primers further upstream would be necessary to use in cr-RT-PCR experiments in order to amplify this larger product.

The breakpoints in homology around *rpl5-rps14* between rice and wheat are -488 upstream of the *rpl5* start codon and +148 downstream of *rps14* (Figure 4.2). The sequence of the P2 promoter is hence also present in rice (there is no indels or point mutation in the 100 bp regions around it) but does not appear to initiate transcription in this plant.

The rice 5' end of *rpl5-Ψrps14* at -82 is not a predominant 5' end in wheat (Table 4.1A). It is to be noted that there is a point mutation between rice and wheat at -85. It is possible that this caused the loss/gain of a regulatory cis-signal (for processing or for transcription initiation) and is responsible for the difference in expression signals used at this site between the two plants.

Circular-RT-PCR clones map the 3' end of the 1.4kb *rpl5-rps14* transcript in rice at +433 downstream of *rps14* (Table 4.1B and Figure 4.2B). No evident secondary structure was found at this position using RNA folding program for the 100nt and 50nt sequence around the 3' end (RNAfold, Genebee, mfold).

The breakpoint in homology downstream of *rps14* between rice and wheat is located at +148 bp. Since the *rps14* 3' end in rice maps downstream of this breakpoint, the 148 bp sequence downstream of *rps14* shared between rice and wheat does not contain signals for 3' end processing.

The *rpl5-rps14-cox3* 3'end in wheat at +58 maps 3 nts downstream of the breakpoint in homology between rice and wheat (Figure 4.2A). Rice *cox3* has undergone a rearrangement and is now located downstream of *atp4*, with which it is co-transcribed (Liu *et al.* 1992). The major 3' end of the *atp4-cox3* co-transcript is at +61, 6 nts downstream of the breakpoint in homology with wheat (Figure 4.2B; Liu *et al.* 1992). Despite new downstream context to *cox3,* the wheat/rice common sequence may be providing signals for 3' end processing or be of importance for the stability of the mRNA. The 56 bp sequence downstream of the *cox3* 3' end has undergone duplication in the wheat mitochondrial genome and is located downstream of the two copies of *atp6* (Gualberto *et al.* 1990). In wheat, the 3' ends of the *atp6* transcripts are located 49 nt upstream of the region of rearrangement, within the recombinational unit (Bonen *et al.*1992). This indicates that the

sequence downstream of the *cox3* 3' end is not involved in the formation of the 3' end. It is to be noted that Gualberto *et al.* also characterized a rearrangement region upstream of *cox3* between *cob* and *cox3,* and *cob* is found downstream of *rpl5-rps14* in eudicot species (Hoffman *et al.*. 1999)

4.6 Minor occurrence of non-encoded nucleotides in cr-RT-PCR products from the *rpl5–Ψrps14* loci in wheat and rice

Only two clones out of 23 in wheat and rice were found to have non encoded nucleotides, namely A and AA (clones for both rice and wheat, one clones for each species, clone no.3 and 23, both from TAP treated samples, table 4.1B). Both of these clones exhibited shorter 5' ends, indicating that these transcripts might be on the path to degradation. Since the relative amount of mtRNA used in each experiment was the same as in experiments for the *ccmFN-rps1* transcripts where non-encoded nucleotides were found in ~50% of the clones, the difference in the presence of non-encoded nucleotides might be due to the difference in relative concentrations of transcripts present *in vivo*, where transcripts present in higher amount would be self-circularized more easily or the nature of the 3' end is more apt to RNA ligation, as was discussed in chapter 3.

4.7 Status of the *rpl5* and *rps14* genes in the mitochondria of other cereal species

4.7.1 The transition state of rpl5 in wheat mitochondria

The *rpl5* gene has undergone several independent gene transfers to the nucleus in the cereal lineage including wheat, and in wheat *rpl5* has a functional copy in both the mitochondria and in the nucleus, a stage in gene transfer known as the transition state (Sandoval *et al.*, 2004). This poses interesting questions regarding the expression of the functional copy in the mitochondria. Northern hybridization analysis showed that the *rpl5-Ψrps14-cox3* co-transcripts in 6d wheat seedlings were less abundant than the transcripts seen in 24hr embryos. The transition state of *rpl5* in wheat might be affecting its expression in different developmental times.

The functionality of a gene can be assessed through its editing status in mitochondria. Previously published work in wheat 24hr embryos showed that the editing sites in *rpl5* and *rps14* demonstrated 50% editing (two non-silent *rpl5* edit sites), 71% (generating *rps14* stop codon in wheat), 21% for the second *rps14* editing sites and 33% for the 3' UTR editing site in RT-PCR clones (Sandoval *et al.*, 2004). We examined the level of editing of *rpl5-rps14-cox3* RNA species population by direct sequencing of RT-PCR products (J. Hardy unpublished results, figure 4.4 A and B). Our results on mt RNA populations are consistent with these results: the four sites showed a degree of editing; ~50% edited in the RNA population (the third editing site in the 3' UTR was not examined in our experiment). Interestingly, *rps14* is also a pseudogene in potato and *Arabidopsis* (albeit arising from different gene transfer events) but *rps14* transcripts were not edited in any position in potato mitochondria (Quiñones *et al.*, 1996), while in *Arabidopsis rps14* transcripts were edited at 3 positions (Aubert *et al.*. 1992). This demonstrates the variation in constraints for editing of pseudogene transcripts in mitochondria. The editing status of *rpl5-Ψrps14-cox3* transcripts in 6d wheat seedlings would be of interest to assess its functionality as well.

4.7.2 Status of rpl5 and rps14 in oats and barley

Work on the status of mitochondrial *rpl5* in cereals was initiated previously by a 4th year honours student in our laboratory, Alain Toutloff. The following results corroborate those from his honours thesis. A Southern blot of wheat, rice, barley and oats mtDNAs was successively hybridized with oligomers and random labelled probes for *rpl5*, *rps14* and *cox3* (Figure 4.5). Notably, Southern analysis revealed the absence of *rpl5* and *rps14* in rye (Figure 4.5A), a close relative of wheat (~5 MYA, Gaut *et al.*, 2002) and the absence of *rps14* in barley (Figure 4.5B). *rpl5* coding sequences are present in oats and barley (Figure 4.5A). Several bands are seen in the EcoRI lanes of oats, rice, wheat and barley with the *rpl5* random-labelled PCR fragment probe due to the presence of an EcoRI site in the sequence spanning the probe (Figure 4.5A). Signals at the top of the wheat lanes are seen for each probe used and due to improper mtDNA electrophoresis (Figure 4.5).

The barley *rpl5* coding sequence was cloned and sequencing revealed that the extreme 3' end of *rpl5* has been disrupted by a 547 bp DNA sequence of chloroplast origin. This insertion has modified the *rpl5* coding sequence at the 3' end by 30 bp and provides a

Figure 4.4: Sequencing gel of rpl5–Ψrps14-cox3 RT-PCR products from 24hr wheat mtRNA

Editing sites are shown by dots, filled representing complete editing and half-filled representing partial editing (A) and blow-ups are shown for the two *rpl5* (top and middle) and the two *rps14* editing sites (bottom) (B). Editing sites are based on Sandoval *et al*, 2004 (*rpl5-rps14*) and Gualberto *et al*, 1990a (*cox3*). Data from J. Hardy (former technician in Bonen lab)

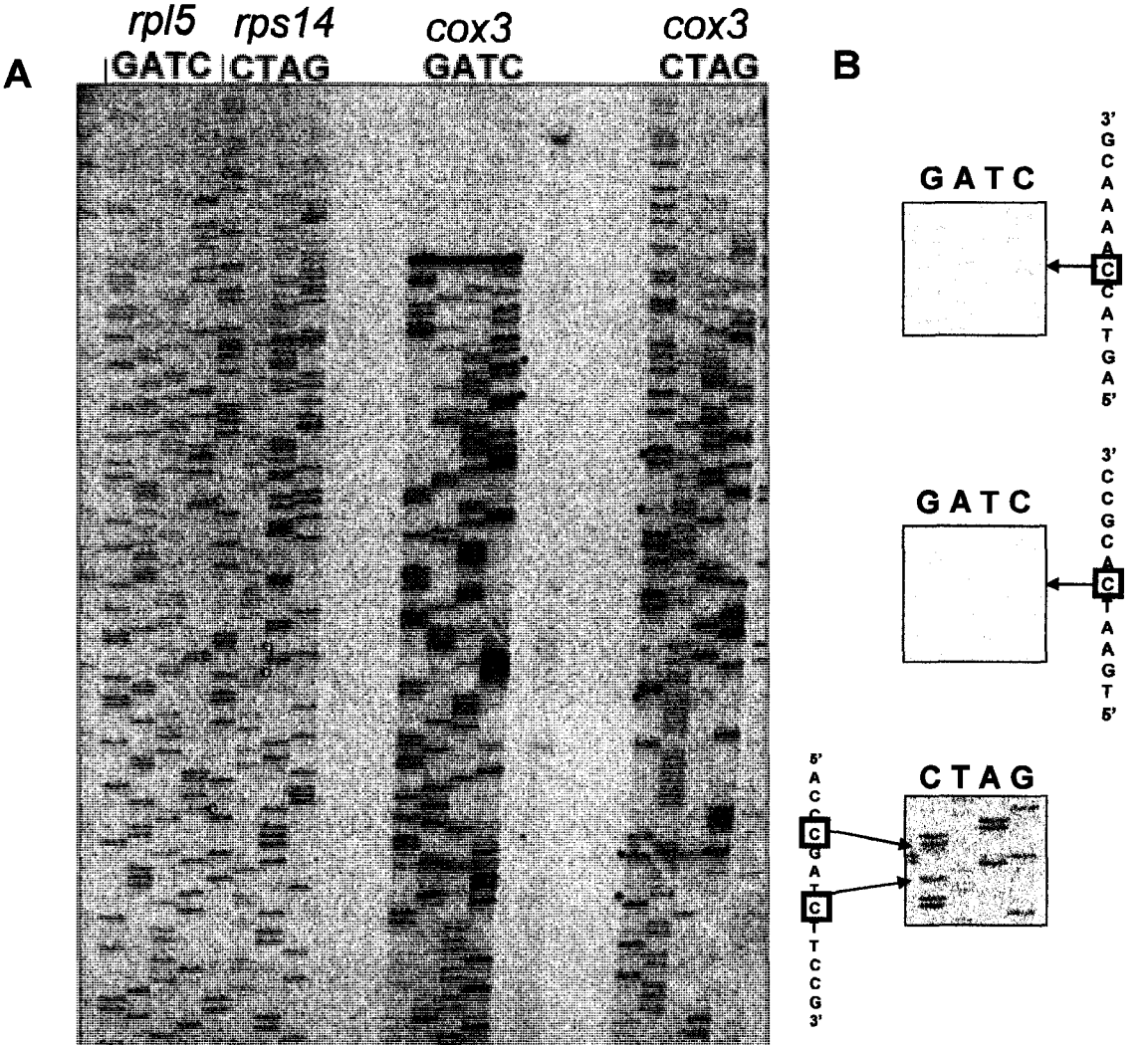
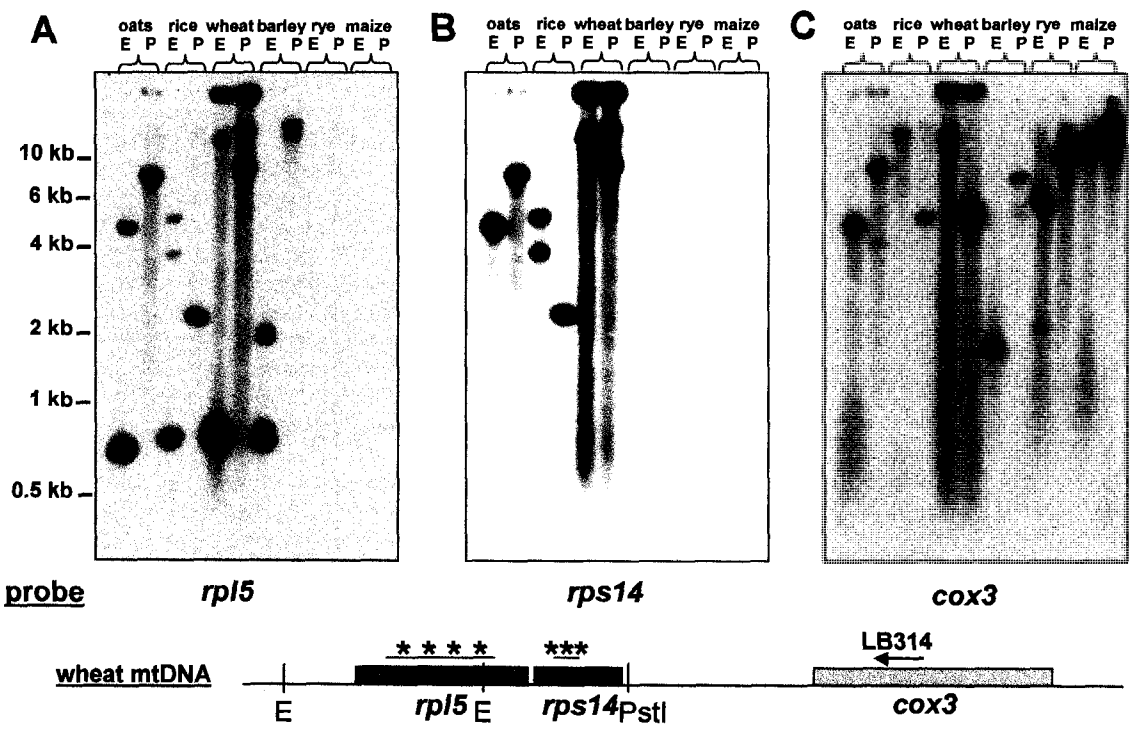

A
rpl5
GATC
rps14
CTAG
cox3
GATC
cox3
CTAG
B
GATC
3' G C A A A A C C A T G A 5'
GATC
3' C C G C A C T A A G T 5'
CTAG
5' A C C G A T C T T C C G 3'

Figure 4.5: Southern blot analysis of the rpl5, rps14 and cox3 genes in cereal species mtDNA

Southern blot was successively probed with oligomers for *rpl5* (A), *rps14* (B) and *cox3* (C), as indicated below. mtDNAs of oats, rice, wheat , barley, rye and maize were digested by EcoRI (E) and PstI(P). Filled boxes below represent gene organization in wheat and show end-labelled oligomer or random- labelled probes position.


A
B
C
oats rice wheat barley rye maize
E P E P E P E P E P E P
10 kb
6 kb
4 kb
2 kb
1 kb
0.5 kb
probe
rpl5
rps14
cox3
wheat mtDNA
* * * *
***
LB314
E
rpl5 E
rps14 PstI
cox3

stop codon (Figure 4.7 A). The translated barley *rpl5* ORF is one amino acid longer than the functional mitochondrial *rpl5* found in rice and wheat (Figure 4.7A, A. Toutloff, 2003). The 547 bp chloroplast sequence corresponds to a portion of the *psaA* gene in chloroplast and has been inserted 821 bp upstream of *rps2* (AB158222, Kubo *et al.*, 2005).

This work has been extended by northern hybridization of rice, wheat, barley, oats and rye mtRNAs. mtRNAs from 24hr embryos and 6d seedlings for these 5 species were successively hybridized with random-labeled probes of the *rpl5* and *rps14* coding sequences and end-labeled *cox3* oligomer probe (Figure 4.6). Rice and wheat displayed transcript profiles as discussed above. Oat and barley mtRNA exhibited hardly any detectable transcripts for *rpl5* (and for *rps14* in oat) (Figure 4.6 A and B).

The oat *rpl5-Ψrps14* sequence revealed that *rps14* is a pseudogene in this species and shares common inactivation frameshifts with rice and wheat *rps14* sequences (K. Mellet unpublished results and S. Calixte). This agrees with previous results demonstrating that a common *rps14* transfer event took place before the divergence of cereals (Kubo *et al.*, 1999, Sandoval *et al.* 2004, Ong and Palmer, 2006). The oat *rpl5* sequence revealed a 4 nt deletion at the 3' end of the coding sequence which translates into a different C-terminus for the *RPL5* protein (Figure 4.7A). In order to assess whether this variation in C-terminus might be expected to have deleterious effects on the *RPL5* protein, an alignment of the coding sequences at the 3'end of the functional *rpl5* gene in rice, *Arabidopsis* and *Marchantia* was done. It shows high conservation of the C-terminal sequence (Figure 4.7B), which argues for a disadvantageous effect of the indels in oats. The oat mitochondrial *rpl5* gene also contains two in-frame deletions in the 5' end of the coding sequence (Figure 4.7A).

Attempts to look at the editing status of *rpl5-rps14* in oats and *rpl5* in barley were done to assess their functionality. Low levels of transcripts in oats and barley made RT-PCR difficult as well as the possibility of DNA contamination high. Three RT-PCR clones in oats and one clone in barley showed no editing (data not shown).

Interestingly in oats, *rpl5 Ψ-rps14* is also followed by *cox3* in the mitochondrial genome, as was seen by PCR amplification (data not shown). Despite sharing this linkage, oat *cox3* transcript profiles do not appear to have higher molecular weight RNA species that could correspond to *rpl5-Ψrps14-cox3* co-transcripts (Figure 4.6C). There is no sequence data available for the sequences upstream of *rpl5* in oats and this region might provide

Figure 4.6: Northern analysis of the rpl5, rps14 and cox3 genes in various cereal species (rice, wheat, rye, oats, and barley) mtRNA

Northern blot was successively probed with oligomers for *rpl5* (A), *rps14* (B) and *cox3* (C). Filled boxes below represent gene organization in wheat and show end-labelled oligomer or random- labelled probes position. mtRNA samples of rice, wheat, rye, oats and barley were taken at 24hr and 6d post-imbibition. D)18SrRNA hybridization is shown as a loading control. Hybridization with 18S rRNA was used as standardization for RNA loading (D)

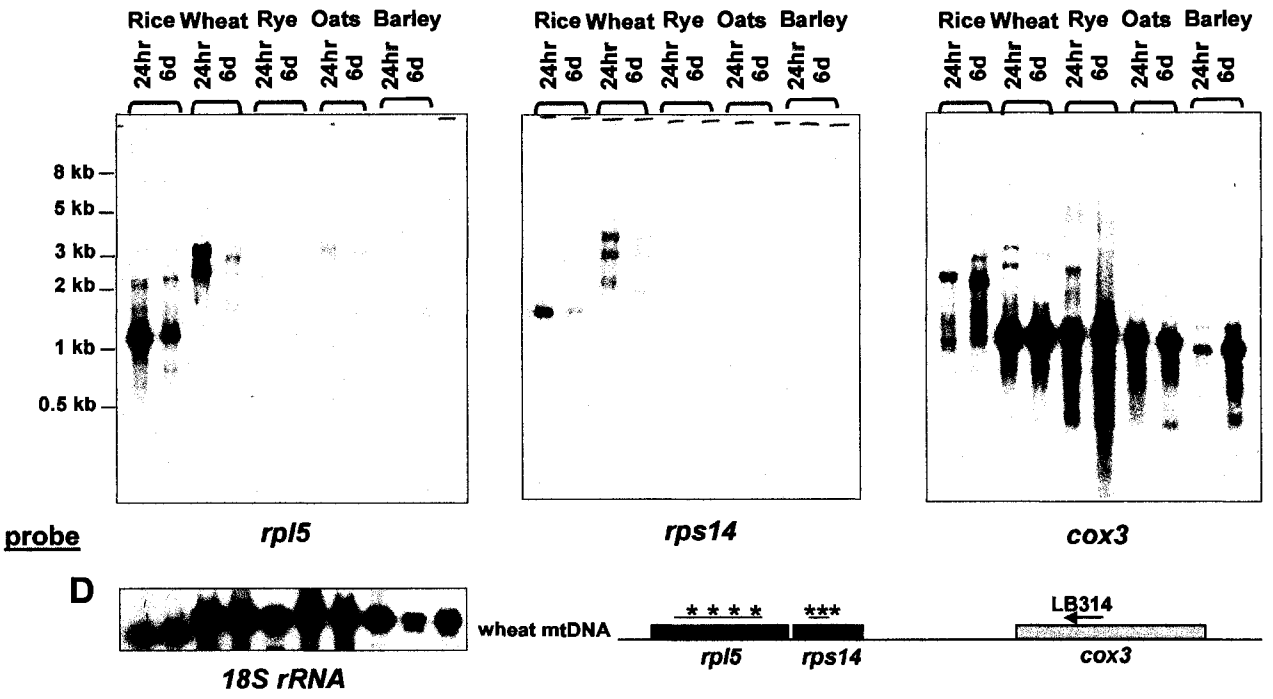

Rice Wheat Rye Oats Barley
24hr 6d
8 kb
5 kb
3 kb
2 kb
1 kb
0.5 kb
probe
rpl5
rps14
cox3
D
18S rRNA
wheat mtDNA
* * * *
***
rpl5
rps14
LB314
cox3

Figure 4.7: Clustal alignment of nuclear and mitochondrial L5 cereal amino acid sequences

Sequences are from the NCBI databank: translated from EST accessions (wheat mt, rice mt, *Marchantia* mt, *Araibidopsis* mt, oats nuc), consensus of translated accessions (Table 2.2) (wheat, barley nuclear) or sequencing of PCR products (oats and barley mt). The pink highlight indicates the common N-terminal extension of *rpl5* nuclear copies in wheat and barley, the yellow highlights indicates the N-terminal extension in the maize nuclear copy. Arrow head show nuclear intron position. Red boxes denote common nuclear-specific indels in wheat, barley and oats. Purple oval show the position of the editing sites based on Sandoval *et al,* 2004. Green highlight represents chloroplast DNA insertion in the barley mitochondrial copy and oat-specific deletion is show in blue box, blue highlight showing the resulting frame shift in the oats mitochondrial amino acid sequence. Alignment of the C-terminus of the rice, *Arabidopsis* and *Marchantia* rpl5 mitochondrial sequences is boxed underneath and teal highlight denotes amino acid conservation.

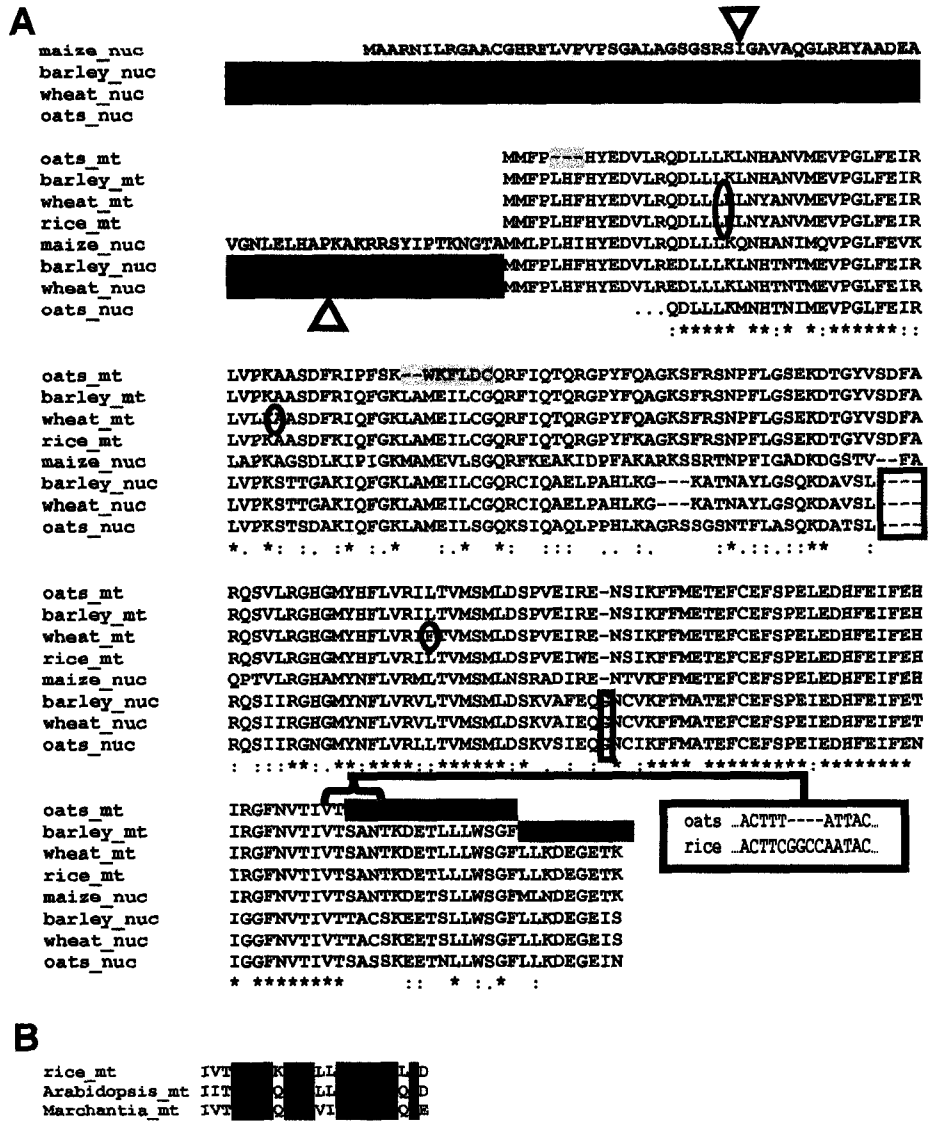
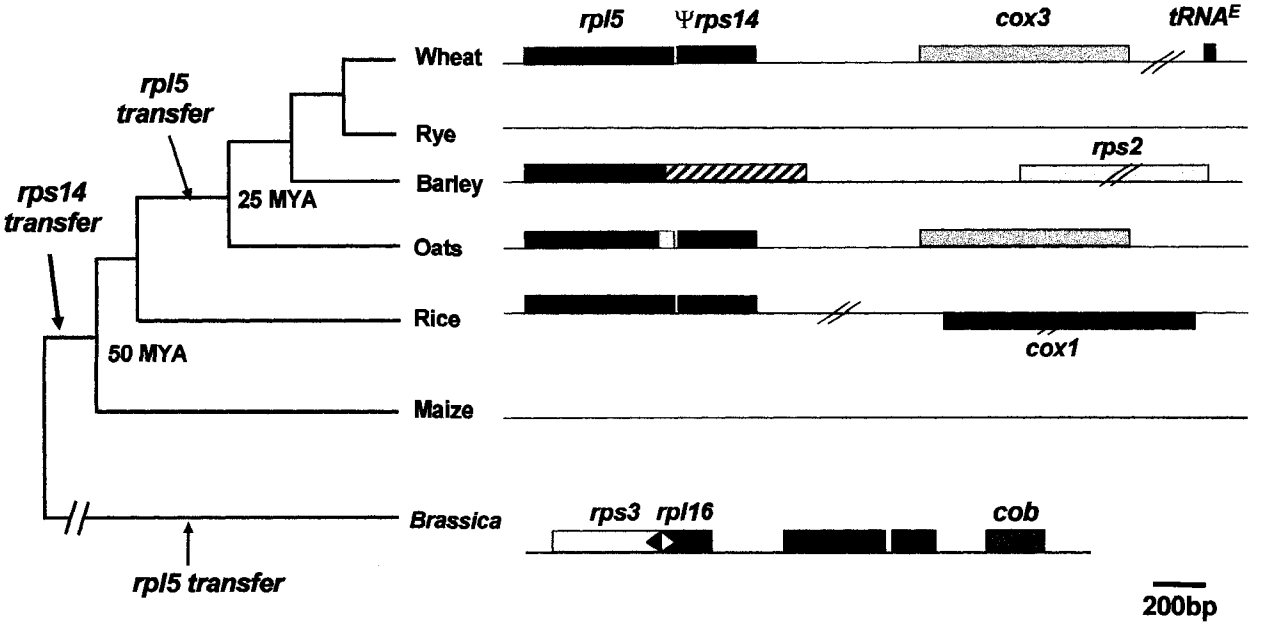

A
maize_nuc MAARNILRGAACGHRFLVPVPSGALAGSGSRSIGAVAQGLRHYAADEA
barley_nuc
wheat_nuc
oats_nuc
oats_mt MMFP---HYEDVLRQDLLLKLNHANVMEVPGLFEIR
barley_mt MMFPLHFHYEDVLRQDLLLKLNHANVMEVPGLFEIR
wheat_mt MMFPLHFHYEDVLRQDLLLKLNYANVMEVPGLFEIR
rice_mt MMFPLHFHYEDVLRQDLLLKLNYANVMEVPGLFEIR
maize_nuc VGNLELHAPKAKRRSYIPTKNGTAMMLPLHIHYEDVLRQDLLLKQNHANIMQVPGLFEVK
barley_nuc MMFPLHFHYEDVLREDLLLKLNHTNTMEVPGLFEIR
wheat_nuc MMFPLHFHYEDVLREDLLLKLNHTNTMEVPGLFEIR
oats_nuc ...QDLLLKMNHTNIMEVPGLFEIR
:***** **:* *:******::
oats_mt LVPKAASDFRIPFSK---WKFLDCQRFIQTQRGPYFQAGKSFRSNPFLGSEKDTGYVSDFA
barley_mt LVPKAASDFRIQFGKLAMEILCGQRFIQTQRGPYFQAGKSFRSNPFLGSEKDTGYVSDFA
wheat_mt LVLKAASDFRIQFGKLAMEILCGQRFIQTQRGPYFQAGKSFRSNPFLGSEKDTGYVSDFA
rice_mt LVPKAASDFKIQFGKLAMEILCGQRFIQTQRGPYFKAGKSFRSNPFLGSEKDTGYVSDFA
maize_nuc LAPKAGSDLKIPIGKMAMEVLSGQRFKEAKIDPFAKARKSSRTNPFIGADKDGSTV--FA
barley_nuc LVPKSTTGAKIQFGKLAMEILCGQRCIQAELPAHLKG---KATNAYLGSQKDAVSL----
wheat_nuc LVPKSTTGAKIQFGKLAMEILCGQRCIQAELPAHLKG---KATNAYLGSQKDAVSL----
oats_nuc LVPKSTSDAKIQFGKLAMEILSGQKSIQAQLPPHLKAGRSSGSNTFLASQKDATSL----
*. *: :. :* :.* :.* *: ::: .. :. :*.::.::** :
oats_mt RQSVLRGHGMYHFLVRILTVMSMLDSPVEIRE-NSIKFFMETEFCEFSPELEDHFEIFEH
barley_mt RQSVLRGHGMYHFLVRILTVMSMLDSPVEIRE-NSIKFFMETEFCEFSPELEDHFEIFEH
wheat_mt RQSVLRGHGMYHFLVRILTVMSMLDSPVEIRE-NSIKFFMETEFCEFSPELEDHFEIFEH
rice_mt RQSVLRGHGMYHFLVRILTVMSMLDSPVEIWE-NSIKFFMETEFCEFSPELEDHFEIFEH
maize_nuc QPTVLRGHAMYNFLVRMLTVMSMLNSRADIRE-NTVKFFMETEFCEFSPELEDHFEIFEH
barley_nuc RQSIIRGHGMYNFLVRVLTVMSMLDSKVAFEQ-NCVKFFMATEFCEFSPEIEDHFEIFET
wheat_nuc RQSIIRGHGMYNFLVRVLTVMSMLDSKVAIEQ-NCVKFFMATEFCEFSPEIEDHFEIFET
oats_nuc RQSIIRGNGMYNFLVRLLTVMSMLDSKVSIEQ-NCIKFFMATEFCEFSPEIEDHFEIFEN
: :::**:.**:****:;******:* . : : :**** *********:********
oats_mt IRGFNVTIVT
barley_mt IRGFNVTIVTSANTKDETLLLWSGF
wheat_mt IRGFNVTIVTSANTKDETLLLWSGFLLKDEGETK
rice_mt IRGFNVTIVTSANTKDETLLLWSGFLLKDEGETK
maize_nuc IRGFNVTIVTSANTKDETSLLWSGFMLNDEGETK
barley_nuc IGGFNVTIVTTACSKEETSLLWSGFLLKDEGEIS
wheat_nuc IGGFNVTIVTTACSKEETSLLWSGFLLKDEGEIS
oats_nuc IGGFNVTIVTSASSKEETNLLWSGFLLKDEGEIN
* ********
oats ...ACTTT----ATTAC...
rice ...ACTTCGGCCAATAC...
B
rice_mt IVT
Arabidopsis_mt IIT
Marchantia_mt IVT

Figure 4.8: Phylogenetic tree showing presence/absence of mitochondrial rpl5 and rps14 genes in cereal species

Filled boxes represent coding sequences, hatched bar is DNA sequence of chloroplast origin and dotted white box represents indels. Dashes on DNA sequence represent a break in distance. Gene names are written above. Divergence times are based on Chaw *et al*, 2004. The eudicot *Brassica* linkage is shown as an outgroup. Red and green arrows represent possible times of gene transfer to the nucleus for *rpl5* and *rps14*, respectively.


rpl5
Ψrps14
cox3
tRNA^E
Wheat
rpl5
transfer
Rye
rps2
rps14
transfer
25 MYA
Barley
Oats
Rice
cox1
50 MYA
Maize
Brassica
rps3
rpl16
cob
rpl5 transfer
200bp

different transcriptional and post-transcriptional signals for the expression of this region. In oats, *cox3* is transcribed into a ~ 1.2 kb RNA species which suggests it shares the transcriptional signals with those present in wheat.

We looked for a nuclear copy of *Rpl5* in barley by BLAST searches against the EST_others (nr) databank. Five expressed sequenced tags (EST) were found for a barley nuclear *Rpl5* (accession numbers in table 2.2). Alignments of these EST sequences with five EST sequences for the wheat *Rpl5* nuclear copy (Sandoval *et al.*, 2004) show that they share a common pre-sequence (amino-terminal sequence acquired after transfer to the nucleus, possibly involved in the targeting of the protein to the mitochondria) indicating a common transfer event (Figure 4.7A). The barley and wheat *Rpl5* pre-sequence originated from a duplication of the pre-sequence upstream of *Rpl4* (Sandoval *et al.* 2004). It should be noted that the maize nuclear *Rpl5* arose from an independent gene transfer event and its presequence is of unknown origin (Sandoval *et al.* 2004).

For oats, one EST sequence for a *Rpl5* nuclear copy was found that does not extend to the N-terminus (Accession number in table 2.2, Table 4.7A). However this sequence shares two indels with the wheat and the barley *Rpl5* nuclear copies and appears to have arisen from the same transfer event. It also contains other in-frame mutations which points towards it not being of wheat or barley origin.

The nuclear copies of *Rpl5* in barley and oats place the *rpl5* gene transfer event after their divergence from rice and as being the same event as in wheat (Figure 4.8), suggesting that the transition stage for *rpl5* persisted for at least 25 MYA (Gaut *et al.*, 2002). Mutations in the oats and barley mitochondrial copies are not the same, indicating the copies have diverged in species-specific ways. The functional status of mitochondrial *rpl5* in oats and barley is in question due to indels in both plants and low transcript levels.

4.8 Future directions

Future work on wheat *rpl5-Ψrps14-cox3* transcripts should be done to definitively map the ends of these transcripts. The slightly smaller co-transcript seen at 6d is of particular interest since it may represent a processing of the larger precursor (from promoter P1) while there appears to be no transcript initiated at the P2 promoter or processed at that site in 6d seedlings. Circular-RT-PCR experiments and guanylyltransferase-capping followed by

nuclease protection assays will help differentiate between promoters and processing sites (see Chapter 3). *In organello* transcription run-on experiments would help determine the relative activity of the promoters in 24hr, 6d and in earlier developmental time points. Preliminary northern analyses at earlier time points show the presence of the higher molecular weight transcripts as early as 6hr and the second molecular weight transcript becomes detectable between ~12 hr and ~24hr. This suggests that transcription is initiated from the P1 promoter during early development and at later stages while the P2 promoter is used between 12hr and 24hr (Appendix 4, J. Li-Pook-Than unpublished data).

In wheat it is likely that there is a processing site as well as a promoter in the *rps14-cox3* intergenic region. The 1.2 kb monocistronic *cox3* transcript could represent a mixed population of processed and primary transcripts, with varying ratios in different developmental stages. Since the 3' end of the co-transcripts are the same as the ones mapped for the monocistronic *cox3* transcript (Gualberto *et al.*, 1990) no further 3' end processing appears to occur. Processing and transcription initiation might be used differentially depending on the developmental stage: the co-factors involved in transcription initiation and 5' end processing are unknown but are imported from the nucleus. Northern analysis of wheat mtRNA from 0 hr to 6d shows that in the earlier stages of development there is a low level of the 3.2 kb *rpl5-rps14-cox3* co-transcript but abundant *cox3* monocistronic mRNAs (Appendix 4, J. Li-Pook-Than unpublished data). These transcripts could be stored or ones arising from the use of the *cox3* promoter. The *cox3* promoter was mapped using 24hr wheat embryos (Covello and Gray 1991), indicating it is used at that time. The lack of *rpl5-rps14* bicistronic transcripts and the presence of a possibly processed *cox3* messenger RNA are reminiscent of the *ccmFN-rps1* transcript profile (chapter 3).

The putative presence of a larger transcript in rice should also be investigated in order to map a promoter upstream in rice. Once again cr-RT-PCR, guanylyltransferase-capping experiments and nuclease protection assays would precisely map the ends of *rpl5-rps14* in rice.

Although there has been no report of such a case, it is possible that *rpl5* in transition state is expressed early in development from the mitochondria and from the nucleus at later times in development. The protein encoded by the nuclear copy has some slight amino acid differences (Figure 4.7A) due to the adaptation to the nuclear genome of the transferred

copy. There might be a specialization of the nuclear Rpl5 protein although it still maintains the same function as the mitochondrial-encoded copy. The expression profile of the nuclear *Rpl5* and direct sequencing of mitochondrial RT-PCR populations in 6d wheat will help us look at the functionality of the *rpl5-rps14-cox3* transcript in later stages of development.

Chapter 5 General discussion

5.1 Expression of the co-transcribed genes *ccmFN-rps1* and *rpl5-Ψrps14-cox3* loci in wheat mitochondria during early embryo to seedling development

The transcript profiles exhibited by the two loci examined here reflect the many post-transcriptional steps involved in the formation of the mature messenger RNAs in plant mitochondria. The presence of relatively abundant precursor RNAs during early stages of embryo-to-seedling development is in accordance with previous observations where the association between transcription and post-transcriptional processes appears to be delayed (Li-Pook-Than and Bonen, 2006).

Events leading to the formation of secondary 5' ends and transcript 3' ends can be hypothesized. We failed to identify t-elements at the 3' ends of the transcripts examined here. Putative single stem loops can be seen just downstream of the *ccmFN-rps1* 3' end (this study) and of *cox3* in wheat (Gualberto *et al.* 1990). If these structures are involved in 3' end formation it would imply a endonucleolytic cut at the 5' end of the structure, possibly by a RNase P-like enzyme, such as the one used in tRNA 5' end formation. However the stem loops cannot be implicated in mRNA stability mechanism since they are not found in the mature transcripts. No evident secondary structure was found at the 5' end processing site -23 upstream of *ccmFN*. Site specific recognition of this processing site presumably is achieved through other mechanisms, possibly through binding of auxiliary factors.

Of the three promoter regions analysed here two contain the motif YRTA: TGTA (P1 *rpl5*) and TATA (P2 *rpl5*) and the third promoter region includes TCTA (-589 *ccmFN*). This variation in promoter motif is not unusual as many mitochondrial promoters mapped in monocots and in eudicots diverge from this consensus (reviewed in Gagliardi and Binder 2007, Zhang and Liu 2006). The transcription machinery recognition component must hence be flexible to recognize transcription initiation site and the varying sequences of promoter regions. It might be that the plasticity of the signals used has evolved to contend with the high rate of rearrangements in plant mitochondrial genomes or vice-versa.

5.2 Rearrangement effect on expression and transcription of genes, especially ones in a transition state

Rice and wheat have diverged about 50 million years ago (Gaut *et al.* 2002). Although their gene content varies, they share a relatively small amount of conserved sequence surrounding genes (Figure 5.2 and Appendix 1) but hardly any common intergenic non-coding sequence (Kubo and Mikami, 2007). Gene linkage conservation is also limited to a few very closely linked genes such as *rpl5-rps14*, *rps3-rpl16*, *nad3-rps12* and *trrn18S-trrn5S*. The exceptional case of *ccmFN-rps1-mat-r-nad1e-nad5c* will be discussed below. Rearrangements in these highly recombinogenic genomes create new gene linkages (e.g. *rpl5-rps14-cox3* in wheat versus *atp4-cox3* and *rps3-rpl16-nad3-rps12* in rice) and gene transfer to the nucleus affects not only gene content but the loss of a gene might disrupt the mitochondrial genomic context of genes close by. Sequence conservation between plants can be due to constraint on specific cis-elements or signals for transcription or translation. For a rearrangement to be successful, functional regulatory signals need to be in place to allow for gene expression. One set of signals should not be preferentially selected for as long as the new signals are functional. In this study we have seen the example of *rpl5-rps14* in rice and wheat, in which the common sequence between rice and wheat upstream of *rpl5* provides a regulatory signal in rice (-82, promoter or processing site) but not in wheat. Wheat-specific upstream sequences have brought in two new promoters that are used. Also a short sequence downstream of *rps14* has been conserved between rice and wheat but does not contain signals for transcription termination or transcript 3'end formation.

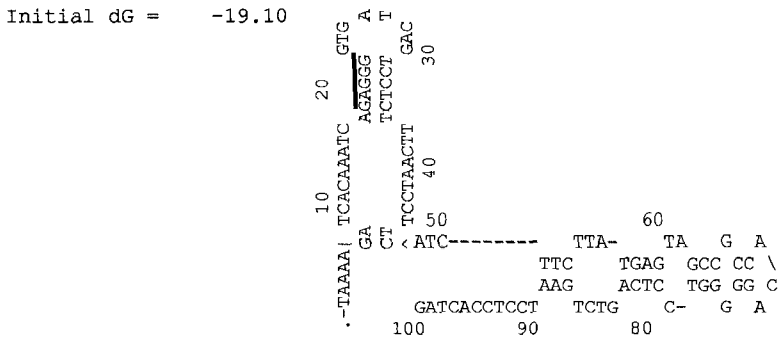
5.3 Secondary structures

Using RNA folding prediction software, the flanking regions and UTRs of the genes were examined for putative secondary structures. The results were analysed to look for inverted repeats. For the sequences analysed, the results seen did not give rise to consistent secondary structures. An example is given in Figure 5.1 for the 100 nt downstream of the *ccmFN* -589 5' end of the 3.2 kb *ccmFN-rps1* transcript. Mfold found a 6 nucleotide stem-loop and RNAfold showed a 9 nucleotide hairpin (in the same region but with different nucleotides) but Genebee did not find either pair. The surrounding sequence was found to be folded in different ways. Other structures generated with the mfold program had similar delta G values between -19 kcal/mol and -18 kcal/mol.

Figure 5.1: Illustrative examples of RNA secondary structures by mfold, RNAfold and GeneBee RNA folding programs.

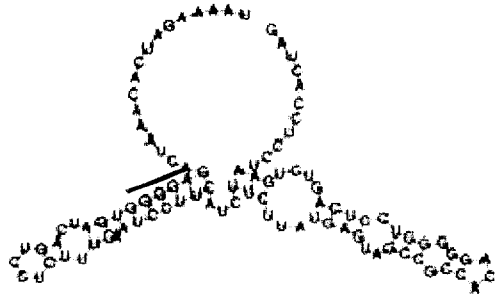
Secondary structures found with RNA folding programs: A, mfold (Zuker, 2003); B, RNAfold (Hofacker et al. 1994) and C, Genebee (http://www.genebee.msu.su) for the 100nt immediately downstream of the *ccmFN* -589 (3.2 kb transcript) 5' end. Results are given as displayed by the programs. A red highlight was added to show the position of a 5nt sequence for comparison purposes.

A Mfold



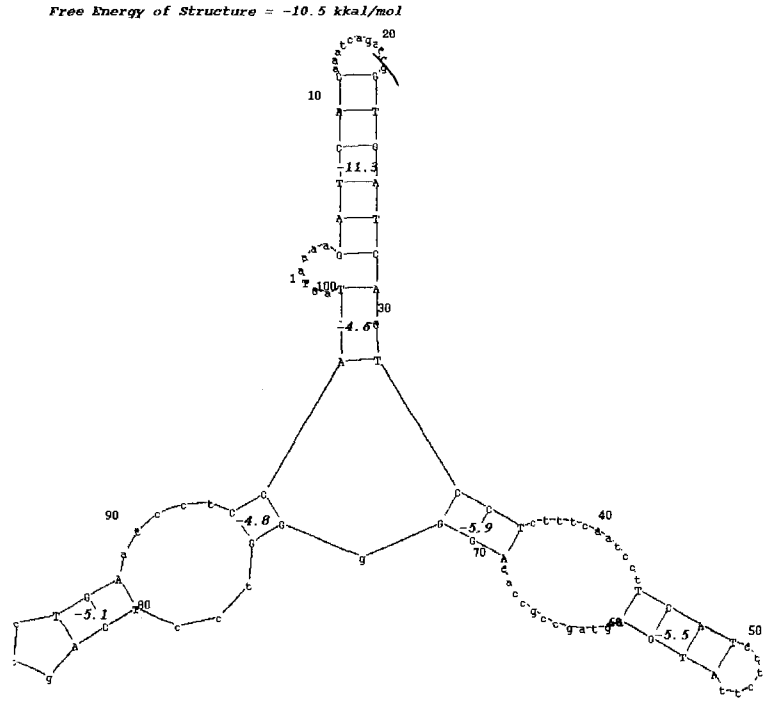
B RNAfold

> s13562
UAAAAGAUCACAAAUCAGAGGGGUGAUCAGUCCUCUUUCAAUCCUUCAUCUUCUUAUGAGUAGCCGCCACAGGGGGUCCUCAGUCUGAAUCCUCCACUAG
.....((((((((..((....)).)).)))))))...(((...((((..(((.((...)).))).))))....)))..........
(-19.00)



C Genebee

```
Stem_ 1 with energy -11.300000 Kkal/mol
   6   11     GATCAC
  28   23     CTAGTG

Stem_ 2 with energy -5.900000 Kkal/mol
  32   34     CCT
  73   71     GGA

Stem_ 3 with energy -5.500000 Kkal/mol
  46   49     TCAT
  59   56     AGTA

Stem_ 4 with energy -5.100000 Kkal/mol
  80   82     TCA
  88   86     AGT

Stem_ 5 with energy -4.800000 Kkal/mol
  75   76     GG
  95   94     CC

Stem_ 6 with energy -4.600000 Kkal/mol
  29   31     AGT
  98   96     TCA
```



5.4 Alignment of plant mitochondrial genomes to search for long conserved regions

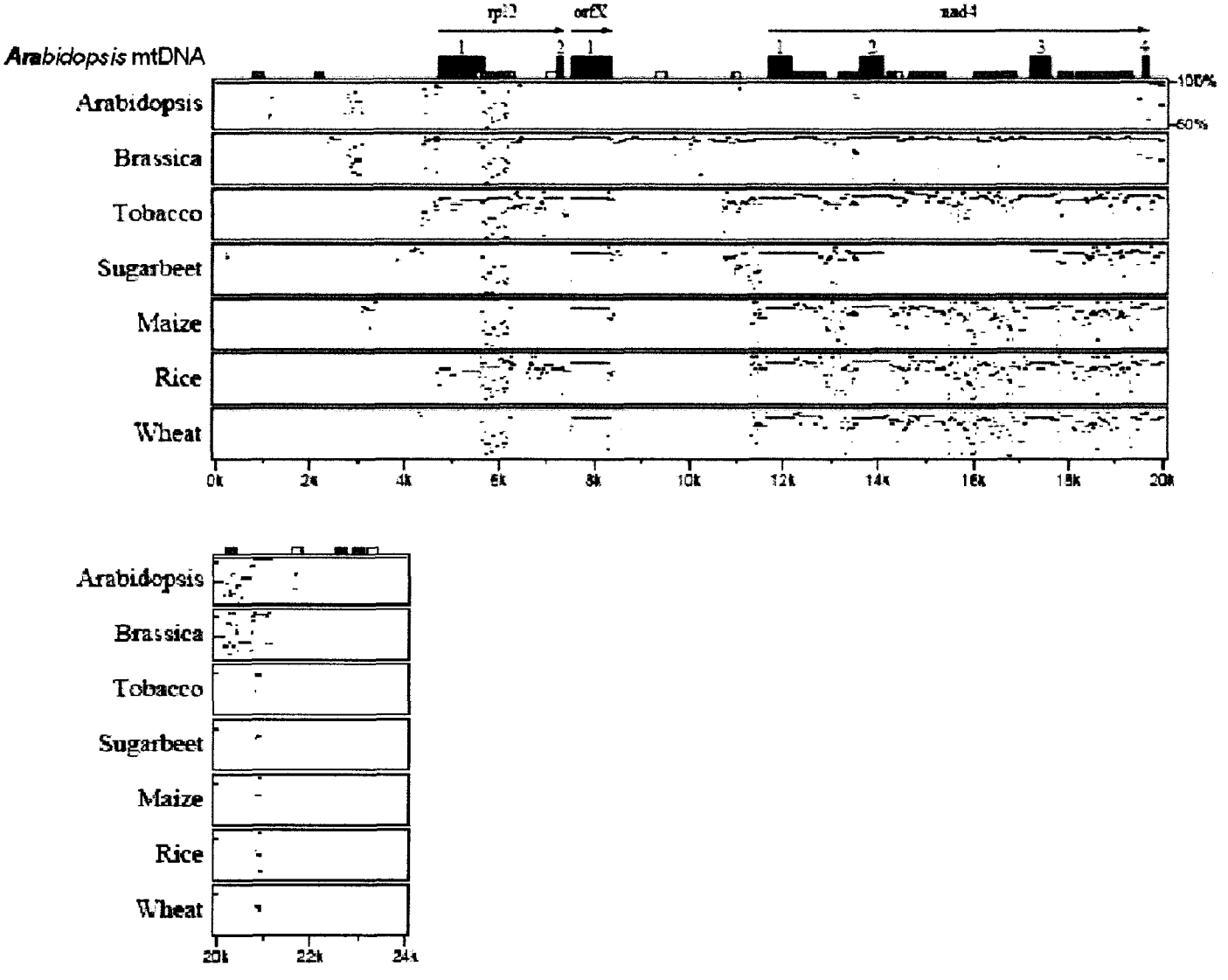
Intergenic sequences are usually species-specific (Kubo and Mikami, 2007). To look for intergenic regions conserved between species, alignments of the seven completely sequenced mitochondrial genomes were performed using the MultipPipmaker program (Schwartz *et al.* 2000, Appendix 1) using each sequence once as input and the rest as query sequences. Alignments of the seven fully sequenced mitochondrial genomes (namely wheat, rice, maize, sugarbeet, tobacco, *Brassica* and *Arabidopsis*; Table 2.2) were done with the MultiPipmaker alignment program (Schwartz *et al*, 2000) (Figure 5.2). This program uses an input sequence with coding sequence coordinates and searches for homologous sequences in query sequences. The program does not however take into account gene linkage in the query sequences, only presence of homologous sequences and one must refer to the original query sequences for gene order. A few repeated regions in intergenic spacers were present in all three species and exceeded a few hundred base pairs (Appendix 1, wheat as query), such as a ~500 bp stretch homologous to the *atpA* sequence of chloroplast origin. This region is also present in tobacco and *Arabidopsis*, which signifies an early duplication and integration of the plastid sequence.

The longest linkage region was found between *Brassica* and *Arabidopsis*, the two most closely related species examined (~20 MYA, based on molecular clock behaviour) (Koch *et al.* 2000): the *rpl2-orfx(mttB)-nad4* loci has a conserved 3.3kb intergenic sequence between *orfx* and *nad4* (Figure 5.2; 95% identity with indels). The mitochondrial *rpl2* copy is co-transcribed with *orfX(mttB)* with no detectable monocistronic transcripts for either gene (Forner *et al.* 2007). It would be of interest to study this region in *Arabidopsis* and *Brassica* to characterize their transcript profiles and examine the intergenic region in detail.

Alignments of the seven fully sequenced plant mitochondrial genomes show that the linkage *ccmFn-rps1-mat-r-nad1e-nad5c* is conserved between maize, rice and wheat but not with the eudicot species. The gene *mat-r* is found in the domain IV of the group II intron4 of *nad1*. This intron is found in cis- (broad bean, *Oenethera*) and trans- (wheat, rice, maize) configuration and when in trans- the intron is split before (wheat) or after (potato) domain IV (Begu *et al.* 1998), which places *mat-r* in the half intron either upstream of *nad1e* or downstream of *nad1d*. A previous study of the *nad1d* and *nad5c* (trans exons 4 and 3 of the

Figure 5.2: MultiPipmaker alignment of the rpl2-orfX*(*mttB*)*-nad4 *locus in Arabidopsis mtDNA with the Arabidopsis, Brassica, tobacco, sugarbeet, maize, rice and wheat mitochondrial genomes*

Lengths of sequences are given below in kilobases (kb). Input sequence is the 24 kb region around the *rpl2-orfX(mttB)-nad4* locus from the *Arabidopsis thaliana* mitochondrial genome (NC_001284) and query sequences are the seven complete mitochondrial genomes (accession numbers found in table 2.2). Filled boxes represent genes, small grey boxes are CpG islands and similarity is scored from 50% to 90% on the vertical axis.


rpl2
orfX
nad4
Arabidopsis mtDNA
Arabidopsis
Brassica
Tobacco
Sugarbeet
Maize
Rice
Wheat
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
Arabidopsis
Brassica
Tobacco
Sugarbeet
Maize
Rice
Wheat
20k
22k
24k

nad1 and *nad5* genes of NADH complex, respectively) transcript profile characterized two promoters: one at -44 upstream of *mat-r* and one at +1404 in the *mat-r* coding sequence (Farre and Araya, 1999) demonstrating the co-transcription of these three coding-sequences.

The percentage identity for the 9833 bp *ccmFn-rps1-mat-r-nad1e-nad5c* conserved region is 95% between maize, rice and wheat (98% between rice and wheat) including the coding sequences and 94% between the three plants (96% between rice and wheat) in the 3.1 kb *rps1-mat-r* intergenic region (Figure 5.3). This linkage was first noted by the mitochondrial genome sequencing of the maize genome (Clifton *et al.* 2004) and comparison with the mitochondrial genomes of wheat and rice. The linkage conservation extends to rye, oats, barley, lolium and orchard grass species as detected by PCR analysis (Figure 5.4). Linkage in brome however seems to have been disrupted (Figure 5.4).

This unusual level of conservation might prove of interest in the search for novel non-coding RNAs or in the comprehension of the function and role of intergenic sequences in genome expression (Figure 5.3). The maintenance of the linkage might be due to the importance of this intergenic region, whether to achieve correct expression of this region or because of the presence of non-coding RNAs. Alternatively, this region might simply be less prone to rearrangements. The G/C composition of the 3132 bp sequence between *rps1* and *mat-r* is 45%, slightly lower than the average of ~44% G/C usually found in flowering plant mitochondrial genomes (Chaw *et al.* 2008), and does not exhibit a pronounced nucleotide composition bias.

5.5 Conservation of an intergenic region between wheat, rice and maize

In preliminary attempts to characterize the expression of *rps1-mat-r* intergenic region, RT-PCR and northern analysis have been done to look for a longer transcriptional unit, perhaps even one containing all 5 coding sequences in wheat mtRNA. Co-transcription of *mat-r* and *nad1e* and *nad5c* has been previously described (Farré and Araya, 1999) and RT-PCR was performed on 24hr wheat mitochondrial RNA using an antisense oligomer in the *mat-r* coding sequence and different primer pairs spanning the *rps1-mat-r* intergenic region up to *rps1* (Figure 5.5). Wheat mt RNA from 24hr imbibed embryos was used since in earlier developmental times precursor transcripts might be present if not processed yet. DNA contamination was unlikely as the control samples without RT showed no products (Figure 5.5) but RT-PCR results must be taken with caution nonetheless, as only sequencing of the

Figure 5.3: MultiPipmaker alignment of the ccmFN-rps1-mat-r-nad1e-nad5c, rpl5-Ψrps14-cox3-tRNAE and atp6-rps13-nad1bc loci in wheat mtDNA with the wheat, rice and maize mitochondrial genomes

Lengths of sequences are given below in kilobases (kb). Input sequence is wheat mitochondrial genome (AP008982) and query sequences are wheat mitochondrial genome, rice mitochondrial genome (BA000029) and maize mitochondrial genome (AY506529). Filled boxes represent genes, grey boxes are CpG islands and sequence similarity is scored from 50% to 90% on the vertical axis.

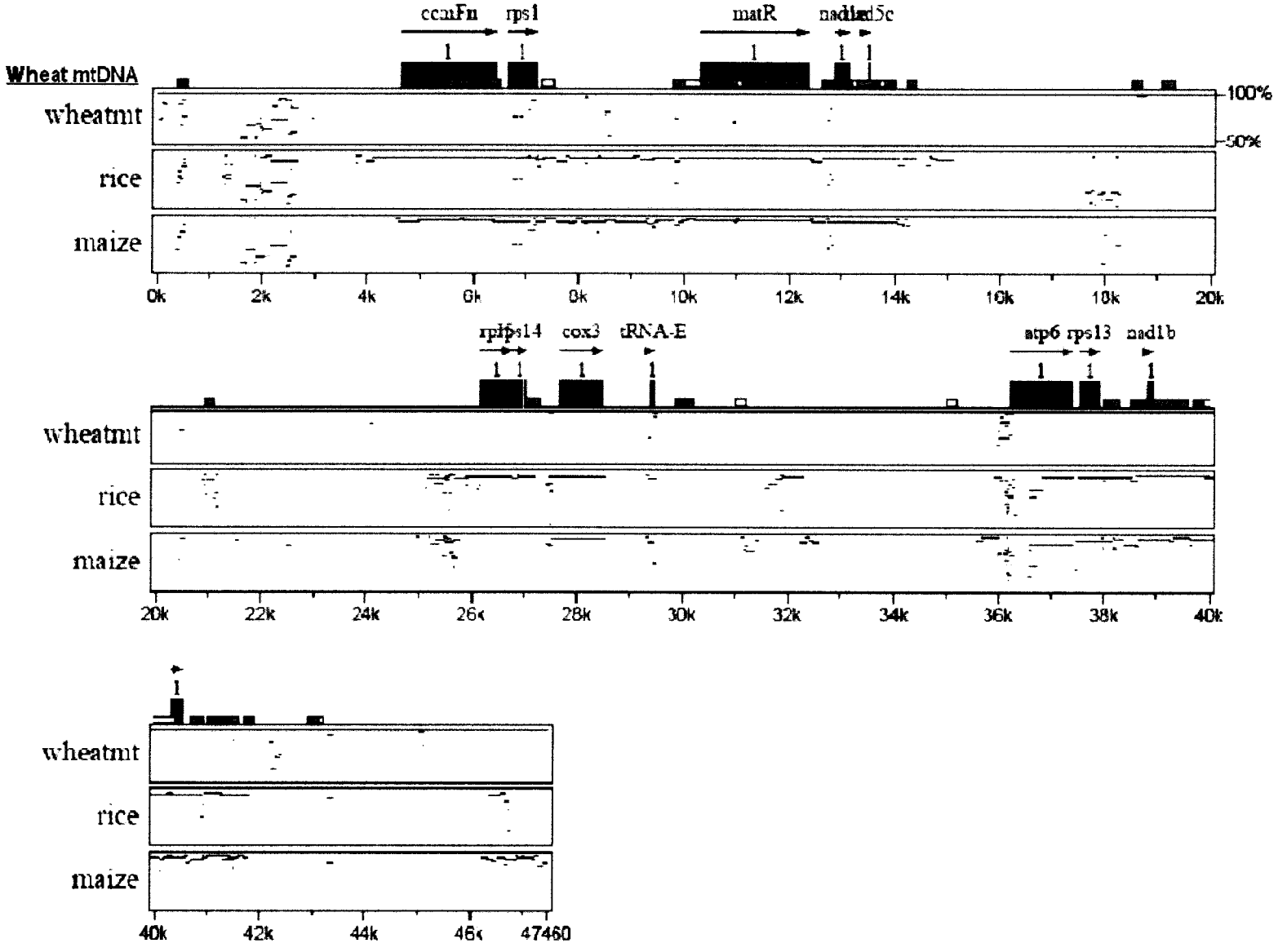
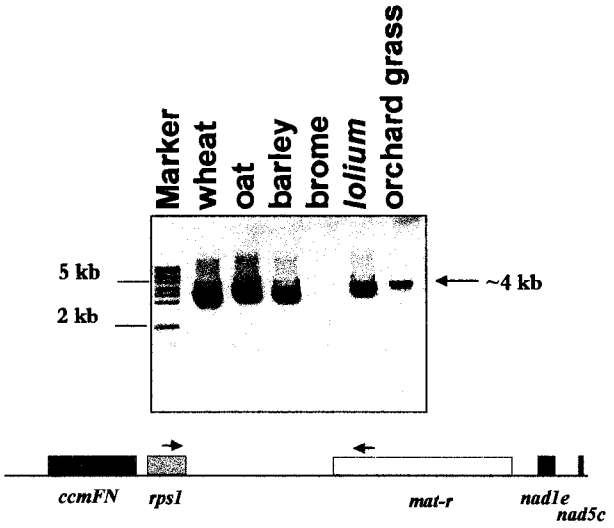
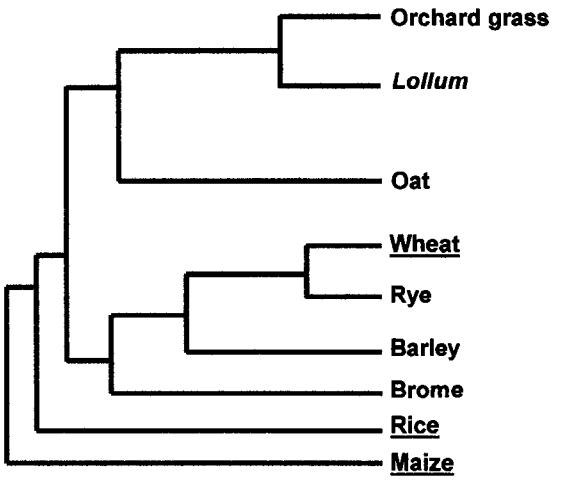

Wheat mtDNA
wheatmt
rice
maize
ccmFn
rps1
matR
nad1
nad5c
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
rpl5
rps14
cox3
tRNA-E
atp6
rps13
nad1b
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
40k
42k
44k
46k
47460

Figure 5.4: PCR analysis of the rps1-mat-r *linkage in cereal species*

Agarose gel analysis of PCR products of wheat, oat, barly, brome, lolium and orchard grass mtDNAs, whose relationship is shown on the phylogenetic tree on the right (Chaw *et al.* 2004). Underlined are species with fully sequenced mitochondrial genomes. Size markers are written in kilobases (kb) on the left and approximate product size on the right. Schematic of gene organization in the wheat mitochondrial genome is show underneath, where filled boxes are genes and arrow heads show the location of the primer used in PCR.


Marker
wheat
oat
barley
brome
lolium
orchard grass
5 kb
2 kb
~4 kb
ccmFN
rps1
mat-r
nad1e
nad5c


Orchard grass
Lolium
Oat
Wheat
Rye
Barley
Brome
Rice
Maize

RT-PCR products and identification of the edit sites can definitely differentiate DNA and RNA (when no intron is present). RT-PCR products were obtained for three primer pairs from the *mat-r* coding sequences up to ~950bp downstream of *rps1*. Primer pairs using a sense oligomer in either *rps1* or *ccmFN* did not give products. This might be due to experimental problems due to the long size of the transcript for which the antisense oligomer RT reaction may not have extended far enough or that biologically there is no transcript including all five coding regions but that transcription is initiated downstream of *rps1*, including at least part of the intergenic region and the downstream genes. In accordance with the relaxed transcriptional system present in plant mitochondria, it is also possible such a transcript represents a low amount of *ccmFN-rps1* co-transcripts that did not terminate downstream of *rps1* but does not necessarily represent a biologically meaningful transcript.

Northern analyses were also done on 24hr and 6d wheat mtRNA (Figure 5.6). A series of long (~1kb) random labelled probes were used together in northern hybridization experiments (Figure 5.6). If long co-transcripts exist for this region their size would exceed 9.8kb and although they would not be accurately sized by northern blotting, they might appear as detectable signals by hybridization if abundant enough. No such transcript was reproducibly found in these preliminary experiments. A long co-transcript might however be unstable or present at very low-levels in the developmental times used here. Control for the northern hybridization was done with a *ccmFN* end-labelled oligomer, however a control for the presence of a long co-transcript (such as the ~9.5 kb *nad4* precursor, N. Niknejad M.Sc. thesis) should also be done to verify the integrity of the northern blot.

A conserved intergenic sequence might also give rise to the transcription of non-coding RNAs. The presence of small non-coding RNAs have long been speculated (Bonnet *et al.* 2006) and recent work using *Arabidopsis* mutants deficient in the PNPase enzyme, involved in RNA degradation (Kuhn *et al.* 2001) has shown the accumulation of seemingly non-coding regions of the mitochondrial genome. The authors hypothesized that transcription can initiate at various locations in the genome and give rise to transcripts which under normal conditions would be quickly turned-over. Such an example is a 500 nt RNA species that accumulated in PNP mutants and demonstrated signs of editing (Kuhn *et al.* 2001). In the search for non-coding RNA species arising from this intergenic region in wheat, northern hybridizations with random-labelled probes spanning the whole 3 kb intergenic region were

Figure 5.5: RT-PCR analysis of the ccmFN-rps1-mat-r-nad1e-nad5c *locus in DNaseI treated wheat 24hr mtRNA*

Filled boxes represent genes, bent arrows are promoter regions (based on this study and Farre and Araya, 1999) and arrow heads show the location of the oligomer used in the RT (black) and PCR (grey) reactions, with predicted products shown underneath (1, 2, 3). Size marker (M) is shown in the right. Samples were treated with (+) or without (-) reverse transcriptase, with no products in the - RT lanes.

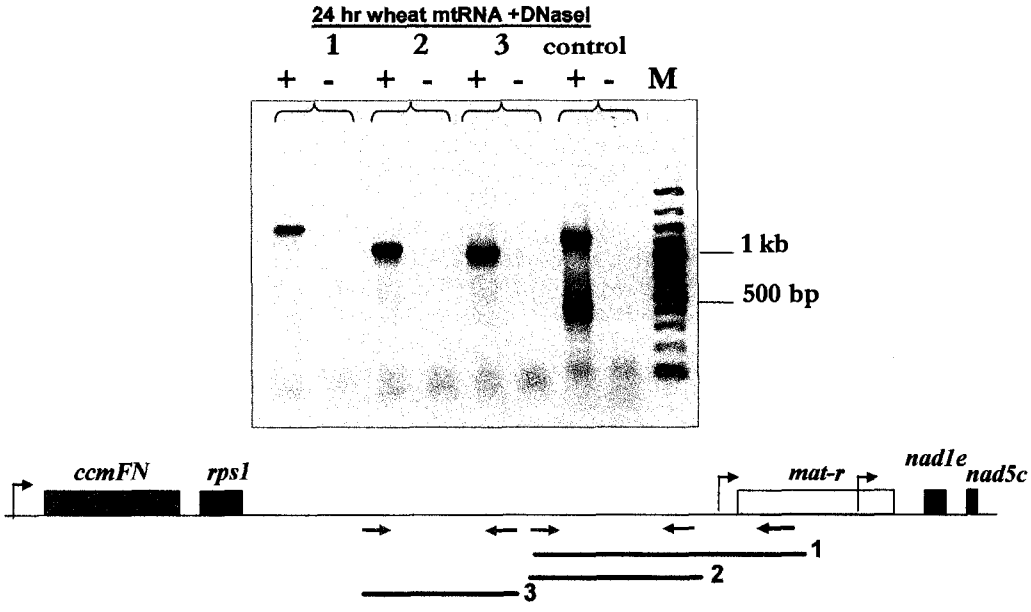
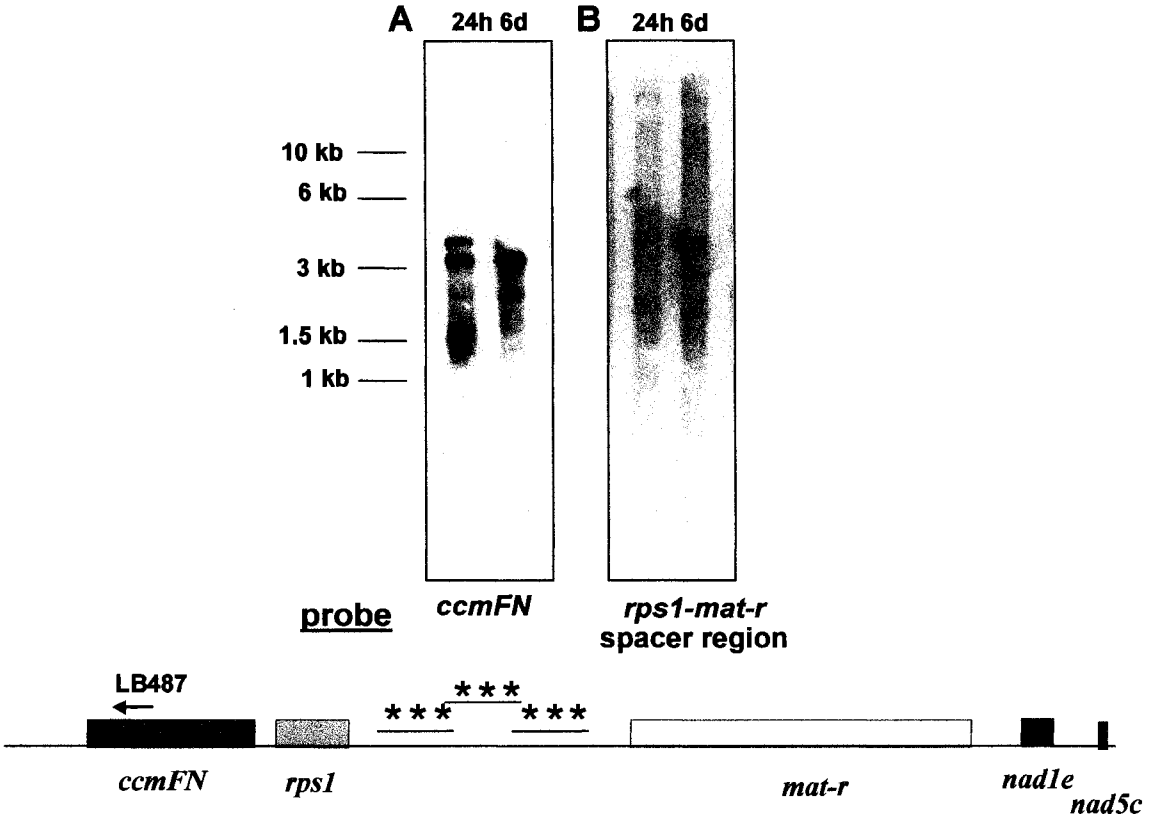

24 hr wheat mtRNA +DNaseI
1
2
3
control
+ - + - + - + - M
1 kb
500 bp
ccmFN
rps1
mat-r
nad1e
nad5c
1
2
3

Figure 5.6: Northern blot analysis of ccmFN and the rps1-mat-r intergenic region in 24hr and 6d wheat mtRNA

Northern blot was successively probes with a *ccmFN* oligomer probe (A) and a pool of three random-labelled PCR probes spanning the *rps1-mat-r* intergenic region (B). Size markers in kilobases (kb) are on the left. Position of the end-labelled oligomer (arrow head) and random-labelled (line with asterisks) probes used are given underneath in the schematic of the loci in the wheat genome, where filled boxes represent genes.


A
24h 6d
B
24h 6d
10 kb
6 kb
3 kb
1.5 kb
1 kb
probe
ccmFN
rps1-mat-r
spacer region
LB487
ccmFN
rps1
mat-r
nad1e
nad5c

also analyzed for small-sized transcripts (Figure 5.6). There was no indication of small-sized transcripts in the northern used here but other RNA gel systems such as polyacrylamide denaturing gel would be more sensitive for small-sized transcripts and give better separation.

Another experimental approach is the *in organello* transcription run-on method, which would detect transcription initiation in this region (Giege *et al.* 2000). Purified mitochondria are incubated with radioactive UTP (usually UTP is used since only RNA being synthesized will incorporate it) which will be incorporated in any newly transcribed RNA. This pool of labeled RNA is then used as a probe on a Southern blot of known DNA regions or in nuclease protection assays. An attempt of this technique was done using a Southern hybridization of clones for the region of interest (*ccmFN-rps1-mat-r-nad1e-na5c*) as well as the control (*18Srrn-5Srrn*) and the *rpl5-rps14-cox3* region. Since rRNAs genes have strong promoters and their transcripts represent up to 90% of the RNA present in plant mitochondria, a strong signal was expected for the control and a weaker signal would be expected for all other transcription initiation sites. Weak signals on the Southern hybridization were seen for the rRNA genes (data not shown) but no signals were seen for other regions. Although the positive control did give results agreeing with previous work (P. Garber, previous Bonen lab member, unpublished results), the low level of signal seen could be due to a low amount of mitochondria isolated and hence low levels of RNA extracted. Attempts with more embryos and more intact mitochondria should be done. The labelled, *in vivo* newly transcribed extracted RNA could be also be used in nuclease protection assays.

5.6 Concluding remarks

Much is still to be learned about transcriptional and post-transcriptional processes in plant mitochondria. The gene expression system seems to exploit a flexible array of signals and the characterization of transcription initiation and processing sites for different genes in different species in the present study is expanding this set. It would be of interest to study the factors potentially regulating transcription initiation in more detail such as the effect of development, light cycle, environment and stress as well as identifying trans-factors involved in this process.

The recent advances in molecular techniques for plant mitochondrial analysis such as the use of electroporation to introduce DNA sequences into isolated mitochondria and study their expression and the availability of a number of *Arabidopsis* mutants (e.g. CMS mutants,

PPR protein mutants) will be a great help in elucidating more of these processes. Work carried out using these tools during the earliest time of embryo development could allow us to catch intermediate states of RNA processing to further understand these events. Work done using tissue cultures is also of great value in pointing out processing events (Forner *et al.* 2007) and revealing interesting patterns of expression. However caution must be used in interpreting these results as they do not reflect *in planta* situations.

The machinery involved in post-transcriptional RNA processing events is not well known but most components are believed to be nuclear-encoded and imported from the cytosol. Proteins such as members of the PPR family have been implicated in mitochondrial RNA processing events (reviewed in Andres *et al.* 2007) and more interestingly as cytoplasm male sterility (CMS) restorers (Wang *et al.* 2006). Identification of PPR proteins as auxiliary factors for transcript specific post-transcriptional events will greatly help us understand the complex interactions taking place in plant mitochondria. Although much hope is placed in nuclear encoded proteins, the role of mitochondrial small RNAs in processing events should also be investigated, as they have been shown to be of great importance in other systems (Bonnet *et al.* 2006)

Plant mitochondrial genomes display unique features in their gene expression system, including a relaxed transcriptional system where post-transcriptional events like processing as well as RNA stability are important control mechanisms (Giege *et al.* 2000, Holec *et al.* 2006, Kuhn *et al.* 2007). Plant mitochondrial RNA level events are extremely complex and expanding knowledge of gene specific processes is a necessary means by which to understand plant mitochondria gene expression as a whole.

References

Adams KL and Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus, Mol Phylogenet Evol. 29:380

Adams KL, Daley DO, Whelan J, Palmer JD (2002b) Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts, Plant Cell 14:931

Adams KL, Ong HC, Palmer JD (2001) Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus, Mol Biol Evol. 18:2289

Adams KL, Qiu YL, Stoutemyer M, Palmer J (2002a) Punctuated evolution of mitochondrial gene content: high variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm, Proc Natl Acad Sci USA 99: 9905

Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD (1999) Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes, Proc Natl Acad Sci USA 96:13863

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool, J Mol Biol. 215:403

André CP, Walbot V (1995) Pulsed-field gel mapping of maize mitochondrial chromosomes, Mol Gen Genet. 20:247

Andres C, Lurin C, Small ID (2007) The multifarious roles of PPR proteins in plant mitochondrial gene expression, Physiol Plant. 129:14

Aubert D, Bisanz-Seyer C, Herzog M (1992) Mitochondrial *rps14* is a transcribed and edited pseudogene in *Arabidopsis thaliana*, Plant Mol Biol. 20:1169

Bégu D, Mercado A, Farré JC, 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

Bendich AJ (1993) Reaching for the ring: the study of mitochondrial genome structure, Curr Genet. 24:279

Bentolila S, Elliott LE, Hanson MR (2007) Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*, Genetics, In press

Bewley JD (1997) Seed germination and dormancy, Plant Cell 9:1055

Binder S and Brennicke A (2003) Review: promoter, transcription RNA polymerases, 5' and 3 maturation, stability, splicing, editing, A Phil Trans R Soc Lond. 358:181

Binder S, Marchenfeld A, Brennicke A (1996) Regulation of gene expression in plant mitochondria, Plant Mol Biol. 32:303

Bonen L (2008) Cis- and trans-splicing of group II introns in plant mitochondria, Mitochondrion 8:26

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

Bonen L and Vogel J (2001) The ins and outs of group II introns, Trends Genet. 17:322

Bonen L, Chapdelaine Y, Garger P (1992) Wheat mitochondrial protein-coding genes and their expression, In Saskuma T and Kinoshita T (Eds.) Nuclear and organellar genomes of wheat species, Kihana Memorial Foundation, Tokyo, Japan, pp 140-148

Bonnet E, Van de Peer Y, Rouzé P (2006) The small RNA world of plants, New Phytol. 171:451

Boore JL (1999) Animal mitochondrial genomes, Nucleic Acid Res. 27:1767

Browning DF and Busby SJW (2004) The regulation of bacterial transcription initiation, Nat. Rev. Microbiol. 2:57

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

Chase CD (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions, Trends Genet. 23:81

Chaw SM, Chang CC, Chen HL, Li WH (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes, J Mol Evol. 58:424

Chaw SM, Shih ACC, Wang D, Wu YW, Liu SM, Chou TY (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites, Mol Biol Evol. 25:603

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs, Nucleic Acids Res. 31:3497

Choury D, Farré JC, Jordana X, Araya A (2004) Different patterns in the recognition of editing sites in plant mitochondria, Nucleic Acids Res. 32:6397

Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, Meyer L, Wilson RK, Newton KJ (2004) Sequence and comparative analysis of the maize NB mitochondrial genome, Plant Physiol. 136:3486

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

de Longevialle AF, Meyer EH, Andrés C, Taylor NL, Lurin C, Millar AH, Small ID (2007) The Pentatricopeptide Repeat Gene OTP43 Is Required for trans-Splicing of the Mitochondrial *nad1 Intron 1* in *Arabidopsis thaliana*, Plant Cell, 19:3256

Dombrowski S, Brennicke A, Binder S (1997) 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators, EMBO J. 16:5069

Eubel H, Heinemeyer J, Sunderhaus S, Braun HP (2004)7 Respiratory chain supercomplexes in plant mitochondria, Plant Physiol Biochem. 42: 93

Farré JC, 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

Farré JC, Choury D, Araya A (2007) *In organello* gene expression and RNA editing studies by electroporation-mediated transformation of isolated plant mitochondria, Methods Enzymol. 424:483

Fey J, Maréchal-Drouard L (1999) Compilation and analysis of plant mitochondrial promoter sequences: An illustration of a divergent evolution between monocot and dicot mitochondria, Biochem Biophys Res Commun. 256:409

Figueroa P, Gomez I, Holuigue L, Araya A, Jordana X (1999) Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulfur subunit of succinate dehydrogenase and expression by alternative splicing, The Plant Journal 18:601

Forner J, Weber B, Thuss S, Wildum S, Binder S (2007) Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5' and 3' end formation, Nucleic Acids Res. 35:3676

Gagliardi D and Binder S (2007) Expression of the plant mitochondrial genome, In Logan D (ed.) Plant mitochondria. Blackwell Publishing, Ames, IA, USA, pp 50-96

Gagliardi D and Leaver CJ (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower, EMBO J. 18:3757

Gagliardi D, Stepien PP, Temperley RJ, Lightowlers RN, Chrzanowska-Lightowlers ZM (2004) Messenger RNA stability in mitochondria: different means to an end, Trends Genet. 20:260

Gaut BS (2002) Evolutionary dynamics of grass genomes, Plant Phyt. 154:15

Geddy R and Brown GG (2007) Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection, BMC Genomics 8:130

GeneBee Molecular Biology Server, http://www.genebee.msu.su/genebee.html (for RNA secondary structure prediction program).

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

Giegé P, Hoffmann M, Binder S, Brennicke A (2000) RNA degradation buffers asymmetries of transcription in *Arabidopsis* mitochondria, EMBO Rep. 1:164

Gonzalez DH, Bonnard G, Grienenberger JM (1993) A gene involved in the biogenesis of c-type cytochromes is co-transcribed with a ribosomal protein gene in wheat mitochondria, Curr Genet. 24:248

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution, Science 283:1476

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

Gualberto JM, Domon C, Weil JH. Grienenberger JM (1990a) Structure and transcription of the gene coding for subunit 3 of cytochrome oxidase in wheat mitochondria, Curr Genet. 17:41

Gualberto JM, Weil JH, Grienenberger JM (1990b) Editing of the wheat *coxIII* transcript: evidence for twelve C to U conversions and for sequence similarities around editing sites, Nucleic Acid Res. 18:3771

Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus L.*): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana, Nucleic Acids Res. 31:5907

Hanic-Joyce PJ, Spencer DF, Gray MW (1990) *In vitro* processing of transcripts containing novel tRNA-like sequences ('t-elements') encoded by wheat mitochondrial DNA, Plant Mol Biol. 15:551

Hazle T and Bonen L (2007a) Status of genes encoding the mitochondrial S1 ribosomal protein in closely-related legumes, Gene 405:108

Hazle T and Bonen L (2007b) Comparative analysis of sequences preceding protein-coding mitochondrial genes in flowering plants, Mol Biol Evol. 24:1101

Hedtke B, Börner T, Weihe A (1997) Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*, Science 277:809

Hedtke B, Börner T, Weihe A (2000) One RNA polymerase serving two genomes, EMBO Rep. 1:435

Herrin DL, Nickelsen J (2004) Chloroplast RNA processing and stability, Photosynth Res. 82:301

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M., Schuster P (1994) Fast folding and comparison of RNA secondary structures, Monatsh. Chem. 125: 167, http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

Hoffmann M, Kuhn J, Däschner K, Binder S (2001) The RNA world of plant mitochondria, Prog Nucleic Acid Res Mol Biol.70:119

Holec S, Lange H, Kühn K, Alioua M, Börner T, Gagliardi D (2006) Relaxed transcription in *Arabidopsis* mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase, Mol Cell Biol. 26:2869

Howell KA, Millar AH, Whelan J (2006) Ordered assembly of mitochondria during rice germination begins with pro-mitochondrial structures rich in components of the protein import apparatus, Plant Mol Biol. 60:201

Ikeda TM and Gray MW (1999a) Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat, Plant Mol Biol. 40:567

Ikeda TM and Gray MW (1999b) Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria, Mol Cell Biol. 19:8113

Khanam SM, Naydenov NG, Kadowaki K, Nakamura C (2007) Mitochondrial biogenesis as revealed by mitochondrial transcript profiles during germination and early seedling growth in wheat, Genes Genet Syst. 82:409

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective, Curr Genet. 46:123

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (*Brassicaceae*), Mol Biol Evol. 17:1483

Kubo N, Salomon B, Komatsuda T, von Bothmer R, Kadowaki K (2005) Structural and distributional variation of mitochondrial *rps2* genes in the tribe *Triticeae* (*Poaceae*), Theor Appl Genet. 110:995

Kubo N. Harada K. Hirai A. Kadawaki K (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins, Proc. Natl. Acad. Sci. USA. 96:9207

Kubo T and Mikami T (2007) Organization and variation of angiosperm mitochondrial genome, Physiol Plant. 129:6

Kubo T and Newton KJ (2008) Angiosperm mitochondrial genomes and mutations, Mitochondrion 8:5

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris L.*) reveals a novel gene for tRNA(Cys)(GCA), Nucleic Acids Res. 28:2571

Kühn J, Binder S (2002) RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria, Nucleic Acids Res. 30:439

Kühn J, Tengler U, Binder S (2001) Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria, Mol Cell Biol. 21:731

Kühn K, Bohne AV, Liere K, Weihe A, Börner T (2007) *Arabidopsis* phage-type RNA polymerases: accurate *in vitro* transcription of organellar genes, Plant Cell19:959

Kühn K, Weihe A, Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*, Nucleic Acids Res. 33:337

Li-Pook-Than J and Bonen L (2006) Multiple physical forms of excised group II intron RNAs in wheat mitochondria, Nucleic Acids Res. 34:2782

Li-Pook-Than J, Carrillo C, Niknejad N, Calixte S, Crosthwait J, Bonen L (2007) Relationship between RNA splicing and exon editing near intron junctions in wheat mitochondria, Physiol Plant.129:23

Li-Pook-Than J, Carillo C, Bonen L (2004) Variation in mitochondrial transcript profiles of protein-coding genes during early germination and seedling development in wheat, Curr Genet. 46 :374

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

Logan DC, Millar AH, Sweetlove LJ, Hill SA, Leaver CJ (2001)Mitochondrial biogenesis during germination in maize embryos, Plant Physiol. 125:662

Lupold DS, Caoile AG, Stern DB (1999a) Genomic context influences the activity of maize mitochondrial *cox2* promoters, Proc Natl Acad Sci U S A. 96:11670

Lupold DS, Caoile AG, Stern DB (1999b) Polyadenylation occurs at multiple sites in maize mitochondrial *cox2* mRNA and is independent of editing status, Plant Cell. 11:1565

Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, Plant Cell 16:2089

Maier RM, Zeltz P, Kossel H, Bonnard G, Gualberto JM, Grieneneberger JM (1996) RNA editing in plant mitochondria and chloroplasts, Plant Mol Biol. 32:343

Malecki M, Jedrzejczak R, Stepien PP, Golik P (2007) *In vitro* reconstitution and characterization of the yeast mitochondrial degradosome complex unravels tight functional interdependence, J Mol Biol. 372:23

Marintchev A and Wagner G (2004) Translation initiation: structures, mechanisms and evolution, Q Rev Biophys. 37:197

Martin W and Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization, Nature, 440:41

Mower JP (2007) Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria, Mol Biol Evol. In press.

Mower JP and Palmer JD (2006) Patterns of partial RNA editing in mitochondrial genes *of Beta vulgaris*, Mol Genet Genomics. 276:285

Mulligan RM, Chang KL, Chou CC (2007) Computational analysis of RNA editing sites in plant mitochondrial genomes reveals similar information content and a sporadic distribution of editing sites, Mol Biol Evol. 24:1971

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

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

Mulligan RM, Maloney AP, Walbot V (1988a) RNA processing and multiple transcription initiation sites result in transcript size heterogeneity in maize mitochondria, Mol Gen Genet. 211:373

Mundel C and Schuster W (1996) Loss of RNA editing of *rps1* sequences in *Oenothera* mitochondria, Curr Genet. 30:455

Nakagawa N and Sakurai N (2006) A mutation in *At-nMat1a*, which encodes a nuclear gene having high similarity to group II intron maturase, causes impaired splicing of mitochondrial NAD4 transcript and altered carbon metabolism in *Arabidopsis thaliana*, Plant Cell Physiol. 47:772

Nakazono M, Nishiwaki S, Tsutsumi N, Hirai A (1996) A chloroplast-derived sequence is utilized as a source of promoter sequences for the gene for subunit 9 of NADH dehydrogenase (*nad9*) in rice mitochondria, Mol Gen Genet. 252:371

Neuwirt J, Takenaka M, van der Merwe JA, Brennicke A (2005) An *in vitro* RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species, RNA 11:1563

Niknejad N (2003) Splicing and editing of unusual group II introns in the wheat mitochondrial *nad4* gene, M.Sc. thesis, University of Ottawa, Ontario, Canada

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002)The complete sequence of the rice (*Oryza sativa L.)* mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants, Mol Genet Genomics 268:434

Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, Takumi S, Murata M, Futo S, Tsunewaki K (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome, Nucleic Acids Res. 33:6235

Okuda K, Myouga F, Motohashi R, Shinozaki K, Shikanai T (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing, Proc Natl Acad Sci USA 104:8178

Ong HC and Palmer JD (2006) Pervasive survival of expressed mitochondrial *rps14* pseudogene in grasses and their relatives for 80 million years following three functional transfers to the nucleus, BMC Evol Biol 6:55

Perrin R, Lange H, Grienenberger JM, Gagliardi D (2004b) *AtmtPNPase* is required for multiple aspects of the *18S rRNA* metabolism in *Arabidopsis thaliana* mitochondria, Nucleic Acids Res. 32:5174

Perrin R, Meyer EH, Zaepfel M, Kim YJ, Mache R, Grienenberger JM, Gualberto JM, Gagliardi D (2004a) Two exoribonucleases act sequentially to process mature 3'-ends of *atp9* mRNAs in *Arabidopsis* mitochondria, J Biol Chem. 279:25440

Quiñones V, Zanlungo S, Moenne A, Gómez I, Holuigue L, Litvak S, Jordana X (1996) The *rpl5-rps14-cob* gene arrangement in *Solanum tuberosum*: *rps14* is a transcribed and unedited pseudogene, Plant Mol Biol. 31:937

Raczynska KD, Le Ret M, Rurek M, Bonnard G, Augustyniak H, Gualberto JM (2006) Plant mitochondrial genes can be expressed from mRNAs lacking stop codons, FEBS Lett. 580:5641

Rogowska AT, Puchta O, Czarnecka AM, Kaniak A, Stepien PP, Golik P (2006) Balance between transcription and RNA degradation is vital for *Saccharomyces cerevisiae* mitochondria: reduced transcription rescues the phenotype of deficient RNA degradation, Mol Biol Cell 17:1184

Salone V, Rüdinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, Szurek B, Small I, Knoop V, Lurin C (2007) A hypothesis on the identification of the editing enzyme in plant organelles, FEBS Lett. 581:4132

Sambrook J. Fritsch E.F. and Maniatis T (1989) Molecular cloning, A laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY, USA.

Sandoval P. León G.Gómez I. Carmona R. Figueroa P. Holuigue L. Araya A. Jordana X (2004) Transfer of RPS14 and RPL5 from the mitochondrion to the nucleus in grasses, Gene 324:139

Sarkar D and Fisher PB (2006) Polynucleotide phosphorylase: an evolutionary conserved gene with an expanding repertoire of functions, Pharmacol Ther. 112:243

Schuster G, Lisitsky I, Klaff P (1999) Polyadenylation and degradation of mRNA in the chloroplast, Plant Physiol. 120:937

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker--a web server for aligning two genomic DNA sequences, Genome Res. 10:577

Shikanai T (2006) RNA editing in plant organelles: machinery, physiological function and evolution, Cell Mol Life Sci. 63:698

Subramanian S, Fallahi M, Bonen L (2001) Truncated and dispersed *rpl2* and *rps19* pseudogenes are co-transcribed with neighbouring downstream genes in wheat mitochondria, Curr Genet. 39:264

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants, Mol Genet Genomics 272:603

Takemura M, Oda K, Yamato K, Ohta E, Nakamura Y, Nozato N, Akashi K, Ohyama K (1992) Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*, Nucleic Acids Res. 20:3199

Toutloff A (2003) Mitochondrial ribosomal protein genes in plants: location and expression of *rpl5* in cereals, 4th biology honors thesis, University of Ottawa, Ontario, Canada

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

Williams MA, Johzuka Y, Mulligan RM (2000) Addition of non-genomically encoded nucleotides to the 3'-terminus of maize mitochondrial mRNAs: truncated *rps12* mRNAs frequently terminate with CCA, Nucleic Acids Res. 28:4444

Wintz H and Hanson MR (1991) A termination codon is created by RNA editing in the petunia *atp9* transcript, Curr Genet. 19:61

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

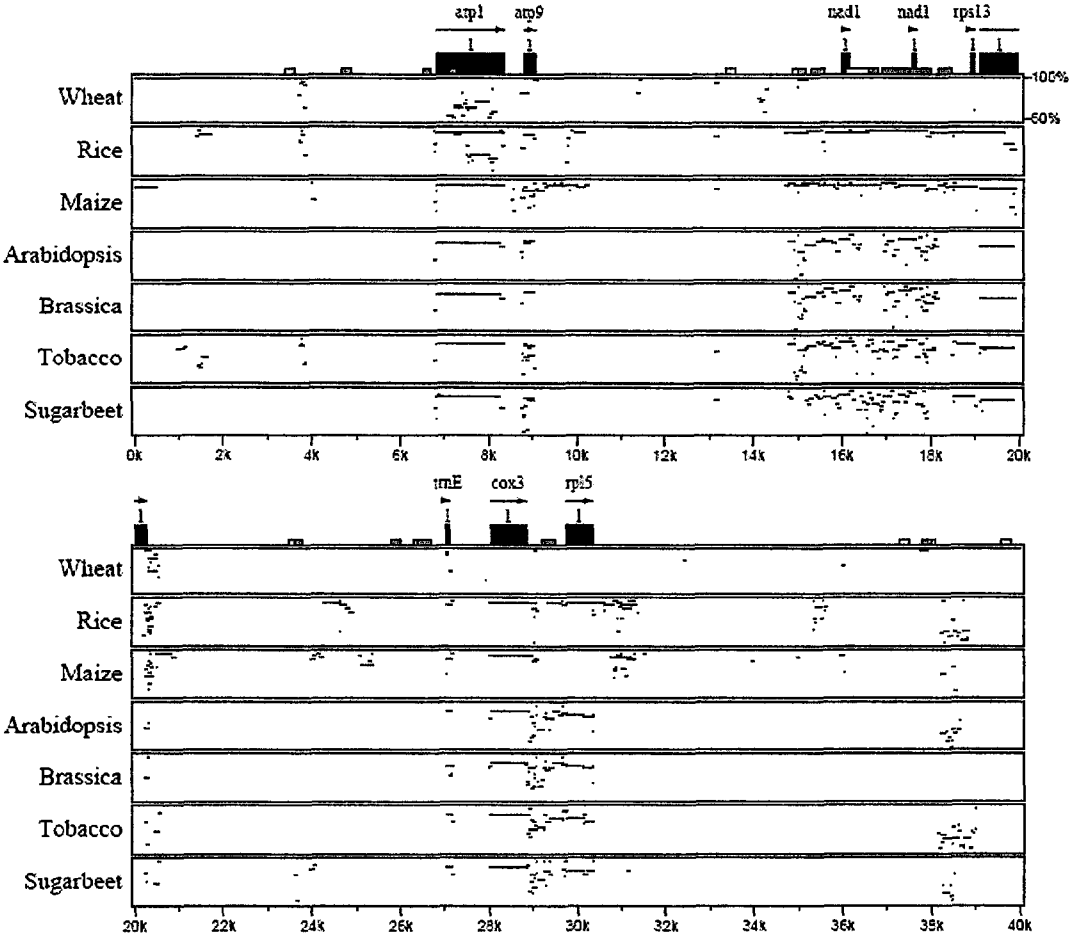
Zhang QY and Liu YG (2006) Rice mitochondrial genes are transcribed by multiple promoters that are highly diverged, J Integrative Plant Biol. 48:1473

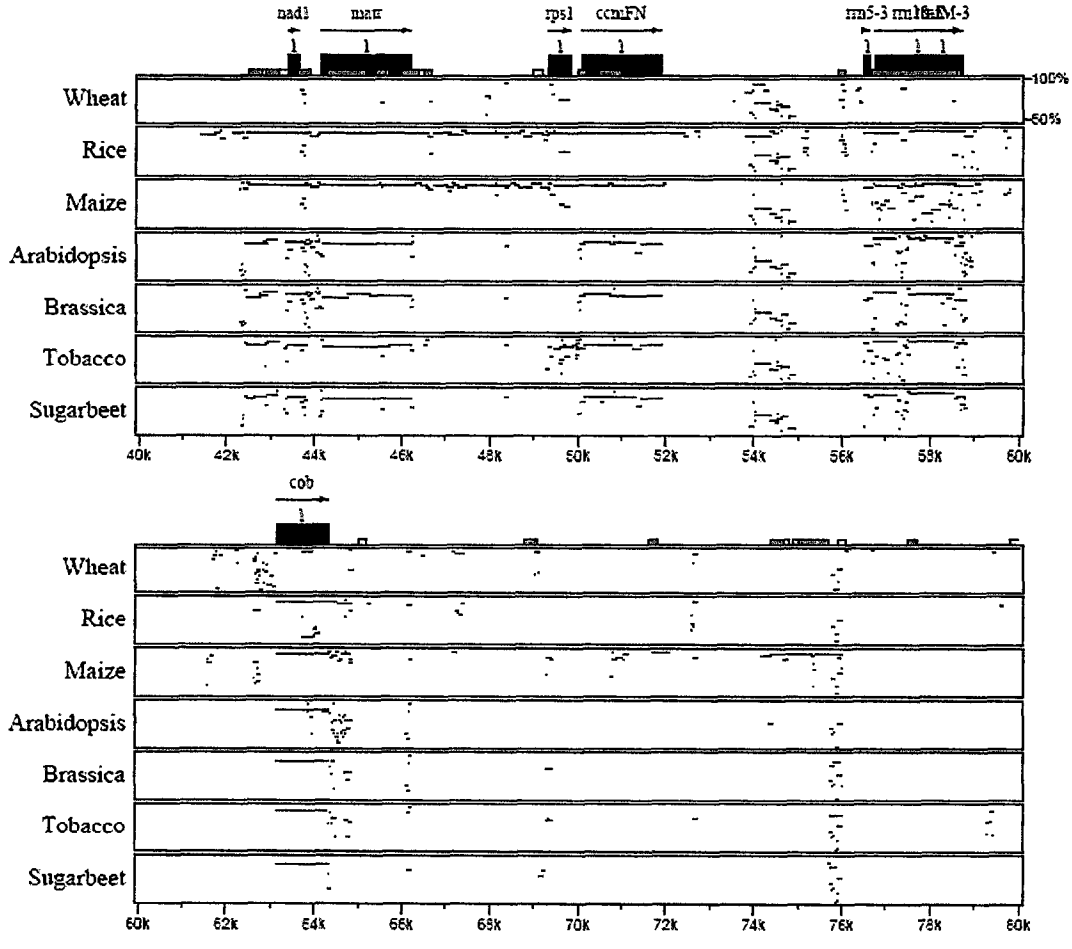
Zuker K (2003) Mfold web server for nucleic acid folding and hybridization prediction, Nuc Acids Res 31: 3406, http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html

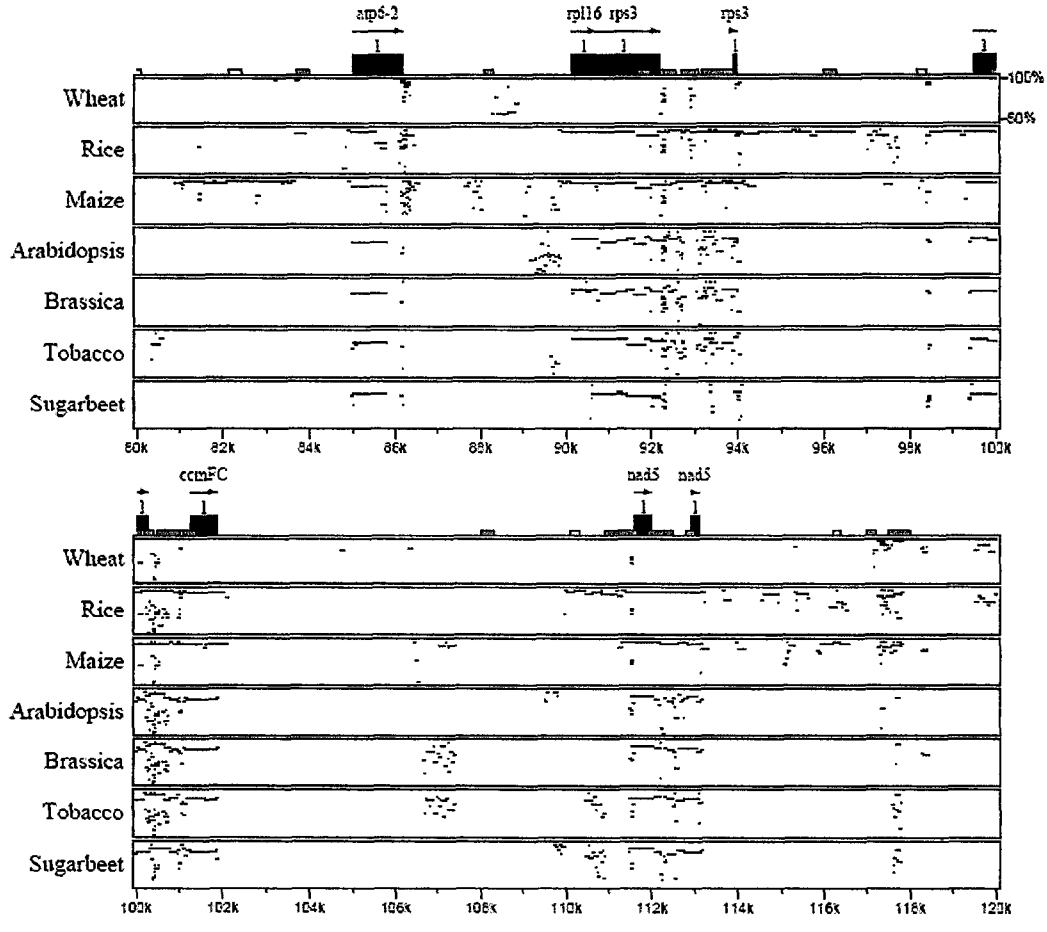
Appendices

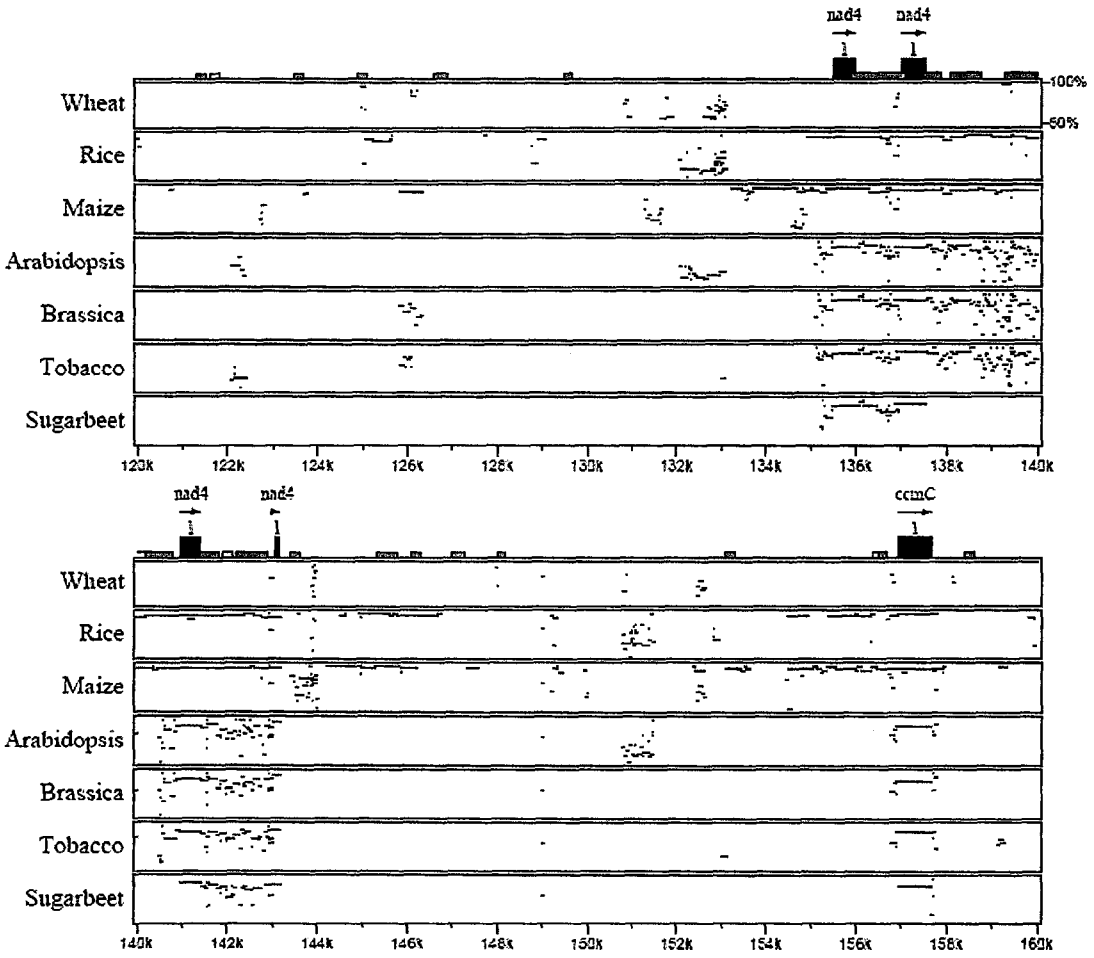
Appendix 1: Multipipmaker alignment of the wheat mitochondrial genome with the seven completely sequenced mitochondrial genomes

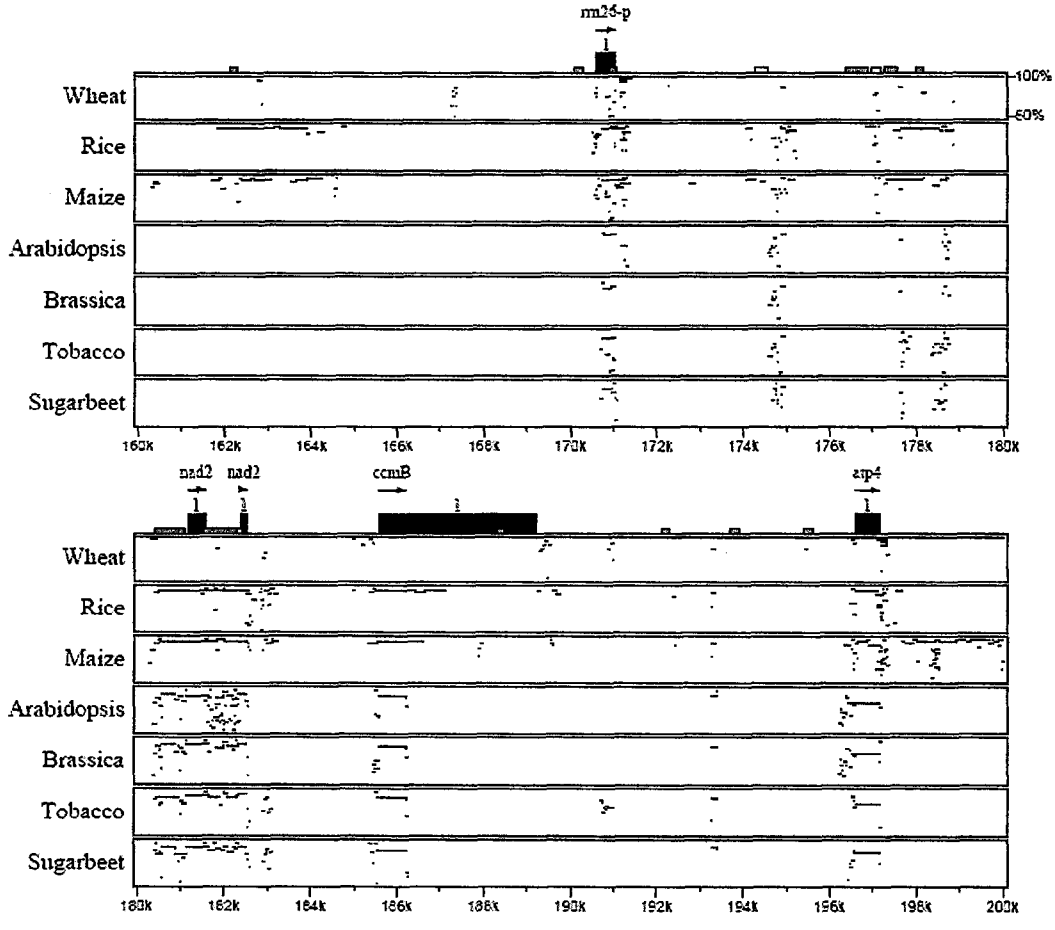
Input sequence is the wheat mitochondrial genome (AP008982, Ogihara *et al.* 2005) and query sequences are wheat (AP008982), rice (BA000029, Notsu *et al.* 2002), maize (AY506529, Clifton *et al.* 2004), *Arabidopsis* (NC_001284, Unseld *et al.* 1997), *Brassica* (AP006444, Handa 2003), tobacco (NC_006581, Sugiyama *et al.* 2005) and sugarbeet (BA000009, Kubo *et al.* 2000). Filled boxes represent genes, with name above and coordinates on the wheat genomes in kilobases (kb) on the horizontal axis. Small grey boxes are CpG islands as defined by the Pipmaker software.

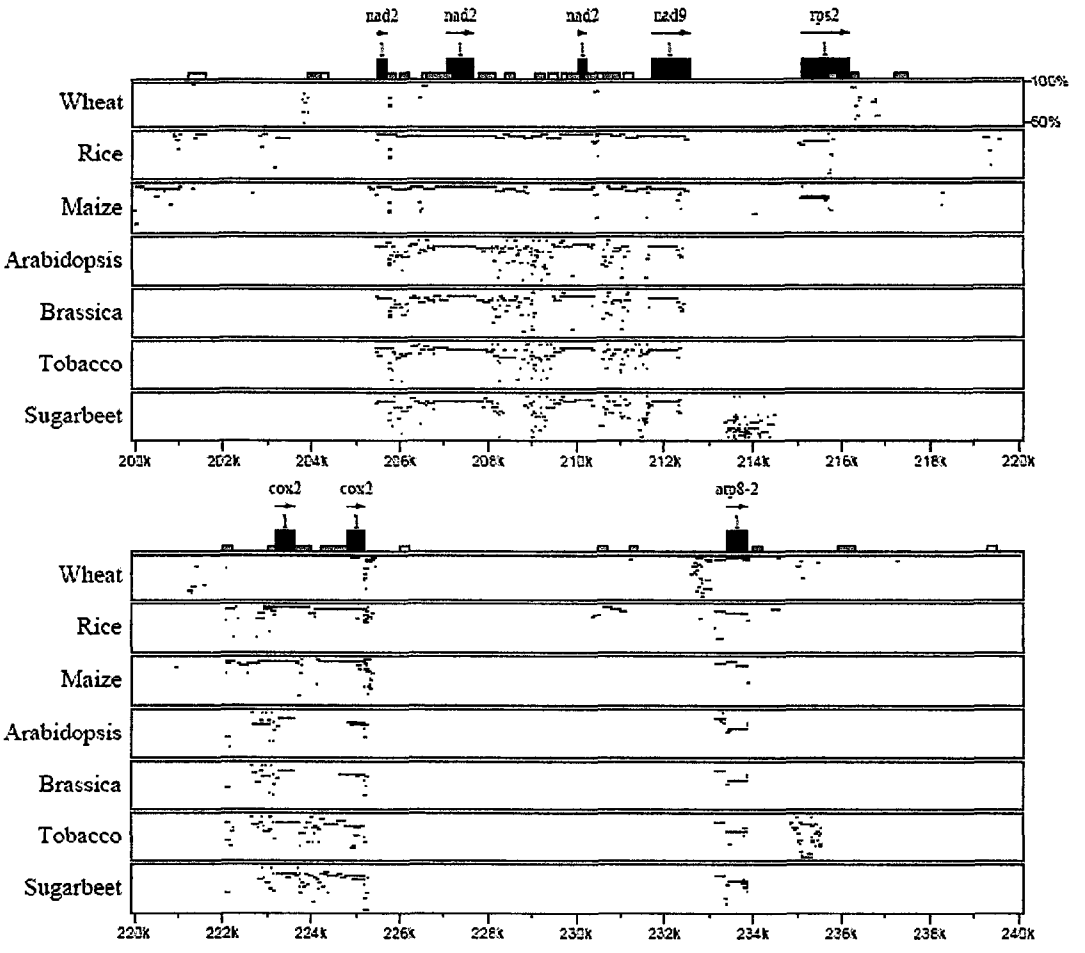


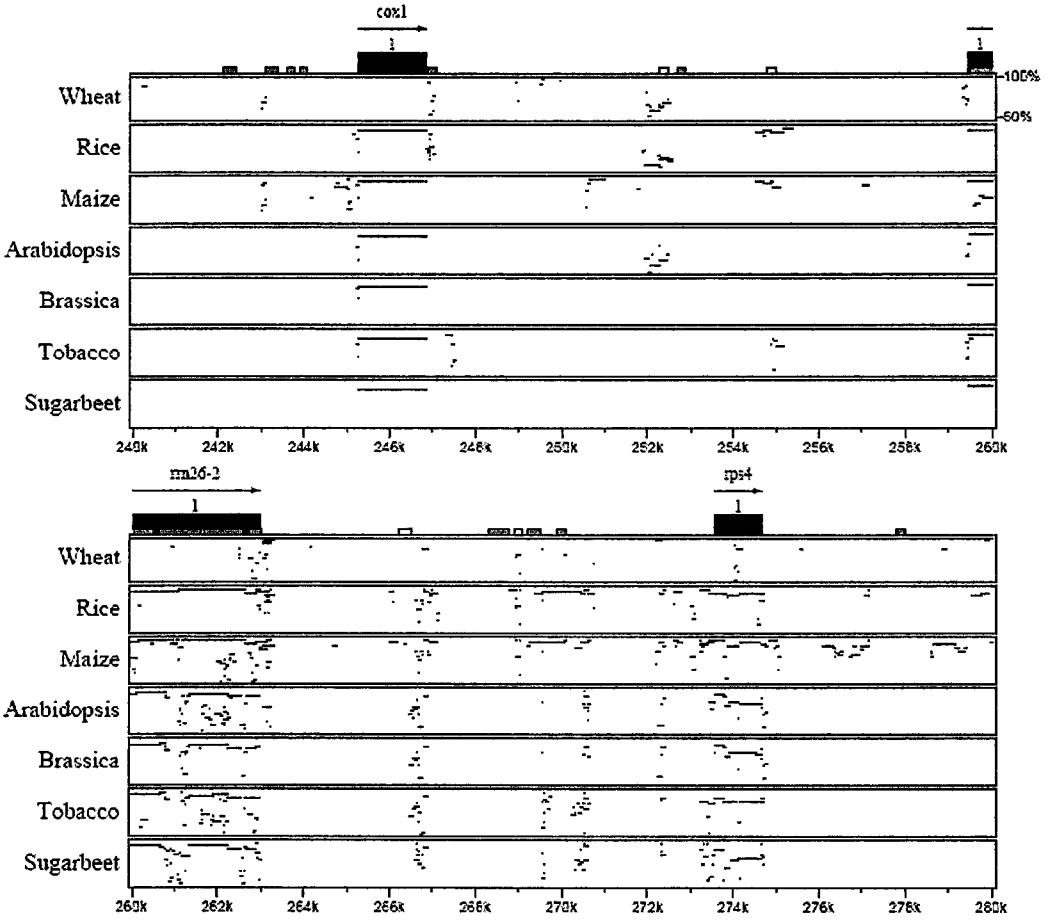

nad1
rps1
ccmFN
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
cob
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k

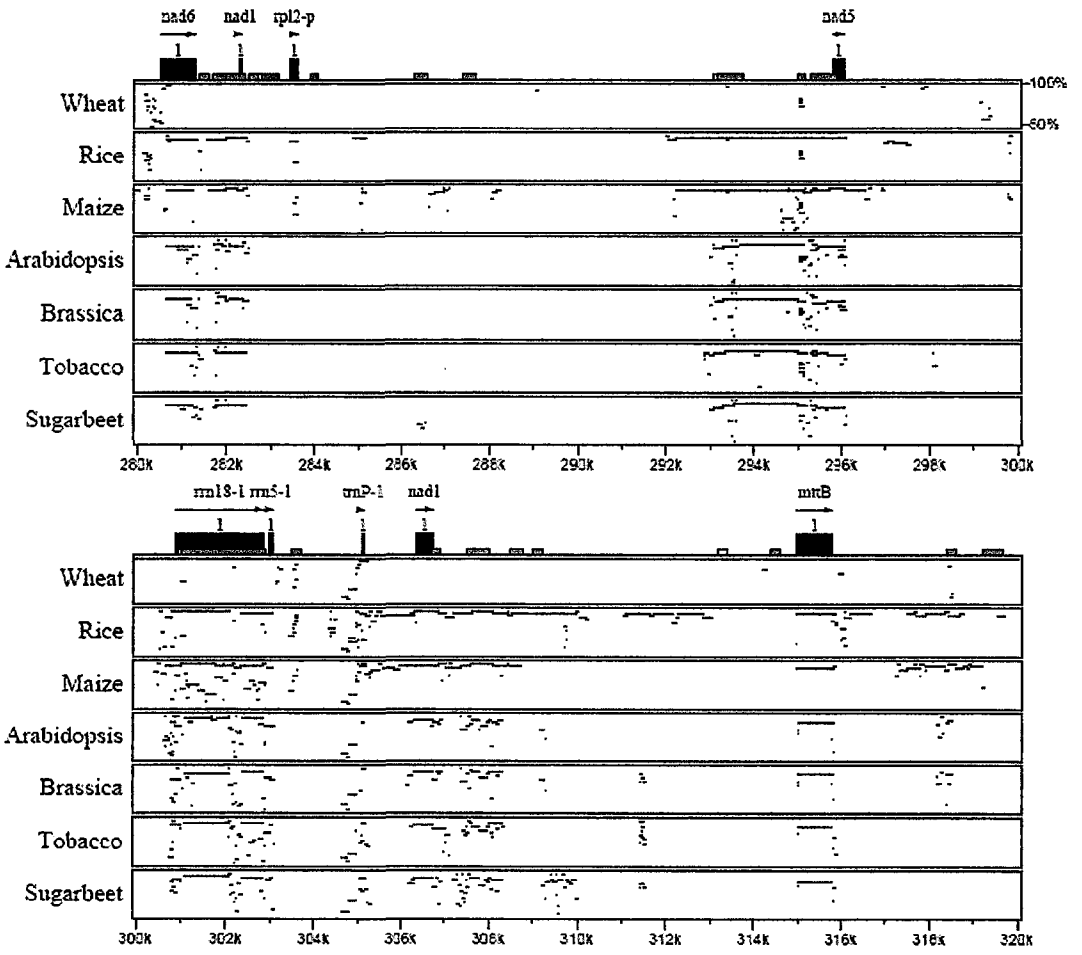

atp6-2
rpl16 rps3
rps3
100%
60%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k
ccmFC
nad5
nad5
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k

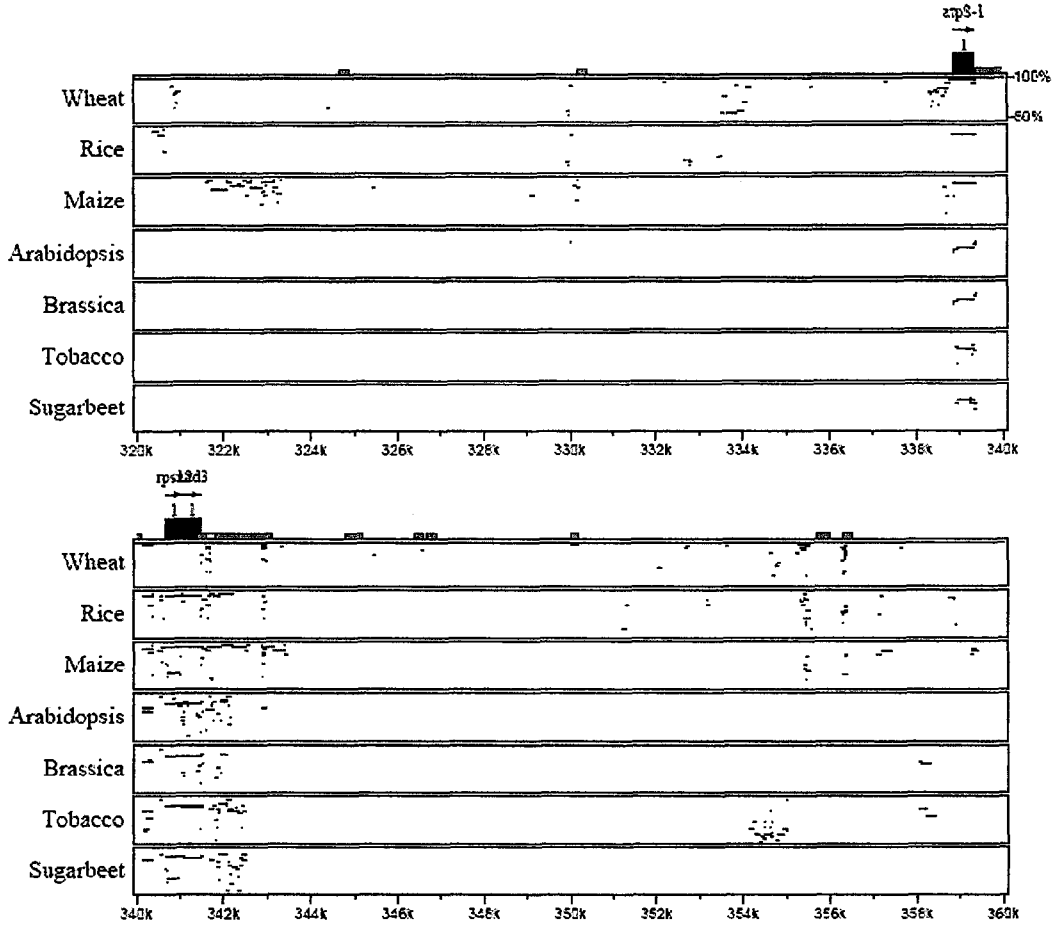

nad4
nad4
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k
nad4
nad4
ccmC
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
140k
142k
144k
146k
148k
150k
152k
154k
156k
158k
160k

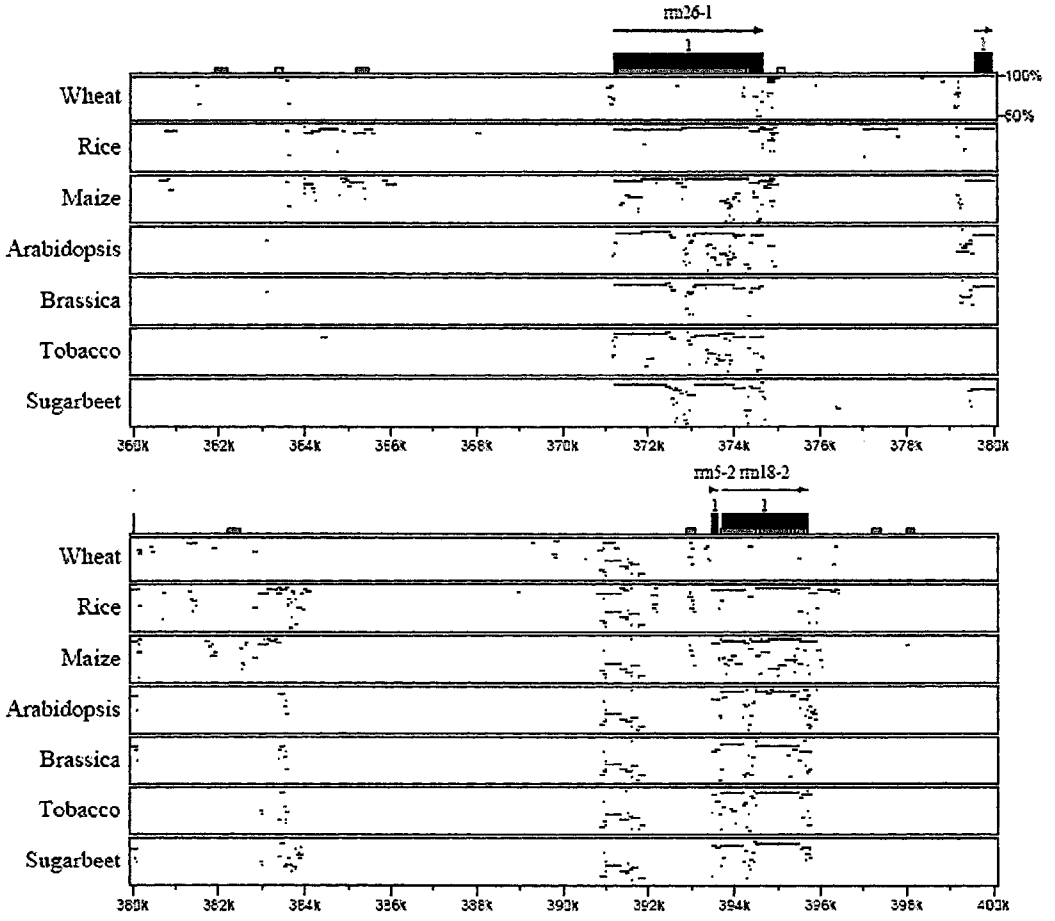

rrn23-p
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
160k
162k
164k
166k
168k
170k
172k
174k
176k
178k
180k
nad2
nad2
ccmB
atp4
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
180k
182k
184k
186k
188k
190k
192k
194k
196k
198k
200k

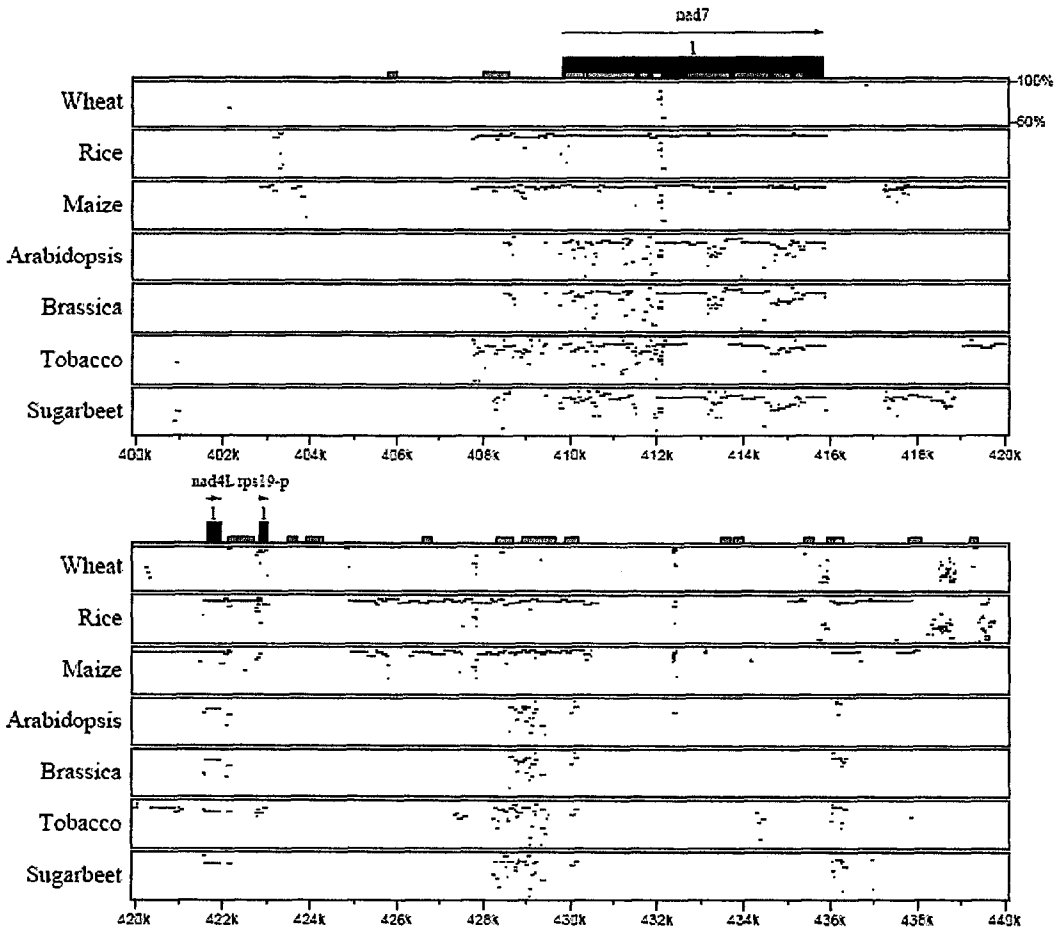

nad2
nad2
nad2
nad9
rps2
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
200k
202k
204k
206k
208k
210k
212k
214k
216k
218k
220k
cox2
cox2
atp8-2
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
220k
222k
224k
226k
228k
230k
232k
234k
236k
238k
240k

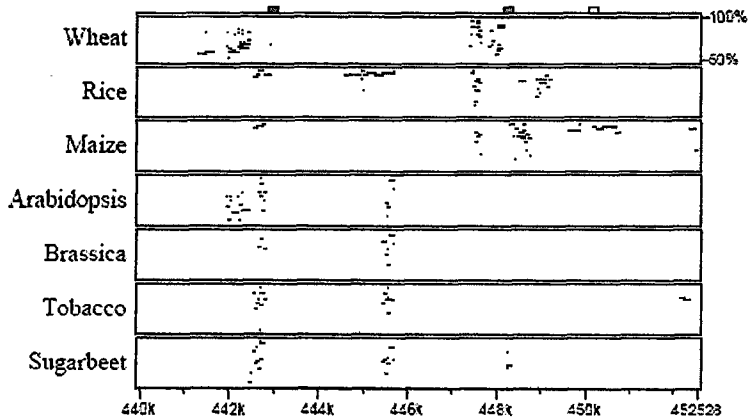

cox1
1
1
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
240k
242k
244k
246k
248k
250k
252k
254k
256k
258k
260k
rrn26-2
rps4
1
1
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
260k
262k
264k
266k
268k
270k
272k
274k
276k
278k
280k


nad6
nad1
rpl2-p
nad5
100%
60%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
280k
282k
284k
286k
288k
290k
292k
294k
296k
298k
300k
rrn18-1 rrn5-1
rnpP-1
nad1
matB
300k
302k
304k
306k
308k
310k
312k
314k
316k
318k
320k


rps8-1
100%
60%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
320k
322k
324k
326k
328k
330k
332k
334k
336k
338k
340k
rps8d3
340k
342k
344k
346k
348k
350k
352k
354k
356k
358k
360k

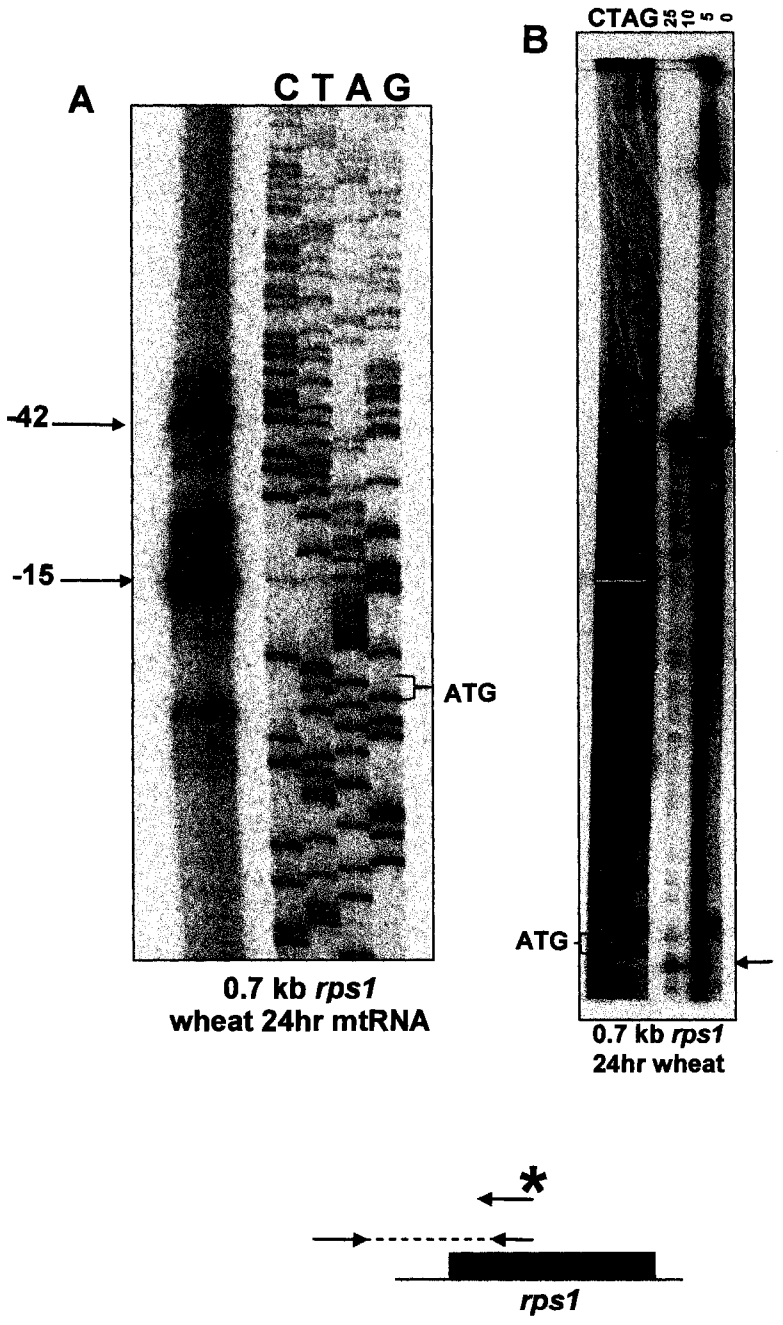

rm26-1
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
360k
362k
364k
366k
368k
370k
372k
374k
376k
378k
380k
rm5-2 rm18-2
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
380k
382k
384k
386k
388k
390k
392k
394k
396k
398k
400k


nad7
100%
50%
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
400k
402k
404k
406k
408k
410k
412k
414k
416k
418k
420k
nad4L rps19-p
Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
420k
422k
424k
426k
428k
430k
432k
434k
436k
438k
440k


Wheat
Rice
Maize
Arabidopsis
Brassica
Tobacco
Sugarbeet
100%
60%
440k
442k
444k
446k
448k
450k
452528

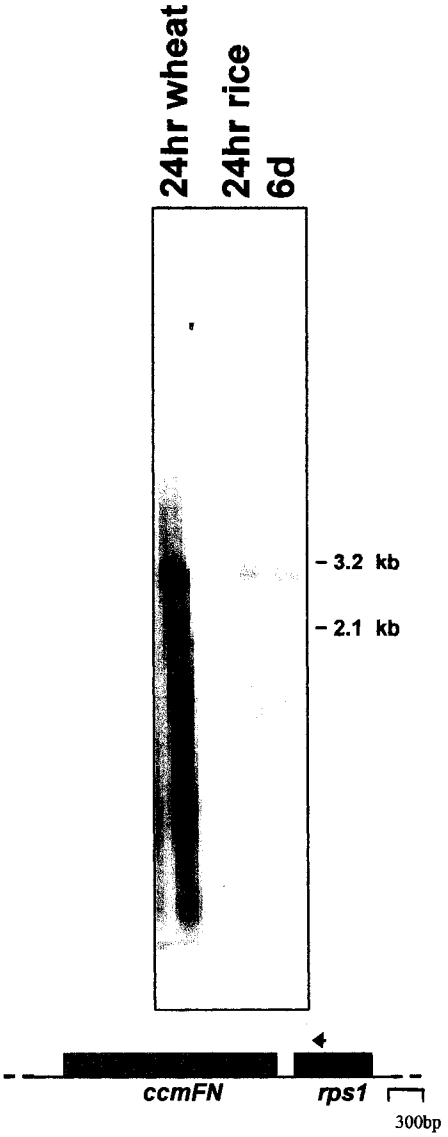
Appendix 2: Primer extension analysis (A) and S1 nuclease protection assay (B) of the rps1 *monocistronic transcript in 24hr wheat mtRNA.*

Filled boxes represent genes, with arrow head showing the location of the labelled oligomer and the labelled PCR probe used in experiment. A) Sizes of extension products relative to the start codon are written in nucleotide on the left and products are shown by arrow head. Sequencing reaction of a *rps1* PCR product is used as a size marker and deoxy-nucleotides used in sequencing reactions at the top.
B) Sizes of protected products relative to the start codon are written in nucleotide on the left and products are shown by arrow head. Sequencing reaction of a *ccmFN* PCR product is used as a size marker and deoxy-nucleotides used in sequencing reactions at the top



Appendix 3: Northern blot analysis of the rps1 *gene in rice mtRNA at 24hr and 6d post-imbibition.*

Filled boxes represent genes, end-labelled *rps1* oligomer used and sizes of putative transcripts are written in kilobases (kb) on the right. Wheat 24hr mtRNA is shown as a control.



Appendix 4: Northern analysis of the rpl5-Ψrps14-cox3 *locus in wheat mtRNA during early seed-to-seedling development (0hr to 6d).*

Filled boxes represent gene and arrow heads denote the position of the end-labelled oligomers used for *rpl5* (A) and *rps14* (B). Sizes of transcripts are written on the right. Mishydridization signals are show with empty arrow heads. Data from J. Li-Pook-Than.