INFORMATION TO USERS

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

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

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

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

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

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

UMI®
800-521-0600

Université d'Ottawa · University of Ottawa

PLANT MITOCHONDRIAL L2 RIBOSOMAL PROTEIN GENES AND THEIR EXPRESSION

Selvi Subramanian

A thesis submitted to the

School of Graduate Studies and Research

University of Ottawa

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in the Ottawa-Carleton Institute of Biology

1999

ABSTRACT

The endosymbiotic hypothesis for the origin of mitochondria invokes the transfer of many genes from the respiring bacterium to the nucleus of the host cell early in the symbiotic association. Recent evidence suggests that such transfer is an ongoing evolutionary process in the plant lineage. This project focuses on the location and expression of genes coding for the mitochondrial (mt) large ribosomal subunit protein L2 in different plants to gain knowledge about their function and evolution. L2 is known to be one of the most conservative ribosomal proteins present in prokaryotes and eukaryotes.

The rice mt genome was found to have a single copy of a functional *rpl2* gene with a group II intron whereas wheat has only a truncated *rpl2* pseudogene corresponding to the extreme carboxyl terminus and no *rpl2* homologous sequences were detected in the soybean mt genome. Nuclear-located transferred mt *rpl2* copies have been identified in wheat and they are found to be transcribed. An RNA-mediated gene transfer is predicted since it lacks the mt intron at the expected splice site. The wheat nuclear-located mt *rpl2* is part of a small multigene family and the two examined copies share 97% identity at the amino acid level. The transferred wheat mt *rpl2* sequences are found to be expressed differentially during development. They have higher transcript accumulation in germinating embryos than seedlings in comparison to the cytosolic *rpl2* mRNA at these stages.

In rice, the mt *rpl2* gene is transcribed, spliced and edited. The mRNA levels of the rice mt *rpl2* are very low compared to typical respiratory chain genes and have higher transcript accumulation in the germinating embryo stage than seedlings. The transcript profile is complex and different between stages indicating different processing pathways in germinating embryos and seedlings. Low steady state *rpl2* mRNA levels compared to excised introns indicate a rapid

turnover of mRNA and a stable excised form of intron. However the excised introns appear to be present as non-lariat forms by RT-PCR experiments. This observation suggests an initial hydrolysis step in the splicing reaction. The rice mt *rpl2-rps19- nad4L* gene cluster is transcribed as a polycistronic molecule and processed to mature mRNAs. In addition, detection of transcribed mt *rpl2* related sequences in the rice nuclear genome by hybridization experiments also suggests a recent gene transfer to the nucleus.

Phylogenetic analyses on derived L2 and S12 protein sequences of mitochondria and other compartments indicate a rapid evolution of mt L2 proteins in comparison to mt S12. The transfer of mt genes to the nucleus points to an accelerated rate of evolution of L2 proteins. The codon usage pattern of the transferred wheat nuclear-located mt *rpl2* sequences is in agreement with an adaptation to the pattern of other wheat nuclear genes. Although plants examined in this study portray various steps involved in a gene transfer event, highly divergent mt *rpl2* sequences in the *Arabidopsis* nuclear genome and the maintenance of a functional gene in the rice mt genome suggest multiple and independent gene transfers.

This study illustrates the recent evolutionary transfer of functional mt genes to the nucleus. It also indicates that the transferred genes may be modified to a certain extent during or after transfer to the new environment. The presence of a mitochondrial as well as a transferred nuclear copy may reflect the transition stage in a gene transfer event or the possibility of maintenance of the mt copy for specialized functional regulation during different developmental stages. The underlying coordinated regulation of mt and nuclear genes to produce a functional mt ribosome is evident from these studies.

ACKNOWLEDGMENTS

I express my sincere thanks to my supervisor Dr. L. Bonen for providing me a laboratory space, financial support and able guidance for a successful accomplishment of my Ph.D. thesis project. I had a very good opportunity to learn most of the plant molecular techniques which will be useful to establish my career advancement in this fast growing field of science. I am also thankful to the Natural Sciences and Engineering Research Council (NSERC) of Canada for providing a research grant to Dr. L. Bonen.

Sincere appreciations are extended to members of my research advisory committee, Drs. D. Johnson, D. Hickey and G. Carmody, for their valuable suggestions throughout the course of my Ph.D. research program.

My special thanks to Karen Williams whose assistance was very helpful in acquiring technical skills. I wish to thank the students of Dr. L. Bonen's lab who provided assistance and companionship in the laboratory with special thanks to Catherine Carrillo, Magid Fallahi and Samir Grover.

Finally, I would like to thank my parents for their moral support and giving me courage to make dreams a reality. I thank my husband Dr. K. S. Subramanian for his constant encouragements and assistance.

TABLE OF CONTENTS

APPENDICES

List of Tables

List of Figures

Abbreviations

AMV	avian myeloblastosis virus
atp	genes for ATP synthase subunits
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
ccb	gene for cytochrome c biogenesis
cox	genes for cytochrome oxidase subunits
cp	chloroplast
ddNTP	dideoxynucleotide triphosphate
dNTPs	all four (G,A,T,C) deoxynucleotide triphosphates
EDTA	ethylenediaminetetraacetic acid
kb	kilobases
L	proteins for large ribosomal subunit
mRNA	messenger RNA
mt	mitochondrial
Myr	million years
nad	genes for NADH dehydrogenase subunits
nt	nucleotides
Oligo dT	oligomer containing only deoxythymidine
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFGE	pulse field gel elecrophoresis
PVP	polyvinyl pyrrolidone
RACE	rapid amplification of cDNA ends
rpl	genes for large ribosomal subunit proteins
rps	genes for small ribosomal subunit proteins
rRNA	ribosomal RNA
RT	reverse transcriptase
S	proteins for small ribosomal subunit
SDS	sodium dodecyl sulphate
SSC	150 mM NaCl, 32 mM sodium acetate, pH 7.0
TE	10 mM Tris-HCl pH 7.5, 1mM EDTA
Tris	Tris (hydroxymethyl) aminomethane
tRNA	transfer RNA

CHAPTER 1

GENERAL INTRODUCTION

1.1 Origin of the eukaryotic cell and its organelles

All living organisms are believed to be evolved from a single primordial cell originating more than 3 billion years ago. The cenancestral cell was very simple in structure and composition and it multiplied and evolved into a rather complex cell to meet the changing atmospheric demands. Approximately 1.5 billion years later, the eukaryotic cell is predicted to have originated by the fusion of a protoeukaryotic (host) and a prokaryotic cell. The host cell did not have the ability to use atmospheric O_2 by aerobic respiration, but gained that activity through symbiotic association with another cell which evolved as a mitochondrion. The eukaryotic cell is highly complex in structure and the information content is chimeric in nature, since many genes from the endosymbiont were transferred to the nucleus during early endosymbiosis. This project addresses the evolution of plant mitochondrial (mt) genes and the ongoing evolutionary transfer of functional genes from mitochondria to the nucleus.

Among the several theories to explain the origin of the present eukaryotic cell and its compartments, the "endosymbiont hypothesis" proposed by Margulis (1970) and summarized above, is the widely accepted and best articulated theory for the evolution of the eukaryotic cell (reviewed by Gray, 1992). Doolittle (1998a) summarizes the endosymbiont theory with the following steps. 1. Some ancestral prokaryotic lineage developed an endomembrane system to engulf and digest other bacteria for survival. 2. Some of the engulfed bacteria were beneficial to the host and they were fixed as an endosymbiont rather than complete digestion. 3. The

endosymbiont, similar to the present day α-proteobacterium lineage, helped the host by providing energy through aerobic respiration and the host in turn provided a niche for the bacterium. 4. Genes required for the independent living of the α-proteobacterium were lost and many genes involved in respiration and other functions were transferred to the host nucleus while a small subset was retained. 5. As a result, the reduced endosymbiont became an organelle (mitochondrion) in the cell.

An alternate hypothesis for the origin of the eukaryotic cell, the "hydrogen hypothesis" has been proposed recently by Martin and Muller (1998). The hydrogen hypothesis and endosymbiont hypothesis are in agreement with the α-proteobacterial symbiont as the ancestor of mitochondrion. The hydrogen hypothesis states that an autotropic, hydrogen dependent, anaerobic archaebacterium (the host) had an association with a eubacterium that generated hydrogen as a waste product in its anaerobic heterotropic metabolism (Martin and Muller, 1998). In the absence of the outside source of hydrogen, the host became dependent on the symbiont and created a tight physical association. Genes from the symbiont were transferred to the host which provided a source for the membrane proteins needed for the import of enzymes and substrates of glycolysis. Finally, the host was able to perform oxidative phosphorylation and the symbiont became a mitochondrion or it was converted to a hydrogenosome. In some cases, the endosymbiont was lost and this hypothesis readily explains the energy metabolism of eukaryotes that lack mitochondria and the presence of bacterial genes in the heterotrophic archaeal lineage. However, the origin of unique eukaryotic structures such as cytoskeleton could not be explained by this hypothesis.

According to the endosymbiotic theory, an initial 'burst of quantum evolution' (Cavalier-

Smith, 1987) was assumed in the eukaryotic cell development. The massive transfer of organellar genes might have been inevitable during early endosymbiosis since the symbiont and the host were dependent on each other. Recent evidence in plant mt systems points to an ongoing evolutionary transfer of mt genes to the nucleus (see section 1.4.5). These transfers can be random events and fixed by drift. In contrast, animal and fungal mt genomes have fixed genes because their genetic code is different from their nuclear genomes. Mitochondrial genomes are present in multiple copies compared to a single nuclear genome in a cell. This increases the possibility of the transfer from mitochondrion to the nucleus being more frequent than from the nucleus to the mitochondrion.

Recently, Doolittle (1998b) has proposed the " food hypothesis" for the evolution of a reduced endosymbiont (mitochondrion) and the replacement of nuclear genes by bacterial genes. He compares the process of gene transfer in one direction to a ratchet model of operation. Accordingly, the necessary steps in a gene transfer process are organelle lysis, genomic incorporation and expression, stochastic fixation and elimination of the resident homologue in the mitochondrion. Though the gene transfer ratchet model was proposed for both the mt gene transfer and replacement of resident nuclear genes determining cytosolic function, gene replacement is expected to be frequent since it involves fewer steps. As discussed in section 1.4.6, mt gene transfer requires the addition of target sequences to direct the protein to mitochondria. This model provides an explanation for mt-type genes in the amitochondriate organisms and predicts when and how genes could have been transferred from the engulfed bacterium and replaced the nuclear genes. In addition, random lysis of mitochondria provides the source of mt DNA for the ongoing evolutionary gene transfer in plant mt systems.

Whether gene transfer from organelles to the nucleus is an early or recent event, the process is believed to be the same. Hence, knowledge about the recent gene transfer events going on in flowering plant evolution may help us to better understand earlier events in the evolution of mt genomes.

1.2 Plant mitochondrial genome and its complexity

1.2.1 *Genome size diversity*

One unusual feature of the plant mt genome is its large size and diversity among species. It varies in size from 208 kb (*Brassica hirta*) to 2400 kb (*Cucumis melo*) (reviewed by Bonen, 1998). There can be a 10 fold difference observed between species belonging to the same family. On the contrary, the animal mt genomes are very small (16-20 kb) with less variability. The unusually large size of the plant mt genome is not reflective of its relative coding capacity (reviewed by Gray et al., 1998). The completely sequenced *Arabidopsis* mt genome is 367 kb (Unseld et al., 1997) and codes an estimated 57 genes, whereas the human mt genome is 16 kb (Anderson et al., 1981) with 37 genes (Table 1). Nearly 62% of the *Arabidopsis* mt genome consists of non-coding sequences. They are contributed by duplications caused by recombination events, imported foreign DNA and the extensive presence of unrecognizable and apparently functionless sequences. When the mt genomes of *Arabidopsis*, human, protists and algae were compared for their size and coding sequences, it was found that the size increase of mt genomes is colinear with an increase in non-coding sequences (Unseld et al., 1997). To date, *Reclinomonas americana*, a flagellate protozoan, has the largest set of genes (97) in its mt genome of 69 kb (Lang et al., 1997). The gene content of several different mt genomes (*Arabidopsis*, liverwort, human, yeast and *Reclinomonas*) are summarized in Table 1. Among the hundreds of gene

products required for the mt function, only a small fraction of them are encoded in the mt genome and the vast majority of them are located in the nucleus. Approximately 5 % (296) of the estimated 6086 genes of the completely sequenced yeast genome codes for mt proteins (Proteome Inc.1999-Proteome Inc. http.WWW.proteome.com YPDCYC7.html).

1.2.2 *Structure of plant mt DNA*

The structure of the plant mt genome is complex and it appears to exist as a collection of multiple circular molecules differing in size and stoichiometry. In contrast, animals contain multiple copies of identical circular DNA molecules (Anderson et al., 1981). The complexity of plant mt genomes depends on the number, location, and orientation of recombinogenic repeats which vary from species to species (reviewed by Gray et al., 1998). The smallest plant mt genome, that of *B. hirta*, does not have any repeat elements (Palmer and Herbon, 1987). In contrast the wheat (430 kb; Bonen, 1995) and rice (492 kb; Zeltz et al., 1996) mt genomes have 7-10 repeat elements based on their cytotypes. The largest circular plant mt DNA molecule (186 kb) observed so far using electron microscopic techniques is that of the nonvascular plant *Marchantia polymorpha* (Oda et al., 1992). According to Bendich, (1996) most plant mt DNAs exist as complex branched structures or linear molecules derived from the complex forms, but not as circular molecules. However the presence of small circular forms of mt DNA has been seen in maize by PFGE in combination with infrequently cutting enzymes (Andre and Walbot, 1995).

1.2.3 *Gene content of mt genomes*

A list of genes present in the completely-sequenced mt genomes of *A. thaliana* (Unseld et al., 1997), *M. polymorpha* (Oda et al., 1992), *Homo sapiens* (Anderson et al., 1981), *R. americana* (Lang et al., 1997) and *Saccharomyces cerevisiae* (Grivell, 1995) is given in Table 1.

Table 1 Compilation of mitochondrial genes in different organisms

	Arabidopsis	Liverwort	Human	Yeast	*R. americana*
Structural RNA genes					
rRNA (16S, 23S)	2	2	2	2	2
rRNA (5S)	1	1	0	0	1
transfer RNA	22	27	22	24	26
Regulatory protein genes					
translation elongation factor	0	0	0	0	1
RNA polymerase (*rpo* A,B,C &D)	0	0	0	0	4
Ribosomal protein genes					
Large subunit	3	12	0	0	15
Small subunit	4	4	0	1[a]	12
Respiratory protein genes					
complex I (*nad*)	9	8	7	0	12
complex II (*sdh*)	1	2	0	0	3
complex III (*cob*)	1	1	1	1	1
complex IV (*cox*)	3	3	3	3	3
complex V (*atp*)	4	4	2	3	5
Protein import/maturation					
cytochrome c biogenesis	5	3	0	0	4
cytochrome oxidase assembly	0	0	0	0	1
protein import (*sec Y*)	0	0	0	0	1
ORFs	86[b]	28[c]	0	0	3
Group1 intronic ORFs	0	2	0	5	0
Group II intronic ORFs	1	8	0	2	0
Protein coding genes	28	37	13	8	62
Total identified genes	56	67	37	34	91
Genome size kb	367	187	17	75	69

[a] -var1 which has no bacterial counterpart. [b] -all open reading frames with >100 codons including *orf25* and *orfx* which were listed as identified genes by Unseld et al. (1997). Note that *orf95* and *orfB* have now been identified as *sdh4* (Giege et al., 1998) and *atp8* (Gray et al., 1998) respectively. [c] -all open reading frames with >60 codons. Table adapted from Palmer (1997).

The rRNA genes coding for the 26S rRNA, 18S rRNA and 5S rRNA are present invariably in all the plants examined including the nonvascular plant *M. polymorpha* and the green alga *Prototheca wickerhamii* (Unseld et al., 1997). However animals, fungi and most protists lack the 5S rRNA gene and they have smaller rRNAs for example the16S & 12S rRNAs in animals are smaller than the bacterial 23S and 18S rRNAs (reviewed by Gray et al., 1998). Animal mt genomes have a complete set of tRNA genes whereas plants do not have a complete set to support the translation (reviewed by Veronica et al., 1996). The completely sequenced *Arabidopsis* mt genome lacks 6 tRNA genes and these tRNAs have to be imported from the cytoplasm (Unseld et al., 1997). Similarly in potato 11 out of 31 tRNA species are found to be imported from the cytosol (reviewed by Bonen, 1998).

At least one or several respiratory chain protein coding genes for all five complexes are present in plants but animal and fungal mt genomes lack genes for complex II (succinate dehydrogenase) (reviewed by Gray et al., 1998). At least ten genes with unidentified functions named ORFs (open reading frames) are conserved between different plant mt genomes (reviewed by Brennicke et al., 1993). Thorough comparative and functional analysis on these ORFs should lead to the identification of the encoded proteins.

Plant mt genomes encode a higher number of ribosomal protein (rp) genes (16 in liverwort) compared to fungi such as yeast (var1) and *Neurospora* (S5) which encode only one rp gene and animals with none (reviewed by Schuster and Brennicke, 1994). In flowering plants at least 14 rp genes have been identified so far in different species which is a subset of the 16 rp genes located in the liverwort mt genome (Takemura et al., 1992). However, the number of rp

Table 2 List of mitochondrial ribosomal protein genes identified in mitochondrial and nuclear genomes of different plants

protein	Liverwort		*Arabidopsis*		*Oenothera*		pea		wheat		rice	
	mt	nu	mt	nu	mt	nu	mt	nu	mt	nu	mt	nu
S1	+		-		+?				+?		+	
S2	+		-				-		+		+	
S3	+		+		+				+		+	
S4	+		+						$+_s$	+		
S7	+		+				ψ	$+_s$	+		$+_s$	
S8	+		-									
S10	+		-	+	-		+		-		-	$+_s$
S11	+		-				-	+	+		ψ	+
S12	+		+		ψ	+			+		+	
S13	+		-	$+_a$	+		-				$+_s$	
S14	+		ψ		+							
S19	+		ψ	+	ψ		-		ψ		+	
L2	+		+?		+?		-		ψ	+	+	$+_{sn}$
L5	+		+		+							
L6	+		-									
L16	+		+		+				+		+	

The mt ribosomal proteins are named according to their *Escherichia coli* counterpart.

+-present based on sequencing information, - -absent based on Southern data, ψ - pseudogene, +? truncated potentially functional, s-Based on Southern data, sn- Based on Southern and Northern data, +a potential gene function replacement

Data from reviews by Schuster and Brennicke, (1994); Kumar, (1995) and Oda et al., (1992) and updated with data from Unseld et al., (1997); Vaitlingom et al., (1998); Wischmann and Schuster, 1995; Kubo et al., (1996): Sanchez et al., (1996); Kadowaki et al., (1996); Kubo et al., (1998); Subramanian and Bonen, (1998); Zhuo et al., (1999).

genes present in different flowering plant mt genomes varies even between closely related plants (Table 2). The completely sequenced mt genome of *Arabidopsis* encodes only seven functional mt rp genes (Unseld et al., 1997). Gene transfer to the nucleus is predicted for missing mt rp genes and several have been identified in the nucleus (see section 1.4.5).

1.2.4 *Mitochondrial gene structure and organization in plants*

The status of individual genes and the organization of different genes in the mt genome varies among plants. Some genes are simple and devoid of introns (*atpA, atp6, atp9, cob, nad3, nad9, rps12, rps13, rps19* and *rpl16*) while others have a very complex structure with 1-4 introns (reviewed by Fauron, et al., 1995). All plant mt introns have been identified as group II category except that the *cox1* gene of plants belonging to the genus *Peperomia* has a group I intron (reviewed by Bonen, 1998), apparently acquired by a horizontal gene transfer event from a fungal source (Adams et al., 1998a). There are 25 group II introns identified in flowering plant mt genomes in which 19 of them are located in *nad* genes. The *nad1, nad2, nad4, nad5 and nad7* genes contain three to four introns. The *nad1, nad2* and *nad5* genes are trans-spliced genes and their coding regions are separated far apart in the mt genome. The precursor mRNA molecules containing the flanking intron sequences fold in a way to bring the exons together. This process is called trans-splicing (reviewed by Bonen, 1993). The organization of trans-spliced genes is an unique feature to the flowering plants mt genomes compared to other mt genomes. Among the ribosomal proteins identified in flowering plants, *rps3, rps10* and *rpl2* are interrupted by single group II introns.

1.2.5 Organization of plant mitochondrial ribosomal protein genes

The mt and chloroplast (cp) rp gene clusters of flowering plants resemble sections of the *E. coli* rp gene operon that is consistent with the bacterial origin of these organelles (Takemura et al., 1992). In *E. coli*, the rp genes are clustered to form 6 operons (Fig. 6.2) α (S13-S11-S4-rpoA-L17), *str* (S12-S7-EF-G-EF-Tu), β (L10-l7/L12-β-β'), L11 (L11-L1), *spc* (S14-L24-L5-S14-S8-L6-L18-S5-L30-L15-sec Y-X) and S10 (S10-L3-L4-L23-L2-S19-L22-S3-L16-L29-S17-Nomura et al., 1980). The nonvascular plant *M. polymorpha* has two rp gene clusters similar to S10 and *str* operons in its mt genome. They are *rps10-rpl2-rps19-rps3-rpl16-rpl5-rps14-rps8-rpl6-rps13-rps11-rps1* and *rps12-rps7*. The *rps2* and *rps4* genes are located separately in the liverwort mt genome (Oda et al., 1992). In flowering plants, segments of the above rp gene clusters as well as rp genes clustered with respiratory protein genes are often observed (Table 3) There are very few rp genes which are located in isolation, for example *rps7* (Zhuo and Bonen, 1993) and *rps2* of wheat (Vaitilingom et al., 1998). The *rps3-rpl16-nad3-rps12* cluster of rice (Nakazono et al., 1995) and *rps3-rpl16-rpl5-rps14-cob* cluster of *B. napus* (Ye et al., 1993) are the largest co-transcribed mt gene clusters found in flowering plants to date. The *E. coli* type S10 operon segment, *rps3-rpl16* is found to be conserved in all plants examined (reviewed by Binder et al., 1996). Similarly, remnants of the *E. coli* α *operon* (S11-S4; Handa et al., 1997) and other parts of the *S10* operon (L2-S19; Kubo et al., 1996) are present in the rice mt genome.

1.2.6 *Structure of plant mitochondrial ribosomal protein genes*

Although the plant mt ribosomal protein genes examined to date show higher similarity to the eubacterial type than the eukaryotic type, there are some unique features like N-or C- terminal extensions or inserts (20-300 a.a) in the derived protein sequence. Out of the 16 mt rp

Table 3 Organization of ribosomal protein genes clustered with respiratory chain protein genes in flowering plant mitochondrial genomes

Plant	Genes	Reference
Wheat	***atp6-rps13-nad1b/c***	(Bonen, 1987; Takvorian et al., 1997)
Arabidopsis	*rpl16-**atp9***	(Sakamoto et al., 1997)
Wheat	***orf585**-rps1*	(Gonzalez et al., 1993)
Oenothera; Arabidopsis	*rps3-**cox1***	(Schuster and Brennicke, 1991)
Rice	*rpl2-rps19-**nad4L***	(this work)
Oenothera; Arabidopsis	*rpl2-**orfX***	(Sunkel et al., 1994)
B.napus	*rps3-rpl16-rpl5-rps14-**cob***	(Ye et al., 1993)
Arabidopsis; potato	*rpl5-ψrps14-**cob***	(Brandt et al., 1993; Quinones et al., 1996)
All plants examined	***nad3**-rps12*	(reviewed by Bonen, 1998)
Oenothera	*rps1-**atp9***	(Mundel and Schuster, 1996)
Brassica	***nad2a**-rps4*	(Handa et al., 1998)
Wheat	***nad9**-rps2*	(Vaitilingom et al., 1998)
Tobacco	***orf1**-ψrpl2-orf2*	(Vitart et al., 1992)
Tobacco	*atp9-rps13-**nad1b/c***	(Bland et al., 1986)
Oenothera	*cox1-rps13-**nad1b/c***	(Wissinger et al., 1990)
Maize	***atp9-nad3**-rps12*	(Hunt and Newton, 1991)

genes identified in liverwort, five (*rps3, rps7, rps8, rpl2, rpl16,*) have additional inserts in flowering plants (Takemura et al., 1992). Gene fusion could be one of the reasons for these extensions as observed in the yeast *Yml8* gene which has sequences homologous to *E. coli rpl17* in the 5'- end and *rps13* in the 3'- end (Kitakawa and Isono, 1991). However, the origin of the inserted sequences in L2 (251 a a), S3 (300 a a) and the N- or C-terminal extensions (~20-80 amino acids in S7, S8, L16) of plant mt rp genes can not be traced, since sequence comparisons do not reveal similarity to any known proteins (Takemura et al., (1992); Kubo et al., (1996); Bock et al., (1994). The inserted sequences found in the *rps4* gene of rice and *Brassica* could be folded into domain 5 and 6 structures of a group II intron, but they are not spliced (Handa et al., 1998). This suggests an insert of intron sequences which are in frame with the coding region followed by a deletion event that impaired splicing. This could be one of the scenarios for the origin of extra segments found in plant mt ribosomal proteins.

1.3 Regulation of gene expression in plant mitochondria

1.3.1 *Regulation at the transcriptional level*

In the previous section, the plant mt gene content and the organization of rp genes were discussed in detail. It is evident that some mt genes are transcribed as monocistronic messengers, while many are co-transcribed with adjacent genes. The presence of many transcriptional units in the plant mt genome implies a corresponding multitude of individual promoters (reviewed by Binder et al., 1996). In contrast, the small animal mt genome contains only two promoters in the displacement loop (D-loop), one each for the transcription of each strand of DNA (Clayton, 1991). The moderately larger mt genome of yeast (*S. cerevisiae*) is found to have 20 promoters (reviewed by Binder and Brennicke, 1993). In plants, apart from the large number of genes and

their promoters, individual genes in some cases are observed to have several promoters. The *cox2* and *atp9* genes of maize are transcribed from multiple promoters which are either adjacent repeat sequences or spread over several hundred base pairs upstream of the coding region (reviewed by Tracy and Stern, 1995).

Plant mt promoters are multipartite in structure and reveal limited similarity. In contrast, fungal mt promoters are well conserved both within and between species (reviewed by Tracy and Stern, 1995). In monocot plants like rice, wheat and maize, an A/T or purine rich region and a tetranucleotide 5'-CRTA-3' motif are the features conserved between potential promoter sequences surrounding transcription initiation sites (Nakazono et al., 1995; reviewed by Binder et al., 1996). In maize, in vitro mutation analysis has shown that the conserved CRTA motif is essential for the promoter activity (reviewed by Tracy and Stern, 1995; Caoile and Stern, 1997). In maize (*Zea perenis*) mitochondria, the *cox2* gene has two promoter sequence motifs. The structural differences observed in these promoters may suggest a possibility of more than one RNA polymerase or a role of additional proteins (Newton et al., 1995). In the chloroplast, two types of RNA polymerases, namely nuclear and chloroplast encoded forms are actively involved in cp transcription (Hedtke et al., 1997). Plant mt genomes do not encode any RNA polymerase genes whereas the *R. americana* mt genome has genes for the four (*rpo* A, B, C &D) components of a eubacterial-type RNA polymerase (Lang et al., 1997). Hence, plants might have lost the eubacterial type RNA polymerases early in evolution and recruited the nuclear-encoded phage type polymerase which has been identified (Gray and Lang, 1998) or also the bacterial type may have transferred to nucleus.

1.3.2 *Regulation of plant mitochondrial genes at the post-transcriptional level*

1.3.2.1 *Plant mt mRNA splicing*

The mRNA levels and maturation processes particularly appear to be controlled by a combination of transcriptional and post-transcriptional events in plant mitochondria. Post-transcriptional processing in plant mt genes include splicing, editing and termini-processing. The group II introns present in mt genes are highly structured and are usually depicted as a central wheel with six radiating helical domains. Splicing of group II introns follow two transesterification steps similar to the nuclear pre-mRNA splicing (reviewed by Michel and Ferat, 1995). The first tranesterification results from the attack on the 5'splice site by a 2' hydroxyl of a bulging adenosine in domain 6 and the second by the 3' end hydroxyl. The catalytic core for the splicing reaction is comprised of domain 5, sequences within domain 1, a bulging adenosine of domain 6 and tertiary interactions between domain 4 and 2 (Chanfreau and Jacquier, 1996). A trans-splicing reaction is required for certain introns in plant mt *nad1, nad2, nad5*, plant cp *rps12* genes and in the cp *psaA* gene of *Chlamydomonas reinhardtii* (reviewed by Bonen, 1993). A tripartite group II intron is observed in *Oenothera nad5* gene (Knoop et al., 1997) in which the intron region between the domains 1 and 4 located in a different part of the genome is transcribed and takes part in the splicing reaction. In addition to the tripartite *nad5* group II intron, the presence of very low levels of bipartite *nad5* intron is also observed in *Oenothera* mt DNA similar to other plants. The above observations suggest that certain disruptions in the group II intron could be tolerated, since the different segments could recognize one another and carry out the splicing reaction properly (Knoop et al., 1997).

1.3.2.2 *Mitochondrial mRNA stability*

Processing of mt mRNA is an important step, since it can influence the stability of mRNA. In certain maize mt protein coding genes, a 20 fold difference observed between the rate of synthesis and the steady state accumulation of mRNA suggests an involvement of post-transcriptional processes that govern stability (reviewed by Binder et al., 1996). The 5' processing of the mt mRNA has not been investigated in as much detail as has 3' processing (Drombrowski et al., 1997). The 3' untranslated regions of certain mt mRNAs have inverted repeats (IR) which are candidate structures for modulation of RNase access to mRNA. Hence, a role of RNA processing and stability is suggested for these IR structures. In pea mitochondria, an *in vitro* transcription system has been developed to analyze the processing of *atp9* mRNAs (Drombrowski et al., 1997). The double stem loop structure present at the 3' end of the mRNA did not inhibit termination of transcription and the 3' end was trimmed exactly as *in vivo* yielding a stable mRNA. From the above study, the IR/double stem loop structures present at the 3' end of mt mRNAs are recognized as processing signals of transcript maturation and most likely are stability signals rather than terminators.

1.3.2.3 *RNA Editing*

RNA editing is defined as a process responsible for any change in the sequence of a RNA that differs from its template DNA. In plant mitochondria and chloroplasts, RNA editing involves the conversion of specific cytidines in the genomic sequence into uridines in the mRNA, whereas the insertion and deletion of uridines occur in trypanosome mitochondrial mRNA (reviewed by Bonnard et al., 1992). RNA editing occurs in the mitochondria of all major groups of land plants examined with the exception of certain bryophytes belonging to Marchantidae (Steinhauser et al.,

1999). Editing usually increases the sequence similarity of the predicted protein to other organisms and it is believed that most changes created in transcripts are essential to produce functional proteins (reviewed by Smith et al., 1997). The magnitude of editing varies between genes. About 3-15% of codons of plant mt protein coding genes are affected by editing. Editing which alters the amino acid occurs at a higher frequency than editing at synonymous sites (reviewed by Hanson et al., 1996). Editing in the mt mRNA in some cases creates a start codon, by changing the ACG to AUG or shortens the predicted open reading frame by creating a stop codon (eg. CAA to UAA). The degree of editing between sites and transcripts varies a great deal (reviewed by Smith et al., 1997). The proteins derived from translation of partially edited maize *rps12* transcripts (Phreaner et al., 1996) were observed in the polysome fraction whereas only the translation products of the completely edited *rps13* transcripts were detected in wheat mt ribosomes (Williams et al.,1998b).

Editing has also been found to modify non-coding transcribed regions, structural RNAs, (rRNAs and tRNAs) and intron sequences but to a much lower extent (reviewed by Maier et al., 1996). Editing events were observed in unspliced mRNAs, hence editing in introns can occur before splicing (reviewed by Hanson et al., 1996). Interestingly, they were predicted to be important to help the proper folding of group II intron and support splicing (reviewed by Hanson et al., 1996). However, Carrillo and Bonen (1997) found that editing is not essential for splicing, since certain excised group II intron transcripts were not found to be edited in positions of domain 5 and 6.

The exact mechanism for editing site selection in plant mitochondria is not known. Wilson and Hanson (1996) have shown that editing is found to vary with mt and nuclear genotypes, not

by the transcriptional context but the sequence context. Observations on editing in chimeric and normal *atp6* genes of rice suggest that 5' flanking sequences are important for the recognition of editing sites and that 3' flanking sequences have little contribution (Kubo and Kadowaki, 1997). In maize, a truncated *rps12* gene is present in addition to the functional one in which a single nucleotide change in the -5 position of an editing site abolished editing, indicating the importance of primary RNA sequence context in editing (Williams et al., 1998a).

1.3.3 *Developmental and differential regulation of plant mt genes*

In addition to transcriptional and post-transcriptional regulation, plant mt genes exhibit some developmental and differential regulation. In wheat, the mRNA accumulation for the *atp6* gene and *nad1b* exon is found to be higher (at least 3 fold) at the seedling stage than in germinating embryos relative to the mt rRNA levels (Bonen, 1987). In contrast, stable transcripts of the cis-spliced *nad7* gene and its introns were found at higher levels (precursors and excised introns) in germinating embryos than in seedlings (Carrillo and Bonen, 1997). Developmentally specific differences in the abundance and complexity of *nad7* intron transcripts might reflect differential stability or RNA processing pathways.

Expression of mt genes during pollen development has been heavily studied, to better understand cytoplasmic male sterility in different plants. This gives an insight into the developmental and tissue specific expression of mt genes. In sunflower, the amounts of transcripts for the mt genes *atpA, atp9, cob* and *rrn26* were found to be increased in young meiotic cells, while nuclear-encoded mt protein genes showed little difference in their transcript levels (Smart et al., 1994). Differences in the amount of mt protein synthesized in various tissues have been observed among leaves, roots and flowers of sugar beet (Lind et al., 1991).

The nuclear genes coding for the mt respiratory chain protein *atpβ* and polypeptides specific to pollen development in tobacco and petunia are found be expressed in a tissue specific manner in leaves and pollen (De Paepe et al., 1993). Similarly, potato and *Arabidopsis* nuclear genes encoding a mt 22 kDa (PSST) subunit of respiratory complex I are also regulated in an organ specific manner. The transcription rate and RNA stability are shown to be higher in flowers compared to leaves and roots (Heiser et al., 1996). The above studies suggest that mt genes involved in pollen development are regulated in a tissue and stage specific manner regardless of their location.

1.4 Transfer of genes between subcellular compartments

1.4.1 *Transfer of mitochondrial respiratory protein genes to the nucleus*

In addition to the developmental and tissue specific regulation discussed above, the presence of copies of a mt gene in two different compartments adds another potential level of regulation. The first observation of both a functional mt copy and a transferred nuclear copy was made by Van den Boogaart et al (1982) in *Neurospora* for the *atp9* gene. Differential expression is observed between them (Bittner-Eddy et al., 1994), in that the mt copy is not expressed in dormant asexual spores but transcripts accumulate rapidly after spore germination, whereas the nuclear copy is expressed at all stages.

The evidence for recent transfer of plant mt genes as opposed to the early endosymbiotic transfer from organelles to the nucleus is mounting (reviewed by Brennicke et al., 1993). The *coxII* gene which was thought to be one of the "classical" mt genes has a functional copy in the nucleus in several legumes (Nugent and Palmer, 1991; Covello and Gray, 1992). The systematic documentation of the location and functioning of mt *coxII* genes in legumes provides a better

understanding of the nature of recent gene transfer events to the nucleus. The mt *coxII* gene is present in the mt genomes of all leguminous plants examined except mung bean and cowpea which have been transferred to the nuclear genome (Nugent and Palmer, 1991). Interestingly, all of the legumes examined have a nuclear copy but their expression varies between species. In soybean the mt copy is silent and the nuclear copy is actively transcribed (Covello and Gray, 1992), in pea the nuclear copy is silent and the mt copy is active. The cowpea nuclear gene has an intron between the transferred mt sequences and acquired targeting sequences. Exon shuffling is proposed for the gain of expression signals. The gene transfer is likely to be an RNA-mediated one, since the transferred sequences correspond to edited sequences. The location and expression of the mt *cox II* gene in the above leguminous plants portrays various stages of a mt gene transfer event. Accordingly, a successful gene transfer event is expected to proceed in the following sequence: 1. a mt gene transfer to the nucleus, 2. nuclear gene activation, 3. loss of function in the mt copy and 4. the complete loss of the mt copy. The legumes illustrate all these four stages.

In another example, the mt cytochrome biogenesis gene *ccb,* which is present in the mt genomes of wheat, *Oenothera* and *B. napus* L. (Itani and Handa, 1998) but in *B. campestris* (Menassa et al., 1997) the gene is split into two but is not spliced properly, hence believed to be a pseudogene. A copy of these sequences is detected by hybridization experiments in the nuclear genome indicating that a recent gene transfer may have happened after the speciation. Additionally, the respiratory protein gene *nad7* is found to be active in the mt genomes of all flowering plants examined but in liverwort it is a pseudogene and the functional copy has been transferred to the nucleus (Kobayashi et al., 1997). Apart from the mt respiratory protein gene transfer events, there are several examples of mt ribosomal protein gene transfer events to the

nucleus and they are discussed in detail in the section 1.5.1.

1.4.2 *Transfer of chloroplast genes to the nucleus*

The gene content of the completely sequenced chloroplast (cp) genomes of photosynthetic unicellular eukaryotes and different plants reveals that among the 210 genes present in them only 45 cp genes are common to all (Martin et al., 1998). The reason could be theoretically either gene loss or the endosymbiotic gene transfers to the nucleus. Phylogenetic studies on the completely sequenced cp genomes in conjunction with 44 cp genes identified in the nucleus so far reveal that parallel and unique losses of cp genes are in 4:1 ratio suggesting a high rate of multiple and independent transfers. The gene content of different flowering plant cp genomes do not vary much, hence the few differences observed may possibly be correlated to recent gene transfers to the nucleus. There are few examples of recent cp gene transfer events to the nucleus including a rp gene. The *rpl22* gene is present in the cp genomes of all flowering plants except legumes (Gantt et al., 1991) and a nuclear-encoded *rpl22* has been identified in pea which contains transferred cp *rpl22* sequences and acquired cp target sequences separated by an intron. Based on phylogenetic analysis the time of the gene transfer event is predicted to be at least 100 million years ago (Gantt et al., 1991).

1.4.3 *Transfer of DNA from the chloroplast genome to the mitochondrion*

The mitochondrial genomes of higher plants have chloroplast (cp) sequences that have been transferred from the cp genome (reviewed by Schuster and Brennicke, 1988). The recombinogenic nature of mt genomes of flowering plants allows the acceptance of extra-chromosomal DNA. The mt genomes of the nonvascular plant *Marchantia* do not have any cp DNA which may be related to the absence of recominogenetic repeats in its mt genome (Oda et

al., 1992; Bonen and Brown, 1993; Schuster and Brennicke, 1994). The rice mt genome has an estimated 22 kb of transferred cp sequences scattered around the genome which is nearly 6% of the size of the mt genome and 19% of the cp genome (Nakazono and Hirai, 1993). The borders of the integrated cp sequences in rice mt DNA do not exhibit any common sequences or structure that might explain their integration. At least 16 cp DNA insertions ranging from 32 bp to 6.8 kb are observed in rice mt genome (Kanno et al., 1997). Nearly 1% of the completely sequenced *Arabidopsis* mt genome (367 kb) is comprised of cp sequences (Unseld et al., 1997). These observations indicate that the mt genomes are dynamic and flexible in accepting "promiscuous" DNA. The cp sequences are expected to have inserted sequentially at various times before and after the divergence of rice and maize (Nakazono and Hirai., 1993).

The transferred cp sequences contain several cp genes, but in most cases they do not encode functional products in mitochondria. However, many cp genes for tRNAs are transcribed and the tRNAs used for the biosynthesis of mt proteins (Joyce and Gray, 1989; Fey et al., 1997). In some cases, native copies of the tRNA genes are found to be lost from the mt genome and the transferred cp genes are the only actively transcribed isoacceptor species (reviewed by Veronica et al., 1996). Recent findings indicate that cp sequences can be recruited for other functions in the mt genome. For example, sequences within the cp gene cluster (*trnS-rps4/3'trn*L-*trn*F-*ndh*J-*adhK*) present upstream of the *nad9* gene of rice mitochondria are found to be used as a promoter (Nakazono et al., 1996).

1.4.4 *Movement of nonfunctional DNA between different compartments*

Although the functional mt gene (RNA-mediated) transfer to the nucleus is particularly interesting, knowledge about the transfer of nonfunctional DNA between compartments may also

provide information for the basis of nucleic acid transfer. Several group II intron sequences specific to mitochondria have been detected in the nuclear genomes of plants by computer searches (Knoop and Brennicke, 1994) and their role in the nuclear genome was examined since they show similarity to the spliceosomal introns. Group II intron sequences were found in the vicinity of the nuclear glucanase genes lectin and photosystem II genes of tobacco and bean. The sequence similarities between the nuclear and mt copies are in the range of 80-97%. The high nucleotide identity observed for sequences which are not under functional constraint suggests a relatively recent transfer. However, the mt intron fragment (*nad5*) inserted upstream of the lectin gene of African bean is predicted to be responsible for the tissue specific expression of the lectin gene (Knoop and Brennicke, 1991). Similar mitochondrial DNA integrations in mammalian and yeast nuclear genomes were detected by extensive comparative analysis of sequences in the databank (Blanchard and Schmidt, 1996). At least seven and five mt DNA insertions were identified in nuclear genomes of mammals and yeast, respectively. Analysis of the integration sites suggests that they are incorporated by an end-joining mechanism common to mammals, yeast and plants.

In yeast, an experimental plasmid system was used to measure the rate of mt DNA transfer (Thorsness and Fox, 1990). Yeast mt DNA within plasmid constructs containing selectable markers were found to escape mitochondria and integrate into the nuclear DNA at a rate of 2×10^{-5} per cell per generation. The vacuole dependent degradation of abnormal mitochondria is predicted to be one pathway by which mt DNA escapes and integrates into the nucleus (Campbell and Thorsness, 1998). At least 10 nuclear genetic loci are found to be associated with the mt DNA transfer and they are called YME genes (yeast mitochondrial DNA escape) which encode

ATP- and zinc dependent proteases that are located in the inner mt membrane. Mutants of YME genes show poorly defined inner membrane and increased DNA escape (Hanekamp and Thorsness, 1996). An electron microscopic study accompanied by in vivo assay of yme1 mutants (Campbell and Thorsness, 1998) revealed the physical association of mitochondria and vacuoles and a higher rate of vacuole dependent turnover of the mitochondrial compartment in mutants compared to the wild type. The mt turnover in yme1 mutants were also associated with the nutrient availability such as fermentable vs non fermentable carbon source.

Transfer of nonfunctional cp DNA to the nucleus has been studied in tobacco by hybridization experiments and by isolation of cp sequences from the nuclear genome (Ayliffe and Timmis, 1992). Sequences corresponding to nearly one third of the size of the tobacco cp genome are predicted to be located in the nuclear genome and are not expected to be expressed because they lack appropriate expression signals (Ayliffe and Timmis, 1992). Digestion of nuclear DNA with methylation-sensitive restriction enzymes followed by Southern analysis was another method used to distinguish promiscuous plastid DNA present in nuclear genomes of cotton, pea, beet, barley and wheat (Ayliffe et al., 1998).

The transfer of nuclear sequences to the mt genomes also appear to occur for example an ORF homologous to the reverse transcriptase is integrated upstream of the mt 18S rRNA gene of *Oenothera* (Schuster and Brennicke, 1987). This might have arisen by a transfer event from the nucleus to mitochondria or a transposition event of a retroelement to the mt genome. About 4% of the *Arabidopsis* mt genome are estimated to be transferred nuclear-derived pieces of retrotransposons (Knoop et al., 1996; Unseld et al., 1997).

1.4.5 *Recent transfer of mitochondrial ribosomal protein genes to the nucleus*

The number of rp genes encoded in the mt genome of flowering plants varies even between closely related plants. This may be due to the random nature of functional mt gene transfer to the nucleus followed by a mt gene inactivation and gene loss by recombination and DNA rearrangements. The successful transfer of ribosomal protein genes seems to be frequent relative to transfers of genes coding for proteins of the respiratory chain (Brandt et al., 1993). A list of mt rp genes varying in their location and expression is given in Table 2. There are several examples of the recent transfer of mt rp genes to the nucleus in the flowering plants namely, *rps12* in *Oenothera* (Grohmann et al., 1992), *rps10* in *Arabidopsis* (Wischman and Schuster, 1995), *rps19* in *Arabidopsis* (Sanchez et al., 1996) and *rps11* in rice (Kadowaki et al., 1996) and pea (Kubo et al., 1998). Duplicated copies of *rps11* gene are observed in the nuclear genomes of rice and pea. Mitochondrial copies of the transferred rp genes have either lost their function and are present as truncated pseudogenes (*rps12-Oenothera, rps19-Arabidopsis, rps11*-rice) or are completely lost (*rps10-Arabidopsis & rps11* pea). However, other plants have a functional copy for example such as wheat *rps12,* petunia *rps19*, and potato *rps10* (reviewed by Kumar, 1995). Certain other mt rp genes are predicted to be transferred to the nucleus based on Southern data showing either the absence of the functional gene in the mt genome in comparison to related plants or its presence in the nuclear genome for example *rps7* in soybean and pea (Zhuo, et al., 1999 in press), maize *rps14* (Wahleithner and Wolstenholme, 1988) and *rps13* (Bland et al., 1986) in maize.

1.4.6 *Expression of transferred ribosomal protein genes located in the nucleus*

When a mt rp gene is transferred to the nucleus, it has to acquire sequences that can be

recognized by the nuclear transcription system and cytoplasmic translation machinery. In addition, the resulting protein has to be targeted back to the mitochondrion in order to assemble and constitute a functional mt ribosome. The N-terminal transit peptide sequence has the ability to form an amphipathic structure with a positively charged side and a hydrophobic side to interact with a number of negatively charged proteins across the mt membranes (reviewed by Pfanner, 1998). The transferred mt genes appear to acquire their expression and transit signals/sequences either from other nuclear-located mt genes or by exon shuffling. The transit peptide sequences and the transferred mt rp coding region are in some cases separated by an intron indicating that exon shuffling might have been involved in the addition of a target peptide signal for example, the mt *rps10* gene of *Arabidopsis* (Wischmann and Schuster, 1995). The duplicated copies of the transferred rice mt *rps11* gene have acquired their mt target sequences by the duplication of target sequences from *atpβ* and *coxVb* which already existed in the nuclear genome (Kadowaki et al., 1996).

1.4.7 *The L2 ribosomal protein genes*

The main focus of this project is on the L2 ribosomal protein genes in plants which are named according to the *E. coli* homologue. In *E. coli,* L2 is the largest ribosomal protein (272 amino acids) of the 50S subunit, and contributes to several aspects of ribosome structure and function. The L2 protein in the bacterial system has been the subject of many investigations to analyze its structure, location in the ribosome, RNA and protein binding domains, and function. Three dimensional structure analysis of the *Bacillus stearothermophilus* L2 protein suggests that it has structural motifs that are often found in RNA- or DNA binding proteins (Nakagava et al., 1999). Single protein omission experiments and mutation analysis have shown that L2 is one of

the key components in the peptidyl transferase centre (reviewed by Green and Noller et al., 1997). The L2 protein protects the domain IV of the 23S rRNA from nucleases and chemicals which leads to the speculation of its direct role in the catalysis of peptide bond formation (reviewed by Noller, 1993). In addition, the bacterial L2 proteins are known to have a role in the binding of elongation factor proteins and in subunit association (Olson et al., 1991).

In plants, ribosomes of all three cellular compartments have distinct but homologous L2 proteins. The size of the cp L2 protein (277 a.a) is comparable to the size of *E. coli* L2 (272 a.a) and its gene is located in the cp genome of all plants examined (Doyle and Doyle, 1995). Mutant analysis has shown that the cp L2 proteins carry out the same function as *E. coli* L2. Non-allelic mutants of barley that showed inefficient splicing of chloroplast *rpl2* transcripts had a severe deficiency for chloroplast ribosomes (Hess et al., 1994). L2 proteins have a conserved amino acid motif which matches all the 50S and 60S L2 proteins and the two residues involved in the peptidyl transferase activity.

The mt *rpl2* gene of liverwort was the only available complete *rpl2* sequence from the land plant mitochondrion when I started this project. It has a coding sequence of 1503 bp with an insert of 750 bp in the middle of the gene coding for a protein with 501 amino acids in contrast to 272 in *E. coli*. Variation in the location of the functional mt *rpl2* genes was assumed in plant species since the presence of pieces of *rpl2* sequences were reported in tobacco (Vitart et al., 1992) and *Oenothera* and *Arabidopsis* (Sunkel et al., 1994). In tobacco, sequences corresponding only to the N-terminal 60 amino acids have been reported (Vitart et al., 1992). In *Oenothera* and *Arabidopsis*, *rpl2* sequences coding for a stretch of 350 amino acids of L2 protein with a 1.4 kb intron are located in the mt genomes (Sunkel et al., 1994). Approximately one third of the coding

region of the gene at the 3' end is missing in these plants. The functional mt *rpl2* genes in these plants are expected to be in the nucleus and the L2 protein is targeted back to mitochondria to produce a functional mt ribosome. In addition, a partial cytosolic mRNA sequence in *Arabidopsis* (450 nt) homologous to the mt *rpl2* was also found in the databank (Ac.No. S49579).

The yeast mt L2 protein is larger (390 amino acids) than *E. coli* L2, but it does not have the plant mt L2 specific insert and it is encoded by a nuclear gene (Mason et al., 1996). In vitro reconstitution experiments in yeast indicate that mt L2 functions like the bacterial L2 (Mason et al., 1996). The mt L2 of the protist *R. americana* is also similar to bacterial one (282 amino acids) without any insert and it is encoded by the mt genome (Lang et al., 1997).

Similar to bacterial, cp and mt ribosomes, cytosolic ribosomes also contain L2 protein (260 a.a) and the corresponding gene is located in the nuclear genome. Complete cytosolic *rpl2* sequences have been characterized in tobacco (Marty and Meyer, 1992), tomato (Fleming et al., 1993), and in *Arabidopsis* (Ac. No. AL022373). Among them, tobacco and tomato were found to be regulated in a tissue and developmental stage specific manner. The young plant tissues and germinating seeds of tobacco have a higher transcript accumulation than adult tissues and dry seeds relative to H4 histone transcripts (Marty and Meyer, 1992). Cytosolic rp genes belong to a multigene family in contrast to the transferred mt rp genes which are usually single copy genes. There are two copies of cytosolic *rpl2* genes observed in tobacco which is an allotetraploid. The yeast cytosolic *rpl2* genes referred in the literature (Presutti et al., 1995) are not plant homologues instead the gene coding for L8.e (Ac.No. S30827) in yeast is homologous to plant cytosolic *rpl2*.

1.5 *Rationale and hypothesis*

This project is aimed at learning more about the plant mitochondrial ribosomal protein genes, their location and expression. The mt L2 protein coding genes are focused on here as there had been interesting differences observed among plant species such as tobacco, *Arabidopsis* and liverwort. The mt rp genes are essential and disruption or deletion of functional rp genes (maize mt *rps3* and *rpl16*) has been shown to be lethal or leading to impaired protein synthesis in mitochondria (Hunt and Newton, 1991). The L2 protein is found in all lineages examined and in all three (chloroplast, cytosol and mitochondrion) compartmental ribosomes. It is one of the important proteins in the assembly and function of the bacterial ribosomes. The highly conserved nature and the functional importance of ribosomal proteins in ribosomes suggest the importance of a functional mt *rpl2* gene for the production of L2 protein in plant mitochondria.

When the mt copy of *rpl2* is absent or truncated, functional gene transfer to nucleus is the simplest model by which the functional mt L2 protein is produced. Gene transfer to the nucleus is not always the case as there are at least three other possible explanations for the absence of a functional *rpl2* gene in the mt genome, since L2 has an essential function. 1. The possibility of an organellar gene loss and replacement by the cytosolic homologue as observed for the spinach chloroplast L23 (Bubenenko et al., 1994). 2. Another possibility is the fusion of genes or segments of genes that could perform more than one function as the *rps19* gene of *Arabidopsis* which has a fused 5'extension in addition to the mt target signals, which is proposed to carry out *rps13* function (Sanchez et al., 1996). It might also be possible for a gene to have both mt rp function and non-translatory functions, for example, yeast MRP-L4, (Graack et al., 1995). 3. A chloroplast counterpart if located in the nuclear genome may be recruited for the mt L2 protein

since there are examples of nuclear-encoded genes targeting its protein to both compartments (Smith et al., 1998). However, this is highly unlikely because all the examined plants have the cp *rpl2* gene in the chloroplast genome (Doyle and Doyle, 1995) and there is no known movement of proteins between mt and cp compartments due to the nature of their membranes.

The above possibilities make this study on mt L2 ribosomal protein genes interesting, with regard to mitochondrial gene location and expression. The expression of plant mt rp genes in the mt genome in relation to other respiratory chain genes at the level of transcription and post-transcription is also examined to determine the impact that differences in location can have. It was hypothesized that the functional gene responsible for the mt L2 protein is located in the nucleus in lineages where a functional copy is not present in the mt genome. In addition, the pattern of evolution of the L2 family of proteins in chloroplast and cytosolic compartments is analyzed to gain an insight into the mt L2 protein.

CHAPTER 2

MATERIALS AND METHODS

2.1 *Isolation of nucleic acids from plants*

2.1.1 *Plant mt DNA isolation*

Plant mt DNA from wheat (*Triticum aestivum* var Frederick), rice (*Oryza sativa* var v20B, v26B), soybean (*Glycine max* var Maple Arrow) and pea *(Pisum sativum* var Thomas Laxton) was isolated from 6 day old etiolated seedlings following the procedure of Wilson and Chourey, (1984) with modifications. Planting of seeds was done in a sterile vermiculite medium after treating the seeds serially with diluted Javex in H_2O (1:10) for 10 min followed by 10 mM HCl for 5 min and repeated washes with sterile distilled H_2O. Seedlings were grown for 6 days at room temperature in the dark. Harvested shoots were weighed, diced and homogenized immediately in a cold mortar and pestle and kept on ice using cold buffer I (0.44 M sucrose; 50 mM Tris pH 8.0; 3 mM EDTA; 1 mM β-mercaptoethanol; 0.1% BSA). The homogenate was filtered through 4 layers of cheese cloth and 1 layer of Miracloth (Calbiochem). The filtrate was centrifuged two times for 5 min at 2000 rpm to remove the plant cell debris. The supernatant was spun again at 10,000 rpm for 25 min to pellet mitochondria. The mitochondrial pellet was resuspended in 20 ml of buffer I and the above three spins were repeated to pellet mitochondria once again. The second set of spins helps to remove remnants of cell debris. The mitochondrial pellet was homogenized in buffer II (50 mM Tris pH 8.0; 20 mM EDTA) and lysed in buffer III (0.2 M Tris pH 8.0; 0.1 M EDTA; 0.2 M NaCl; 2% SDS; 0.2 M β-mercaptoethanol). The lysate was incubated at 65°C for 20 min and 5 M potassium acetate was added to a final concentration of 1.25 M. This mixture was incubated on ice for 45 min and centrifuged for 3 min at 12,000 rpm

at room temperature. The mt DNA was precipitated using 5 M ammonium acetate and isopropanol. The DNA pellet was obtained by spinning at 12,000 rpm for 5 min and the pellet was washed with 70% ethanol to remove the salt. The DNA was resuspended in buffer II and precipitation was repeated again with isopropanol and sodium acetate. The DNA pellets were dried under vacuum and resuspended in TE buffer (10 mM Tris-HCl pH 7.5; 1 mM EDTA) to a final concentration of 0.5 to 1.0 μg/μl. Fifty grams of 6 d old etiolated seedling yielded about 100 μg of crude mitochondrial DNA.

2.1.2 *Plant mitochondrial RNA isolation*

Mitochondrial RNA from 24 h germinating embryos and 6 d etiolated seedlings of rice and wheat were isolated using a method adapted from Bonen (1987). The 24 h germinating embryos were prepared by dissecting the embryos from dry seeds which were allowed to germinate in the dark at room temperature on wet filter papers containing 1% glucose in petri dishes. The 6 day etiolated seedlings were prepared as described in section 2.1.1. The mitochondrial isolation was as described earlier for the DNA. The mitochondrial pellet was homogenized in buffer IV (10 mM Tris HCl pH 7.5; 50 mM KCl; 10 mM $MgCl_2$) and lysed by adding an equal volume of buffer IV with 8% Triton- X100 and three volumes of 2X detergent mix (2% Tri-isopropylnapthalene sulfanate (Sigma), 12% Sodium p-amino salicylate (Sigma) 0.1 M NaCl, 20 mM Tris HCl pH 7.5). The lysate was incubated on ice with occasional vortexing and spun at 12,000 rpm for 5 min. The supernatant was extracted twice with phenol and the mitochondrial RNA was precipitated by 2.0 vol of 95% ethanol and 0.1 vol of 5 M NaCl.

2.1.3 *Isolation of total DNA*

Total DNA from rice, wheat, soybean and pea was isolated from 6 d etiolated seedlings by

the method described by Hattori et al. (1987). The harvested shoot tissues were frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. Twenty grams of frozen powder was transferred to a 250 ml polypropylene bottle and 200 ml of extraction buffer (0.1 M Tris-HCl; 0.05 M EDTA; 1% SDS, pH 8.0) was added. This mixture was extracted with 20-40 ml of phenol (phenol was equilibrated with TE pH 7.5 buffer containing 0.1% hydroxyquinoline) and the phenol phase was back-extracted with 20 ml of the extraction buffer. The combined aqueous phases were serially extracted with 40 ml of phenol, 40 ml of phenol/Sevag (1:1 mixture, Sevag is a 24:1 mixture of chloroform and iso-amyl alcohol) and 10 ml of ether. Ether was removed by placing the centrifuge bottles at 60°C for 15 min. An extraction with PVP (polyvinylpyrrolidone) slurry (2 g PVP washed in 0.1 M HCl and neutralized with 50 mM Tris-HCl, pH 8.0 and filtered through Whatman No. 1 paper), was done to remove the phenolic compounds present in the plant tissue. The extraction was carried out by the addition of PVP slurry to the plant extract followed by an incubation at room temperature for 10 min and centrifugation for 10 min at 7000 rpm. The supernatant was passed through Whatman No. 1 paper and then NaCl and PEG 8000 (Sigma) were added to the filtrate to achieve a final concentration of 0.5 M and 10% (W/V) respectively. This mixture was transferred to polypropylene tubes and incubated over-night at 4°C. The DNA pellet was collected by centrifugation at 8000 rpm for 10 min at 4°C. The supernatant was removed completely and the DNA was resuspended in 4 ml of TE. The RNA present in the DNA was removed by DNase-free RNase (Boehringer Manheim) treatment (50 μg/ml) for 30 min at 37°C. A phenol/Sevag extraction was carried out as above and the DNA was precipitated at -20°C in 2 vol of 95% ethanol, and ammonium acetate to a final concentration of 2.5 M. The DNA pellet was obtained by spinning at 10,000 rpm for 30 min at 4°C and washed twice with 70%

ethanol, dried at 37°C and resuspended in 2 ml of TE buffer. Approximately 10-50 μg of DNA per gram of plant tissue was recovered.

2.1.4 ***Isolation of total RNA***

Total RNA from plant tissue was isolated by the hot phenol method as described by Verwoerd et al. (1989). Plant tissues were harvested from 6 d etiolated seedlings, frozen quickly in liquid nitrogen and ground to a fine powder using mortar and pestle. The frozen plant tissues (10 g) were transferred to sterile polypropylene bottles and 50 ml of hot (80°C) extraction buffer [phenol-0.1 M LiCl; 100 mM Tris-HCl pH 8.0; 10 mM EDTA; 1% SDS (1:1)] was added. The above mixture was vortexed for 30 sec, and 25 ml of chloroform isoamyl alcohol (24:1) was added and mixed well. RNA was precipitated by spinning at 10,000 rpm for 5 min. The supernatant was mixed with one volume of 4 M LiCl and the RNA was allowed to precipitate over-night at 4°C. Then RNA was collected by centrifugation at 10,000 rpm at 4°C for 20 min. The pellet was dissolved in 1 ml of water and 2 volumes of 95 % ethanol and 0.1 vol of 3 M sodium acetate pH 5.5 were added and the RNA was stored as an ethanol-precipitate at -20°C. Approximately 2.0 mg of total RNA were obtained from 10 g of plant tissue.

2.1.5 ***Purification of poly A^+ RNA***

The poly A^+ mRNA was purified from total RNA using oligo (dT) cellulose (Sambrook et al., 1989). Approximately 0.5 g of oligo (dT) cellulose type 7 (Pharmacia) was used to purify mRNA from 1 mg of total RNA. The oligo (dT) cellulose was first dissolved in 1x binding buffer (0.01 M Tris -HCl pH 7.5; 0.5 M NaCl; 1 mM EDTA; 0.5% SDS) and saturated with 500 μl of 0.1 N NaOH by repeated spinning (8-10 times, 3 min each) in a 10 ml disposable tube at 4000 rpm in a table top, swinging bucket centrifuge. The oligo (dT) cellulose was then rinsed in 1x

binding buffer by repeated spinning (8-10 times). Approximately 1.0 mg of total RNA stored in 95% ethanol was centrifuged at 12,000 rpm for 20 min in a microtube centrifuge at 4°C and washed two times with 70% ethanol to remove the salt. The RNA pellet was dried under vacuum and dissolved in 500 µl of water and 500 µl of 2x binding buffer and denatured by incubation for 10 min. at 65°C and then cooled on ice. This RNA and 1 ml of oligo (dT) slurry in 1x binding buffer were mixed gently at room temperature in a sterile 10 ml tube. Two sterile 1 ml syringes were plugged with glass wool and the RNA/oligo (dT) mixture was loaded and centrifuged at 4000 rpm for 5 min. The filtrate was discarded and the unbound RNA was washed off the column by repeated spinning (3-4 times) for 2 min each after adding 500 µl of wash buffer (0.01 M Tris-HCl pH 7.5; 0.1 M NaCl; 1 mM EDTA). The poly A^+ RNA was eluted by spinning at 4000 rpm for 5 min after adding 150 µl of elution buffer (0.01 M Tris-HCl pH 7.5; 1 mM EDTA) for three times. The filtrates were pooled and 0.1 vol 3 M sodium acetate and 2.5 vol. 95 % ethanol were added and left at -20°C overnight. Approximately 5 µg of poly A^+ RNA was recovered from 1 mg of total RNA.

2.2 Polymerase Chain Reaction (PCR) techniques

2.2.1 *Polymerase Chain Reaction on DNA templates*

PCR amplifications were carried out on both mt and total DNA templates from rice, wheat, pea and soybean. The amount of template used in amplification was approximately 50 ng for mt DNA and 100 ng for total DNA with 100 ng of each primer. The Taq polymerase from Promega (1.25 units per reaction) and the manufacturer's buffer (Promega) were used with 1.5 mM $MgCl_2$ and 0.2 mM dNTPs. The annealing temperatures ranged from 45 to 55°C based on the GC content of the oligomers used. The program for the DNA thermal cycler (Perkin-Elmer

Cetus) used was segment 1: 94°C 3 min 1 cycle; segment 2: 94°C 1 sec., 94°C 45 sec., 45 to 55°C 30 sec., 45 to 55°C 1 min 30 sec., 72°C 1 sec., 72°C 2 min 30 cycles; segment 3: 72°C 5 min 1 cycle. Amplified products were separated on agarose gels and purified for further manipulations.

The list of oligomers used in various PCR and RT-PCR experiments for rice mt *rpl2*, and nuclear-encoded wheat mt *rpl2*, cytosolic *rpl2*, and mt *atpB* genes are given below.

Name	**Oligo # in thesis**	**Sequence (5'-3')**	**Databank Accession number**	**Coordinates**
LB 56	1	GGTCACCTAGTCGGAGTACCTC	Ac.No. D78336	3980-4001
LB 55	2	CTGTTTTTCACCGAGGGGGTGG	Ac. No.M68929	60854-76
LB 72	3	CTAGTACGATGGATCGAAGG	Ac.No. D78336	1332-1351
LB 63	4	GCAGATAACCGTAACACCTT	Ac.No. D78336	1273-1292
LB 77	5	CTCGCATGTTTCAGCTGGCA	Ac.No. D78336	1930-1949
LB 78	6	AGTCACGATTGGAAGCGGG	Ac.No. D78336	3455-3474
LB 61	7	ATAGATTCCCGATGCCGAGC	Ac.No. D78336	3851-3840
LB 62	8	GTTACAGACTCACTTATCCC	Ac.No. D78336	3085-3104
LB 166	9	ATTGGTTCATCGTGCTATGC	Ac.No. D78336	2152-2171
LB 167	10	CCCGAATCACACTG	Last of rice mt *rpl2* intron	
LB 161	11	GACTAGTGAGGCTATAGTAC	App. N0.2.	446-465
LB 162	12	TGTATTTGGTAACTCTCCTG	App. N0.2.	1007-1026
LB 163	13	TGGAAGGGAAGTTTTGTTGATGC	Ac.No. D78336	4130-4152
LB 164	14	AGCAAACTCTCCAAATTTATGACC	Ac.No. D78336	4307-4325
LB 171	15	TGAGGAGCAGGAAAATTTGG	Ac.No. D78336	4188-4207
LB 97	16	TTGAGTACGGGGAAGTCCGC	Ac.No. D78336	1161-1180
LB 89	17	GTTGGAGACGATGCCGATGG	Fig. 4.1.	513-532
LB 79	18	AAGGATCCAGCATACAAGGC	Fig. 4.1.	31-50
LB 80	19	TGCACGCTCGGCCTTGTATG	Fig. 4.1.	41-60
LB 98	20	ATGGGTCGTTTGATCAGAGC	Fig. 4.6.	1-20
LB 160	21	CTCCTTCCTGCAATAAGACC	Fig. 4.6.	706-725
LB 73	22	GTGTGCCAGGTCATCGGTGC	Ac.No. X02868	666-685

LB 75	23	TTTTTTTTTTTTTTTTTTTT	Oligo dT 20mer	
LB 83	24	AAGCTTTTTTTTTTG	Oligo dT with G at 3' end	
LB 84	25	AAGCTTTTTTTTTTC	Oligo dT with C at 3' end	
LB 85	26	AAGCTTTTTTTTTTA	Oligo dT with A at 3' end	
LB 174	27	GAAGTCCGCAGGGAGGAATT	Ac.No. D78336	1172-1191
LB 109	28	CATCCATACCCAAAATAGCAAT	Ac.No.X02868	3463-3484

2.2.2 *RT-PCR (Reverse transcription- polymerase chain reaction)*

Reverse transcription reactions were carried out using mt and cytosol transcripts to generate cDNA in order to amplify their sequences by PCR methods (Krug and Berger, 1996). Approximately 1 μg of mt RNA and 5 μg of total RNA or 1 μg of poly A^+ RNA were used to generate cDNA for the mt and cytosolic transcripts respectively. The primer concentration for mt gene specific transcripts was 20 ng and 100 ng for cytosolic transcripts. The oligo (dT) primer concentrations ranged from 100 to 500 ng. Primer annealing was performed by incubating the RNA primer mix at 70°C for 3 min and quick cooling on ice. The reverse transcriptase reactions were carried out either with 15 units of AMV (Promega) or 100 units of Superscript II (Gibco BRL) reverse transcriptases using the manufacturer's buffer. Other ingredients used in the reaction were 0.5 mM dNTPs, 10 units of RNase inhibitor (Promega) and 4 mM sodium pyrophosphate. The reactions were carried out at 42°C for AMV or up to 50°C for Superscript II for about 2 h in a 50 μl reaction volume. The reverse transcriptase activity was stopped by incubating the samples at 85°C for 15 min. The cDNAs were stored at -20°C. The PCR reactions on cDNA template were carried out using standard procedures.

2.2.3 *3' and 5'RACE (Rapid Amplification of cDNA Ends)*

Rapid amplification of cDNA ends (RACE) techniques were employed to amplify the 3'

end of the transcript as described by Frohman et al. (1988). Oligo dT primers which anneal to the poly A^- tail of the transcripts of interest were designed with an additional nucleotide G, C or A at its 3' end. The extra nucleotide at the 3' end of the primer helps to increase the specificity of reverse transcription and amplification of transcripts that end with any of these three nucleotides (Liang et al., 1993). Reverse transcription reactions were carried out as described earlier with 500 ng of primers and 1 μg of poly A^- RNA. PCR amplification experiments with a gene specific 'reverse' primer and oligo dT primer with one extra nucleotide at the 3' end were carried out. Only one primer set would amplify the product depending on the 3' nucleotide of the transcript preceding the poly A^- tail. The 3' RACE products of interest were gel purified and either sequenced directly or cloned into plasmid vectors.

To amplify the 5' end of a particular transcript, gene specific oligomers were first used to generate cDNA (Frohman et al., 1988). The reverse transcription was carried out as described in section 2.3.2. Excess primers in the cDNA pool were removed by spinning through a Sephadex-G50 column and quick precipitation with isopropanol and ammonium acetate. The 3' ends of cDNAs were extended with adenosine residues (0.2 mM dATP) with the help of 10 units of terminal deoxynucleotidyl transferase (Gibco BRL) using the manufacturer's buffer. This tailed cDNA was employed as a template in the PCR experiments with an oligo (dT) primer having a specific nucleotide in the 3' position and a gene specific primer.

2.3 Cloning and Screening methods

2.3.1 *Cloning plant mitochondrial DNA*

Plant mitochondrial (mt) DNA (approximately 5 μg) was digested with the restriction enzyme of interest and separated by electrophoresis on agarose gels with TBE buffer (50 mM

Tris-HCl, pH 8.0; 2 mM EDTA; 20mM boric acid). The DNA fragments of interest were sliced out of the gel and purified by the filter tip (BIO-RAD) method (Dean and Greenwald, 1995). DNA purification was carried out by spinning the filter tip carrying the gel slice at 4000 rpm for 15 min. The filtrate was collected, and precipitated by 2.5 vol. of 95% and ethanol 0.1 vol 3 M sodium acetate pH 5.5. DNA was pelleted by centrifugation at 12,000 rpm for 20 min at 4°C. The pellet was washed with 70% ethanol, dried under vacuum and the DNA was resuspended in low TE buffer (10 mM Tris pH 7.5; 1 mM EDTA). DNA fragments were ligated into dephosphorylated pUC (Pharmacia) vectors (15 ng) using T4 DNA ligase (Promega). *E. coli* TB1 strains were used in transformations. Clones of interest were selected by colony screening methods and DNA was isolated by boiling method. These cloning methods were carried out as described by (Sambrook et al., 1989).

Cloned mitochondrial DNA was either sequenced directly or the restriction fragments of interest were subcloned into M13 vectors for sequencing. *E. coli* JM 101 strain was used for transformation and the M13 single stranded DNA was prepared by the polyethyleneglycol (PEG) method (Sambrook et al., 1989).

2.3.2 *Cloning PCR products*

Because Taq polymerase adds an adenosine residue to the 3' end of the DNA strand during the PCR reactions, this residue has to be removed in order to clone the DNA fragment into a blunt end vector. The blunting reaction was followed by the method described by Maniatis et al.(1982), using DNA polymerase (Large fragment Klenow - Gibco BRL), which has a 3' exonuclease activity and T4 poly nucleotide kinase (T4 PNK- Pharmacia), which adds a phosphate group to the 5' ends of the PCR amplified fragments. The blunting reaction contained

approximately 1 μg of gel purified PCR product (using filter tip method), 10 units of Klenow, 10 units of T4 PNK, 1 mM ATP and DNA polymerase buffer (0.5 M Tris pH 7.5; 0.1 M $MgCl_2$; 10 mM dTT; 20 μM dNTPs; 0.5 mg/ml BSA). The reaction was carried at 37°C for 60 min in a 50 μl volume. This DNA was purified by Gene Clean II (BIO/CAN Scientific), prior to ligation into commercial dephosphorylated pUC Sma1 vector (15 ng). T4 DNA ligase (Pharmacia-2.5 units) was used in the ligation reaction as described by Sambrook et al., (1989). *E. coli* TB1 strains were used in the transformation and the clones of interest were identified by colony screening (Sambrook et al., 1989).

2.3.3 *Screening a soybean genomic lambda (λ) library*

A soybean genomic library (*Glycine max* var Resnik) from Clontech was used to screen for a nuclear-located mt *rpl2* gene. Soybean nuclear DNA partially digested with Sau3A had been cloned into λ EMBL3 SP6/T7 vectors and the inserts ranged from 8-20 kb with an average size of 15 kb. The library was amplified at least once by Clontech. *E. coli* strain LE 392 was maintained on Mg-free LB plates containing the appropriate medium additives as described by Sambrook et al. (1989). The titer (pfu/ml) for the soybean genomic library was determined by the following method. The library stock was diluted 10^{-2} to 10^{-6} times using 1x dilution buffer (35 mM Tris-HCl pH 7.5; 10 mM Mg SO_4; 0.2 M NaCl; 0.2% gelatin). Different dilutions of the genomic library in 200 μl of SM buffer (10 mM Tris pH 7.5; 50 mM NaCl; 8 mM Mg SO_4) were allowed to infect an overnight bacterial (LE 392) culture (200 μl) for 15 min at 37°C. Infected cells were plated on top agar in a prewarmed plate and grown for 6-8 h at 37°C. The titer value of the soybean genomic library was determined to be 3.4×10^9 pfu/ml (Pfu/ml= No. of plaques/ amount of DNA used x dilution factor) using a *E. coli* strain LE 392. Approximately 30,000 plaques were plated

in each plate (150mm dia) and transferred to nylon membranes (Amersham) by the procedure described by Sambrook et al., (1989). Phage membranes were hybridized with the DNA of interest and the potentially positive plaques identified were removed from the plate with agar and stored in 1 ml of SM buffer. These phages were plated again after determining the titer value and a second round of screening was carried out.

Recombinant λ DNA was isolated as described by Sambrook et al. (1989). Each positive plaque were grown in 5 ml of LB broth,10 mM Mg SO_4 and 50 µl of an overnight culture LE 392 *E. coli* cells. The λ phage was grown for 6-7 h at 37°C until the lysis of *E. coli* cells was complete and then spun at 8,000 rpm for 20 min at 4°C after adding few drops of chloroform. The supernatant (2 ml) and 10 µl of DNase 1 (Boehringer Manheim-2 mg/ml in 0.1 M Tris HCl pH 7.5) were incubated at 37°C for one hour. NaCl and PEG (8000 Sigma) were added to a final concentration of 0.5 M and 10% (W/V) respectively and incubated on ice for an hour. The phage were pelleted by centrifugation at 5000 rpm for 10 min at 4^0C and were slowly resuspended in 0.2 ml of SM buffer (10 mM Tris pH 7.5; 50 mM NaCl; 8 mM Mg So_4) on ice for one hour. The resuspended phage was incubated with 0.3 ml of chloroform. The aqueous phase was treated with 20 mg /ml of proteinase K and incubated at 65°C for 20 min with 3 µl EDTA pH 8.0 and 6 µl SDS 10%. Extractions were carried out sequentially in phenol, phenol/chloroform: isoamyl alcohol (24:1) and chloroform. DNA was precipitated with 2.5 vol. of 95% ethanol and 0.6 vol. of 5M NaCl. The DNA pellet was obtained by centrifugation, washed with 70%, ethanol then dried and resuspended in TE. RNA present in the λ phage DNA was removed by treating with 5µg of RNase at 37°C for 30 min.

2.4 Generation of ^{32}P radiolabelled probes for hybridization

2.4.1 *^{32}P-5' end labelling of oligonucleotides*

Synthetic oligomers (100 ng) were end labelled using γ -^{32}P- ATP (3000 Ci/mmole, Amersham 50 µci) at their 5' termini with T4 polynucleotide kinase (Pharmacia, 7 units) and kinase buffer (50 mM Tris-HCl pH 9.5; 10 mM $MgCl_2$; 5mM dithiothreitol). The reaction was carried at 37°C for 30 min. Labelled oligonucleotides (50 µl) were spun through a Sephadex G-50 (Pharmacia) column in a 1 ml syringe to remove the free radiolabel. The columns were previously equilibrated in TE buffer (Sambrook et al. 1989). The second or third eluates were collected for further use.

2.4.2 *^{32}P- labelling by second strand synthesis using M13 DNA templates*

Probes specific to M13 DNA inserts were obtained by synthesizing a radiolabelled complementary strand using the universal sequencing primer. Approximately 50 ng of single strand M13 DNA was used in the reaction, which contained 1.5 ng of the universal primer, 10 mM NaCl, 1.7 µM dNTPs (G,C,T), α-^{32}P- ATP; 50µci (3000 Ci/mmole, Amersham) and 2 units of Klenow (Gibco BRL). The primer annealing was carried out by heating the DNA and the primer to 65°C for 3 min and slow cooling to room temperature. The labelling reaction was carried out in 10 µl for 4 h at room temperature. After adding 40 µl TE, the probe was spun through a Sephadex G-50 column to remove the unincorporated ^{32}P. Probes which had at least 25% or more incorporation of radioactivity in comparison to the radiolabel remaining on the column were used for hybridizations after denaturation by boiling.

2.4.3 *^{32}P-labelling DNA fragments with random nonamers*

Double stranded DNA fragments were labelled by the random priming method using

nonamers (Megaprime labelling kit, Amersham). Approximately 25 ng of purified double stranded DNA was denatured and labelled with 50 μci α-^{32}P- ATP (3000 Ci/mmole, Amersham) using Klenow large fragment polymerase (Amersham), random nonamer mix and dNTPs (G,C,T) according to manufacturers protocol and spun through a Sephadex G-50 column. The specific activity of the probes were approximately $1.5x10^8$ dpm/μg. The probe was denatured by boiling and quick cooling on ice before use.

List of PCR product probes used in this study

Name	Oligomers	Template	Size
RPP1	Oligo # 1 & 3	Rice mt DNA	2750 bp
WPP1	Oligo # 1 & 3	Wheat nuclear-located mt *rpl2* copy I	640 bp
WPP2	Oligo # 27 & 28	Wheat mt *atpβ* cDNA	1300 bp
WPP3	Oligo # 1 & 21	Wheat cytosolic *rpl2* cDNA	635 bp
WPP4	Hpa2 fragment	Wheat nuclear-located mt *rpl2* copy I	360bp

2.4.4 *^{32}P-labelling DNA using PCR*

The PCR method of labelling DNA with α-^{32}P- ATP was done to generate uniformly labelled probes for S1 nuclease mapping and hybridization experiments. The PCR reaction mix was prepared as described in section 2.3.1, with the dNTPs concentration being 0.01 mM instead of 0.2 mM and an additional 15 μci α-^{32}P- ATP (3000 ci/mmole, Amersham). The program for the thermal cycler (Perkin-Elmer-Cetus) used was (Segment 1: 94°C, 3 min, 1 cycle; Segment 2: 94°C 1 min, 47°C 1 min, 72 °C 1 min, 20 cycles, Segment 3: 72 °C 7 min, 1 cycle). The PCR labelled probes were spun through a Sephadex G-50 column to remove the free radiolabel and denatured by boiling before use.

2.5 Southern and Northern hybridization experiments

Southern blots were prepared by standard procedures (Sambrook et al. (1989). The mitochondrial DNA separated in 0.75- 1% gels was first partially hydrolyzed for 15 min (0.25 N HCl), denatured (1.5 M NaCl: 0.5 M NaOH) for 30 min and neutralized (3 M sodium acetate pH 5.5) for 30 min. The gel was transferred to a nylon membrane (HyBond-N; Amersham) by capillary action in 20 x SSC (3M NaCl; 0.65 mM sodium citrate pH 7.0) overnight. The membrane was air dried and the DNA was cross-linked by UV light (254 nm) for 5 min.

For RNA gels, the dried RNA pellet was resuspended in 1 µl TE, 1 µl 10x MOPS buffer (Sambrook et al., 1989), 2 µl 37% formaldehyde (BDH), 5 µl deionized formamide and 1µl 4 ng/µl ethidium bromide. The samples were incubated at 65°C for 5 min and sterile RNA loading buffer (50% glycerol, 1mM EDTA, bromophenol blue) was added. The denatured RNA samples along with an RNA ladder (Gibco BRL) were run on 1.2% formaldehyde gels. Electrophoresis was performed at 100 volts for the first h and increased to 200 volts for the remaining 3-4 hours. After photography the RNA was transferred from the gel to a nylon membrane by capillary action overnight in 20 x SSC buffer.

The hybridizations using M13, random primed and PCR labelled probes were carried out in 50% deionized formamide (BDH), 5 x SSC, 0.5 % SDS, 50 mM $NaPO_4$ pH 7.0, 250 µg/ml sheared and denatured herring sperm DNA and 5 x Denhardt's solution (Sambrook et al., 1989). The temperature for pre-hybridization, hybridization and washings were 42°C except for heterologous probes (39°C or 37°C). Pre-hybridization was carried out for at least 3 h and hybridization was overnight. Membranes were washed at the respective hybridization temperatures twice in 2 x SSC, 0.1% SDS for 15 min each and then with 0.2 x SSC, 0.1% SDS

for 15 min. The end labelled oligonucleotide probes were hybridized in 5% deionized formamide (BDH), 5x SSC, 0.1% SDS, and 50 μg/ml yeast tRNA at 39°C overnight. The washes were done at the same temperature, but the more stringent wash was omitted. Membranes were exposed on Kodak XAR films for autoradiography at -80°C with intensifying screens for the required amount of time.

2.6 DNA sequence analysis

DNA sequencing was performed by the dideoxynucleotide termination method (Sanger et al., 1977). Sequencing reactions for single stranded M13 DNAs were carried out using Sequenase version 2.0 sequencing kit (US Biochemicals) and α-^{35}S- dATP (Amersham). Sequencing products were separated in 7% polyacrylamide sequencing gels.

Direct sequencing of PCR products were carried out with slight modifications in which 75 ng of primer was annealed to 500 ng of template (purified by filter tip method see section 2.3.1) in the presence of 0.5% Nonidet P-40. Samples were boiled for 3 min, quick-cooled in a -80°C ethanol bath, then gradually warmed to room temperature. The labelling reaction had four times more labelling mix, 0.5 % NP-40, Mn buffer and 1.5 times higher α-^{35}S- dATP.

2.7 Primer extension analysis

Primer extension analyses were carried out to map the 5' termini of transcripts of interest. Approximately 25 μg mt RNA was used with ^{32}P -5' end labelled oligonucleotides (25 ng) complementary to the RNA. Annealing (20 μl volume in TE pH 7.5) was carried out at 70°C for 3 min followed by quick cooling on ice. Reverse transcription was performed as discussed in section 2.2.2. with AMV (Promega, 15 units) or Superscript II (100 units) RT enzymes according to standard procedures (Krug and Berger, 1996). A DNA template of same template was used as a

sequencing ladder when possible. Reverse transcription was stopped as described earlier and products were resolved on sequencing gels to determine their size.

2.8 S1 nuclease mapping analysis

The method described by Sambrook et al., (1989) was used in S1 mapping experiments to determine the 5' or 3' terminus of transcripts. Uniformly labelled DNA fragments (PCR generated), spanning the expected position of the terminus of the transcript of interest was allowed to hybridize to RNA. Approximately 25 μg RNA was dissolved in 5 μl of water and 3 μl of 5 x PIPES buffer (0.2 M piperazine-N,N' -bis (2 ethanesulfonic acid) disodium salt pH 6.5, 5 mM EDTA, 2 M NaCl) and dried. The RNA pellet was dissolved in 12 μl of deionized formamide (BDH) and 3 μl of uniformly labelled DNA template (approximately 10 ng) was added and incubated at 85°C for 15 min. The mixture was cooled to 49°C and left overnight for hybrid formation. The RNA/DNA hybrids were slowly cooled to 30°C and 300 μl of S1 buffer (0.25 M NaCl; 30mM Sodium acetate pH 5.5; 1 mM Zn SO_4) with 20 μg/ml carrier salmon sperm DNA was then added. This mix was divided into four tubes containing 0, 50, 200, and 400 units of S1 nuclease (Boehringer Mannheim) and incubated at 30°C for 30 min. The reaction was stopped by adding 25 μl of stop solution (4 M ammonium acetate, 50 mM EDTA and 50 μg/ml carrier tRNA) and 240 μl 95% ethanol. The protected hybrids were precipitated at -20°C for one hour, centrifuged at 4°C for 20 min, washed twice with 70% ethanol, dried under vacuum and resuspended in 6 μl of TE. Typically one third of the product per lane was electrophoresed on 7% polyacrylamide sequencing gels.

2.9 Computer analyses of sequences

The sequence data were analyzed using Microgenie (Beckman) and BLAST programs

(Altschul, et al., 1990) to compare the sequences from Genbank or NBRF databanks. The multiple sequence alignments were carried out in GDE (genetic data environment) program using a Sun SPARC workstation.

2.9.1 *Sequence alignment*

The amino acid and nucleotide sequences were aligned using the ClustalW (Thompson et al., 1994) program. The settings used for the protein alignments were a weighting matrix of PAM 250 and the fixed and floating gap penalty of 10. Sequence alignments were corrected by visual inspection. Genbank accession numbers of the sequences used in the alignments are given in Appendix 5.

2.9.2 *Phylogenetic tree construction*

Protein distance matrix analysis was carried out by PHYLIP v 3.52c program (Felsenstein, 1993). A distance matrix of aligned sequences was generated using the program PROTDIST and the distances were then converted to phylogenetic trees using FITCH and neighbor joining methods. The Dayhoff PAM matrix was used as a model of substitution. Bootstrap resampling was performed with 100 replications using SEQBOOT, and a consensus unrooted tree was constructed using CONSENSE (Felsenstein, 1993). The PUZZLE v 4.01 program by Strimmer and von Haeseler, (1999) http://www.zi.biologie.uni-muenchen.de/~strimmer/puzzle.html was also used in tree construction. The model of substitution used was JTT and the model of heterogeneity was γ distribution. Appendix 5 has the Genbank accession numbers of protein sequences used in the analysis.

2.9.3 *Codon usage analysis*

The CODONS 1.4 program by Lloyd and Sharp (1992) was employed to determine the

nucleotide content and the frequency of codon usage for genes in the species examined. The names and accession numbers of genes used in the analysis are given in Appendix 6. The codon usage table is given in Appendix 7.

2.9.4 *Analysis of nucleotide substitution rates*

The rates of synonymous and nonsynonymous substitutions between the nuclear- and mt-located genes were estimated by the Li 93 program (Li, 1993). These substitution rates could be used to calculate the divergence time in conventional phylogenetic analysis using the molecular clock hypothesis. The formula used was $T = K/2r$ where, T is divergence time, r is rate of nucleotide substitution per site per year and K is number of substitutions per site.

CHAPTER 3

STATUS OF THE RIBOSOMAL PROTEIN L2 GENE IN VARIOUS PLANT MITOCHONDRIAL GENOMES

3.1 Status of the ribosomal protein L2 (*rpl2*) gene in the wheat mitochondrial genome

3.1.1 *Identification of rpl2 homologous sequences in the wheat mt genome*

During experiments to identify elements involved in the expression of the trans-spliced *nad1* gene encoding subunit 1 of NADH dehydrogenase, I identified a short stretch of mt *rpl2* related sequences upstream of the fourth *nad1* exon. The *nad1* gene requires special features for its expression since its exons are located in different parts of the genome and are spliced at the RNA level to produce the mature mRNA (Chapdelaine and Bonen, 1991). In order to gain knowledge about sequences responsible for the transcription of the separate *nad1* exons, sequences located in the upstream regions were analyzed and northern experiments were carried out to determine the presence of any stable transcripts. The wheat mt DNA upstream of *nad1d* was sequenced and used as probes in northern hybridizations. Interestingly, sequence analyses indicated the presence of a short (152 bp) reading frame with high similarity (86%) to the 3' terminus of the liverwort mt *rpl2* gene (Oda et al., 1992). This short *rpl2* reading frame does not have a start codon and is located approximately 1 kb upstream of the *nad1d* exon, which in turn is upstream of the *nad6* gene (Fig. 3.1.A & B). The size of the wheat mt *rpl2* sequences constitutes only 10% of the liverwort mt *rpl2* coding sequence (1500 bp). The derived L2 amino acid sequences share 78% and 56% identity with liverwort mt and *Escherichia coli* counterparts respectively (Fig. 3.1.C).

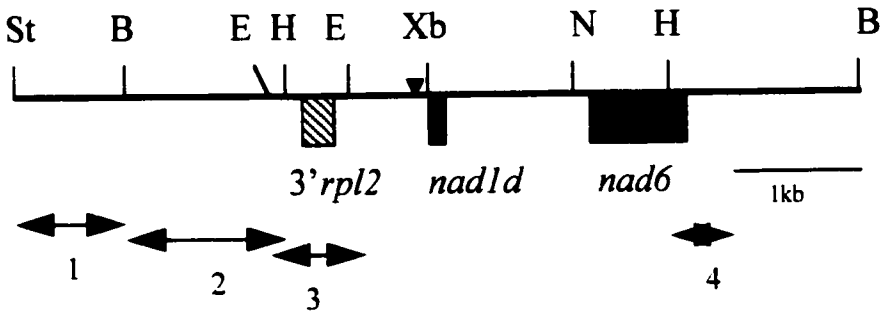
Figure 3.1 Characterization of the wheat mt *rpl2* 3'segment and comparison with liverwort mt and *Escherichia coli* L2 homologues.

A: Restriction map of the *rpl2* 3' segment, *nad1d* (exon 4) and *nad6* genes in the wheat mitochondrial genome. Selected restriction sites are shown, BamH1 (B), EcoR1 (E), Hind III (H), Xba I (Xb), Nar1 (N) and Sst 1 (St). The triangle indicates the breakpoint in homology relative to *Oenothera* (Wissinger et al., 1990), which is 135 nt upstream of *nad1* intron sequences. Arrows and numbers shown below the map indicate probes used in hybridization experiments.

B: Nucleotide sequence of the wheat mt *rpl2* 3' segment aligned with that of liverwort mt *rpl2* (Oda et al., 1992). The oligomer used in primer extension experiments is underlined and numbered (oligo # 1). The transcript 5' terminus is denoted by an arrow pointing downwards. The conserved promoter element with the CRTA motif (Covello and Gray, 1991) is shown as bold letters underlined. Asterisks indicate identical nucleotides. The stop codon is indicated by a bold asterisk. The sequences of the 0.5 kb EcoR1 fragment shown in the restriction map (A) are given here.

C: Amino acid alignment of derived mt L2 protein sequences from (1) the wheat *rpl2* 3'segment, (2) liverwort mt *rpl2* (Oda et al., 1992) and (3) *E. coli rpl2* (Zurawski and Zurawski, 1985). Identical amino acids are indicated with asterisks below. The position corresponding to the termination codon are indicated by bold asterisks.

A



B

wheat mt *rpl2*

GAATTCTTATCTATTTCTCGAGGCAATTATACTCATATCGACAATATGAACTGCATGAGAATTA**GCATAT** 70

R C

AGAAAGAAAGCTTTACCGAATACAGGTCTAGTAGGTATAGAGCCGTAAGCGTGGTGGGGGTGCCGGTGCT 140

liverwort mt *rpl2* AAGTTTAACAAAGCAGGACAAAGCCGGTGGT

** * oligo # 1 * *******

L G R C P I V H D V A M N P V D H P H G G G E G

TAGGCAGATGCCCCATTGTTCATGATGTTGCAATGAATCCAGTGGATCATCCTCATGGAGGAGGTGAGGG 210

TAGGCAGACGCCCCATTGTTCGGGGTGTTGCTATGAATCCAGTTGATCACCCTCATGGAGGCGGTGAAGG

******** ************ * ********* ******** ***** ********** ***** **

R T K G G R P S V S P W G K L T K A G F R A V

GCGCACGAAAGGAGGTAGACCTTCGGTGTCACCTTGGGGAAAGCTCACCAAAGCAGGATTTCGGGCAGTA 280

ACGCACTAAAGGAGGTAGACCTTCGGTATCACCTTGGGGCAAGCCTACCAAAGGTGGATTTAAAACAGTA

***** ********************* ********** **** ******* ****** *****

V G V V K L *

GTAGGGGTGGTGAAACTTTAGATGTCAAATGCCACCAATTTTTCTGCTCAAGTTAGAACGTAGTTTTTCG 350

GTAAGAAAACGCAGAAATTAG

*** * * * ****

AAGTGAATTGGCCAGTTCCTTTTAAAAGGAGACTGGTTTCTTTTTATAAGCGATTGGCTTCACTTTAGAT 420

TTTTTATCGTAATTTCTCTTTGATTTATGGGTTGTTATTTATTATTTCGATTTTTGATAAACGAAACGAA 490

TCCAAGCCCGACAAGTA**GAATTC** 513

C

```
Wheat:       RCLGRCPIVHDVAMNPVDHPHGGGEGRTKGGRPSVSPWGKLTKAGFRAVGKVGVVKL*
Liverwort:   RWLGRRPIVRGVAMNPVDHPHGGGEGRTKGGRPSVSPWGKPTKGGFKTVVRKRRN*
E. coli:     RWRGVRPTVRGTAMNPVDHPHGGGEGRNFGKHP-VTPWGVQTKGK-KTRSN-RTDKFIVRRRSK*
             *  *  * *   ***************  *  * * **
```

3.1.2 *Transcript analysis of the wheat mt rpl2 3' segment*

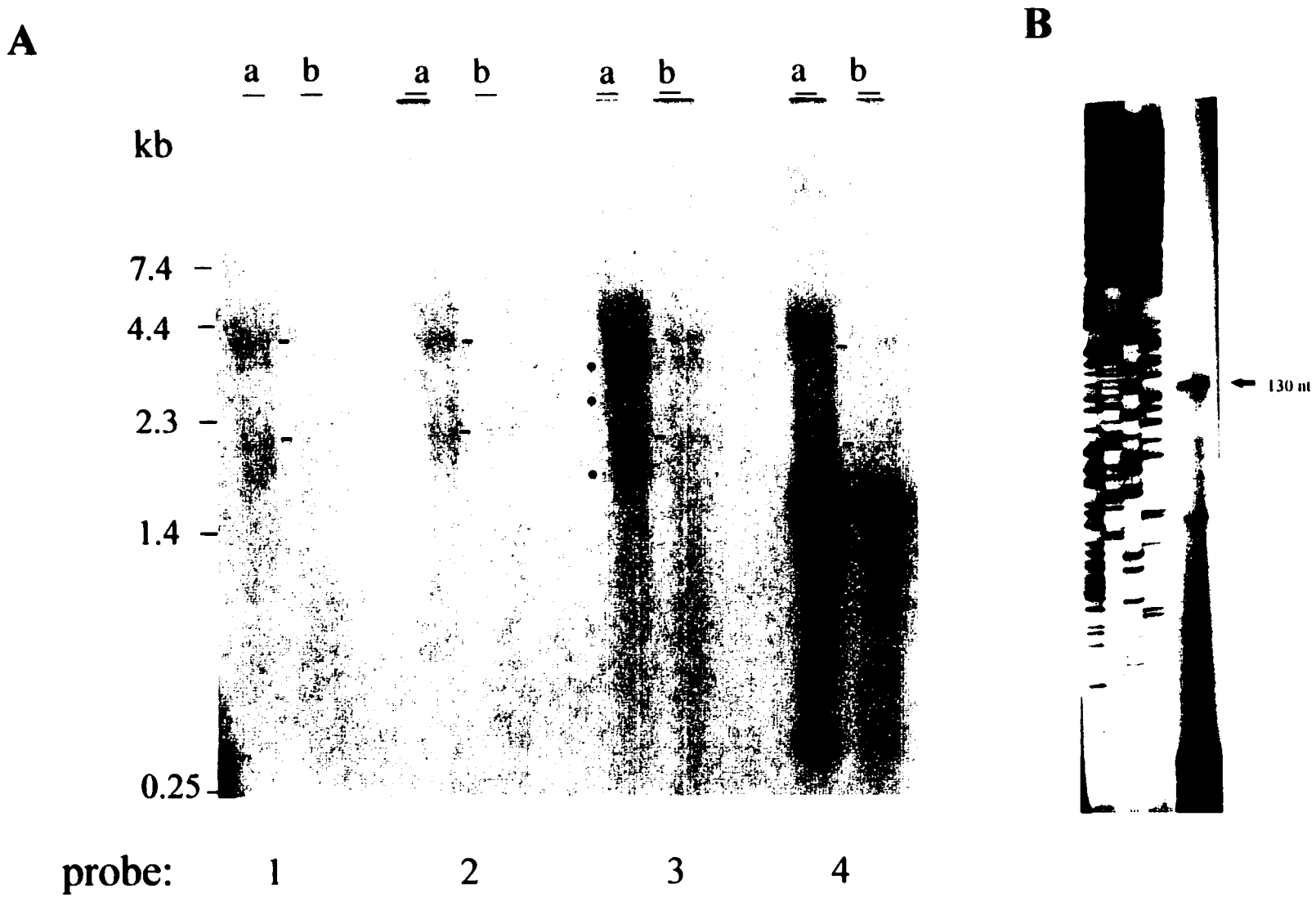
Northern hybridization was carried out for these *rpl2* homologous sequences using wheat mt RNA isolated from both germinating embryos (24 h) and etiolated seedling (6 d) stages, because transcript accumulation is found to vary considerably at these stages for certain wheat mt genes like *atp6* (Bonen, 1987). Only low levels of heterogeneous high molecular weight transcripts (approximately 2.1, 3.5, 4.0 kb) were detected at the 24 h stage and none were detected at the 6 d stage (Fig 3.2.A.probe #3 black dots). The other signals (1.9 kb and 3.5 kb) correspond to the nonspecific binding of probes to the rRNAs which are indicated by dashes. The 2.1 kb and 3.5 kb transcripts correspond to *nad1d-nad6* co-transcripts as reported by Haouazine et al. (1993) and Chapdelaine (1992). The *nad6* gene showed an abundant (approximately1.2 kb) transcript at both stages of development (Fig. 3.2.B probe # 4) and again the larger minor transcripts correspond to rRNA signals. The large transcripts observed for wheat mt *rpl2* 3'segment are hardly detected by the *nad6* probe possibly due to the shorter size (250 bp vs 500 bp) of the probe used (Fig. 3.1.A).

In order to determine the position of the 5'ends of wheat mt *rpl2* transcripts, primer extension analysis was carried out on wheat mt RNA (6 d). The wheat mt *rpl2* 3'segment specific primer (oligo # 1) generated an extension product with a major transcript terminus at 61 nt upstream of the *rpl2* homologous sequences (Fig. 3.2.B) consistent with the co-transcription of the mt *rpl2* 3'segment and the downstream *nad1d - nad6* region to yield transcript sizes of 3.5 kb and 4.0 kb. The different sizes of transcripts (2.1, 3.5 and 4.0 kb) may indicate differences at the 3' terminus which is consistent with the northern analysis. The region upstream of the *rpl2* 3'segment has no stable detectable transcripts at either stage (Fig. 3.2.A. probe#1 & 2), only a

Figure 3.2 Transcript analysis of the wheat mt region containing the *rpl2* 3'segment

A: Northern blot hybridization analysis: Wheat mt RNA isolated from 24 h germinating embryos (a) and 6 d etiolated seedlings (b) were used with probes 1-4 as indicated in Fig.3.1.A. Probes 1 and 2: Wheat mt 3'*rpl2* upstream region, probe 3: Wheat mt 3'*rpl2* sequences and probe 4: *nad6* sequences. The RNA size marker is shown in the left side of the panel. The 2.1, 3.5 and 4.0 kb transcripts are indicated by dots and the rRNAs by dashes. The amount of RNA loaded for 24 h and 6 d stages were standardized based on equivalent 18S rRNA amounts (data not shown).

B: Primer extension analysis of wheat mt *rpl2* transcripts from 6 d seedlings: The location of the primer (oligo # 1) is shown in Fig. 3.1.B. The size of the primer extension product (130 bp) indicated by an arrow based on the DNA sequencing ladder on left side indicates a major 5'terminus 61 nt upstream of the mt *rpl2* homologous stretch.


A
a b a b a b a b
kb
7.4
4.4
2.3
1.4
0.25
probe: 1 2 3 4
B
130 nt

non-specific binding of rRNA was observed.

The sequences immediately upstream of the 5'terminus were examined and a conserved motif (GAATTAGCATATAGAAAGA) that resembles a monocot mt gene promoter (Covello and Gray, 1991-rAaannGCRTAtArtragt), which is present adjacent to the 5'termini of several wheat mt gene transcripts (*cox1, cox2, atpA* and *orf 25*) was identified (Fig. 3.1.B). Together the northern and transcript terminus analyses suggest that the wheat mt *rpl2* sequences are co-transcribed with *nad1d* exon from this promoter although guanylyl transferase capping experiments would be needed to determine this. The transcription of *nad1* exons is controlled by different promoters since they are located in distant locations in the wheat mt genome. The transcription control also varies between plants since the location of exons are different. For example, the homology between wheat and *Oenothera nad1d* exon ends 135 nt upstream of the *nad1d* exon. The first trans-spliced *nad1* exon is located downstream of one copy of the 18S rRNA gene (Chapdelaine, 1992) and the cis-spliced second and third exons are located downstream of the *atp6/rps13* genes in wheat (Chapdelaine and Bonen, 1991).

3.2 Status of the *rpl2* gene in the mitochondrial genomes of wheat, rice and soybean

To further investigate whether this *rpl2* segment was part of a functional gene or if another full copy was present elsewhere in the wheat mt genome, Southern hybridization experiments were carried out. Results indicated that this *rpl2* 3' segment is a single copy sequence (Fig. 3.3.A & B and data not shown). While carrying out this analysis in wheat, I also isolated and characterized the rice mt *rpl2* gene which is complete (2990 bp) and has a group II intron (see section 3.2.1) and used those clones as probes to examine the wheat mt genome for the presence of the rest of the mt *rpl2* gene. When a rice mt *rpl2* probe (RPP1 see section 2.4.3. for the list of

PCR product probes) having a 70 nt overlap with the wheat 3'mt *rpl2* segment (Fig.3.3.A) was used to probe wheat mt DNA digests, a single signal was observed corresponding to the mt *rpl2* sequences located upstream of the *nad1d* exon. This suggests that the mt *rpl2* 5' coding sequences are not present anywhere in the wheat mt genome (Fig. 3.3.A). This observation was supported by other Southern hybridization experiments. For example, the rice mt *rpl2* 5'exon (650 bp) alone showed no signal (Fig. 3.3.B).

Southern analyses of mt DNA from a monocot (rice) and a dicot plant (soybean) for the presence of the mt *rpl2* gene indicated a single copy of mt *rpl2* sequences in rice mt genome and none in soybean (Fig. 3.3.A. probe RPP1). The rice mt genome also has distantly related intron sequences present elsewhere (data not shown). The hybridization conditions were at reduced stringency and were able to detect the mt *orf25* sequences in soybean mt DNA using wheat mt *orf25* sequences sharing approximately 90% identity. The soybean mt DNA blot used in Fig. 3.3.A. was reprobed with wheat *orf25* probe which detects the two soybean *orf25* copies (Fig 3.3.C).

Since the mt DNAs of flowering plants are made up of multiple copies of subgenomic molecules in different stoichiometric amounts (reviewed by Bonen, 1998), it might be possible that such molecules contain *rpl2* sequences and could be detected by PCR experiments on wheat mt DNA. Mitochondrial *rpl2* gene specific oligomers were designed from regions that are conserved among the three plant *rpl2* sequences, namely the wheat mt 3' *rpl2* segment (Fig. 3.1.B oligo # 1) and mt *rpl2* genes of liverwort (oligo # 2) and rice (oligo # 3-8). The list of mt *rpl2* oligomers and their sequences are shown in the Materials and Methods chapter. PCR amplification experiments using oligomer pairs (oligo # 1 & 2, 1 & 3, 1 & 8, 1 & 6, 3 & 5) with

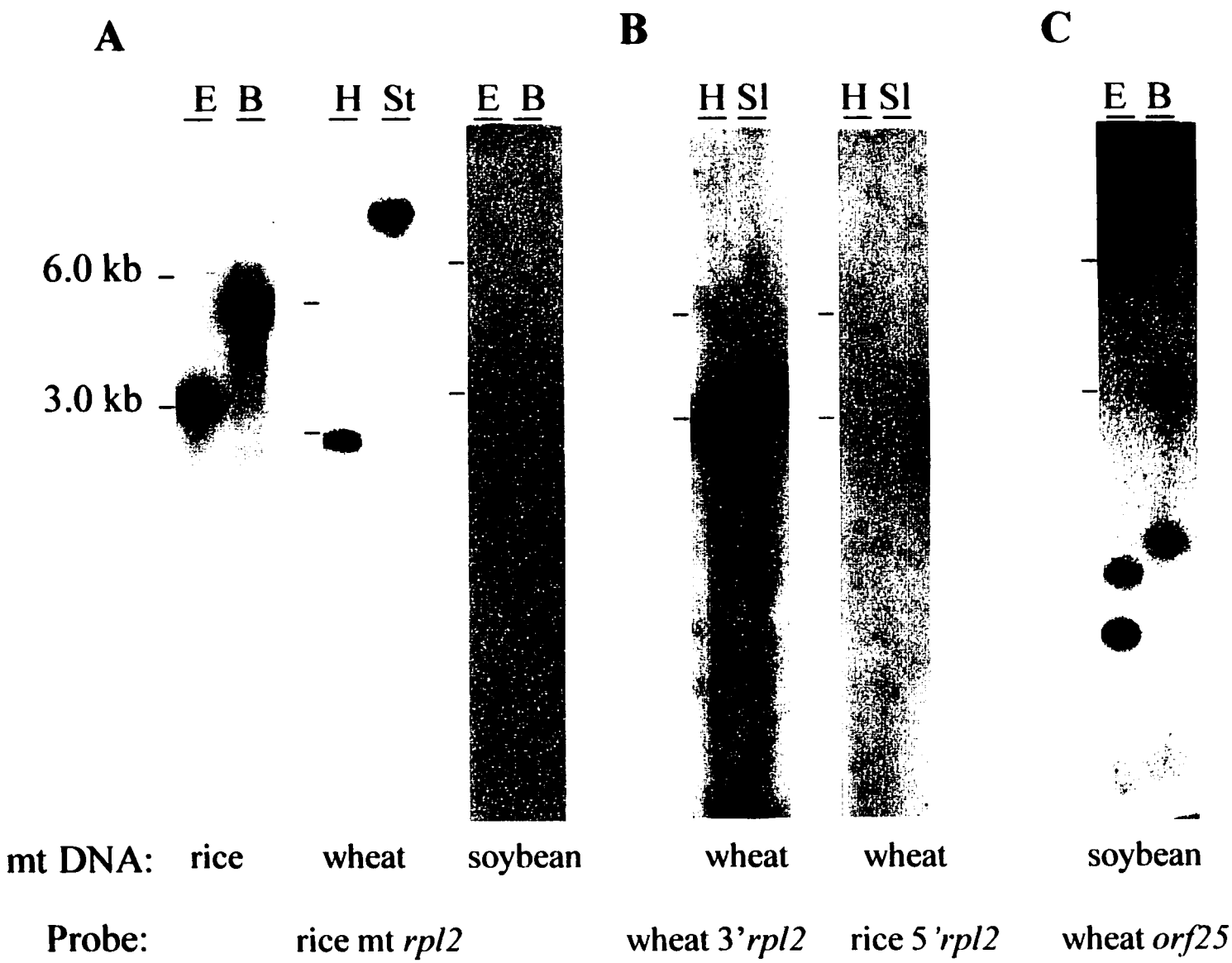
Figure 3.3 Southern hybridization analysis of mt *rpl2* to find out its status in wheat, rice and soybean

A: DNA blots of rice, wheat and soybean mt DNA digests probed with the rice mt *rpl2* gene (RPP1)

B: Wheat mt DNA blots hybridized with the wheat *rpl2* 3' segment (left lane) and rice mt *rpl2* exon1 probe of 650 bp (right lane).

C: Soybean blot shown in (A) reprobed with the wheat *orf25* gene.

Size markers 6.0 kb and 3.0 kb are shown by horizontal lines. Restriction enzymes used are E (EcoR1), B (BamH1), H (Hind III), St (Sst 1) and Sl (Sal1).


A
B
C
E B
H St
E B
H Sl
H Sl
E B
6.0 kb
3.0 kb
mt DNA:
rice
wheat
soybean
wheat
wheat
soybean
Probe:
rice mt rpl2
wheat 3'rpl2
rice 5'rpl2
wheat orf25

wheat mt DNA template generated no products consistent with the Southern results. Plant mt gene specific primers can almost always be successfully used for both monocot and dicot plants, since they have high sequence similarity (90-95%). Hence, negative PCR experiments on wheat mt DNA is consistent with the absence of 5'sequences or substoichiometric DNA molecules carrying the *rpl2* gene. The region upstream of the wheat mt *rpl2* 3'segment was sequenced (Appendix 1), to find out whether there might be an intron linking it to a divergent 5'*rpl2* exon. The plant mt introns are classified as group II intron and are typically 1 to 2 kb and do not exceed 3 kb. The absence of mt *rpl2* related sequences and the conserved group II intron domain 5 and 6 sequences (reviewed by Michel and Ferat, 1995) in the upstream region (~ 4 kb) suggests that wheat mt *rpl2* piece is not a part of a cis-spliced gene. RT-PCR experiments using oligomers (oligo # 1 & 2 and 1 & 3) on wheat mt RNA were negative, again suggesting the absence of mt *rpl2* mRNAs spliced either in cis or in trans.

The wheat mt *rpl2* 3'segment has higher non-synonymous substitutions than synonymous when compared with the rice mt *rpl2* gene (8 non- synonymous vs 2 synonymous), consistent with a lack of functional constraint. Besides, the wheat mt *rpl2* 3' segment does not have an initiation codon and has only 10% of the mt *rpl2* coding region. In addition, only low levels of high molecular weight transcripts were detected and there was no evidence for the presence of cis- or trans-spliced *rpl2* 5' sequences. Based on these data, the wheat mt *rpl2* 3' segment is believed to be a pseudogene.

3.2.1 ***Rice mt genome has an intact mt rpl2 gene***

3.2.1.1 *Characterization of the rice mt rpl2 gene*

The rice mt *rpl2* sequences were obtained by PCR, using the gene specific 5'oligomer

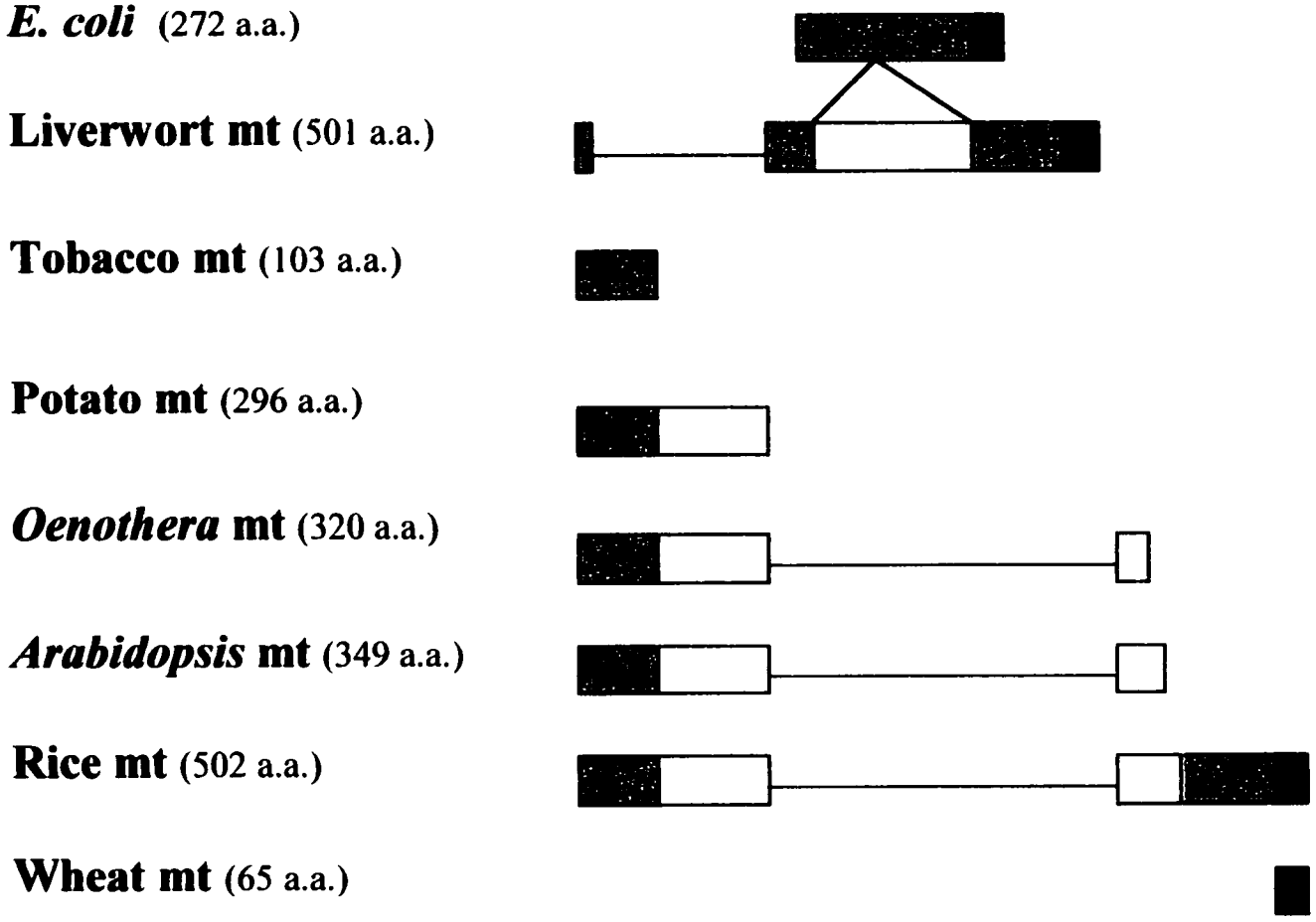
(oligo # 2) designed from liverwort mt *rpl2* sequences, and the 3'oligomer (oligo # 1) from wheat mt *ψrpl2*. Sequence analysis of the 2.7 kb product revealed the presence of mt *rpl2* sequences with an intron. Inverse-PCR techniques were carried out to obtain the extreme termini of the coding sequences and the flanking regions. Based on Southern data (Fig. 3.3.A) that the rice mt *rpl2* gene is located on a 3.0 kb EcoR1 fragment, the rice mt DNA was digested with EcoR1, ligated into circular molecules (Fig. 3.6.C), and used as template for inverse- PCR experiments (oligo # 4 & 7). Direct sequencing of the inverse-PCR products gave the remainder of the *rpl2* coding region and revealed the presence of the mt *rps19* gene in the downstream region. A partial rice mt DNA pUC plasmid library was also screened with the rice mt *rpl2* probe (RPP1) and a 5.0 kb BamH1 clone containing both *rpl2* and *rps19* was isolated, subcloned into M13 vectors and sequenced (Appendix 2). The sequencing of part of the rice mt *rpl2* intron was carried out by S. Grover (a summer student in the lab).

During the isolation and characterization of the rice mt *rpl2* gene in this study, Kubo et al., (1996) simultaneously characterized and published similar work. The rice mt *rpl2* gene consists of two exons (840 bp and 669 bp) separated by a group II intron of 1481 bp. The position of the rice mt *rpl2* intron (Fig. 3.4) is identical to the ones located in the truncated mt *rpl2* genes of *Oenothera* (Sunkel et al., 1994) and *Arabidopsis* (Unseld et al., 1997). The sections of mt *rpl2* sequences present in different plants are shown in Figure 3.4. The black blocks represent the highly conserved C-terminus region present in wheat, but not in *Oenothera* and *Arabidopsis*. A truncated mt 5' *rpl2* segment in tobacco (Vitart et al., 1992) and complete 5' exon in potato (Ac.No. AF095278) were identified in their mt genomes but it is not known whether they also have the complete mt *rpl2* gene. The predicted

Figure 3.4 Schematic of *rpl2* sequences identified in mitochondrial genomes of various plants and comparison with *E. coli*

Data are given for *E. coli* (Zurawski and Zurawski, 1985), liverwort (Oda et al., 1992), *Arabidopsis* (Unseld et al., 1997), *Oenothera* (Sunkel et al., 1994), tobacco (Vitart et al., 1992), potato (Ac. No. AF095278), rice (this work & Kubo et al., 1996) and wheat (this work). The *E. coli* homologous regions are shown as gray blocks, the plant mt specific internal insert as white blocks and the highly conserved carboxyl terminus present in wheat as black blocks. Lines represent introns. The predicted numbers of amino acids are given in brackets. The wheat mt *rpl2* segment does not have a start codon.

rpl2 gene segments in mitochondrial genomes of different plants

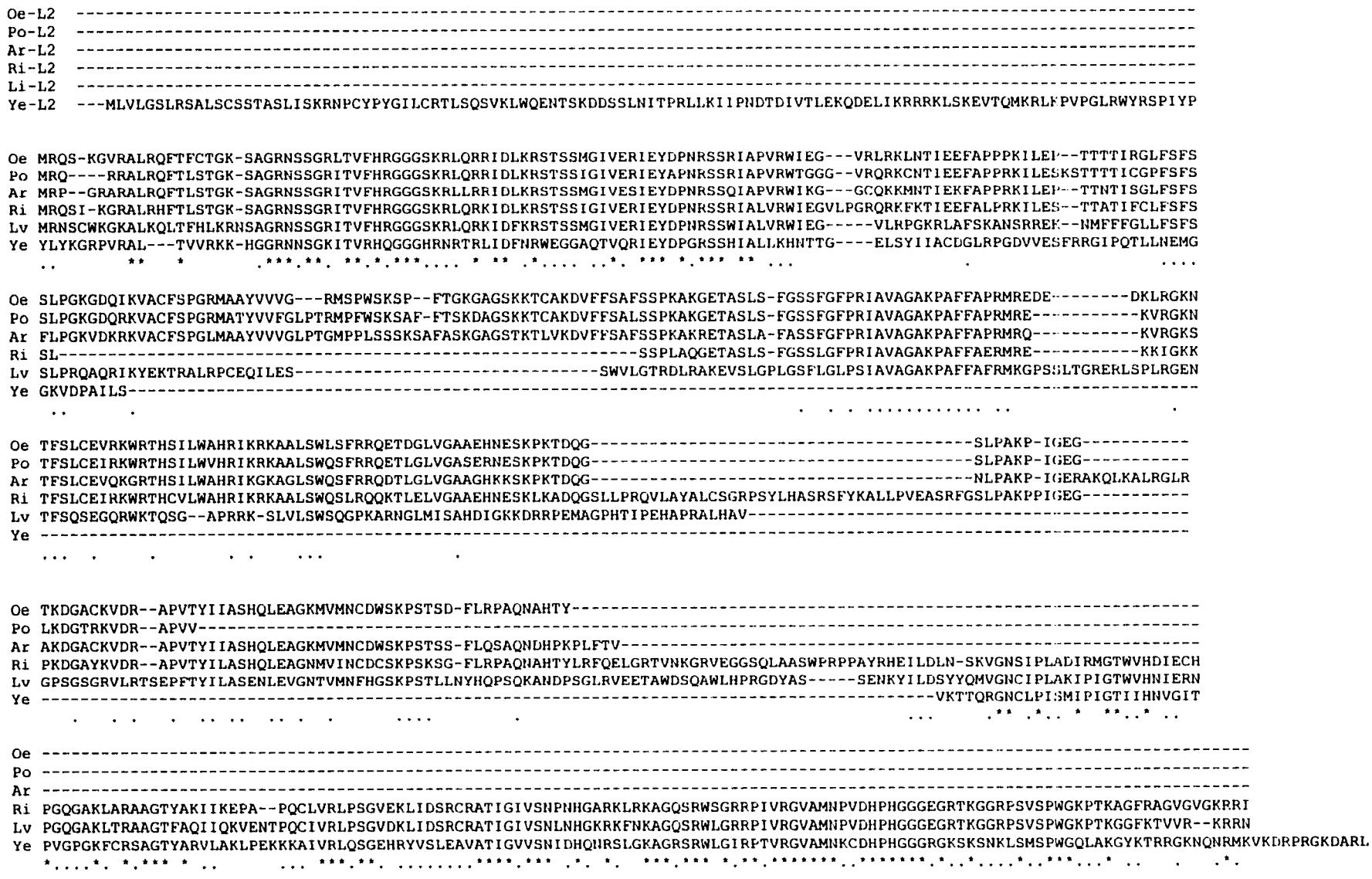


rice and liverwort L2 proteins have 502 and 501 amino acids respectively, with a 250 amino acid inserted sequence in the middle of the protein (Fig. 3.4) when compared with the *E. coli* counterpart (Zurawski and Zurawski, 1985; Takemura et al., 1992). The liverwort and rice mt L2 derived proteins share 33% identity at the amino acid level in the insert region in contrast to 72% in the rest of the protein (Fig. 3.5). The size of the mt L2 protein of the protist *R. americana,* is similar (282 amino acids) to the size of bacterial ones without any inserted sequences (Lang et al., 1997). In contrast, the nuclear-encoded mt L2 homologue of yeast (Accession No.S46947) has additional sequences (390 amino acids) that include unrelated inserted sequences and target peptide sequences when compared with plants (Fig.3.5).

The complete rice mt *rps19* gene (93 codons) located three base pairs downstream of the mt *rpl2,* was cloned and sequenced. The organization of the rice mt *rpl2* and *rps19* genes are found to be conserved both in the *E. coli* S10 operon *(rps10-rpl3-rpl4-rpl23-**rpl2-rps19**-rpl22-rps3-rpl16-rpl27-rps17)* (Zurawski and Zurawski, 1985) and the rp gene cluster (Fig. 6.2) of the liverwort mt genome (*rps10-**rpl2-rps19**-rps3-rpl16-rpl5-rps14-rps8-rpl6-rps13-rps11-rps1*) (Takemura et al., 1992). Further sequence analyses indicated the presence of a respiratory chain gene *nad4L*, coding for the NADH dehydrogenase subunit 4L located 861bp downstream of the *rps19* gene. The *nad4L* gene encodes a protein of 100 amino acids (Appendix 3) and is also a single copy gene like *rpl2* and *rps19* in rice mitochondria (data not shown). The ACG at the position of the initiation codon of *nad4L* undergoes editing as do nine other positions, which cause conservative amino acid changes (Appendix 3). This was determined by RT-PCR experiments. A 450 bp region, located 100 bp downstream of the *nad4L* stop codon does not match with any published sequences in the databank. In the region upstream of the rice mt *rpl2*

Figure 3.5 Alignment of plant and yeast mt L2 derived proteins

The derived L2 proteins of the intact mt *rpl2* gene of rice (this work and Kubo et al., 1996 D78336), liverwort (P26859) and yeast (S30827) and the truncated *rpl2* genes of *Arabidopsis* (S49579), *Oenothera* (S46947) and potato (AF095278) were aligned. The location of these genes are in the mitochondrion except for yeast which is in the nucleus with target peptide sequence. Asterisks and dots indicate the conserved amino acid positions among all proteins or among only the plant mt L2 sequences respectively.


Oe-L2
Po-L2
Ar-L2
Ri-L2
Li-L2
Ye-L2 ---MLVLGSLRSALSCSSTASLISKRNPCYPYGILCRTLSQSVKLWQENTSKDDSSLNITPRLLKIIPNDTDIVTLEKQDELIKRRRKLSKEVTQMKRLKPVPGLRWYRSPIYP
Oe MRQS-KGVRALRQFTFCTGK-SAGRNSSGRLTVFHRGGGSKRLQRRIDLKRSTSSMGIVERIEYDPNRSSRIAPVRWIEG---VRLRKLNTIEEFAPPPKILEI--TTTTIRGLFSFS
Po MRQ----RRALRQFTLSTGK-SAGRNSSGRITVFHRGGGSKRLQRRIDLKRSTSSIGIVERIEYAPNRSSRIAPVRWTGGG--VRQRKCNTIEEFAPPRKILESKSTTTTICGPFSFS
Ar MRP--GRARALRQFTLSTGK-SAGRNSSGRITVFHRGGGSKRLLRRIDLKRSTSSMGIVESIEYDPNRSSQIAPVRWIKG---GCQKKMNTIEKFAPPRKILEI--TTNTISGLFSFS
Ri MRQSI-KGRALRHFTLSTGK-SAGRNSSGRITVFHRGGGSKRLQRKIDLKRSTSSIGIVERIEYDPNRSSRIALVRWIEGVLPGRQRKFKTIEEFALPRKILES--TTATIFCLFSFS
Lv MRNSCWKGKALKQLTFHLKRNSAGRNSSGRITVFHRGGGSKRLQRKIDFKRSTSSMGIVERIEYDPNRSSWIALVRWIEG-----VLRPGKRLAFSKANSRREK--NMFFFGLLFSFS
Ye YLYKGRPVRAL---TVVRKK-HGGRNNSGKITVRHQGGGHRNRTRLIDFNRWEGGAQTVQRIEYDPGRSSHIALLKHNTTG----ELSYIIACDGLRPGDVVESFRRGIPQTLLNEMG
Oe SLPGKGDQIKVACFSPGRMAAYVVVG---RMSPWSKSP--FTGKGAGSKKTCAKDVFFSAFSSPKAKGETASLS-FGSSFGFPRIAVAGAKPAFFAPRMREDE--------DKLRGKN
Po SLPGKGDQRKVACFSPGRMATYVVFGLPTRMPFWSKSAF-FTSKDAGSKKTCAKDVFFSALSSPKAKGETASLS-FGSSFGFPRIAVAGAKPAFFAPRMRE----------KVRGKN
Ar FLPGKVDKRKVACFSPGLMAAYVVVGLPTGMPPLSSSKSAFASKGAGSTKTLVKDVFFSAFSSPKAKRETASLA-FASSFGFPRIAVAGAKPAFFAPRMRQ----------KVRGKS
Ri SL----------SSPLAQGETASLS-FGSSLGFPRIAVAGAKPAFFAERMRE----------KKIGKK
Lv SLPRQAQRIKYEKTRALRPCEQILES----------SWVLGTRDLRAKEVSLGPLGSFLGLPSIAVAGAKPAFFAFRMKGPSSLTGRERLSPLRGEN
Ye GKVDPAILS----------
Oe TFSLCEVRKWRTHSILWAHRIKRKAALSWLSFRRQETDGLVGAAEHNESKPKTDQG----------SLPAKP-IGEG----------
Po TFSLCEIRKWRTHSILWVHRIKRKAALSWQSFRRQETLGLVGASERNESKPKTDQG----------SLPAKP-IGEG----------
Ar TFSLCEVQKGRTHSILWAHRIKGKAGLSWQSFRRQDTLGLVGAAGHKKSKPKTDQG----------NLPAKP-IGERAKQLKALRGLR
Ri TFSLCEIRKWRTHCVLWAHRIKRKAALSWQSLRQQKTLELVGAAEHNESKLKADQGSLLPRQVLAYALCSGRPSYLHASRSFYKALLPVEASRFGSLPAKPPIGEG----------
Lv TFSQSEGQRWKTQSG--APRRK-SLVLSWSQGPKARNGLMISAHDIGKKDRRPEMAGPHTIPEHAPRALHAV----------
Ye ----------
Oe TKDGACKVDR--APVTYIIASHQLEAGKMVMNCDWSKPSTSD-FLRPAQNAHTY----------
Po LKDGTRKVDR--APVV----------
Ar AKDGACKVDR--APVTYIIASHQLEAGKMVMNCDWSKPSTSS-FLQSAQNDHPKPLFTV----------
Ri PKDGAYKVDR--APVTYILASHQLEAGNMVINCDCSKPSKSG-FLRPAQNAHTYLRFQELGRTVNKGRVEGGSQLAASWPRPPAYRHEILDLN-SKVGNSIPLADIRMGTWVHDIECH
Lv GPSGSGRVLRTSEPFTYILASENLEVGNTVMNFHGSKPSTLLNYHQPSQKANDPSGLRVEETAWDSQAWLHPRGDYAS-----SENKYILDSYYQMVGNCIPLAKIPIGTWVHNIERN
Ye ----------VKTTQRGNCLPISMIPIGTIIHNVGIT
Oe ----------
Po ----------
Ar ----------
Ri PGQGAKLARAAGTYAKIIKEPA--PQCLVRLPSGVEKLIDSRCRATIGIVSNPNHGARKLRKAGQSRWSGRRPIVRGVAMNPVDHPHGGGEGRTKGGRPSVSPWGKPTKAGFRAGVGVGKRRI
Lv PGQGAKLTRAAGTFAQIIQKVENTPQCIVRLPSGVDKLIDSRCRATIGIVSNLNHGKRKFNKAGQSRWLGRRPIVRGVAMNPVDHPHGGGEGRTKGGRPSVSPWGKPTKGGFKTVVR--KRRN
Ye PVGPGKFCRSAGTYARVLAKLPEKKKAIVRLQSGEHRYVSLEAVATIGVVSNIDHQNRSLGKAGRSRWLGIRPTVRGVAMNKCDHPHGGGRGKSKSNKLSMSPWGQLAKGYKTRRGKNQNRMKVKDRPRGKDARL

gene, that is 334 bp upstream of start codon, there are sequences homologous to the cp *rpoB* gene in the opposite orientation and they presumably have arisen from chloroplast DNA integrated into the rice mt genome (see Fig 3.7.A).

3.2.2 *Expression of the rice mt rpl2-rps19-nad4L gene cluster*

3.2.2.1 *The rice mt rpl2 is transcribed, spliced and edited*

Northern hybridization experiments on rice mt RNA using mt *rpl2* sequences as probes indicated lower levels of stable transcripts than for the *nad4L* gene which is located downstream (see Fig. 3.7 B). The splicing of the rice mt *rpl2* mRNA was detected by RT-PCR experiments using the mt *rpl2* gene specific oligomers (oligo # 1 & 2 or 1 & 3) (see Fig. 4.1). The size of the amplification products (2.7 kb and 1.2 kb) was consistent with the presence of precursor RNAs or DNA and spliced exon sequences devoid of intron. Direct sequencing of the 1.2 kb product confirmed the proper splicing of exons. The sequences around the splice junction are shown in Fig. 3.6.A. In addition a single RNA editing site in the second exon which changes a proline codon to a conservative serine one was also identified (Fig. 3.6.B) when a cloned Hind III fragment (0.6 kb) of the RT-PCR product was sequenced. Kubo et al. (1996) have also detected this single editing site in all 10 cDNA clones they examined.

3.2.2.2 *Transcript analysis of the mt rpl2-rps19-nad4L genes in rice embryos and seedlings*

Northern hybridization experiments were conducted for the rice mt *rpl2* gene region using rice mt RNA isolated from germinating embryos (24 h) and etiolated seedlings (6 d). A complex transcript profile with at least 7 different transcript sizes 7.0, 5.0, 3.5, 3.0, 2.6,1.7 and 1.5 kb (Fig. 3.7.B. probe 3) with differences between the two stages relative to rRNA levels being observed. Specific breakdown of transcripts may also contribute in a minor way to the complexity

Figure 3.6 Splicing and editing of the rice mt *rpl2* transcripts

A: Sequence gel of RT-PCR products showing the properly spliced exons of the rice mt *rpl2* gene. Arrow shows the splice junction and the corresponding nucleotides are given on the left side.

B: Sequence gel of the rice mt *rpl2* DNA (left panel) and cDNA (right panel) showing the single RNA editing site. The arrow points the position of the C>U change. These sequences are obtained from a cloned RT-PCR product (Hind III fragment Fig. 3.6.C).

C: Schematic showing the RT-PCR and inverse-PCR strategies. The oligomer positions (oligo # 1 & 2 , 4 & 7) and the expected size of the product from DNA or precursor mRNA and spliced mRNA are indicated. The position of the Hind III sites used in cloning the RT-PCR product is also given.

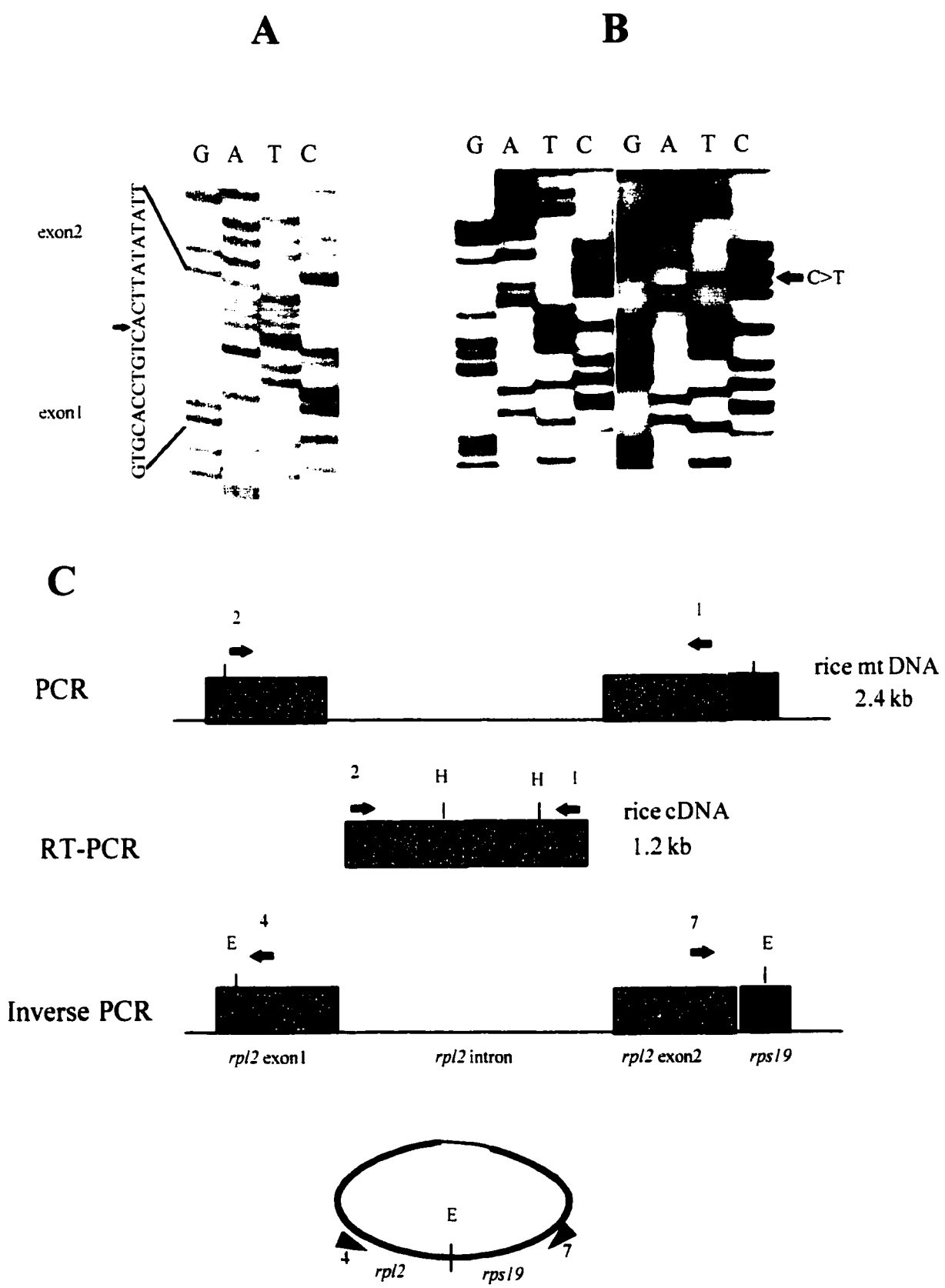
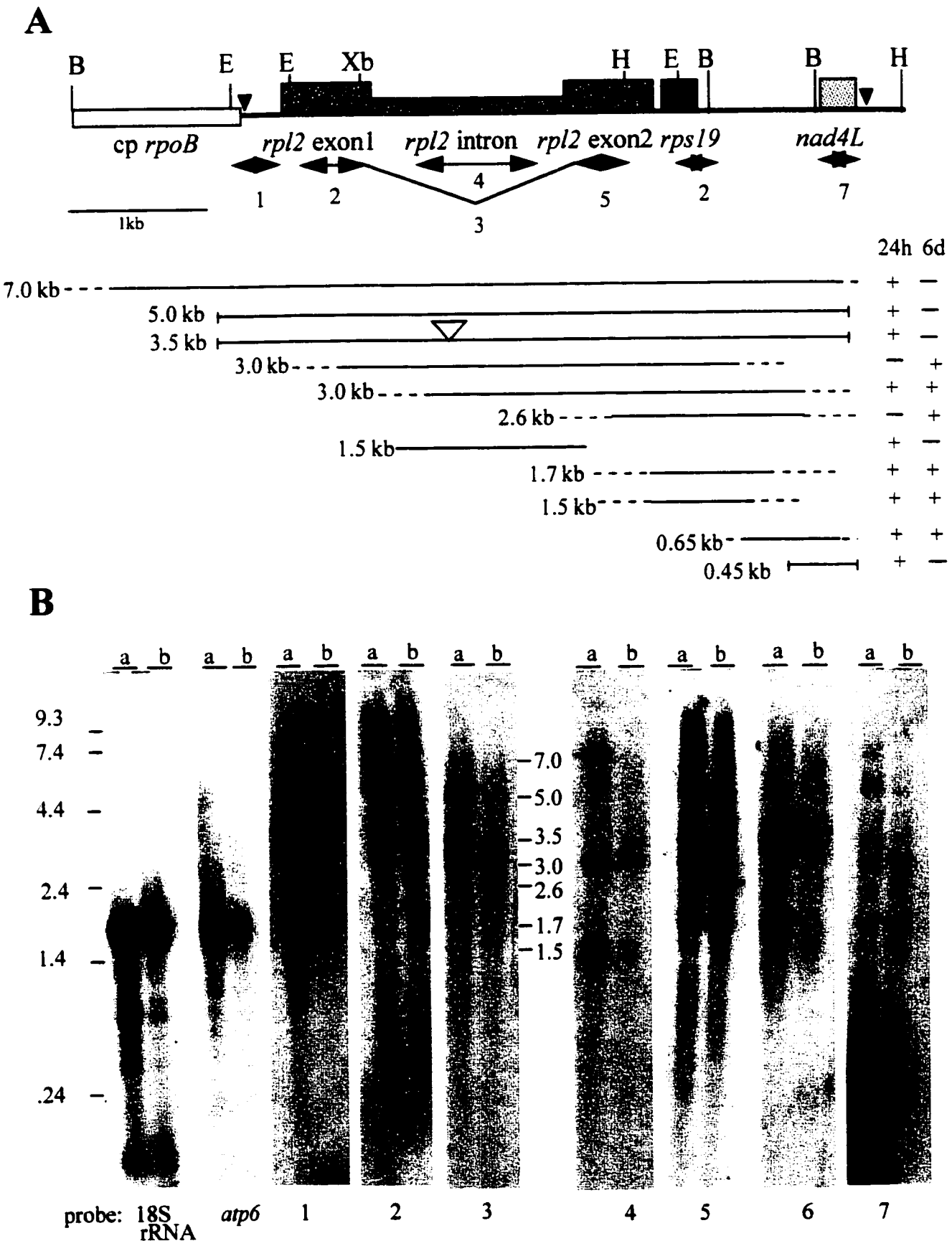

A
B
G A T C
G A T C G A T C
exon2
exon1
GTGCACCTGTCACTTATATATT
C>T
C
2
1
PCR
rice mt DNA
2.4 kb
2
H
H
I
RT-PCR
rice cDNA
1.2 kb
4
E
7
E
Inverse PCR
rpl2 exon1
rpl2 intron
rpl2 exon2
rps19
E
4
7
rpl2
rps19

Figure 3.7 RNA blot analysis of the rice mt *rpl2 -rps19-nad4L* gene region

A: Restriction map showing the cp *rpoB*-mt *rpl2-rps19-nad4L* gene cluster in the rice mt genome. The *rpoB* is part of a chloroplast DNA segment integrated into the mt genome. Selected restriction sites are shown as landmarks, EcoR1 (E), BamH1 (B), Xba1 (Xb), and Hind III (H). All the Hind III sites are not shown here. Arrows and numbers below the map represent probes used. The transcript sizes, and their terminii with their relative abundance at 24 h and 6 d stages are indicated as + (high) and - (low). The dotted lines indicate the absence of terminus mapping information.

B: Northern blot hybridization analysis of rice mt RNA isolated from (a) 24 h germinating embryos and (b) 6 d etiolated seedlings. Probes are indicated below the RNA blots. The RNA size marker is indicated on the left side of the panel. Sizes of *rpl2* transcripts are indicated beside the panel with probe 3 (*rpl2* spliced cDNA sequences). RNA blots were from a single gel and blots were reused after testing for the complete decay of radioactivity. The blot used by probe number 7 was reused for probe 2 and 5 respectively. Similarly the blot probed by number 1 was reused for probes 6 and *atp6* respectively. The 18S rRNA and *atp6* blots were exposed for approximately 4 hrs in contrast to 4 days for rest of the blots. These observations were confirmed by other independent northern experiments.


A
B
E
E
Xb
H
E
B
B
H
cp rpoB
rpl2 exon1
rpl2 intron
rpl2 exon2
rps19
nad4L
1
2
4
5
2
7
3
1kb
24h 6d
7.0 kb
5.0 kb
3.5 kb
3.0 kb
3.0 kb
2.6 kb
1.5 kb
1.7 kb
1.5 kb
0.65 kb
0.45 kb
B
a b a b a b a b a b a b a b a b a b
9.3
7.4
4.4
2.4
1.4
.24
7.0
5.0
3.5
3.0
2.6
1.7
1.5
probe: 18S rRNA
atp6
1
2
3
4
5
6
7

of the profile since the 18S rRNA hybridization panel shows some discrete breakdown products. The 3.5 kb transcript is believed to be the *rpl2* mRNA since it was the only transcript detected by all exon probes (Fig 3.7. B. probes 1,2,3 and 5) and not by the intron probe (probe 4). The mRNA levels are observed to be higher in germinating embryos than seedlings. The 5.0 kb and 7.0 kb transcripts are consistent with being precursors since they also hybridize to intron probes. The 7.0 kb transcript contains cp *rpoB* sequences since it was also detected by the upstream probe (data not shown). However the relative stochiometry of these transcripts is different between probes 3 and 4 suggesting a further complexity perhaps due to additional processing intermediates. The intron probe 4 detected two transcripts (3.0 and 1.5 kb) along with these precursors. The 3.0 kb signal is expected to be a mixture of two different species of the same size since it was detected by all probes even though probe 1 and probe 7 are more than 3.0 kb apart (Fig. 3.7). One of them may contain the intron, the 3'exon and part of the 5'exon and the other may have the *rps19* and *nad4L* sequences (Fig. 3.7.B). These 3.0 kb transcripts are approximately equal in levels at both stages compared to the excised intron (1.5 kb), which is higher in the embryo stage than the seedling stage. Similarly, higher steady state levels of excised introns are observed for the wheat *nad7* gene in germinating embryos than in seedlings (Carrillo and Bonen, 1997). The processing intermediates containing 3'exon and downstream gene sequences (2.6 kb, 1.7 kb and 1.5 kb) are found to be either more abundant in seedlings than in embryos or in equal levels in both stages. Higher amounts of 2.6 kb and 1.5 kb transcripts in seedlings than embryos were detected by probes 3, 5, and 7. These differences in transcript profiles suggest that the splicing and processing machinery of rice mt *rpl2* transcripts are differentially regulated at these developmental stages.

Northern experiments by Kubo et al. (1996) on rice mt *rpl2* RNA detected only three transcripts (designated as 3.8, 2.9 & 2.5 kb). They used RNA from young seedlings which may be the reason for the failure to detect the high molecular weight transcripts which are relatively more abundant in germinating embryos. In summary, *rpl2* precursors, spliced mRNA and excised intron levels are higher in germinating embryos than seedlings in contrast to the processing intermediates. This could mean either that the transcription and splicing of *rpl2* mRNA is higher in germinating embryos or that there is a rapid turnover of transcripts in seedlings or a combination of both. The variations in transcript accumulation at the embryo and seedling stages may suggest a developmental regulation of the mt RNA processing machinery, which is presumably nuclear-encoded.

3.2.2.3 *Rice mt rpl2, rps19 and nad4L genes are co-transcribed*

Co-transcription of the rice mt *rpl2* and *rps19* genes is expected, since they are only three nucleotides apart. Consistently, the *rps19* probe in the northern hybridization experiment detected transcripts of 7.0, 5.0, 3.5, and 1.7 kb sizes (Fig. 3.7 probe 6). However, longer exposures detected all transcripts (data not shown) arising from the mt *rpl2* gene in both rice seedling and embryo stage mt RNAs. The 1.5 kb detected by probe 6 and 7 is expected to be different from the one detected by other upstream probes since these probes span more than 1.5 kb. This processing intermediate may correspond to the one which was detected by primer extension analysis of *rps19* transcripts (section 3.2.2.5). The 3.5 kb transcript is more abundant than other transcripts and observed in higher amounts in embryos than seedlings. The 3.5 kb mRNA is consistent with a bicistronic transcript of the spliced mt *rpl2* and downstream *rps19* sequences. Accordingly, the RT-PCR experiments with a reverse 5' *rpl2* oligomer (oligo # 3) and a forward 3' *rps19*

oligomer (oligo # 14) gave products that are in agreement with the co-transcription of these genes (data not shown). Kubo et al. (1996) have also shown the co-transcription and editing of *rpl2* and *rps19* by RT-PCR experiments, but their northern experiments detected only two transcript sizes for the *rps19* probe (2.9 and 2.5 kb). According to my transcript size analysis, these RNA species do not contain the full *rpl2* coding sequences as they are not detected by probe 1 (Fig. 3.7.B)

The *nad4L* gene located downstream of *rps19* has an abundant 0.45 kb transcript in addition to the larger ones observed with the *rps19* and *rpl2* probes (Fig. 3.7.B.probe 7). The small monocistronic mRNA could suggest an additional independent promoter in the spacer region downstream of *rps19* gene. Nevertheless, the 3.5 kb species common to *rpl2-rps19* mRNA and other high molecular weight transcripts (3.0, 2.6, 1.7, 1.5 kb) were also detected by the *nad4L* probe implying a co-transcription of *rps19* and *nad4L*. The 2.6 and 1.5 kb transcripts observed in higher levels in seedlings than embryos may be cleavage products of the 3.5 kb mRNA which is higher in embryos than seedlings. The level of *nad4L* transcript (0.45 kb) was observed to be higher in the embryos than seedlings. The relative abundance of the 0.45 kb transcript of *nad4L* at embryos when compared to other larger transcripts, suggests a high mRNA stability and/or an independent promoter for *nad4L*.

3.2.2.4 *Mapping of the transcript termini of rpl2, rps19 and nad4L*

To get more precise information of transcript sizes, primer extension and S1 nuclease mapping experiments were carried out to map the transcript termini of *nad4L, rps19* and *rpl2* genes. The rice mt *rpl2* gene primer (oligo # 27) was used in primer extension experiments (Fig.

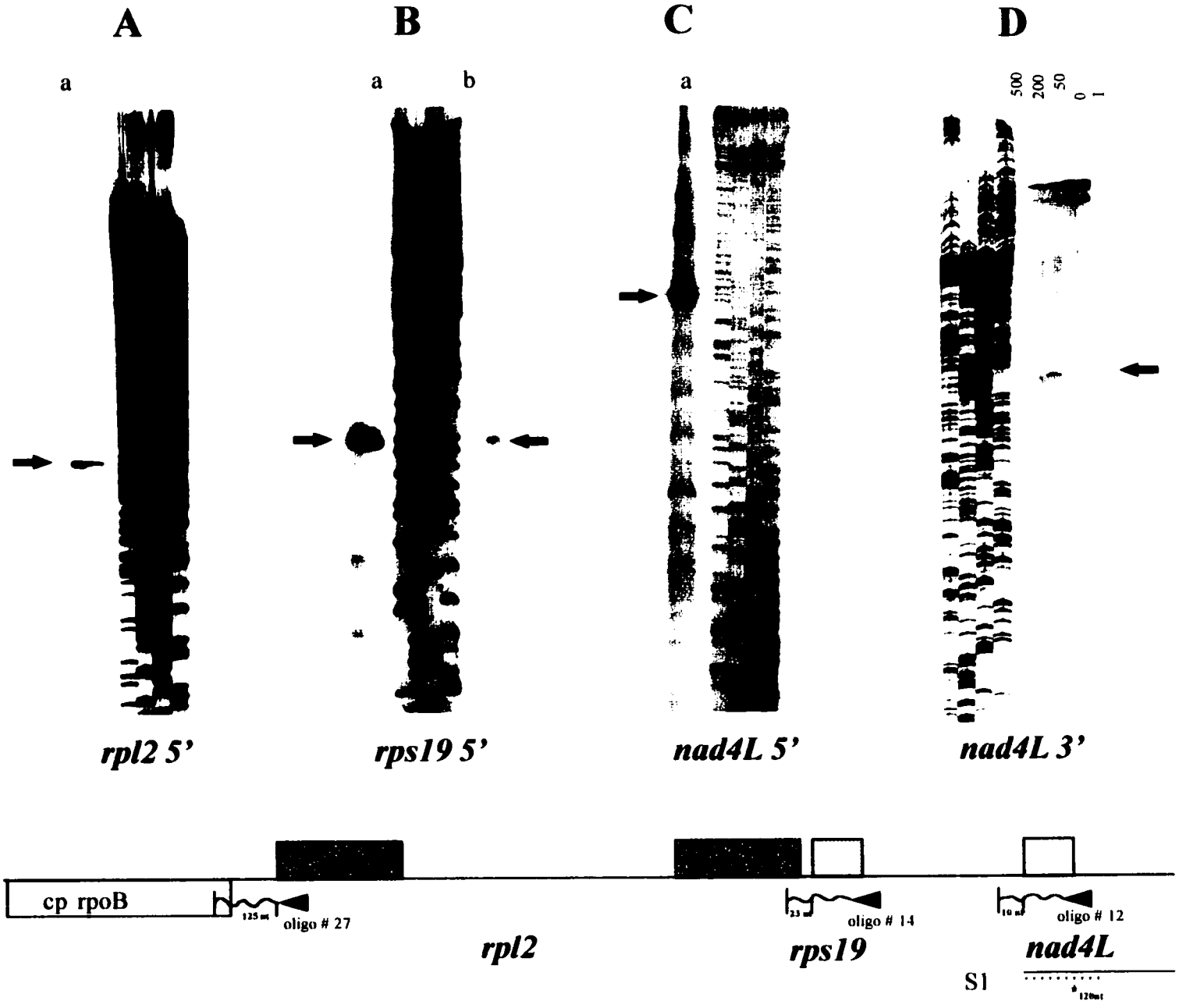
Figure 3.8 Transcript termini mapping for the rice mt *rpl2*, *rps19* and *nad4L* genes

A: Primer extension analysis to map the 5' termini of rice mt *rpl2* mRNAs from 24 h germinating embryos (a). The arrow indicates the extension product. The position of the oligomer (oligo # 27) used in the experiment and the 5' terminus position of the RNA is shown in the schematic below (not drawn to scale).

B: Primer extension analysis to map the 5' termini of rice mt *rps19* mRNAs from 24 h (a) and 6 d (b) stages. The oligo # 14 position is shown in the schematic below.

C: Primer extension analysis to map the 5' transcript termini of *nad4L* mRNAs from 24 h germinating embryos (a). The position of the oligomer (oligo # 12) and the major terminus (schematic), and a conserved motif identified upstream are given in Appendix 3.

D: S1 nuclease mapping experiment to map the 3' end of *nad4L* mRNAs from 24 h (a) mt RNA. The concentrations of S1 nuclease (500, 200, 50, 0 units) used in different lanes and the input DNA (I) are indicated at the top. The arrow indicates the protected fragment (approximately 420 nt). The line below the schematic (not drawn to scale) shows the labeled DNA fragment (750 nt) and the protected region (dotted line) with the position of the stop codon (asterisk). Position of oligomers and transcript termini and the conserved motif present upstream of rice mt *nad4L* gene are given in Appendix 3.


A
B
C
D
a
a
b
a
500
200
50
0
rpl2 5'
rps19 5'
nad4L 5'
nad4L 3'
cp rpoB
125 nt
oligo # 27
rpl2
23 nt
oligo # 14
rps19
10 nt
oligo # 12
nad4L
S1
120nt

3.8.A); the 5' transcript termini of *rpl2* correspond to a location within the transferred chloroplast *rpoB* sequences. Consistently, northern hybridization experiments using a probe upstream to this major terminus did not detect the 3.5 kb *rpl2* mRNA in either stage (data not shown). Sequences immediately upstream of the *rpl2* transcript terminus were analyzed for promoter elements, but no conserved motif that matches with the rice consensus promoter RRaa/ta/tNN**CRTA**Na/tNaaa/t (Nakazono et al., 1995) could be detected. Although northern results are consistent with the co-transcription of rice mt *rpl2-rps19*, primer extension analysis on *rps19* transcripts (oligo # 14) also indicated a major terminus at 23 nucleotides upstream of the *rps19* start codon (Fig. 3.8.B) in both germinating embryo and seedling stages of RNA. High molecular weight extension products consistent with the co-transcript are also seen and their representation is very low because the size of the products is too big for the RT to give complete extension products that can be visualized. The transcript terminus within exon 2 of *rpl2* suggests additional cleavage or processed *rps19* transcripts in comparison to *rpl2* transcripts (Fig. 3.7.B probe 5 and 6). As mentioned earlier the terminus of the1.5 kb transcripts which can not accommodate the full *rpl2* may correspond to this 5' terminus observed in both stages.

Primer extension analysis, to determine the 5' end of the *nad4L* transcript (oligo # 12), indicated a major transcript terminus 10 nucleotides upstream of the start codon in addition to number of less prominent larger ones (Fig. 3.8.C). The minor extension products may correspond to *rps19-nad4L* co-transcripts based on northern data. Sequences upstream (19 nt) of the 5' end of the transcript contained a conserved sequence AGCCTCACA**TAATTTAG** with the CRTA motif which has 13 nucleotides identical to the 17 nt rice consensus promoter element (Nakazono et al., 1995). S1 nuclease mapping experiments to find the 3' transcript termini of the *nad4L* gene

revealed that a major 3' end was located at 120 nucleotides downstream of the stop codon (Fig. 3.8.D). The S1 experiments also indicated low levels of larger protected products suggesting additional termini in the downstream region. The primer extension and S1 nuclease mapping analysis reconfirmed the size of the major *nad4L* transcript as 0.45 kb. The presence of a potential promoter sequence suggests the possibility of independent transcripts but guanylyl capping experiments are needed to distinguish between primary transcripts and processed monocistronic mRNAs.

3.2.2.5 *Physical form of excised rice mt rpl2 intron and editing status of domain 5 and 6*

Classical group II introns are highly structured and are spliced as lariat molecules with the help of accessory proteins (reviewed by Michel and Ferat, 1995). The plant mt group II introns show several deviations from the classical group II intron structures (Carrillo and Bonen, 1997), hence it is interesting to find out the nature of excised introns to gain an insight into the splicing reaction. The rice mt *rpl2* intron in the group II category has a typical domain 6 structure with an adenosine residue at the branch site. There is an A-C mispair at the base of the helix which could potentially be edited to improve the folding of this domain (Fig 3.9. F). The domain 5 is unusual in that it is short (30 vs 34 nt) compared to other group II introns and has pyrimidine in the tetraloop. It also has an A-C mispair distal to the dinucleotide bulge (Fig 3.9. F). In addition, northern blot analysis indicated that the 24 h stage RNA has higher stable levels of excised introns compared to 6 d stage which is also observed as a doublet signal (Fig 3.9. B & 3.7 probe 4). These observations suggest a possibility of different forms of excised introns. Hence, a comparative analysis was carried out with the rice *nad7* intron 3 since in wheat, the *nad7* intron 3 was found to be excised as lariat molecules (C. Carrillo unpublished result). The rice *nad7*

Figure 3.9 The rice mt *rpl2* excised intron levels and nature in comparison to rice *nad7* intron3

A: Rice mt RNA blots probed with *nad7* intron 3 and *nad7* exon 5. The amount of RNA loaded for both stages 24 h (a) and 6 d (b) were standardized based on equivalent 18S rRNA signal. The excised introns in germinating embryos are indicated by a dot. RNA size markers are given in the left side and as horizontal lines. The *nad7* exon blot was exposed for 4 h in contrast to 3 d for the *nad7* intron3.

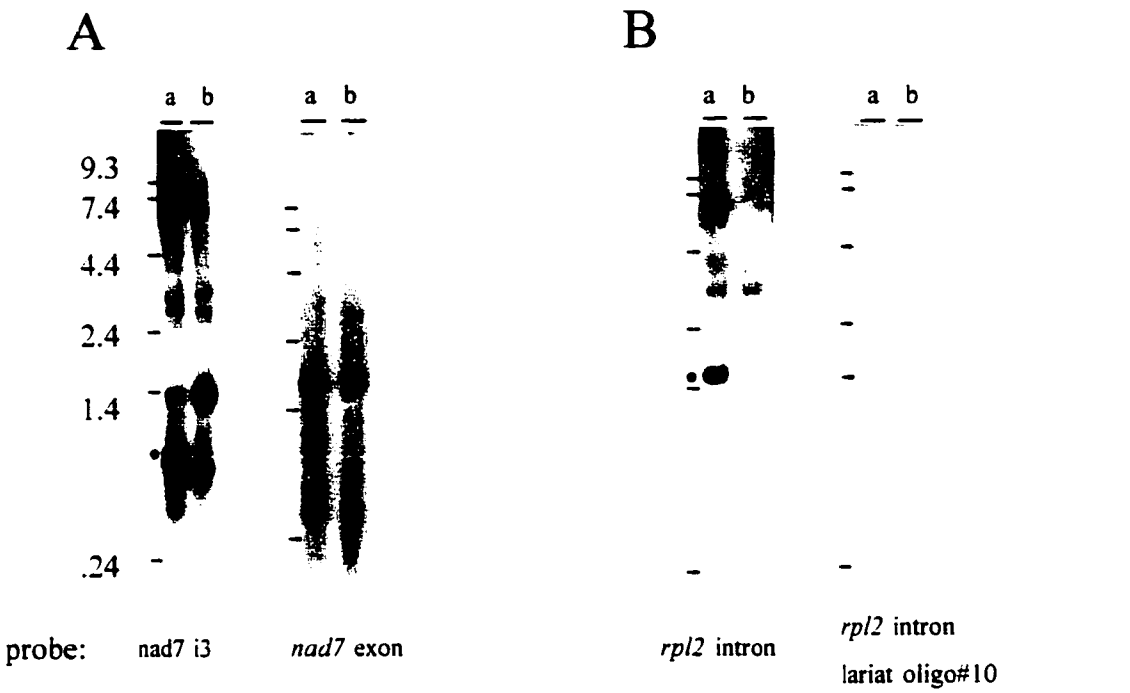
B: Rice mt RNA blots probed with *rpl2* intron, and *rpl2* lariat oligomer (oligo #10) for 24 h (a) and 6 d (b) stages. The exposure time for *rpl2* intron blot is 2 days and 1 week for the lariat oligomer probe. The rpl2 intron data agrees with that shown in Fig. 3.7 probe 4.

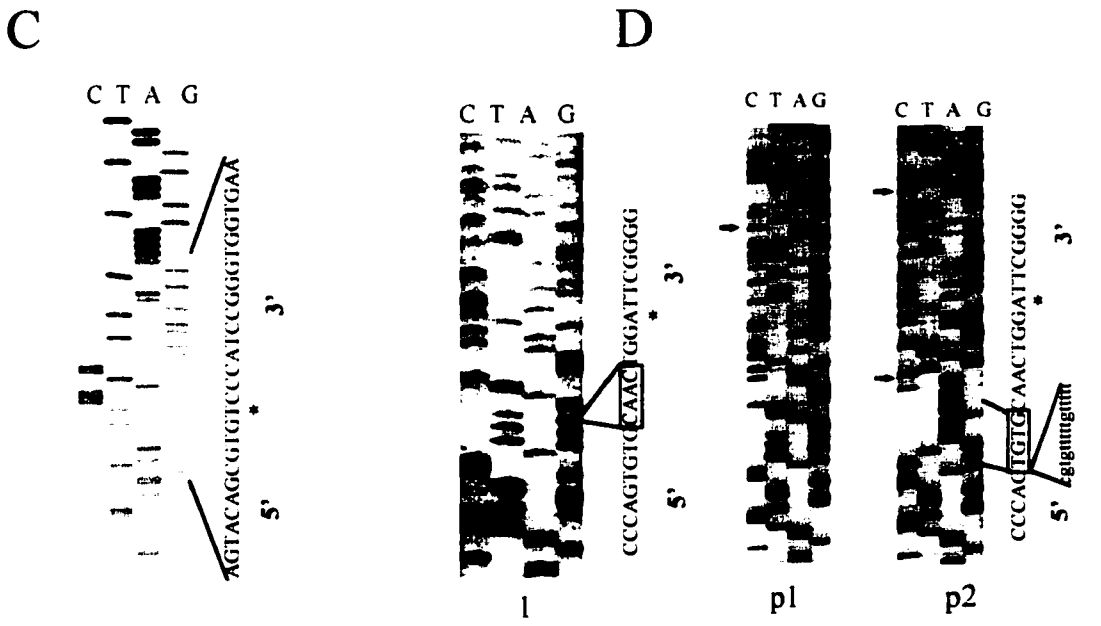
C: Sequence gel for the RT-PCR products from excised rice mt *nad7* intron 3 sequenced directly using oligomer located near the 3' end of the intron. Sequences are consistent with lariat form of excised intron molecules lacking the nucleotides 3' to the branch site. The thymidine corresponding to the adenosine residue at the branch site is shown as an asterisk (Figure kindly provided by C. Carrillo).

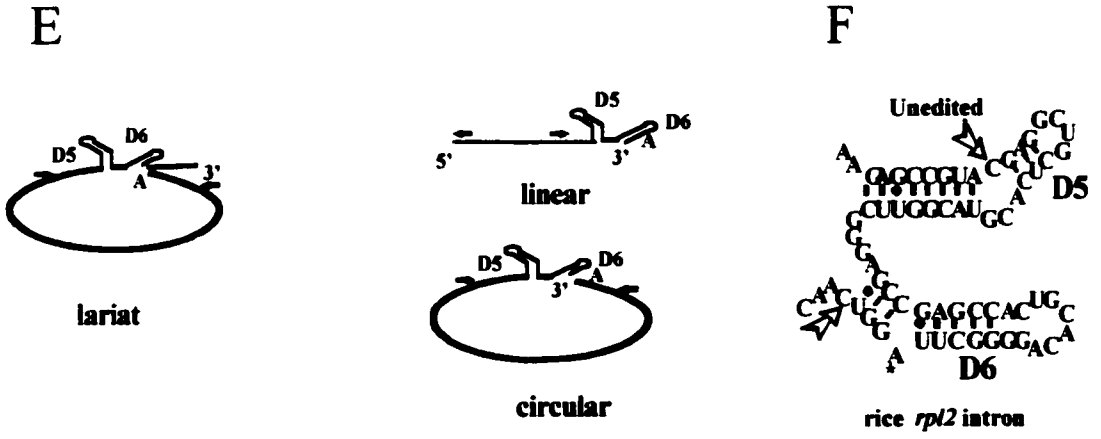
D: Sequence gel for the RT-PCR products from excised intron of rice mt *rpl2* directly sequenced using oligo # 9 (panel 1). The 4 nucleotides (block) missing at the 5' end of the intron and the intron sequences near the junction are indicated in the right side of the panel. The adenosine residue corresponding to the branch site is shown with an asterisk. Sequence gel of one clone of RT-PCR product of excised intron (panel p1) sequences. The A-C mispairs which are not edited are indicated by solid arrows (see schematic E open arrows). Sequence gel of another independent excised intron clone (panel p2) which lacks 4 nucleotides (block) at the 3' end of the intron and has an insert of 15 nucleotides (lower case letters).

E: Schematic showing the RT-PCR strategy, oligomer positions and different forms of excised introns (lariat, linear and circular).

F: The domains 5 and 6 structures of the rice *rpl2* intron with the bulging adenosine (asterisk) and A-C mispairs (open arrows).


A
a b
a b
9.3
7.4
4.4
2.4
1.4
.24
probe: nad7 i3
nad7 exon
B
a b
a b
rpl2 intron
rpl2 intron
lariat oligo#10


C
C T A G
AGTACAGCGGTGTCCCATCCGGGTGGTGAA
3'
5'
D
C T A G
CCCAGTGTCCAACTGGATTCGGGG
3'
5'
1
C T A G
p1
C T A G
CCCAGTGTCCAACTGGATTCGGGG
3'
5'
gtgttttgttt
p2


E
D5
D6
A
3'
lariat
D5
D6
5'
3'
A
linear
D5
D6
3'
A
circular
F
Unedited
D5
D6
rice rpl2 intron

intron 3, probe showed a higher level of excised introns (1.0 kb) and high molecular weight precursors at 24 h compared to the 6 d stage. It was also noticed that seedlings had a slightly faster migrating species of excised intron (Fig. 3.9. A) and higher levels of a potential splicing intermediate (1.4 kb) than embryos (Fig. 3.9 A). In contrast, the mature mRNA (1.7 kb) was observed to be approximately equal in amounts at both stages (Fig. 3.9. A). Differences observed in the abundance of the excised introns of both rice mt *nad7* i3 and *rpl2* between stages suggest that they have different forms of excised introns that differ in stability implying different RNA processing pathways at these stages.

In order to find out the physical nature of the excised introns, RT-PCR experiments were carried out on excised intron RNA template. The oligomers were designed in such a way that they specifically amplify only the excised introns (Fig. 3. 9. E). The reverse transcriptase enzyme (Superscript II) is capable of reading through the non-typical bonds (2'-5') at the branch site (Vogel et al., 1997). The size of the RT-PCR products corresponded to the excised introns as expected and they were either directly sequenced or cloned into plasmids and then sequenced. The rice *rpl2* excised intron sequences appear not to be lariats since sequences 3' to the branch were also present in the PCR products (Fig.3.9 D). In contrast, the *nad7* i3 excised introns were lariat molecules (Fig. 3.9 C) and they lack the 6 nucleotides present downstream of the branch site. The branch site adenosine residue is copied as A instead of T as also observed by Vogel et al., (1997). The direct sequencing of the rice mt *rpl2* intron (Fig. 3.9 D1) as well as cloned RT-PCR product plasmids (Fig. 3.9. D. p1 & p2) indicated the lack of a few 5' or 3' end nucleotides of the intron (boxes) which is not consistent with it being a lariat. Instead it is possible that there are linear molecules in vivo, shortened by exonuclease activity. These observations, in addition to

the insertion of 15 nucleotide between the ends of the intron in one of the plasmids (p2), suggest that these introns may be excised as linear molecules. The RT enzyme would have played a role in adding the extra nucleotides leading to circular PCR product amplifications. If that is the case the intron ends must be still in very close proximity in vitro. However, it is also possible that the excised introns could have been ligated in vivo to give circular molecules.

In wheat, RNA blot analysis using oligomers designed across the lariat have successfully detected the excised *nad7* i3 introns (S. Macdonald and C. Carrillo unpublished results). A lariat oligomer for rice mt *rpl2* (oligo # 10) was designed and used in northern analysis (Fig. 3.9.B) but it failed to detect the excised introns even with longer exposure and under different hybridization and wash conditions. This observation reconfirms the absence of lariat molecules of rice mt *rpl2* intron. Primer extension analysis using rice mt *rpl2* intron specific oligomers may be informative to determine whether they are linear molecules, since the extension would stop at the 5' end of the intron. In addition, RNase protection assays using the circular plasmid molecules may resolve whether they are circular or linear molecules.

The secondary structure of the domains 5 and 6 which play important roles in the splicing of plant mt group II introns could potentially be improved by editing in the A-C mismatch positions. Carrillo and Bonen, (1997) have compiled the A-C mismatches of domains 5 and 6 of group II introns in flowering plants in comparison to the liverwort introns. The base pair distal to the dinucleotide bulge in the domain 5 of group II introns often has a relaxed base pairing. The rice *rpl2* intron also has an A-C mismatch at that position (open arrow schematic Fig. 3.9.F) in addition to the fourth position from the 3' end of the intron in domain 6. Hence it is interesting to analyze the editing status especially because the domain 5 is unusually short. The RT-PCR products sequenced directly (Fig. 3.9.D.1) and cloned plasmid sequences (Fig. 3.9.D. p1, p2)

showed no editing at the expected sites suggesting that Watson-Crick base pairing here is not crucial for domain 5 and 6 function in keeping with the observation in several plant mt introns (Carrillo and Bonen, 1997).

In summary, the status and expression of the mt *rpl2* gene is found to differ among wheat, rice and soybean. In wheat, the 3' *rpl2* segment located upstream of the *nad1d* exon is found to be a single copy pseudogene lacking 90% of the mt *rpl2* coding region. However, it is transcribed as a part of the large *nad1d* precursors involved in trans-splicing events to generate *nad1* mRNA. In contrast, the rice mt genome has an intact *rpl2* gene with a group II intron. The rice mt *rpl2* gene is expressed as a complex set of transcripts with a 3.5 kb mRNA showing higher steady state levels in germinating embryos than seedlings. However, its steady state levels are much lower than for other rice respiratory chain genes such as *atp6* and *nad4L*. The excised group II intron appears to be non-lariat form since excised intron-RT-PCR products revealed the presence of intron sequences downstream of the branch site. This observation suggests that the splicing might start with a hydrolysis rather than with a transesterification step. The mt *rps19* and *nad4L* genes that are located downstream of rice mt *rpl2* are co-transcribed with *rpl2* and processed subsequently. In contrast, the presence of very abundant small transcripts for *nad4L* and sequences resembling the promoter motif upstream of the 5' transcript terminus may suggest an additional potential independent promoter for the *nad4L* gene. Moreover, the observation of several heterogenous transcript termini in primer extension products points to the possibility of multiple promoters. The accumulation of rice mt *rpl2* mRNAs and other transcripts differs between germinating embryos and seedling stages suggesting differences in RNA stability and processing mechanisms at these stages. Unlike rice, the soybean and wheat mt genomes do not encode an intact mt *rpl2* gene raising the possibility that the functional gene is in the nucleus.

CHAPTER 4

MITOCHONDRIAL L2 RIBOSOMAL PROTEIN GENE SEQUENCES IN THE NUCLEAR GENOMES OF WHEAT AND RICE

4.1 Isolation and characterization of mt *rpl2* sequences from the wheat nuclear genome

The presence of a functional *rpl2* gene in rice and liverwort mitochondria and strong L2 sequence conservation among bacterial, organellar and cytosolic counterparts suggest the importance of the L2 protein in plant mt ribosomes. Hence, the simplest explanation for the absence of the *rpl2* gene in wheat and soybean mt DNA is that the functional gene is in the nucleus having been transferred there sometime during evolution. In order to test this hypothesis, PCR and RT-PCR experiments using rice mt *rpl2* gene specific oligomers (oligo # 1 & 2, 1 &3) were carried out on wheat total DNA and poly A^+ RNA respectively. A PCR amplicon (600 bp) which gave a positive Southern hybridization signal to the rice mt *rpl2* probe (RPP1) was obtained from both wheat DNA and cDNA (Fig. 4.1.A. lanes1 & 2). The position of oligomers relative to the rice mt *rpl2* and sizes of PCR products amplified on different templates are shown in a schematic (Fig. 4.1.D). Amplification of a 1.2 kb product for rice mt *rpl2* cDNA (Fig 4.1.A lanes 3 & 4) and a 2.7 kb amplicon for rice mt DNA (lane 5) for the same primer set are also shown in Fig.4.1.A. Direct sequencing of the wheat PCR products (600 bp) from both DNA and RNA templates was carried out and revealed the presence of mixed sequences of mt *rpl2*-homologous and other nonspecific amplification products. The RT-PCR products were cloned into plasmid vectors and colony screening using the wheat mt *rpl2* fragment without primer sequences (probe: WPP 4; see section 2.4.3. for more details about PCR product probes) was carried out to obtain the wheat mt *rpl2*- homologous clones. DNA from ten positive plasmids

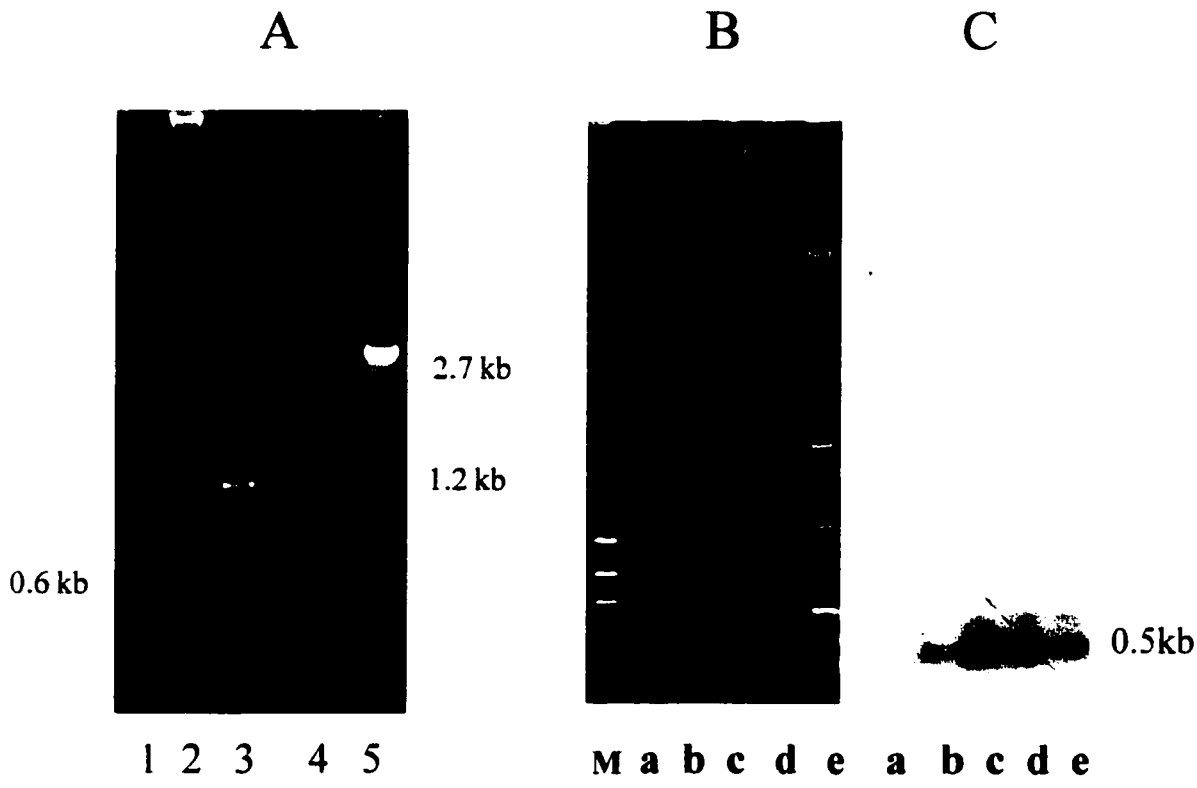
Figure 4.1 PCR amplification of wheat mt *rpl2* sequences located in the nucleus

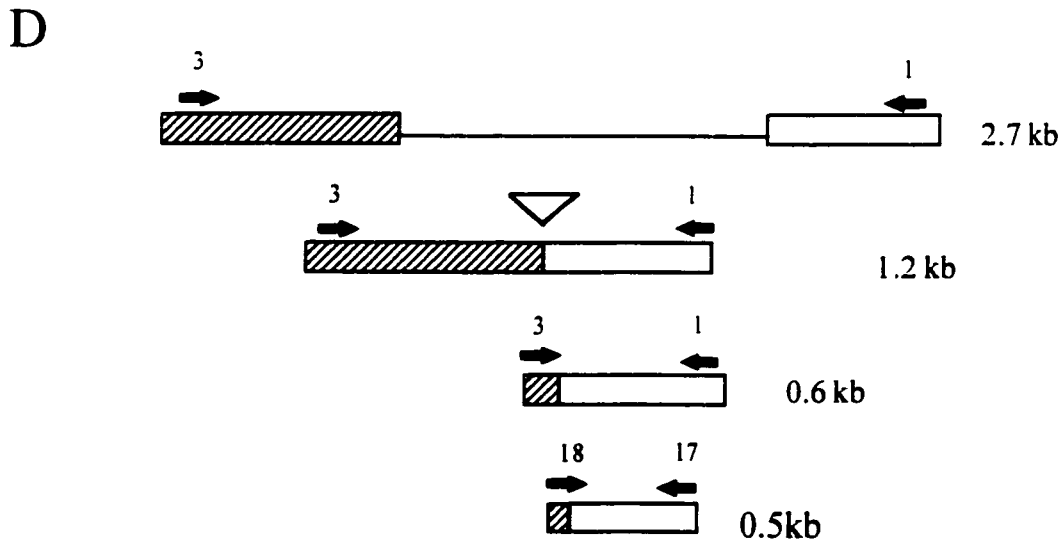
A: RT-PCR products from wheat poly A^- RNA using two different concentrations of cDNA (lanes1 and 2), rice mt *rpl2* cDNA two different concentrations (lanes 3 and 4) and rice mt DNA (lane 5). Sizes of the products are indicated. Primer positions (oligo #1 & #3) are given in the schematic below.

B: PCR amplification products (oligo # 17 & 18) using primers specific to wheat nuclear-located mt *rpl2* sequences on wheat total DNA templates obtained by different DNA isolation procedures (lanes a,b,c,d and e) along with the DNA size marker λ Hind III + ϕx174 Hae III (M)

C: Southern blot of the gel shown in the panel B hybridized with the wheat nuclear-located mt *rpl2* probe (WPP1). The size of the mt *rpl2* related product is indicated.

D: Schematic showing the position of oligomers (black arrow) and different size products amplified on different template DNAs. Open and striped boxes correspond to the exon II and exon I regions of the rice mt *rpl2* gene respectively. Picture is not drawn to scale.


A
B
C
2.7 kb
1.2 kb
0.6 kb
0.5kb
1 2 3 4 5
M a b c d e a b c d e


D
3
1
2.7 kb
3
1
1.2 kb
3
1
0.6 kb
18
17
0.5kb

was isolated, sequenced and all were found to be identical. Sequences (600 bp) of the positive clones were found to be related (70% identity at nt. level) to rice mt *rpl2* sequences (Fig.4.2.A) however, the size was only 50% of the expected length of 1.2 kb. Hence, different PCR conditions such as broad range of temperatures ranging from 45^0C to 55^0C and various Mg^{2+} concentrations and from different DNA isolation procedures of wheat total DNA templates (Fig. 4.1.B) were used to obtain the expected 1.2 kb product, but they were not successful. In addition, PCR amplification experiments using another oligomer (oligo # 16) upstream of the oligomer # 3 (derived from rice mt *rpl2* 5' conserved sequences) were also employed and were negative.

The wheat nuclear-located mt *rpl2* sequences from the 600 bp product that was obtained by PCR share 70% identity at the nucleotide level with rice mt *rpl2* (Fig. 4.2.A) and 69% identity at the amino acid level (Fig. 4.2.B). They contain a highly conserved region (84%) homologous to *E. coli* L2 and a less conserved region (54%) related to the plant mt L2 specific inserted sequence (Fig. 4.2.B & 3.4). The presence of plant mt specific *rpl2* sequences and high similarity to the rice mt *rpl2* gene confirm that the isolated wheat nuclear *rpl2* sequences are part of a mitochondrial homologue, which appears to have been transferred from the mitochondrion.

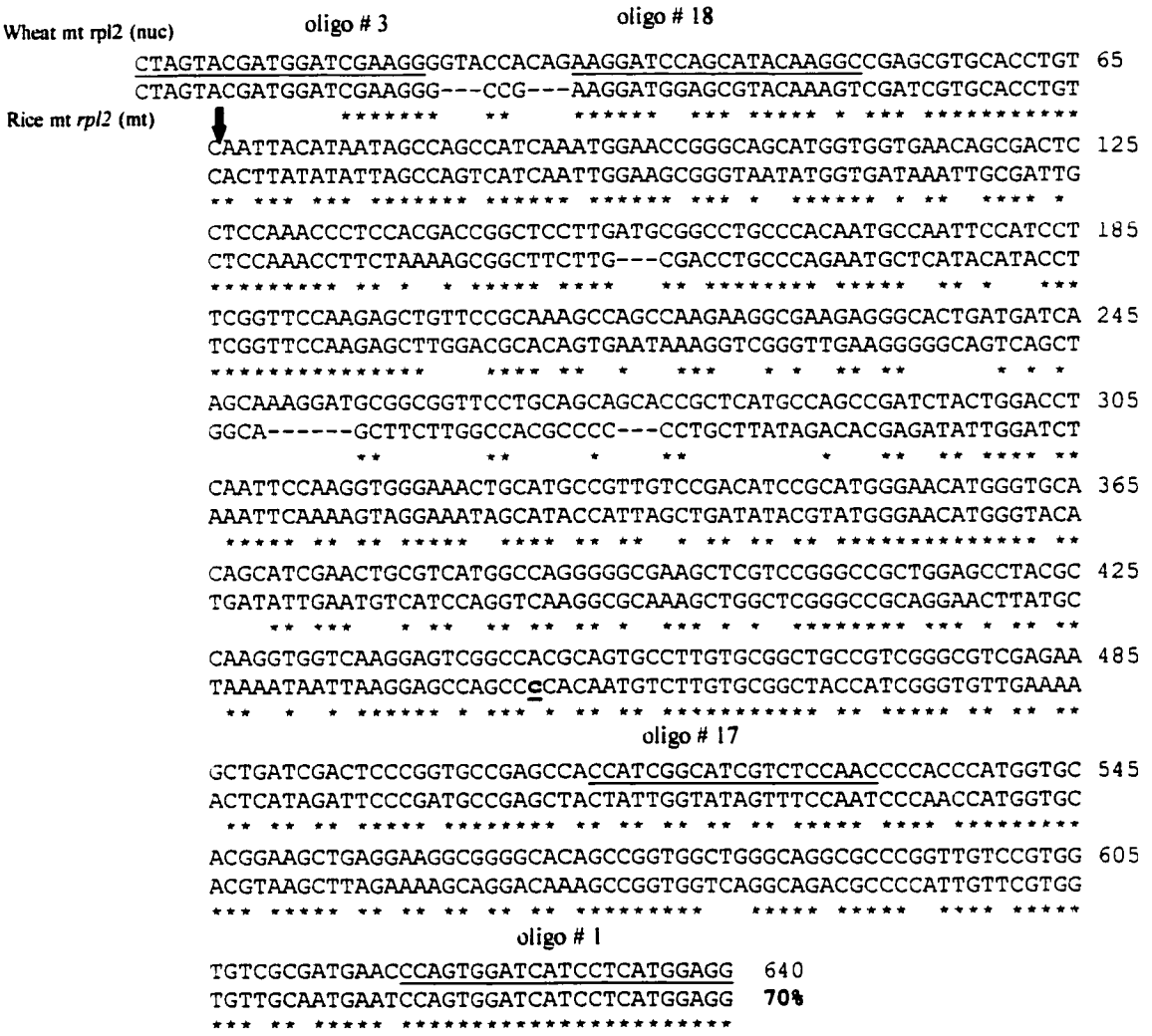
When the wheat nuclear-located mt *rpl2* and rice mt-located *rpl2* sequences were examined, more than 60% of the nucleotide substitutions were at synonymous sites. There are five insertions in triplets or multiples of three base pairs relative to the rice mt *rpl2* gene and they are located in the less conserved plant-specific mt *rpl2* insert region (Fig. 4.2.A). These observations are consistent with the wheat nuclear-located mt *rpl2* sequences being under functional constraint. The wheat nuclear-located mt *rpl2* sequences do not have an intron at the position corresponding

Figure 4.2 Comparison of the wheat nuclear-located mt *rpl2* sequences with the rice mt *rpl2* homologue

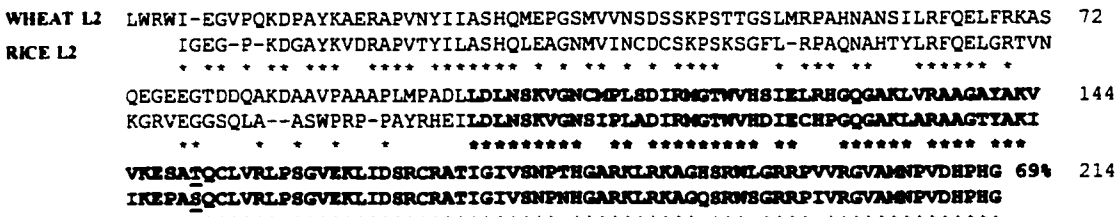
A: Alignment of the wheat nuclear-located mt *rpl2* sequences with the rice mt-located *rpl2* sequences (this work and Kubo et al., 1996, Ac. No. D78336-Appendix 2). The asterisks indicate the identical nucleotides and dashes indicate deletions. Oligomers are numbered and their positions underlined. The percentage identity of nucleotides is given at the end of mt sequences. The arrow indicates the intron position in the rice mt *rpl2* gene. The single editing site in the rice sequence is shown as a bold lowercase letter underlined.

B: Comparison of the derived amino acid sequences of the nuclear-located wheat mt *rpl2* and the rice mt *rpl2* gene. The conserved *E. coli* L2 homologous region is indicated as bold letters compared to the less well conserved plant mt specific L2 insert region. Asterisks denote amino acid identities. The amino acid change (Ser) corresponding to the editing site in rice and the conservative change (Thr) in the wheat mt *rpl2* are shown as bold letters underlined.

A


Wheat mt rpl2 (nuc)
oligo # 3
oligo # 18
Rice mt rpl2 (mt)
oligo # 17
oligo # 1
70%

B


WHEAT L2
RICE L2
69%

to the group II intron of the rice mt *rpl2* or truncated *Oenothera* and *Arabidopsis* mt *rpl2* genes (Fig. 4.2.A arrow). This observation is consistent with the transfer having occurred through a spliced RNA intermediate assuming that the original wheat mt *rpl2* copy had an intron at the same position. RNA-mediated transfer events have been proposed for several other mt rp genes (reviewed by Schuster and Brennicke, 1994). For example the *Arabidopsis* nuclear-located mt *rps10* (Wischmann and Schuster, 1995) gene does not contain an intron at the position corresponding to the one in the potato *rps10* gene located in the mitochondrion (Zanlungo et al., 1994). The mode of transfer is even more convincingly predicted based on RNA editing information of mt copies and by comparing the corresponding positions in the nuclear sequences, for example, *Oenothera rps12* (Grohmann et al., 1992), *Arabidopsis rps10* (Wischmann and Schuster, 1995) and *rps19* of *Arabidopsis* (Sanchez et al., 1996). In the case of *rpl2*, only one editing site is observed in rice mt transcripts (this work and Kubo et al., 1996). The amino acid residue (Ser- TCA) corresponding to the edited sequence in rice mt *rpl2* is genomically encoded by a conservative amino acid (Thr-ACG) in the wheat nuclear copy (Fig. 4.2.A). Hence the low degree of editing is not of much value to predict the mode of transfer, however the absence of an intron is a strong argument.

4.2 Wheat mt *rpl2* sequences located in the nucleus are actively-transcribed

Northern hybridization experiments carried out on wheat 6 d poly A^+ RNA using wheat nuclear-located mt *rpl2* sequences (WPP1) as a probe revealed an abundant transcript of 1.2 kb (Fig. 4.3.A). The mRNA levels were compared with another nuclear-located gene encoding a mt protein ATPase subunit β (WPP2) using probes with similar specific activity. The wheat *atpβ* cDNA sequences (1.3 kb) were amplified by RT-PCR techniques using primers designed from the

tobacco *atpβ* sequences (oligo # 22 & 28) on wheat total RNA. A 2.1 kb transcript was detected for wheat mt *atpβ* cDNA (Fig. 4.3.A).

The size of the wheat mt *rpl2* mRNA is only 1.2 kb. Although, this is larger than the bacterial and protist *rpl2* coding region (850 bp), it is not large enough to accommodate the entire plant mt *rpl2* coding region (1.5 kb) which contains a large internal coding insert. In addition, nuclear-located sequences must have sequences that target the protein back to mitochondria. However, the rice mt *rpl2* 5'exon sequences, that were upstream of the wheat nuclear-located mt *rpl2* sequences obtained by PCR did hybridize to the same size (1.2 kb) transcript (data not shown) suggesting the presence of at least part of that region. The abundant wheat poly A^- transcripts detected by the wheat mt *rpl2* sequences are consistent with the nuclear-located sequences being a part of a functional gene.

4.3 Multiple copies of mt *rpl2* sequences in the wheat nuclear genome

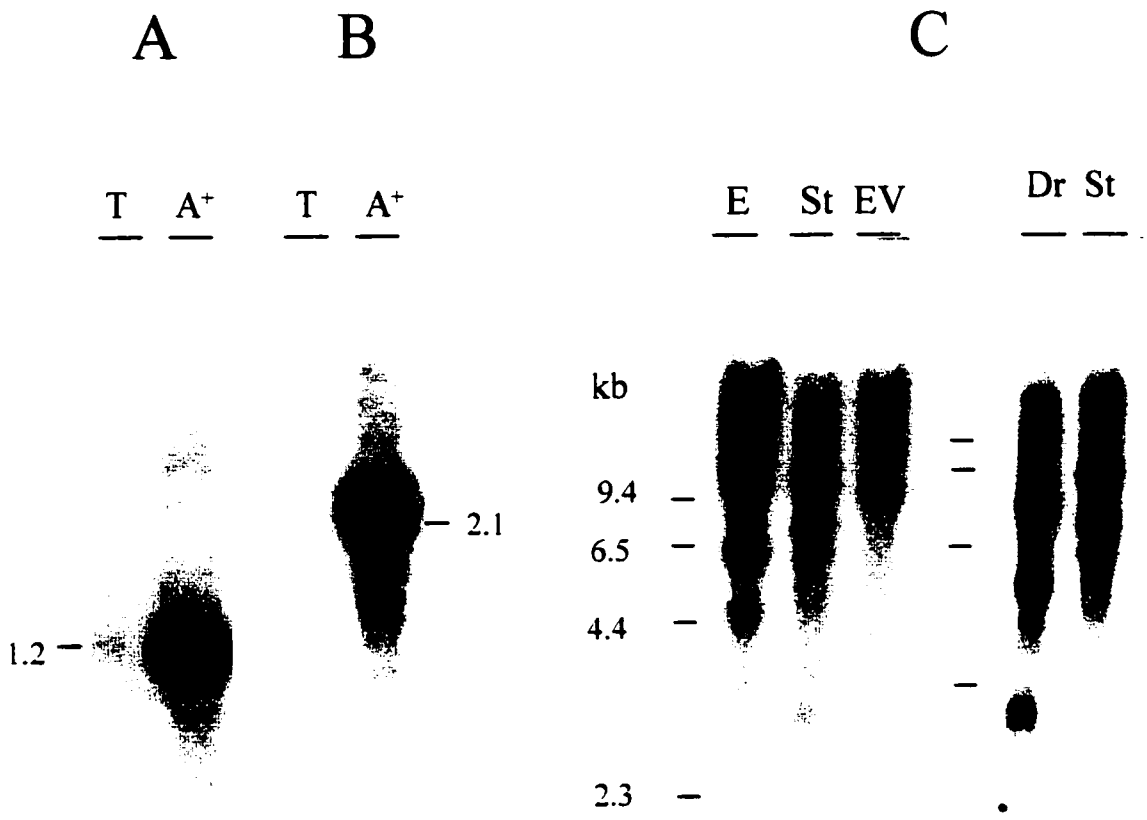
Wheat mt *rpl2* sequences isolated from nuclear DNA were further analyzed by Southern experiments using wheat nuclear-located mt *rpl2* probes (WPP1). Three hybridizing bands, ranging from 3 kb to 9 kb in size were detected in EcoR1, Sst1, EcoRV and Dra1 restriction digests (Fig. 4.3.C). Although the three hybridizing bands of mt *rpl2* sequences differ in their relative intensity they may correspond to the homeologous copies from the A, B & D genomes since wheat is a hexaploid plant (discussed below). Attempts to get the other copies by PCR amplification experiments on wheat nuclear DNA using oligomers (oligo # 17 &18) that are specific to the nuclear-located wheat mt *rpl2* led to the amplification of a 500 bp product (Fig. 4.1 Panels B & C). Direct sequencing revealed the presence of a mixture of *rpl2* and other non-specific amplification products. The PCR products were cloned into plasmid vectors and three out

Figure 4.3 Southern and northern analyses of the wheat nuclear-located mt *rpl2* gene sequences

A: Northern analysis: Wheat total (T) and poly A^+ (A^+) RNA (6 d seedlings) blot with ~10 μg of RNA in each lane was probed with wheat nuclear-located mt *rpl2* sequences (WPP1).

B: A comparable RNA blot probed with wheat *atpβ* cDNA (WPP2). The transcript sizes are indicated at the side.

C: Southern analysis: Wheat total DNA blot in which ~10 μg of total DNA digested with EcoR1 (E), Sst1 (St), EcoV (EV) and Dra1 (Dr) restriction enzymes was probed with wheat nuclear-located mt *rpl2* probe (WPP1). DNA marker sizes are given in the left side of the blot.



Wheat Poly A+RNA

Wheat total DNA

probe: Wheat mt *rpl2* (nuc)

WPP1

Wheat *atpB*

WPP2

Wheat mt *rpl2* (nuc)

WPP1

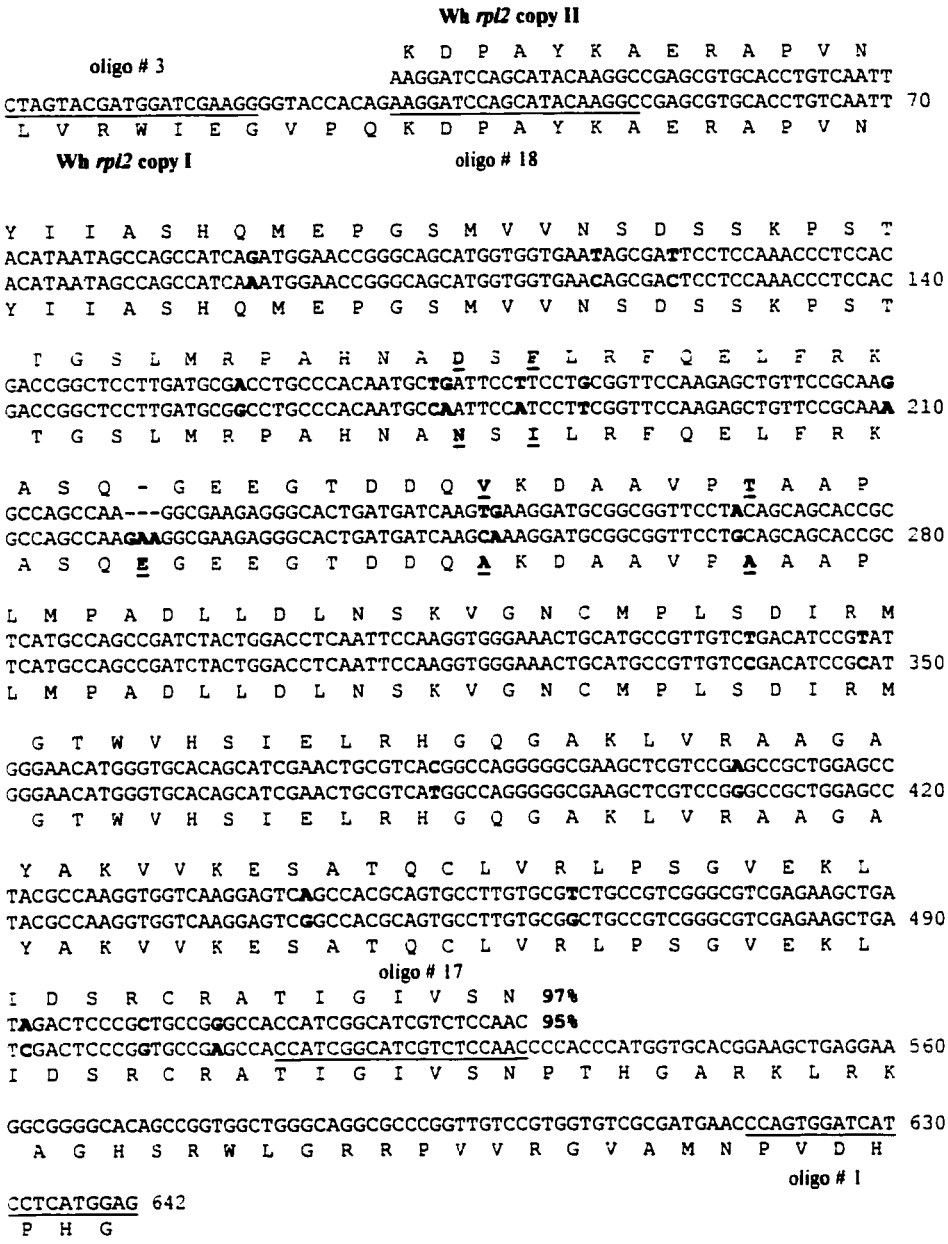
of one hundred clones were detected to be mt *rpl2* clones by a colony screening method (probe: WPP4). The first copy (600 bp) is called copy I and the second one (500 bp) is called copy II. Sequence analysis revealed that two of the clones were identical to the first copy and the third one (copy II) had 95% identity at nucleotide level and 97% identity at the amino acid level (Fig. 4.4). More than 75% of nucleotide substitutions were synonymous ones and among those, 12 out of 16 are transitional changes. A deletion of three base pairs in the second copy was observed in comparison to the first (Fig. 4.4) in the highly variable mt specific L2 insert region. These observations are consistent with both sequences being under functional constraint suggesting more than one functional mt *rpl2* gene copy in the wheat nuclear genome. Wheat is a hexaploid plant with three different chromosomes A, B and D, which are different in their origins (allohexaploid). The degree of amino acid and nucleotide identity observed between the two wheat nuclear-located mt *rpl2* copies is in agreement with the possibility that they are either homeologous or recently-duplicated paralogous copies. For example, homeologous copies of the chloroplast specific acetyl-CoA carboxylase gene of wheat share 98 -99.4% identity at the nucleotide level (Gornicki et al., 1997).

4.4 Attempts to isolate the complete cDNA of the wheat nuclear-located mt *rpl2* genes

The mt *rpl2* mRNA of wheat is approximately 1.2 kb so that the RT-PCR amplicon of 600 bp obtained constitutes around 50%. In order to obtain the rest of the sequences 5' and 3' RACE and inverse-PCR experiments on DNA templates were carried out. The wheat mt *rpl2* copy I isolated from cDNA and the copy II isolated from DNA revealed the absence of any intron in the amplified region. However, transferred mt sequences and acquired target sequences are often separated by intron sequences, for example in the case of *Arabidopsis* mt *rps10* (Wischmann and

Figure 4.4 Comparison of the sequences of the two *rpl2* homologous regions in the wheat nuclear genome

Wheat nuclear-located mt *rpl2* copy I (bottom) and copy II (top) are aligned and their derived amino acid sequences are given. Dashes indicate the deletion in copy II. The percent identities of nucleotides and amino acids are given at the end. The oligomer (oligo # 1, 3,17&18) sequences are underlined. Changes in nucleotide positions are indicated by bold letters and amino acid changes underlined by bold letters.


Wh rpl2 copy II
oligo # 3
Wh rpl2 copy I
oligo # 18
oligo # 17
oligo # 1
97%
95%

Schuster, 1995) an intron is found between transferred mt and acquired nuclear sequences. In order to amplify the 3' end of the wheat mt *rpl2*, cDNAs were made using oligo dT-clamp primers that had an extra nucleotide G, C, or A at the 3' end (oligo # 24, 25 & 26) and a gene specific reverse primer (oligo # 18) was used in PCR amplifications. Among the several product amplifications in these 3' RACE experiments, three different ones (850, 750 & 600 bp) showed hybridization to wheat nuclear-located *rpl2* sequences. When these products were cloned and screened with wheat mt *rpl2* sequences (probe WPP4), a 730 bp positive clone was identified. Sequence analysis revealed that it had an open reading frame with only part of mt *rpl2* sequences (240 bp) and an unrelated (115 bp) region preceding the stop codon that showed no match in the databank or in the 360 bp preceding the poly A tail (Appendix 4). An artifact may have occurred during the reverse transcription with the fusion of different cDNAs due to less stringent conditions (42°C) used in the cDNA synthesis because of oligo (dT) primers. Consistently an AT rich region is noted in sequences flanking the stop codon (Appendix 4).

3' RACE experiments with a higher temperature (45 °C) for reverse transcription led to the preferential amplification of a larger product (850 bp). These products gave a positive signal with nuclear-encoded *rpl2* probe, hence they were cloned. However, the sequence analysis indicated that these sequences were not related to *rpl2*, but they shared 92% nt identity to the maize DNA binding protein gene and its 3' UTR (Ac. No. X79086; Werr, 1994). A 60% identity was observed between the mt *rpl2* primer sequences and the maize DNA binding protein gene sequence at the comparable positions. PCR amplification conditions used were 10 °C less than the optimal annealing temperature for the gene specific primer. This may explain a preferential amplification of the wheat DNA binding protein gene because the homologous region in wheat

may have even higher identity to the mt *rpl2* primers.

Attempts to amplify the 5' end of the mt *rpl2* cDNA were carried out by reverse transcription of the mt *rpl2* mRNA using a gene specific oligomer (oligo # 17) and subsequently tailing the cDNA ends using terminal deoxynucleotidyl transferase and dATP. These tailed cDNAs were used in PCR amplification experiments using the gene specific oligomer (the one used in cDNA synthesis oligo # 17 or a nested gene specific oligomer oligo # 19) and oligo dT clamp primers. Southern hybridization experiments on the heterogenous PCR products (WPP4) at low stringency conditions revealed a weak signal reflecting the presence of *rpl2* sequences only in very low amounts or the presence of distantly related sequences that hybridized under these conditions. Attempts made to clone the 5' RACE products by colony screening experiments were unsuccessful. Hence, a DNA titration experiment was carried out to estimate the level of mt *rpl2* sequences in 5' RACE product mixture and a ratio of approximately 1:1000 molecules was predicted. Tailing of cDNA extensions with dGTPs or adding linkers to the cDNA would enable to a higher annealing temperature and could lead to increased specificity in PCR amplifications.

In addition, inverse-PCR experiments were carried out on the wheat nuclear DNA using oligomers specific to mt *rpl2* that were designed to amplify the flanking sequences (oligo # 19 & 7). Southern data had indicated a relatively small Hind III (2 kb) fragment containing *rpl2* (data not shown), hence wheat nuclear DNA was restricted with Hind III enzyme, circularized with DNA ligase and used for PCR amplification. However, the inverse-PCR experiments on nuclear DNA were not successful even though the amplification of the rice mitochondrial *rpl2* flanking sequences by this approach had been successful (section 3.2.1.1). No PCR products were obtained even when DNA size fraction corresponding to one of the copies instead of total DNA

digests were used. The large size of the wheat hexaploid genomic DNA may have caused difficulty in amplifying specific flanking sequences.

4.5 Characterization of the cytosolic ribosomal L2 coding sequences in the wheat nuclear genome

Since all cytosolic ribosomal proteins are encoded by the nuclear genome it was important to determine that the *rpl2* sequences isolated from wheat nuclear DNA encode the mitochondrial L2 protein and not the cytosolic L2 one. The chloroplast *rpl2* gene of wheat was not a concern because it is located in the chloroplast genome and also in all plants examined (Doyle and Doyle. 1995). The tobacco, tomato and *Arabidopsis* cytosolic *rpl2* sequences were available in the databank, and tobacco was used to design oligomers since the mt *atpβ* oligomers designed from tobacco *atpβ* had been successful in amplifying the wheat *atpβ* cDNA. Wheat cytosolic *rpl2* sequences were isolated using RT-PCR techniques on wheat total RNA using primers (oligo # 20 & 21) designed from tobacco cytosolic *rpl2* sequences (Marty and Meyer, 1992). cDNA was synthesized using a cytosolic *rpl2* specific primer (oligo # 21) and an internal *rpl2* primer from wheat mt *rpl2* (oligo # 1) was used in the PCR amplification and it gave a 635 bp product. The PCR products were tested by Southern blot hybridization (probe: WPP3) and a colony screening method was employed to isolate the wheat cytosolic *rpl2* clone and it was sequenced. The tobacco and wheat sequences share 85% and 95% identity at the nucleotide and amino acid levels respectively (Fig. 4.5). The mt and cytosolic *rpl2* sequences of wheat are highly divergent and they can be aligned at the amino acid level only within a conserved carboxyl domain of approximately 120 amino acids, where they share 38% identity. This contrasts with the 84% identity between the rice mt L2 and the wheat L2 sequences derived from the nuclear-located mt

Figure. 4.5 Comparison of wheat and tobacco cytosolic *rpl2* sequences and their derived amino acid sequences

Wheat cytosolic *rpl2* sequences (top) were compared with the complete tobacco cytosolic *rpl2* gene (bottom) (Marty and Meyer, 1992 Ac. No. X62500) at nucleotide and amino acid levels. The percentage identity is indicated as bold letters at the end of the sequence. Changes in nucleotides are indicated by bold letters and amino acids by underlined bold letters. The oligomer used in the wheat cytosolic *rpl2* cDNA (oligo # 21) synthesis and the oligomers used in PCR (oligo # 20 & 1) are underlined.

Wheat cyt *rpl2*

```
 M  G  R  L  I  R  A  Q  R  K  G  A  G  S  V  F  K  S  H  T  H  H  R
ATGGGTCGTTTGATCAGAGCACAACGTAAGGGAGCAGGGTCCGTTTTCAAATCCCACACTCACCACCGTA
ATGGGTCGTTTGATCAGAGCACAACGTAAGGGAGCGGGTTCAGTCTTCAAGTCCCATACCCACCACCGTA  70
 M  G  R  L  I  R  A  Q  R  K  G  A  G  S  V  F  K  S  H  T  H  H  R
```

Tobacco cyt *rpl2* oligo # 20

```
 K  G  P  A  R  F  R  T  L  D  F  G  E  R  N  G  Y  L  K  G  V  I  T  E
AAGGCCCGGCCCGATTCCGTACCCTCGATTTCGGCGAACGTAATGGTTACCTAAAGGGAGTAATCACAGA
AAGGCCCTGCCCGTTTCCGTTCCCTCGACTTCGGCGAACGGAACGGTTACCTCAAGGGTGTCATAACGGA  140
 K  G  P  A  R  F  R  S  L  D  F  G  E  R  N  G  Y  L  K  G  V  I  T  E

  V  I  H  D  P  G  R  G  A  P  L  A  R  V  T  F  R  H  P  F  R  Y  K
AGTGATTCACGATCCCGGTAGAGGAGCACCACTCGCGCGCGTGACATTCCGTCACCCGTTCCGTTACAAG
GGTTATTCATGACCCAGGTAGGGGTGCTCCGTTGGCACGTGTCACTTTCCGTTACCCGTTCCGTTACAAC  210
  V  I  H  D  P  G  R  G  A  P  L  A  R  V  T  F  R  Y  P  F  R  Y  N

 H  Q  K  E  L  F  V  A  A  E  G  M  Y  T  G  Q  F  V  Y  C  G  K  K
CACCAAAAAGAGTTGTTCGTTGCTGCTGAAGGGATGTACACTGGTCAGTTCGTTTACTGTGGTAAAAAAG
CACCAGAAGGAACTGTTCGTGGCTGCTGAGGGTATGTACACTGGCCAGTTCGTTTACTGTGGGAAGAAGG  280
 H  Q  K  E  L  F  V  A  A  E  G  M  Y  T  G  Q  F  V  Y  C  G  K  K


A  T  L  M  V  G  N  V  L  P  L  R  S  I  P  E  G  A  V  V  C  N  V  E
CTACTCTTATGGTTGGTAATGTGTTGCCGCTCAGATCTATACCAGAAGGAGCTGTTGTATGTAACGTGGA
CTACTCTCATGGTTGGGAATGTGCTCCCACTCAGATCTATCCCTGAAGGAGCTGTCGTTTGCAATGTGGA  350
A  T  L  M  V  G  N  V  L  P  L  R  S  I  P  E  G  A  V  V  C  N  V  E

  H  K  V  G  D  R  G  V  F  A  R  C  S  G  D  Y  A  I  V  I  S  H  N
GCATAAAGTAGGAGATCGTGGTGTTTTTTGCTAGATGCTCTGGTGATTATGCCATTGTTATCAGTCACAAC
ACACAAAGTTGGTGACCGTGGTGTTTTTTGCCAGGTGCTCTGGTGCTTATGCTATTGTCATTAGTCACAAC  420
  H  K  V  G  D  R  G  V  F  A  R  C  S  G  A  Y  A  I  V  I  S  H  N

 P  D  N  G  T  T  R  V  K  L  P  S  G  A  K  K  I  V  P  S  G  C  R
CCTGATAATGGTACCACTAGGGTTAAGCTTCCATCAGGAGCCAAAAAGATTGTGCCCAGTGGATGTCGTG
CCCGATAACGGAACCACCAGGATTAAGCTGCCTTCAGGATCCAAGAAGATTGTCCCCAGTGGGTGTCGTC  490
 P  D  N  G  T  T  R  I  K  L  P  S  G  S  K  K  I  V  P  S  G  C  R

A  M  I  G  Q  V  A  G  G  G  R  T  E  K  P  M  L  K  A  G  N  A  Y  H
CAATGATTGGTCAAGTTGCTGGAGGAGGACGTACCGAGAAACCAATGCTCAAAGCCGGTAATGCTTACCA
GTATGATTGGTCAAGTTGCTGGAGGAGGACGTACTGAGAAACCAATGCTTAAGGCCGGTAACGCATACCA  560
R  M  I  G  Q  V  A  G  G  G  R  T  E  K  P  M  L  K  A  G  N  A  Y  H
                                                    oligo # 1

  K  Y  R  V  K  R  N  C  W  P  K  V  R  G  V  A  M  N  P  V  D  H  P
CAAGTATCGGGTGAAGAGGAACTGCTGGCCTAAGGTCCGTGGTGTTGCTATGAATCCAGTGGATCATCCT
CAAGTACCGGGTAAAGAGGAACTGCTGGCCCAAGGTTCGTGGTGTTGCCATGAACCCTGTGGAGCATCCT  630
  K  Y  R  V  K  R  N  C  W  P  K  V  R  G  V  A  M  N  P  V  E  H  P

 Q    95%
CAGGG 85%
CATGGTGGTGGTAACCATCAACATATTGGTCATGCTAGTACTGTCGCTCGCGATGCACCACCTGGTCAAA  700
 H  G  G  G  N  H  Q  H  I  G  H  A  S  T  V  A  R  D  A  P  P  G  Q

AGGTCGGTCTTATTGCAGGAAGGAGGACTGGTCGTCTTCGTGGTCAAGCTGCTGCTACAGCTGCCAAGGC  770
K  V  G  L  I  A  G  R  R  T  G  R  L  R  G  Q  A  A  A  T  A  A  K  A
          oligo # 21
TGACAAAGCTTAA    783
  D  K  A
```

rpl2 (see Fig. 5.1). The above observation clearly indicates that the nuclear-located mt and cytosolic *rpl2* are two distinct genes coding for different L2 proteins.

4.6 Expression of the wheat mt *rpl2* sequences at different developmental stages

The expression of the wheat nuclear-located mt *rpl2* gene at germinating embryo and seedling stages was interesting to analyze since the rice mt *rpl2* gene located in the mitochondrial genome was shown to have higher *rpl2* mRNA accumulation in germinating embryos than 6 d seedlings. Similarly, the expression of several other mt genes, for example wheat *atp6* (Bonen, 1987) and wheat *nad7* (Carrillo and Bonen, 1997) differs between these developmental stages relative to 18S rRNA. Northern analysis of the wheat mt *rpl2* gene was carried out on wheat poly A^+ RNA isolated from germinating embryos and seedlings, which were loaded equally based on cytosolic *rpl2* mRNA levels (Fig. 4.6.B). The hybridization experiment with the wheat cytosolic *rpl2* sequences (WPP3) detected a 1.4 kb mRNA which co-migrates with rRNA breakdown products. In addition, a non specific hybridization to rRNAs and its breakdown products were also observed (Fig. 4.6.A). The mt *rpl2* hybridization profile showed higher levels of stable transcripts (Fig. 4.6.C) in germinating embryos compared with etiolated seedlings whereas, cytosolic *rpl2* transcripts are approximately equal in level at both stages. The wheat nuclear *atpβ* gene coding for a mitochondrial protein also had a higher level of transcript accumulation (Fig. 4.6.D) during germination than 6 d seedling stage. Similarly, it had been determined that the transcript accumulation starts early on in seed germination and growth for certain chloroplast ribosomal protein genes (spinach *rps1, rps22, rpl21 and rpl40*) located in the nuclear genome compared to other nuclear-located cp genes like chlorophyll a\b-binding protein gene (reviewed by Mache et al., 1997).

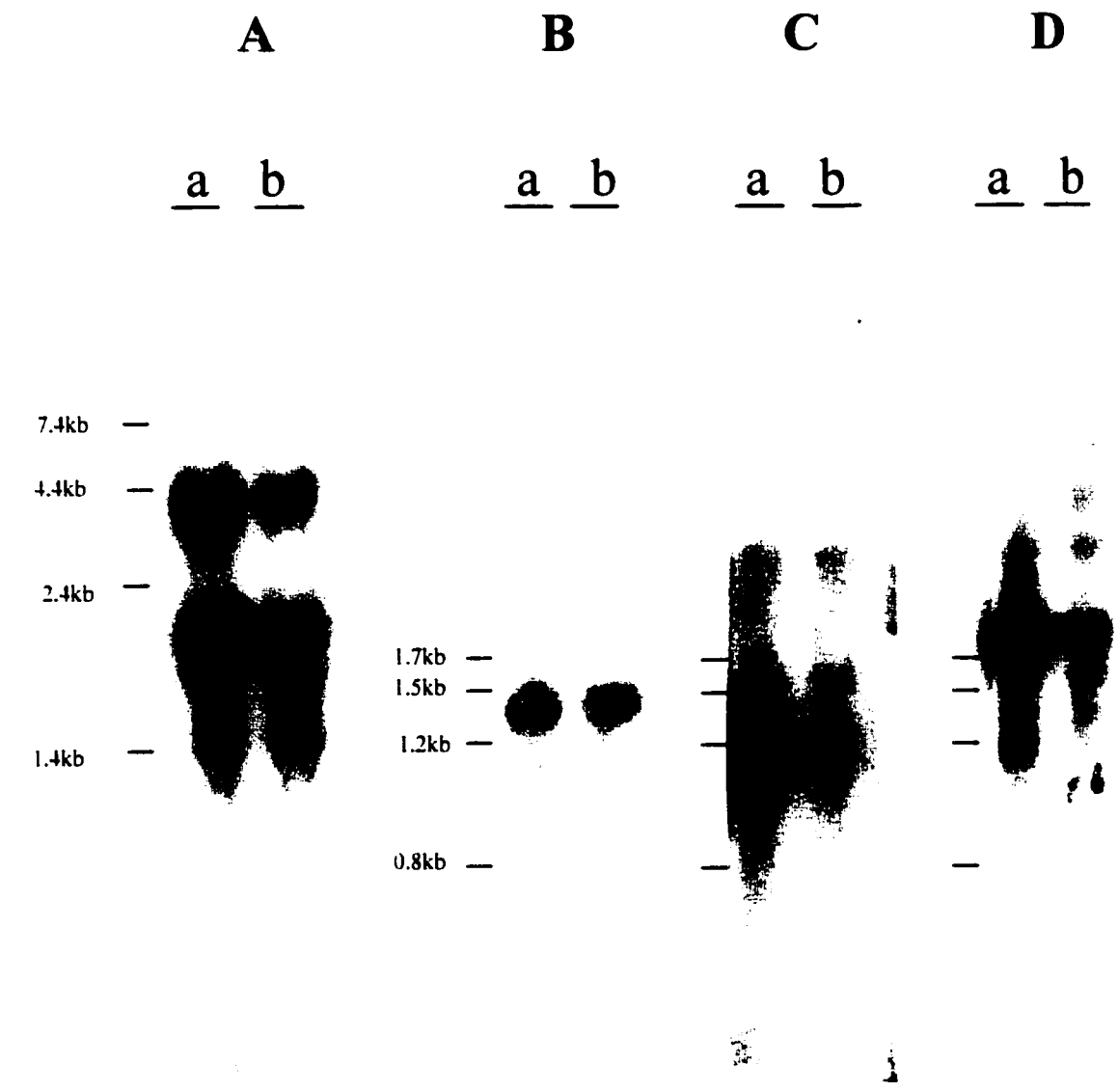
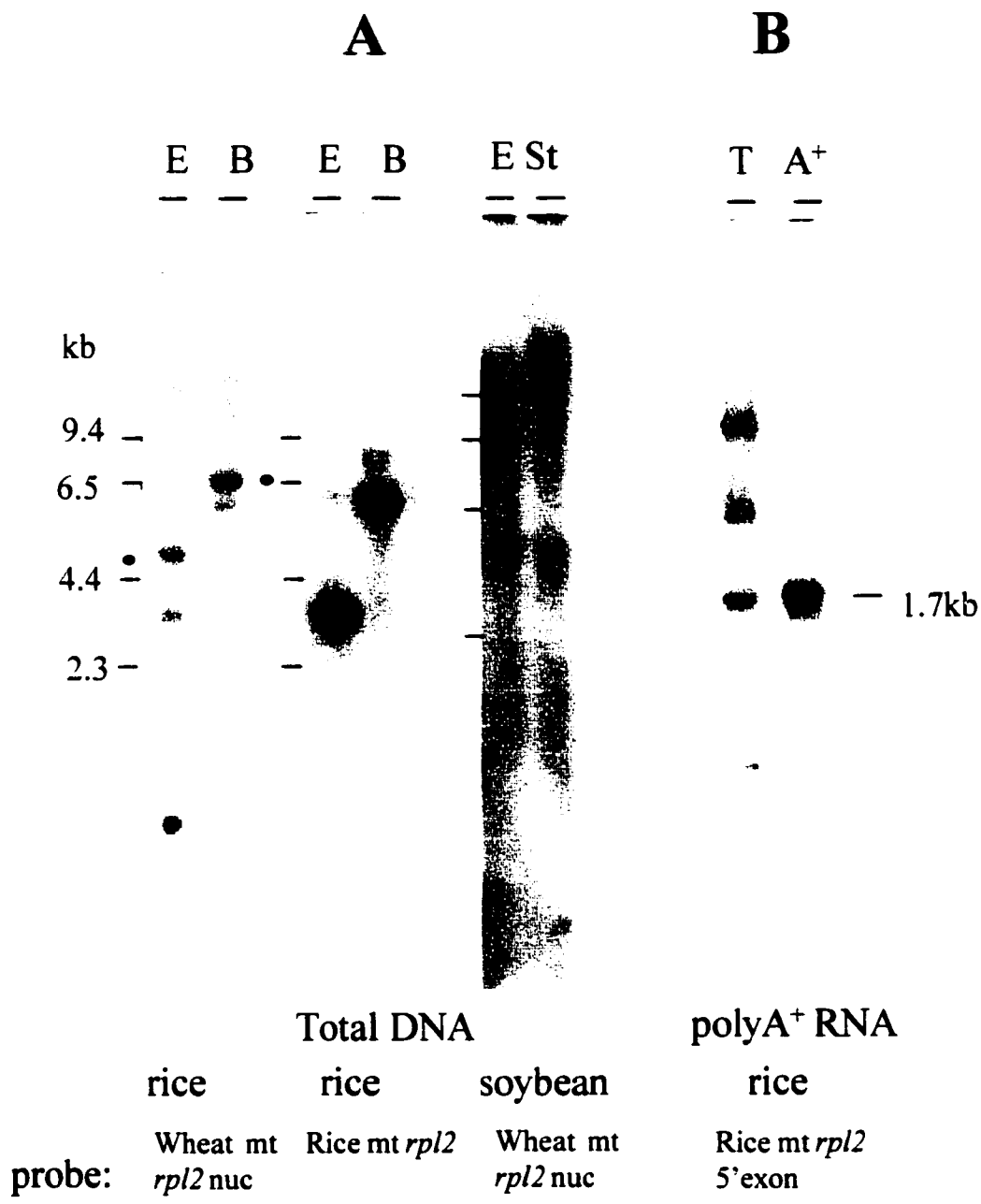

A
B
C
D
a b
a b
a b
a b
7.4kb
4.4kb
2.4kb
1.4kb
1.7kb
1.5kb
1.2kb
0.8kb
Wheat cyt rpl2
Wheat cyt rpl2
Wheat mt rpl2
Wheat atpB
Probe:
WPP3
oligo # 21
WPP1
WPP2

Figure 4.6 RNA blot analysis of wheat mt *rpl2*, cytosolic *rpl2* and mt *atpβ* at germinating embryos and seedlings

Wheat poly A^- RNA isolated from 24 h germinating embryos (a) and 6 d etiolated seedlings (b) were probed with wheat cytosolic *rpl2* (635 bp WPP3) sequences (**A**), cytosolic *rpl2* gene specific oligomer (oligo # 21) (**B**), mt *rpl2* copy 1 (WPP1) (**C**) and *atpβ* cDNA (WPP2) (**D**) probes. Panel B, C and D are the same blot serially hybridized and the exposure time was the same in all cases (3 days). Hybridizations were carried consecutively without stripping.


A
B
E B E B E St
T A+
kb
9.4
6.5
4.4
2.3
1.7kb
Total DNA
polyA+ RNA
rice rice soybean rice
probe:
Wheat mt rpl2 nuc
Rice mt rpl2
Wheat mt rpl2 nuc
Rice mt rpl2 5'exon

4.7 Functional mitochondrial *rpl2* sequences in the nuclear genome of rice

The low steady state levels of the rice mt-encoded *rpl2* mRNA relative to certain other rice mt protein coding genes (*atp6* and *nad4L*) and differential accumulation at two developmental stages (Fig. 3.7.B) raises questions about its functionality at different developmental stages and raises the possibility of another copy in the nucleus. When the wheat nuclear-located mt *rpl2* sequences were used as probe (WPP1) under low stringency hybridization conditions (37°C), rice total DNA showed two hybridizing bands for each restriction digest (Fig. 4.7.A). One of them (the lighter band) corresponds to the mt copy and the other nuclear copy as confirmed by reprobing the membrane with rice mt *rpl2* sequences (Fig 4.7.B). To determine whether these mt *rpl2* related sequences are expressed, northern analyses were performed using wheat mt *rpl2* sequences (WPP1) in rice poly A^- RNA. A non-specific hybridization to the rRNA (data not shown) was observed. However, when the rice mt *rpl2* 5' exon sequences (650 bp) were used, a 1.7 kb transcript was detected (Fig. 4.2.C) under low stringency conditions (37°C). The 1.7 kb signal is expected to be specific to poly A^- transcripts since mt RNA hybridization with this probe showed the mitochondrial signals of 3.5 kb and larger (5.0 kb and 7.0 kb) (Fig. 3.7.B probe 2) .

Because information about the rice nuclear-located mt *rpl2* sequences could provide insight into the mt *rpl2* gene transfer event(s), attempts were made to isolate the rice nuclear copy. PCR amplification experiments using wheat nuclear-located mt *rpl2* specific primers (oligo # 17 & 18) were carried out using rice total DNA as template, but the amplification of the mt copy apparently impeded the amplification of the nuclear copy. To avoid the contaminating mt DNA, gel purified nuclear DNA fragments (7.0 kb BamH1 and 5.0 kb EcoR1, see Fig. 4.7) corresponding to the nuclear copy were used in PCR experiments but they did not give any specific amplification. In

Figure. 4.7 Southern and northern analyses of rice nuclear-located mt *rpl2* sequences

A: Panel **1**: Rice total DNA blot probed with wheat nuclear-located mt *rpl2* copy 1 sequences (WPP1). The signal indicated by dots correspond to nuclear signal. The restriction enzymes EcoR1 (E) and Bam H1 (B) used in the digestion are indicated at the top. The DNA size marker is given at the left. Panel **2**: Rice total DNA blot (A) reprobed with rice mitochondrial *rpl2* sequences (RPP1). The additional faint signal corresponds to distantly related intron sequences located elsewhere in the rice mt genome. Panel **3**: Soybean total DNA blot (EcoR1 (E) and Sst1(St) probed with WPPI.

B: Rice total RNA and polyA$^-$ RNA blot probed with rice mt *rpl2* sequences (650 bp 5'exon). The faint signals in the total RNA lane corresponds to the migration position of rRNAs.

addition, cDNA synthesis was attempted for the nuclear-located mt *rpl2* copy of rice using poly A^- RNA. PCR and RT-PCR experiments were not successful in amplifying the nuclear copy. The oligomers used in the PCR experiments may be too divergent to amplify the rice nuclear copy. When tested in Southern hybridization experiments with rice total DNA, one of the oligomers (oligo # 17) hybridized to many positions (data not shown) so that might have hindered specific amplification in PCR experiments. Screening a rice cDNA or genomic library with the rice mt *rpl2* cDNA sequences and PCR experiments with degenerate primers might be more appropriate to obtain the rice nuclear-located *rpl2* sequences.

4.8 Search for mt *rpl2* sequences in the nuclear genomes of soybean and pea

The soybean and pea mt genomes lack *rpl2* homologous sequences and Southern hybridization results (Fig. 4.7.A panel 3) on total DNA indicated the presence of mt *rpl2* related sequences in the soybean nuclear genome. In an attempt to isolate these sequences, a soybean nuclear DNA library (soybean var Resnick) from Clontech was screened with wheat nuclear-located mt *rpl2* sequences under low stringency (37°C) hybridization conditions. The average insert size in the library were 8-20 kb and the size of the soybean genome is approximately 1-2 x 10^9 bp (Goldberg, 1978) so that screening 7 x 10^5 clones should give the desired clone with 99% confidence. Approximately a total of 8.8 x 10^5 lambda clones were screened in three different experiments. In the first screening 2, 6 and 8 potential positives were identified and carried over for the second screening. A single potential positive clone was identified in the second screening. This clone was examined for mt *rpl2* related sequences by subcloning the fragments, which showed hybridization to the nuclear-encoded wheat mt *rpl2* sequences. However, the sequences turned out to have no homology to *rpl2* but to be identical to *E. coli* sequences (Ac. No. U14003

sequences from chromosome K12). The transferred mt *rpl2* sequences may be too divergent between a monocot and dicot plant to be easily detected by hybridization. Similarly, genomic screening experiments to find the mt *rps2* gene in the *Arabidopsis* nuclear genome using wheat mitochondrial sequences were not successful (Vaitlingom et al., 1998).

Preliminary Southern data on pea nuclear DNA indicated the presence of mt *rpl2* related sequences (data not shown). Attempts were made to screen a pea cDNA library (Clontech) that was available in our lab using PCR amplification methods. The total pea cDNA pool ligated into lambda vectors was used as a template to amplify the mt *rpl2* sequences using wheat nuclear-located mt *rpl2* specific oligomers (oligo # 17 & 18) in PCR experiments. A single PCR product (400 bp) was amplified, but when it was directly sequenced it turned out to be vector sequences.

4.8.1 *PCR experiments to obtain the mt rpl2 gene from soybean total DNA or cDNA*

Soybean, unlike rice, does not have any *rpl2* sequences in the mt genome, hence low stringency conditions can be used for PCR amplification experiments. PCR and RT-PCR amplification techniques were tried on both soybean nuclear DNA and cDNA. Rice mt *rpl2* primers (oligo # 1 & 3) as well as primers specific to the wheat nuclear-located mt *rpl2* sequences (oligo # 17 & 18) were used. Low stringency temperatures (10°C lower than T_a of the primers) in PCR experiments led to the amplification of many nonspecific products that did not hybridize to wheat *rpl2* sequences in Southern experiments and no amplification was observed at higher temperatures including the temperature for successful wheat nuclear *rpl2* amplification.

Soybean sequences corresponding to the oligomer positions might be too divergent from the wheat mt *rpl2* to be amplified in PCR experiments even though selected regions of higher conservation were detected by Southern experiments. The nuclear-located mt *coxII* genes in the

dicot species soybean and cowpea belonging to the same family (Fabaceae) share 87.5% nucleotide identity which was sufficient for obtaining the soybean nuclear copy by screening (Covello and Gray, 1992). The wheat nuclear-located and rice mitochondrial-located mt *rpl2* copies share 70% identity over 600 bp, however, the oligomer regions might have had higher identity leading to the amplification of the wheat nuclear copy.

In summary, the nuclear genome of wheat has multiple copies of *rpl2* sequences of mitochondrial origin. These DNA sequences detect an abundant poly A^+ mRNA suggesting that at least one copy represents a functional gene. An RNA-mediated gene transfer is predicted, since the nuclear sequences lack an intron corresponding to the one in the rice *mt rpl2* gene. The rice nuclear genome also has a single copy of transcribed mt *rpl2* related sequences in addition to the functional mt copy.

CHAPTER 5

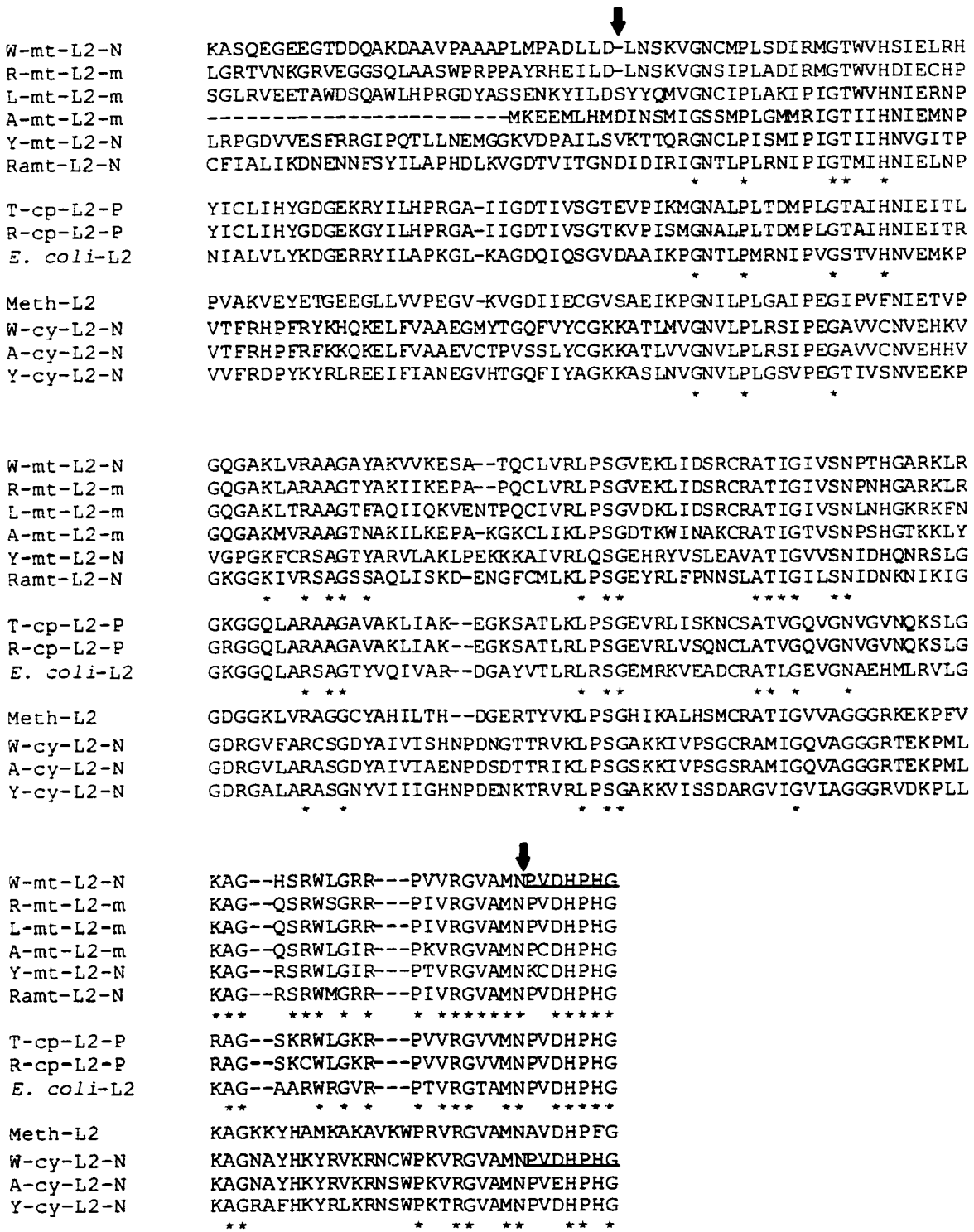
COMPARATIVE SEQUENCE ANALYSIS OF MITOCHONDRIAL L2 PROTEINS

5.1 Phylogenetic analysis of L2 ribosomal proteins

The L2 protein is one of the few ribosomal proteins that is sufficiently conserved for homologues in all three cellular compartments (chloroplast, mitochondrial and cytosol) to have been identified and also in all lineages of organisms examined (from eubacteria, archaea and eukarya). A conserved carboxyl domain of 18 amino acids corresponding to *E. coli* L2 protein positions 220-237 is a diagnostic characteristic of the L2 family of proteins (Marty and Meyer et al., 1992). A phylogenetic analysis of L2 sequences from different compartments and organisms was carried out taking advantage of our knowledge of the evolutionary history of species to examine whether the mode of evolution of mitochondrial L2 proteins was in agreement. The organellar genes included in the analysis are the chloroplast *rpl2* of rice and tobacco and the mitochondrial-located *rpl2* genes of a protist (*Reclinomonas americana*), liverwort and rice as well as the nuclear-located yeast, wheat and *Arabidopsis* mt *rpl2* sequences. In addition, the cytosolic *rpl2* genes of wheat, *Arabidopsis* and yeast, the archaeal *(Methanococcus jannaschii) rpl2* and the bacterial *(E. coli) rpl2* were also included in the analysis. The rice mt, wheat mt and wheat cytosolic L2 sequences were characterized in this work and the rest were retrieved from the Genbank/EMBL databases. The cDNA sequences reported for the *Arabidopsis* L2 protein as a databank submission were very short (Fig. 5.1) but included an AUG codon (see details below). Representatives of both monocot and dicot plants in each compartment were included in the analysis for comparing the available nuclear-located mt *rpl2* sequences of a monocot (wheat) and a dicot (*Arabidopsis*) and eubacterial, protist and fungal sequences were also included. A 115

Figure 5.1 Comparison of L2 sequences within the conserved carboxyl terminus from different organisms

The carboxyl-terminal sections of L2 proteins (approximately 150 a.a) from mitochondria (mt), chloroplast (cp) and cytosolic (cy) compartments were aligned. The location of the gene, that is nuclear (N), mitochondrial (m) and plastid (P) are indicated. The name of the sequences; their accession number; and coordinates used in the analysis are indicated in brackets for all sequences analyzed. Wheat (W-mt-L2 Fig 4.2.B 71-213), rice (R-mt-L2- D78336, 315-467), liverwort (L-mt-L2- P26859, 312-468), *Arabidopsis* (A-mt-L2- S49579, 1-122), yeast (Y-mt-L2-S30827, 180-346), *Reclinomonas americana* (Ra -mt-L2- O21247 94-239), chloroplast L2 of tobacco (T-cp-L2-P06379, 88-231), rice (R-cp-L2-P17351, 88-231), cytosolic L2 of *Arabidopsis* (A-cy-L2-1173018 61-211), Wheat (Fig. 4.5. 61-211), yeast (Y-cy-L2-S50243, 61-211), *Methanococcus jannaschii* (*Archaea*-L2-D64322,62-210) and *E. coli* (*E. coli*-L2-P02387, 90-233). Asterisks indicate the conserved positions among all of the sequences positioned above the asterisk. The mitochondrial, chloroplast & bacterial, archaeal and the cytosolic L2s are separated categorically to illustrate their relative similarities to the mt proteins. The a.a residues corresponding to oligomer # 1 in wheat mt and cytosolic L2 are underlined. The arrows indicate the region used in the phylogenetic tree construction. The rice and wheat mt and wheat cytosolic L2 sequences are from this work and the rest retrieved from Genbank/EMBL entries. Their accession numbers are given in Appendix 5.


W-mt-L2-N KASQEGEEGTDDQAKDAAVPAAAPLMPADLLD-LNSKVGNCMPLSDIRMGTWVHSIELRH
R-mt-L2-m LGRTVNKGRVEGGSQLAASWPRPPAYRHEILD-LNSKVGNSIPLADIRMGTWVHDIECHP
L-mt-L2-m SGLRVEETAWDSQAWLHPRGDYASSENKYILDSYYQMVGNCIPLAKIPIGTWVHNIERNP
A-mt-L2-m ----------MKEEMLHMDINSMIGSSMPLGMMRIGTIIHNIEMNP
Y-mt-L2-N LRPGDVVESFRRGIPQTLLNEMGGKVDPAILSVKTTQRGNCLPISMIPIGTIIHNVGITP
Ramt-L2-N CFIALIKDNENNFSYILAPHDLKVGDTVITGNDIDIRIGNTLPLRNIPIGTMIHNIELNP
T-cp-L2-P YICLIHYGDGEKRYILHPRGA-IIGDTIVSGTEVPIKMGNALPLTDMPLGTAIHNIEITL
R-cp-L2-P YICLIHYGDGEKGYILHPRGA-IIGDTIVSGTKVPISMGNALPLTDMPLGTAIHNIEITR
E. coli-L2 NIALVLYKDGERRYILAPKGL-KAGDQIQSGVDAAIKPGNTLPMRNIPVGSTVHNVEMKP
Meth-L2 PVAKVEYETGEEGLLVVPEGV-KVGDIIECGVSAEIKPGNILPLGAIPEGIPVFNIETVP
W-cy-L2-N VTFRHPFRYKHQKELFVAAEGMYTGQFVYCGKKATLMVGNVLPLRSIPEGAVVCNVEHKV
A-cy-L2-N VTFRHPFRFKKQKELFVAAEVCTPVSSLYCGKKATLVVGNVLPLRSIPEGAVVCNVEHHV
Y-cy-L2-N VVFRDPYKYRLREEIFIANEGVHTGQFIYAGKKASLNVGNVLPLGSVPEGTIVSNVEEKP
W-mt-L2-N GQGAKLVRAAGAYAKVVKESA--TQCLVRLPSGVEKLIDSRCRATIGIVSNPTHGARKLR
R-mt-L2-m GQGAKLARAAGTYAKIIKEPA--PQCLVRLPSGVEKLIDSRCRATIGIVSNPNHGARKLR
L-mt-L2-m GQGAKLTRAAGTFAQIIQKVENTPQCIVRLPSGVDKLIDSRCRATIGIVSNLNHGKRKFN
A-mt-L2-m GQGAKMVRAAGTNAKILKEPA-KGKCLIKLPSGDTKWINAKCRATIGTVSNPSHGTKKLY
Y-mt-L2-N VGPGKFCRSAGTYARVLAKLPEKKKAIVRLQSGEHRYVSLEAVATIGVVSNIDHQNRSLG
Ramt-L2-N GKGGKIVRSAGSSAQLISKD-ENGFCMLKLPSGEYRLFPNNSLATIGILSNIDNKNIKIG
T-cp-L2-P GKGGQLARAAGAVAKLIAK--EGKSATLKLPSGEVRLISKNCSATVGQVGNVGVNQKSLG
R-cp-L2-P GRGGQLARAAGAVAKLIAK--EGKSATLRLPSGEVRLVSQNCLATVGQVGNVGVNQKSLG
E. coli-L2 GKGGQLARSAGTYVQIVAR--DGAYVTLRLRSGEMRKVEADCRATLGEVGNAEHMLRVLG
Meth-L2 GDGGKLVRAGGCYAHILTH--DGERTYVKLPSGHIKALHSMCRATIGVVAGGGRKEKPFV
W-cy-L2-N GDRGVFARCSGDYAIVISHNPDNGTTRVKLPSGAKKIVPSGCRAMIGQVAGGGRTEKPML
A-cy-L2-N GDRGVLARASGDYAIVIAENPDSDTTRIKLPSGSKKIVPSGSRAMIGQVAGGGRTEKPML
Y-cy-L2-N GDRGALARASGNYVIIIGHNPDENKTRVRLPSGAKKVISSDARGVIGVIAGGGRVDKPLL
W-mt-L2-N KAG--HSRWLGRR---PVVRGVAMNPVDHPHG
R-mt-L2-m KAG--QSRWSGRR---PIVRGVAMNPVDHPHG
L-mt-L2-m KAG--QSRWLGRR---PIVRGVAMNPVDHPHG
A-mt-L2-m KAG--QSRWLGIR---PKVRGVAMNPCDHPHG
Y-mt-L2-N KAG--RSRWLGIR---PTVRGVAMNKCDHPHG
Ramt-L2-N KAG--RSRWMGRR---PIVRGVAMNPVDHPHG
T-cp-L2-P RAG--SKRWLGKR---PVVRGVVMNPVDHPHG
R-cp-L2-P RAG--SKCWLGKR---PVVRGVVMNPVDHPHG
E. coli-L2 KAG--AARWRGVR---PTVRGTAMNPVDHPHG
Meth-L2 KAGKKYHAMKAKAVKWPRVRGVAMNAVDHPFG
W-cy-L2-N KAGNAYHKYRVKRNCWPKVRGVAMNPVDHPHG
A-cy-L2-N KAGNAYHKYRVKRNSWPKVRGVAMNPVEHPHG
Y-cy-L2-N KAGRAFHKYRLKRNSWPKTRGVAMNPVDHPHG

amino acid sequence positioned near the carboxyl terminus of the L2 protein was chosen for the analysis because it is the only region that can be accurately aligned between cytosolic and organellar L2 derived proteins (Marty and Meyer, 1992). The wheat nuclear-located mt *rpl2* copy I sequences were used since the sequenced region was longer than for copy II. Alignment of sequences was performed using the ClustalW program and then minor adjustments by visual inspection. The L2 protein alignment is shown in Fig. 5.1 and a subset of 115 aa was used in constructing unrooted distance matrix phylogenetic trees of L2 sequences using the Fitch and neighbor joining methods in PHYLYP (Felsentein, 1993) and PUZZLE programs. The numerical values for percentage amino acid identity between pairwise comparisons of the L2 and S12 sequences used in the analysis are given in matrixes in Table 4 and 5 respectively. Although the use of a relatively short sequence (115 a.a) and specific region (conserved carboxyl terminus) of the protein is less preferable than fuller information, there are examples of even shorter ribosomal proteins of less than 100 amino acids (S17, S19) which have given phylogenetic trees comparable to the larger proteins (Muller and Whitman-Labeled, 1997). Similarly, conserved regions of other ribosomal proteins such as S10, S12 and S7 have given the same topology as that of the full length ribosomal proteins (Muller and Whitman-Labeled, 1997).

The conventional phylogenetic analysis where such data can be used to estimate relative divergence times of different species makes the assumption of a constant molecular clock over time, although this may not actually be the case. This analysis of plant mitochondrial L2 proteins has some representatives which have been transferred to a different compartment (nucleus) during evolution. In the new environment, it is expected that there will be an adaptation period of accelerated evolution, as proposed for the nuclear *cox2* gene in legumes (Nugent and Palmer,

1991) and in the nuclear *rps12* in *Oenothera* (Grohmann et al., 1992). In addition, the nucleotide substitution rates of nuclear compartment is estimated to be approximately 12 times higher than the mitochondrial compartment at synonymous substitution positions (Wolfe et al., 1989). Hence, the phylogenetic tree of the mt L2 proteins is expected to reflect these two potential factors that may have an influence on the branching pattern and length.

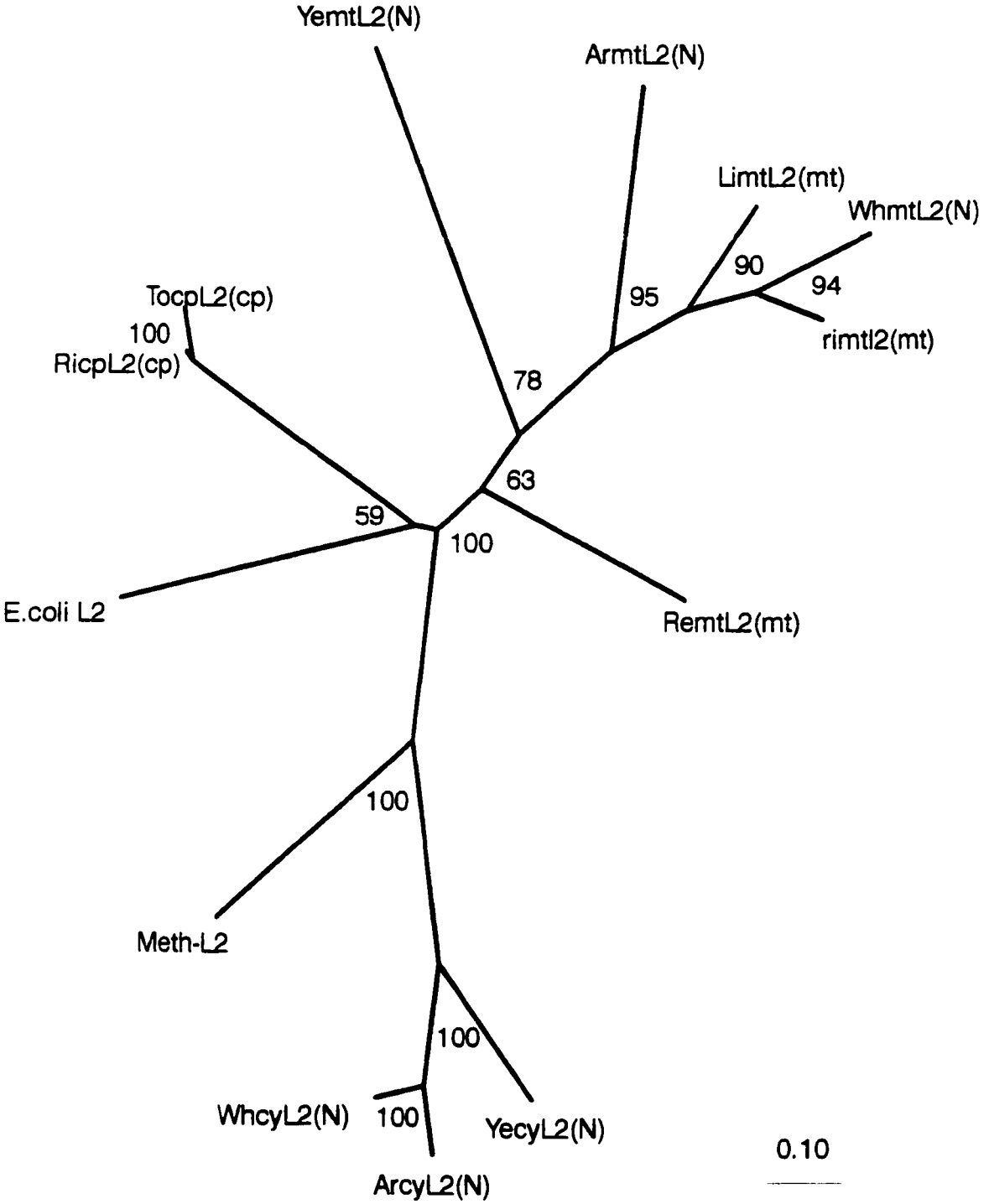
The phylogenetic trees constructed using different programs show similar topology with minor differences (Fig 5.2). It is clear that proteins from chloroplast and cytosolic compartments each form coherent and well resolved non-overlapping phylogenetic domains compared to mitochondrial L2 proteins. The chloroplast L2 proteins, *E. coli* L2 and the *Reclinomonas* mt L2 share approximately equal percentage identities between each other which is reflected in all treeing methods as nearly equal nodal distance between these units. The number of substitutions observed between the organellar and eukaryotic type cytosolic proteins may have reached the saturation point which is shown by the PUZZLE tree as unresolved nodes for *E. coli*, cp, cytosol and *R. americana* L2 proteins. The neighbor joining and Fitch trees have a different branching order for the *Reclinomonas* sequence. The Fitch tree groups the mitochondrial type proteins together, whereas the neighbor joining tree is consistent with the organismal phylogeny of protists branching off early within the eukaryotic lineage, that is relatively close to the eukaryotic and eubacterial split (Roger et al., 1999).

The eukaryotic-type cytosolic L2 proteins present in yeast and wheat share a higher identity of 68% than their mt counterparts which show only 49%. The greater divergence may reflect independent gene transfer events in lineages leading to wheat and yeast. The branching order of wheat and rice mt L2 proteins clearly indicates that they are closely related even though their

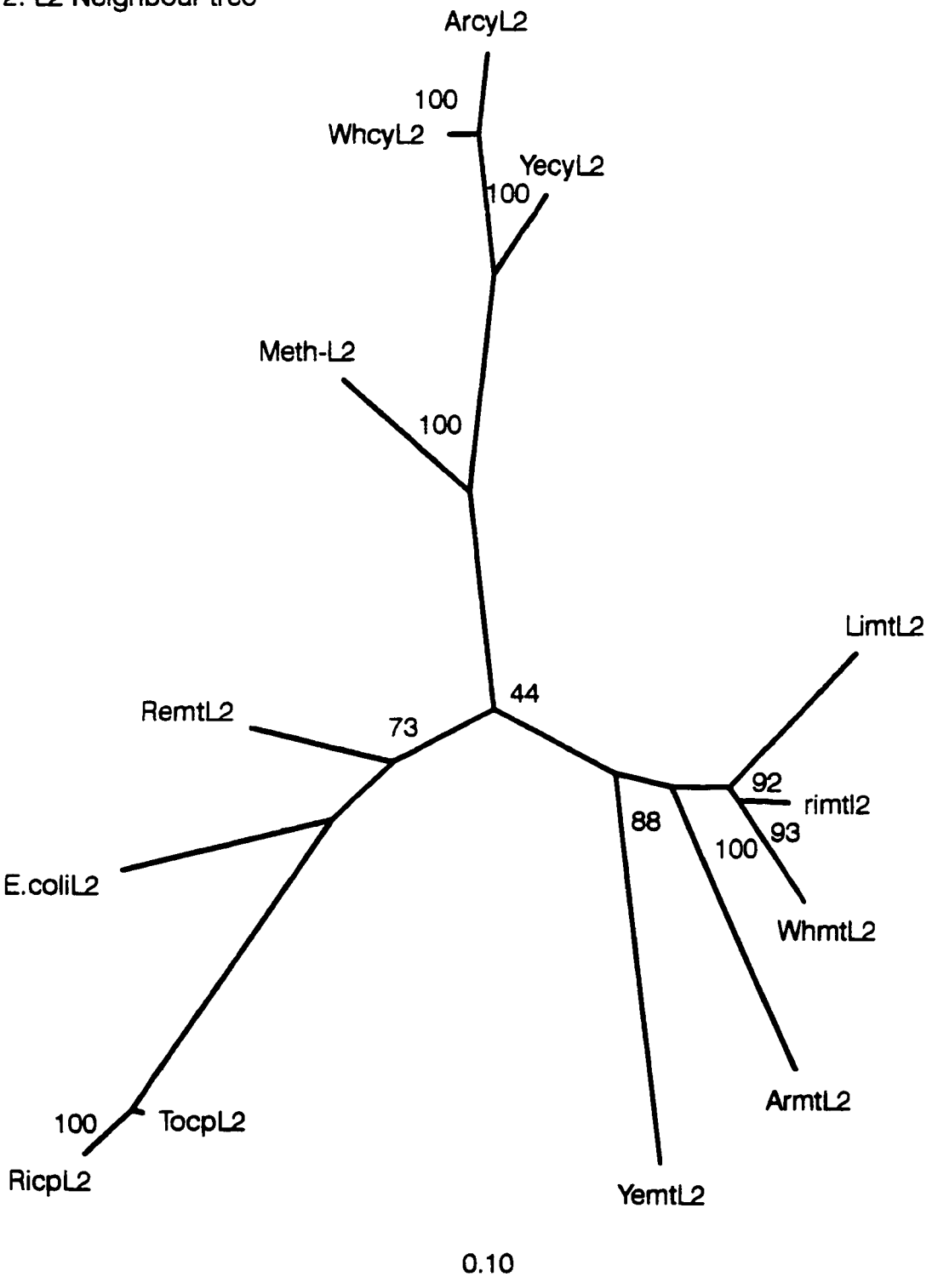
Figure 5.2 Phylogenetic trees of L2 protein sequences

L2 phylogenetic trees using Fitch (1) Neighbor joining (2), and PUZZLE (3) methods. A 115 amino acid region indicated by triangles from the conservative carboxyl-terminus of different L2 proteins was used to construct unrooted distance matrix phylogenetic trees. The sequence alignment is shown in Fig. 5.1 Numbers at the nodes are bootstrap values based on 100 replications. The accession numbers of the sequences used in this tree are given in Appendix 5. The scale for the branches are given below and note that they are different between trees. The different sequences used are from rice (ri), wheat (wh), liverwort (li), *Oenothera* (Oe), yeast (ye), tobacco (to), *Reclinomonas americana* (Ra), *Arabidopsis* (Ar), from mitochondrial (mt), chloroplast (cp) and cytosolic (cy) compartments and from *Methanococcus jannaschii* (Arch) and *E. coli*.

1: L2 Fitch tree



2: L2 Neighbour tree



3: L2 Puzzle tree

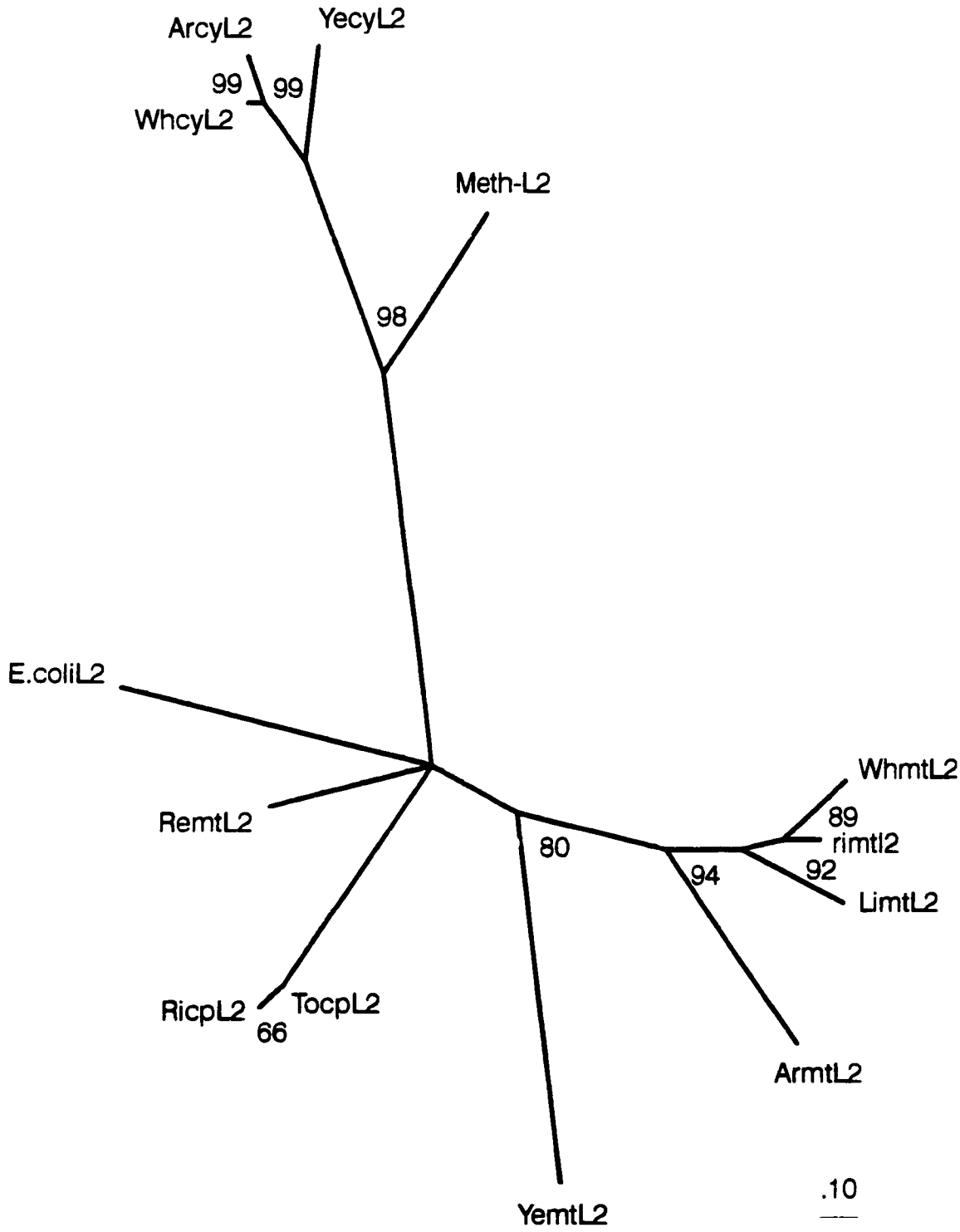


Table 4 pairwise amino acid identities between L2 sequences of different organisms and compartments used in the phylogenetic analysis

	W-mt	R-mt	A-mt	L-mt	Y-mt	Ra-mt	T-cp	R-cp	Meth	*E.coli*	W-cy	A-cy	Y-cy
W-mt	100												
R-mt	84.1	100											
A-mt	62.6	65.4	100										
L-mt	69.1	76.6	60.6	100									
Y-mt	48.7	50.0	52.4	49.1	100								
Ra-mt	50.5	51.5	51.9	57.8	51.2	100							
T-cp	52.4	49.5	51.0	47.1	49.0	51.7	100						
R-cp	51.5	48.5	48.6	46.2	50.0	50.0	92.1	100					
Meth	43.0	43.0	41.5	46.8	36.8	42.2	41.7	40.0	100				
E. coli	44.9	47.7	44.9	46.1	49.0	47.0	49.1	48.2	43.7	100			
W-cy	38	39.8	36.7	39.4	36.6	37.2	34.7	35.6	52.1	41.9	100		
A-cy	39.8	41.7	36.4	40.9	35.7	37.4	35.6	36.4	53.4	41.9	89.7	100	
Y-cy	41.2	44.7	38.9	42.2	37.4	35.2	36.9	37.7	49.6	39.3	67.8	66.7	100

L2 protein sequences of mitochondrial (mt), chloroplast, (cp), cytosolic (cy) compartments from wheat (W), rice (R), *Arabidopsis* (A), *Oenothera* (O), tobacco (T), liverwort (L), yeast (Y), *Reclinomonas americana* (Ra), *Methanococcus jannaschii* (Meth) and *E. coli* were used. The length of different L2 sequences compared here are approximately 115 amino acids. The accession numbers are given in Appendix 5.

genes are located in different compartments. The rice mt L2 has a shorter branch length compared to wheat which is in congruence with the slower evolution rate in plant mt genomes than in nuclear genomes (Wolfe et al., 1989). The mitochondrial L2 proteins of the monocots, rice and wheat share 84% identity. In contrast those from the chloroplast and cytosol compartments share approximately 90% identity between a dicot and monocot and this is are reflected as a long branch length between rice and wheat in contrast to shorter branches between monocot and dicot in cp and cytosol counterparts. This may be due to the accelerated evolution expected in the transferred genes immediately after a change in compartment. The wheat mt L2 protein is encoded by a nuclear gene which has been transferred from the mitochondrion relatively recently. In the nucleus, the transferred gene must acquire sequences for its expression and it has to adapt to the nuclear transcription and translation machinery, hence it may undergo a period of rapid evolution immediately after transfer. The branching order of wheat and rice L2 in relation to liverwort is keeping with the traditional concept of plant phylogeny (Cho et al., 1998).

The *Arabidopsis* L2 sequences were obtained from an unpublished databank entry (Accession No. S49579) which contains only 169 amino acids with a stretch of 150 amino acids related to L2 sequences in contrast to the 500 amino acids for plant mt L2 proteins situated in the mt genome. A lower amino acid identity is shared between wheat and *Arabidopsis* L2 (63%) than wheat and liverwort L2 (69%), so that the *Arabidopsis* nuclear-encoded L2 sequence branches earlier than that of the mt-located liverwort L2 in all three L2 phylogenetic trees (Fig. 5.2). The branch length of *Arabidopsis* L2 sequence is also very long compared to those of liverwort and wheat. The nucleotide substitution rate which is expected to be approximately twelve times higher in the nuclear compartment relative to the mt compartment (Wolfe et al., 1989) can not completely

account for the divergence of *Arabidopsis* L2, since wheat mt L2 is also encoded by a nuclear-encoded gene. The cytosolic genes are eukaryotic type derived from the genome of the host cell (protoeukaryote) and they are in the nucleus from the time of eukaryotic cell development whereas the mt L2 proteins are eubacterial type transferred to the nucleus. An escalated evolution rate immediately after the transfer of a gene to a nucleus is the most likely reason for the divergence observed between *Arabidopsis* and rest of the mitochondrial members. Besides, two independent gene transfer events may also show changes for a transient period of time. The integration sites of independent transfer events would be different, relating to the sequence context in the new environment. The transferred gene could have gained or lost certain domains that have a different substitution rate because they may be under different functional constraint than L2. For example, in *Arabidopsis*, the transferred mt *rps19* gene has acquired sequences from an RNA binding protein (Sanchez et al., 1996). It is also known that long branches in a phylogenetic tree can cluster together giving false representation of the topology (Moreira et al., 1999). Hence, it is possible that certain long branches in the tree, such as those of yeast and *Arabidopsis* L2 proteins were drawn together artefactually.

The highly divergent nature of the *Arabidopsis* nuclear-located sequences raise some other possibilities. A new functional constraint is likely since the *Arabidopsis* sequences do not have the plant mt specific L2 inserted sequences. It is possible that an additional copy is present in the *Arabidopsis* nuclear genome which may be responsible for the protein functioning in the ribosome. In addition, the truncated mt copy in the mt genome which is transcribed and spliced may produce a protein corresponding to the amino terminal region and the carboxyl terminus of the protein by the nuclear copy. Nonetheless, the mt and nuclear sequences together would not

constitute a complete protein because an internal 50 amino acid stretch would be missing. These internal 50 amino acid sequences are present in both copies of the wheat nuclear- located mt *rpl2* sequences. Regardless of the unusual nature of *Arabidopsis* L2 protein, the accelerated rate of evolution and multiple independent transfer events appear to be important factors in explaining the high divergence of mt *rpl2* genes located in the nuclear genome.

5.2 Prediction of time of *rpl2* gene transfer to the nucleus

According to the molecular clock hypothesis, genes that are under similar functional constraint accumulate nucleotide substitutions at an approximately constant rate in all evolutionary lineages (Li and Graur, 1991). The molecular data for the available mt *rpl2* homologues of rice and wheat located in different compartments may theoretically be used to predict the divergence time if we take into consideration the differences in substitution rate between compartments and the apparent accelerated rate of evolution upon transfer. A time estimation was carried out for the wheat nuclear-located *rpl2* sequences considering just the change in substitution rate between compartments using the formula T=K/2r where T is divergence time, r is rate of nucleotide substitution per site per year and K is number of substitutions per site. Between wheat and rice, the number of substitutions per synonymous site was estimated to be approximately 1.46 and the number of substitutions at the nonsynonymous sites was 0.2 using the Li 93 program (Li, 1993). The synonymous nucleotide substitution rate of approximately $6x\ 10^{-9}$ substitutions per site per year which corresponds to the average rate of plant nuclear genes (Wolfe et al., 1989) was used in the calculation. The divergence time of these sequences were therefore estimated to be approximately 120 Myr, that is after the divergence of monocots and dicots which is approximately 200 Myr, but before the divergence of lineages

leading to rice and wheat which is 70 Myr as estimated by Laroche et al., (1995). Considering an initial burst of evolution when the gene is transferred to a new compartment, 120 Myr may even be an overestimate. The wheat mt *rpl2* gene transfer event could have happened relatively recently or immediately after the split from lineage leading to rice. In that case the *rpl2* sequences in the rice nuclear genome would be the result of an independent transfer event. Accordingly, the presence of a rice mitochondrial copy may suggest that the gene transfer in rice is very recent and the mt copy is yet to be inactivated. On the other hand, if there was a single transfer event, the rice mt copy might also have been maintained for specialized functions.

In addition, the available information for mt *rpl2* genes from this study is consistent with nuclear copies in both monocots and dicots. This can suggest two possibilities, either a single event before the monocot-dicot split or multiple independent events within different lineages. Multiple independent events appear more likely at this time because the *Arabidopsis* and wheat nuclear sequences show a very high degree of substitutions.

Data on the location of homologous genes in different compartments and phylogenetic analyses have been used as a tool to predict the possible time of gene transfer in several other cases. For example, for the cowpea mt *cox2,* transfer to the nuclear genome was predicted to be 60-200 Myr (Nugent and Palmer, 1991). In the L2 comparative analysis, additional flowering plant mt L2 sequences encoded by either mitochondrial and/or nuclear genomes would provide information on the evolution rate between L2 proteins located in both genomes which may help understand the evolutionary pressure or constraint on mt L2 sequences in different compartments.

5.3 Rapid evolution of the mt *rpl2* gene in comparison to the mt *rps12*

To gain further insight into whether the mode of evolution of the L2 protein is similar to that

of other ribosomal proteins, a comparative analysis was carried out using another mt ribosomal protein S12. The mt S12 protein was selected because both L2 and S12 proteins are located in the mt genome of liverwort and rice, and in each case they share 77% identity suggesting rather a similar evolution rate. There is also a nuclear-located mt *rps12* gene in *Oenothera* (Grohmann et al., 1992) available for comparison. The size of S12 (125 a.a) is comparable to the length of the L2 sequences used in the above comparative analysis, both have a common function and they bind to each other in the functional ribosome (Wittman, 1983). Although the mt *rps10* (Wischmann and Schuster, 1995) and *rps11* (Kadowaki et al., 1996) genes are located in the nucleus in *Arabidopsis* and rice respectively, there is very little sequence information for mt copies from different plants. In wheat, the gene coding for the S12 protein is in the mt genome whereas the mt *rpl2* gene is in the nucleus. This allows us to compare the evolution of rp genes in different compartments.

The amino acid alignment of different S12 proteins used in the phylogenetic analysis are given in Fig. 5.3 and the pairwise comparisons of mt S12 proteins are given in Table 5. When the genes are located in different compartments a dramatic difference is observed in the percentage identity. The wheat mt S12 protein shares 99% identity with rice, whereas the L2 protein shares only 84% identity. This observation may partly be due to the difference in substitution rates between compartments (nucleus vs mitochondria). Analogously, a 78% identity between wheat and liverwort S12 proteins, in contrast to just 69% in L2 was also observed. The synonymous and nonsynonymous substitution rates of plant nuclear genes are approximately 12 and 6 times higher respectively than for mt genes (Wolfe et al., 1989). Moreover an accelerated evolution is expected when a mt gene is transferred to a new compartment as discussed earlier.

Figure 5.3 Comparison of S12 sequences used in the phylogenetic analysis

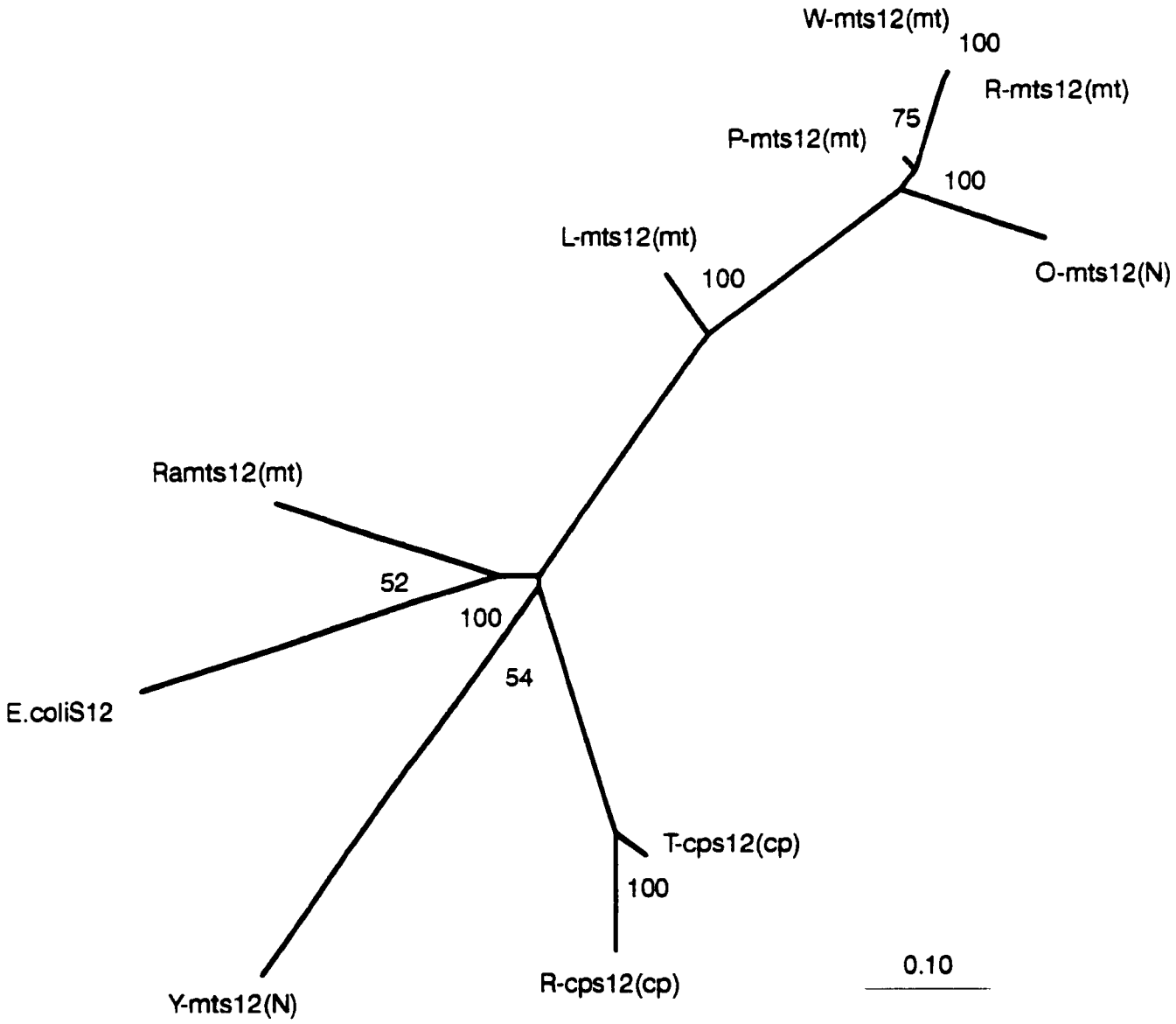
S12 proteins from mitochondria (mt), and chloroplast (cp) compartments were aligned. The *Oenothera* and yeast mt proteins are encoded by nuclear genes while the rest by genes in the mt genome. The names given for sequences in the alignment and their accession number are given in brackets. Wheat mt S12 (W-mts12; P10851), rice mt S12 (R-mts12; P28520), liverwort mt S12 (L-mts12; P26871), petunia mt S12 (P-mts12; P49195) *Reclinomonas* mt S12 (Ramts12; AA11866), yeast mt S12 (Y-mts12; P53732), *Oenothera* mt S12 (O-mts12; Q01607), *E. coli* S12 (*E. coli* s12; P02367), rice cp S12 (r-cps12; P12149) and tobacco cp S12 (t-cps12; R3NT12).

```
W-mts12       ---------- ---------- ---------- ---------- ----------
R-mts12       ---------- ---------- ---------- ---------- ----------
L-mts12       ---------- ---------- ---------- ---------- ----------
P-mts12       ---------- ---------- ---------- ---------- ----------
Ramts12       ---------- ---------- ---------- ---------- ----------
Y-mts12       ---------- ---------- ---------- ------MLSR FMSNTWCTPL
O-mts12       MSGGRWISNL RQVFGRSRAS ISSSLLPFPP TSISSGFHVP RLLDAWKNPV
E.coliS12     ---------- ---------- ---------- ---------- ----------
R-cps12       ---------- ---------- ---------- ---------- ----------
T-cps12       ---------- ---------- ---------- ---------- ----------

W-mts12       ---------- ---MPTKNQ- LIRHGREEKR RTDRTRASDQ CPQKQGVCLR
R-mts12       ---------- ---MPTKNQ- LIRHGREEKQ RTDRTRASDQ CPQKQGVCLR
L-mts12       ---------- ---MPTMNQ- LVRKGRESKR RTKRTRALNK CPQKQGVCLR
P-mts12       ---------- ---MPSLNQ- LIRHGREEKR RTDRTRALDQ CPQKQGVCPR
Ramts12       -------MPT INQLLKINK- KRKADKNERR KFSKAPALES CPQKKGVCTR
Y-mts12       RQAQRLFSST TTMQATLNQI KRGSGPPRRK KISTAPQLDQ CPQRKGVVLR
O-mts12       AAAAAFRLPQ SSGFATMNQ- LIRHGREEKH RSDRKRALGK CPQKQGVCLR
E.coliS12     ---------- ---MATVNQ- LVRKPRARKV AKSNVPALEA CPQKRGVCTR
R-cps12       ---------- ---MPTVKQ- LIRNARQPIR NARKSAALKG CPQRRGTCAR
T-cps12       ---------- ---MPTIKQ- LIRNTRQPIR NVTKSPALRG CPQRRGTCTR

W-mts12       VSTRTPKKPN SALRKIAKVR LSNRHDIFAH IPGEGHNSQE HSIVLVRGGR
R-mts12       VSTRTPKKPN SALRKIAKVR LSNRHDIFAH IPGEGHNSQE HSIVLVRGGR
L-mts12       VSTRSPKKPN SALRKIAKVR LTNRNEIIAY IPGEGHNLQE HSVVMVRGGR
P-mts12       VSTRTPKKPN SAPRKIAKVR LSNRHDIFAH IPGEGHNLQE HSMVLIRGGR
Ramts12       VFIASPKKPN SAKRKVARVR LTNGQEVTAY IPGEGHSLQE HSVVLIRGGR
Y-mts12       VMVLKPKKPN SAQRKACRVR LTNGNVVSAY IPGEGHDAQE HSIVYVRGGR
O-mts12       VSTRTPKKPN SALGKIAKVR LSNRNDVFAY IPGEGHNLQE HSMVLVRGGR
E.coliS12     VYTTTPKKPN SALRKVCRVR LTNGFEVTSY IGGEGHNLQE HSVILIRGGR
R-cps12       VYTINPKKPN SALRKVARVR LTSGFEITAY IPGIGHNLQE HSVVLVRGGR
T-cps12       V-TITPKKPN SALRKVARVR LTSGFEITAY IPGIGHNLQE HSVVLVRGGR

W-mts12       VKDSPGVKSH RIRGVKDLLG IPDRRKGRSK YGAERPKS-K *
R-mts12       VKDSPGVKSH RIRGVKDLLG IPDRRKGRSK YGAERPKS-K *
L-mts12       VQDLPGVKYH CIRGVKDLQG IPGRRRGRSK YGTKKPKDYI *
P-mts12       VKDSPGVKSH CIRGVKDLLG IPDRRRGRSK YGAEKPKS-I *
Ramts12       VKDLPGVKYH IVRGTFDLQG VQNRRQGRSL YGTKKQKD-L *
Y-mts12       CQDLPGVKYH VIRGAGDLSG VVNRISSRSK YGAKKPSK-S *
O-mts12       VKDLPGVKFH CIRGVKDLLG IPDRRRGRSK YGAEKPKS-T *
E.coliS12     VKDLPGVRYH TVRGALDCSG VKDRKQARSK YGVKRPKA-- *
R-cps12       VKDLPGVRYR IIRGALDAVA VKNRQQGRSK YGVKKPKK-- *
T-cps12       VKDLPGVRYH IVRGTLDAVG VKDRQQGRSK YGVKKPK--- *
```

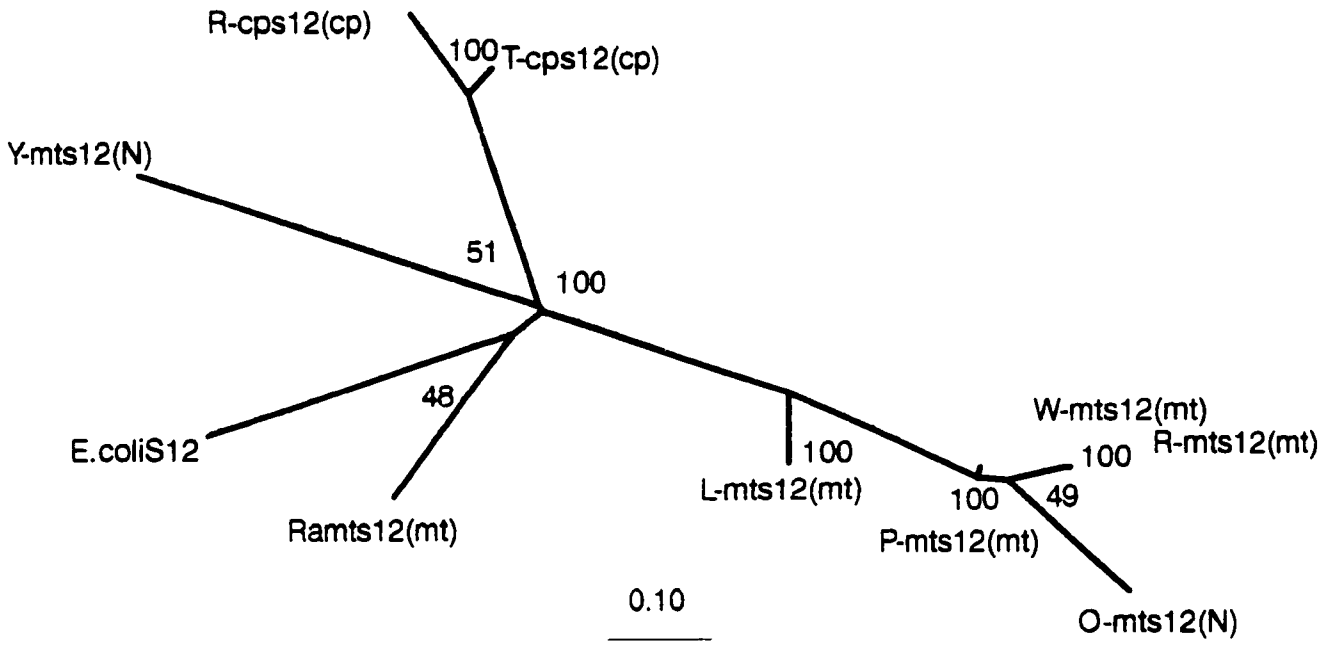
Figure 5.4 Phylogenetic trees of S12 protein sequences

S12 phylogenetic trees using Fitch (1) and neighbor joining (2), and PUZZLE (3) methods. The mitochondrial and nuclear-located genes coding for the mt S12 ribosomal protein, chloroplast, cytosolic and *E. coli* S12 were used in the analysis to construct unrooted distance matrix trees. Numbers at nodes indicate the bootstrap values. The accession numbers of the protein sequences are given in Appendix 5. The different sequences used are rice (ri), wheat (wh), liverwort (li), *Oenothera* (Oe), yeast (ye), tobacco (to), *Reclinomonas americana* (Ra), *Arabidopsis* (Ar), Petunia (Pe) from mitochondrial (mt), chloroplast (cp) and cytosolic (cy) compartments and *E. coli*.

1: S12 Fitch tree



2: S12 neighbour tree



3: S12 Puzzle tree

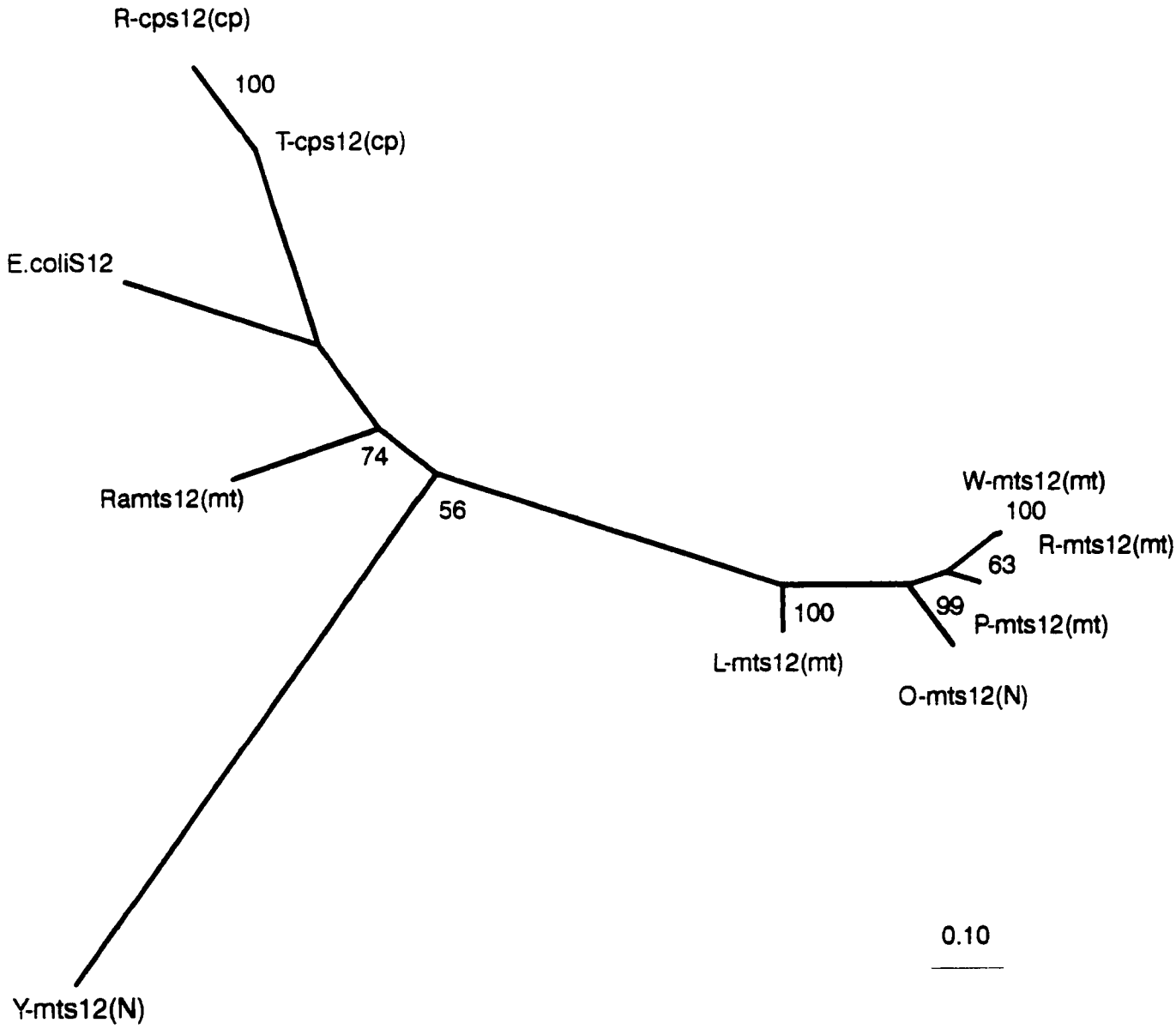


Table 5 Pairwise amino acid identities between S12 sequences of different organisms and compartments used in the phylogenetic analysis

	W-mt	R-mt	P-mt	L-mt	O-mt	Y-mt	Ra-mt	T-cp	R-cp	*E. coli*
W-mt	100									
R-mt	99.2	100								
P-mt	91.1	90.3	100							
L-mt	78.0	77.2	78.9	100						
O-mt	85.2	85.2	88.3	81.8	100					
Y-mt	56.8	56.0	56.5	61.5	55.8	100				
Ra-mt	58.1	57.3	58.9	65.4	59.8	59.2	100			
T.cp	59.3	58.5	56.9	64.5	61.5	55.4	68.2	100		
R-cp	56.9	57.9	56.9	62.6	59.8	56.1	64.3	88.6	100	
E. coli	56.9	56.9	59.3	61.8	59.8	58.4	66.7	72.8	68.8	100

S12 protein sequences of mitochondrial (mt), chloroplast, (cp), cytosolic (cy) compartments from wheat (W), rice (R), *Arabidopsis* (A), *Oenothera* (O), tobacco (T), Petunia (P), liverwort (L), yeast (Y), *Reclinomonas americana* (Ra) and *E. coli* were used. The length of different S12 sequences compared are approximately 125 amino acids. The accession numbers are given in Appendix 5.

In the phylogenetic tree comparisons (Fig. 5.4), even though the percentage identity of L2 mt and nuclear sequences (84%) and S12 mt and nuclear sequences (85%) were approximately same they (Fig. 5.2) have different branch lengths compared to other mt proteins suggesting a difference in the evolution rate. In contrast, the topology of chloroplast protein branches are similar in both S12 and L2 phylogenetic trees as expected. When the topology of just the mitochondrial S12 and L2 are examined, the plant mt S12 proteins are seen to cluster more tightly than do L2 proteins. For example, the wheat and rice mt S12 share 99% identity and their branches are not resolved, and the dicot petunia which shares 91% identity with wheat also has a very short branch suggesting that the mt S12 is a highly conserved protein. In addition, the branches leading to chloroplast nodes are shorter in S12 than L2 suggesting that eubacterial type S12 protein are conservative.

L2 proteins are constrained enough to detect the protein homologues in all lineages (Fig. 5.1) whereas S12 cytosolic partners have not been identified in the completely sequenced yeast nuclear genome or in other eukaryotes. The cytosolic protein which is named as S12 in *Xenopus* and yeast actually has no detectable homology to the *E. coli* counterpart (Seery et al., 1994). The differences observed between branch lengths of yeast mt counterparts of L2 and S12 are in keeping with the rapid evolution of the mt L2 especially when both genes have been transferred to the nucleus during early endosymbiosis. The amino acid identity observed between *R. americana* mitochondrial S12 and *E. coli* S12 is 66% identity whereas only 50% was observed in L2 between them which was reflected in the branching order. The overall topologies of the trees with regard to protist, fungal and plants are as expected. In summary, the evolution of the mitochondrial-type L2 regardless of the compartment where the gene is located, appears to be rapid compared to

S12 even though the L2 family of proteins as a whole are conserved. Hence the absence of mt *rpl2* or pseudogene nature of the mt genome can very well correspond to a functional gene in the nucleus.

5.4 Codon usage pattern of the wheat nuclear-located mt *rpl2* gene in comparison with other wheat mitochondrial and nuclear genes of wheat

The "genome hypothesis" proposed by Grantham et al. (1980) postulates that genes in any given genome use a similar coding strategy with respect to choices among synonymous codons. In addition, selection for translation efficiency may also play an important role in the choice of synonymous codons (Sharp et al., 1988). The wheat mt *rpl2* gene has been transferred from the mt genome to the nuclear genome. The change in compartment suggests an evolutionary pressure on the gene to switch over to the optimal type of codon usage pattern for translation in the cytoplasm. The chloroplast genes that were transferred to the nucleus during early endosymbiosis show a codon usage pattern of nuclear genes rather than genes encoded by the cp genome (Oliver et al., 1990).

To get more insight into the evolution of the transferred-wheat mt *rpl2* gene with regard to codon usage pattern, a comparative analysis was carried out using the codon usage pattern of other wheat mt- and nuclear-located genes of wheat. Available wheat mt rp genes and respiratory protein genes (as of October, 1998) were included in this analysis for a total of approximately 5500 codons. The mt genes were categorized into respiratory chain and ribosomal protein genes based on their function. However, no difference was observed for the two categories of the mt genes (Table 6). The nuclear genes totaling 6100 codons were divided based on the compartment to

Table 6. Codon usage pattern of the wheat nuclear-located mt *rpl2* gene in comparison to wheat mt and nuclear genes

Genes	Third position of codons %						Overall GC content of genes
Wheat nuclear genes	G	C	A	T	AT	GC	
cytosolic proteins	23.4	32.2	17.3	27.1			
mitochondrial proteins	24.3	37.7	14.2	23.8			**58.0**
chloroplast proteins	29.0	50.2	7.2	13.7			
					34.4	**65.6**	
Wheat mitochondrial genes							
respiratory chain proteins	13.8	20.7	29.3	36.2			**48.5**
ribosomal proteins	18.1	16.8	31.2	34.0			
					65.4	**34.7**	
Wheat mt *rpl2* (nuclear)	30.5	40.5	15.3	13.7	**28.9**	**71.1**	**61.0**

The names and accession numbers of the genes used in the analysis are given in Appendix 6. The codon usage table is given in the Appendix 7. Wheat nuclear genes are categorized according to the compartment to which the proteins are targeted and the mt genes according to function.

which their products are targeted (chloroplast ~1400 codons, mitochondrial ~1500 codons and cytosol~ 3300 codons). The chloroplast-targeted protein coding genes showed a difference in their codon usage pattern compared with mt specific nuclear genes (Table 6) in that the percentage of codons ending with A or T in mt genes is 38% in contrast to only 20% for cp genes. Unlike wheat, maize nuclear genes do not show any discrepancy in their codon usage pattern between genes coding for the mitochondrial and chloroplast proteins (Fennoy and Bailey-Serres, 1993). This suggests a species-specific variation for codon usage pattern in these plants. The whole set of all three types of wheat nuclear genes tends to show a preference for codons with G or C in the third position (66 %) compared with genes in the mt genome which have a bias (65%) for A or T at the third position of codons. Similarly, analysis of maize nuclear genes also has revealed that 40% of the variation in codon usage were due to bias towards the GC third position versus AT (Fennoy and Bailey-Serres, 1993) when nuclear genes specifying proteins for all three compartments were considered as a unit. The GC content of the wheat nuclear genes is found to be 58% in the analysis whereas the mt genes has 49% which may have influenced their codon preferences. The cp-encoded genes in higher plants typically do not show a strong codon bias (Mortan, 1998) with some exceptions such as the *psbA* and *rbcL* genes. The mt *rpl2* sequences that are located in the nucleus have a codon usage pattern similar to the nuclear genes (Table 6) suggesting the adaptation of the transferred genes in the new environment. This observation is consistent with the nuclear type codon usage observed for transferred cp genes (Oliver et al., 1990). However, these cp genes appear to have been transferred very early on in evolution at the time of plastid origin (Palmer and Delwiche, 1996) unlike the wheat mt *rpl2* which has been predicted to be relatively recent. This suggests that the mode of evolution of the

transferred genes and the degree of selective constraints in the nuclear environments have led to a codon usage pattern similar to the nuclear genes. This is in agreement with the transferred genes adapting to the new environment to express and target the protein back to mitochondrion.

CHAPTER 6

GENERAL DISCUSSION

6.1 Hypothetical steps in gene transfer from the mitochondrion to the nucleus

The transfer of a functional mt gene to the nucleus during evolution is expected to proceed through specific stages as inferred from various lines of evidence. A hypothetical model is depicted in Figure 6.1 In plants, mt gene transfer to the nucleus appears to be an ongoing evolutionary process unlike animal and fungal systems, where the use of non-universal genetic codes for mt genes precludes functional gene transfer (reviewed by Bonen, 1998). The random breakage of mitochondrial membranes could lead to the movement of genetic information in the form of DNA or RNA or reverse transcribed cDNA molecules and subsequent integration into the nuclear genome. The movement of mt genetic material can be theoretically tolerated, since numerous copies of mt genomes are present in each mitochondria as well as many mitochondria in every cell. However, it is presumably a rare event when compared to the early endosymbiotic period, since mitochondria are well established with specific function in the eukaryotic cell. Theoretically, the enormously large size of the plant nuclear genome and the presence of extensive amount of non-coding sequences can easily accommodate mt DNA integrations. On the other hand, integrated sequences are found to be only activated if they are integrated in transcribed regions of the genome possibly because of the higher probability of acquiring expression sequences to achieve nuclear transcription, cytosolic translation and import of the protein into the mitochondrion. During this period, the mt copy must remain active and produce the functional protein. Once the nuclear copy is "activated" and it performs equivalent to the original mt copy the possibility of gene inactivation is theoretically equal in both genomes. The rapid

Figure 6.1 Hypothetical steps in functional gene transfer from the mitochondrion to the nucleus

Panels A,B,C and D depict various stages of a gene transfer event.

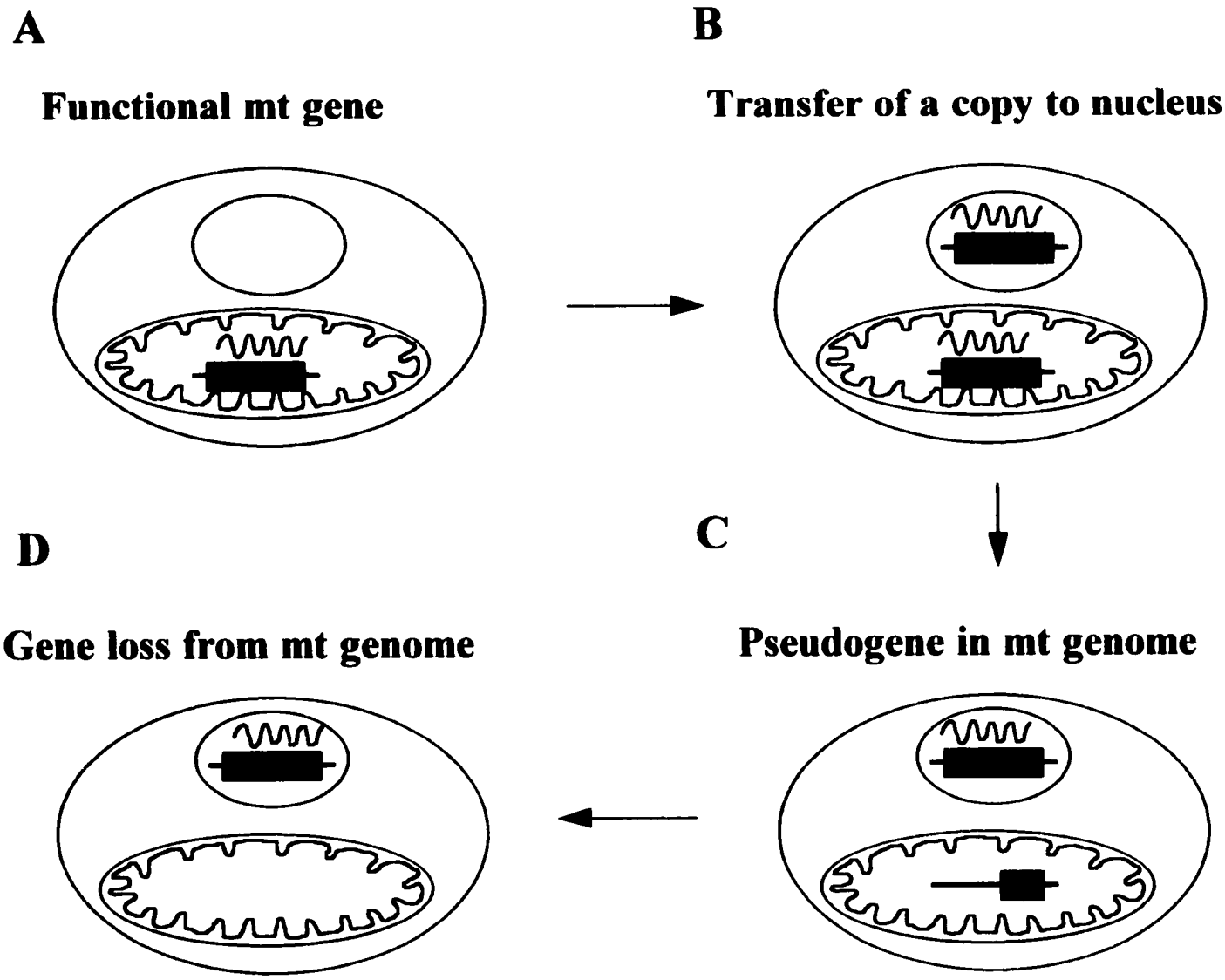
A: Presence of a functional mt gene in the mt genome. Although multiple copies of DNA and RNA are present in the mitochondrion only one is shown here for simplicity.

B: A transferred mt gene is integrated into the nuclear genome and the mt copy is still active in the mt genome. The nuclear copy can either be activated or degenerated based on the acquisition of sequences responsible for its expression.

C: Activation of the nuclear copy is followed by the loss of function in the mt copy which can occur either by point mutation or truncation and deletion caused by DNA rearrangements.

D: Complete loss of the mt copy after the activation of the transferred nuclear gene.

In the figure, different sized ellipses represent the cell, the mitochondrion with cristae and the nucleus. The black block represent a gene. The wavy line on the top of the blocks indicate the RNA.


A
Functional mt gene
B
Transfer of a copy to nucleus
C
Pseudogene in mt genome
D
Gene loss from mt genome

rearrangements occurring in the mt genome may be one reason for more frequent degeneration of mt genes than nuclear ones. In addition, the control and regulation of a gene may be more advantageous in the nuclear genome than in the mitochondrion.

In this project, the status of the mt *rpl2* gene was examined in both the mt and nuclear genomes of several different plants (rice, wheat, soybean and pea) to gain knowledge about the mode of production of the L2 protein and the effects of gene migration to the nucleus in flowering plants. In wheat, the mt genome has only a *rpl2* pseudogene and the transferred mt *rpl2* sequences in the nuclear genome are found to be actively expressed. In contrast, the rice mt genome has an intact *rpl2* gene which is found to be transcribed at low levels with a complex transcript profile. In addition, a single copy of the mt *rpl2* related sequences were detected in the rice nuclear genome and corresponding stable poly A^+ transcripts by hybridization experiments. The soybean and pea mt genomes do not have any *rpl2* sequences but related sequences were detected in the soybean nuclear genome by Southern analysis. The location and expression of mt *rpl2* sequences in these plants illustrate certain of the hypothetical evolutionary steps involved in a functional gene transfer event.

6.2 Mechanisms of functional gene transfer and adaptation

Inactivation of a mitochondrial gene can occur if the gene is not constrained to produce the functional protein, for example if an active copy is present elsewhere in the mt or nuclear genome. The co-existence of a functional mt gene and truncated copies is observed occasionally in mt genomes (eg. wheat *cox2*; Bonen et al., 1984). However the presence of only a pseudogene or the absence of a mt gene in comparison to other plant mt genomes is consistent with the presence of a nuclear copy having been transferred from the mt genome and this has been observed in several

instances (Table 2).

The wheat mt *rpl2* homologous sequences that have been identified in the wheat nuclear genome are believed to be transferred mt sequences based on their similarity (70% and 69% identity at both nt and amino acid levels respectively) to the rice mt *rpl2* gene located in the mt genome and they do not have an intron at the corresponding rice *rpl2* site, hence they are believed to be transferred via an RNA intermediate. The nuclear-located mt *rps10* gene of *Arabidopsis* (Wischmann and Schuster, 1995) also does not have an intron corresponding to the position in potato mt *rps10* (Zanlungo et al., 1995). RNA editing information in plant mt transcripts can also be used for identifying the mode of gene transfer since there is no evidence for the presence of RNA editing machinery for the nuclear genes. Unlike DNA-mediated transfers, ones that involve RNA intermediates lead to functional nuclear genes, since they have the complete coding region that is spliced and edited. However the transferred mt sequences must be modified to a certain extent during the process of gene activation in the nuclear genome.

The wheat mt *rpl2* transcript size is consistent with the absence of a portion of the mt *rpl2* counterpart. This suggests a possible deletion of *rpl2* sequences during or after transfer. Because northern hybridization experiments using the rice mt *rpl2* 5' sequences which are not present in the isolated wheat nuclear *rpl2* sequences detected transcripts of the same size, a deletion in the plant specific *rpl2* inserted sequences is predicted. Moreover, the N-terminal region of the L2 protein has a conserved region that binds to the peptidyl transferase center studied from bacterial system (*E. coli* L2, 5-186 amino acids; Olson et al., 1991), so that a loss of the 5'end of the gene may not be tolerated unless there is a separate gene present in the nuclear genome that codes for N-terminal protein sequences. The nuclear-located mt *rpl2* sequences of *Arabidopsis* were

sufficient only to code for 150 amino acids in contrast to 500 amino acids for the plant mt L2. In this regard, the nuclear-located mt *rps19* gene of *Arabidopsis* encodes a short protein with a deletion of 15 out of 95 amino acids in comparison to the mt encoded S19 of rice (Sanchez et al., 1996). This suggests that deletions which are not detrimental to the function of the gene could happen during or after the transfer of a mt gene to the nuclear compartment. The mt L2 specific inserted sequences are highly divergent (33% between rice and liverwort) and they may have a different functional constraint such as a non-translatory role specific to plants. In addition, a specific function may be attributed to the inserted sequences present in plant mt L2 proteins since they have been conserved between nonvascular and flowering plants over ~400 Myr but not present in *Reclinomonas* mt and *E. coli* L2 proteins. In addition, the presence of at least part of the plant mt L2 specific sequences in both copies of the nuclear-located mt *rpl2* in wheat suggests a potential role for them in the ribosomal function. However, they are divergent and there are insertions and deletions observed between the two nuclear copies and between rice mt *rpl2* sequences. Similar deletions in triplets were observed in the nuclear-located *Arabidopsis* mt *rps10* in comparison to the potato mt copy (Wischmann and Schuster, 1995). These amino acid changes may favor the protein in folding and/or transport of the protein through mt membranes which is in keeping with the idea of an accelerated evolution of the transferred sequences.

Additional functional domains besides the mt target signals could also be added to the transferred mt genes as observed in *Arabidopsis rps19* (Sanchez et al., 1996). An RNA binding domain is attached to its N terminus which has been proposed to replace the function of S13. Similarly N-terminal extension (NTE) sequences are observed in the chloroplast S18 protein of rye (Weglohner et al., 1995). They consist of tandem repeats of a basic heptapeptide motif. The

number of repeats are different between species and they are presumed to have RNA binding capacity. The L2 inserted sequences have no such repetitive nature that can be correlated to any function and they do not match with any known proteins. Yeast mt rp proteins for example, L23, L5 and L2 are bigger and have extra stretches (~ 100 amino acids) compared to the *E. coli* counterpart (reviewed by Graack and Wittman-Liebold, 1998). The extra domain present in plant mt L2 proteins may have multi functional nature similar to ribosomal proteins of other lineages. The yeast *mrp4* is a cytosolic rp gene that has been analyzed to have a non-respiratory role in addition to the ribosomal function by mutation studies (Gracck et al., 1995). The L32 protein of yeast influences three different elements of RNA processing, which include processing of pre-rRNA, splicing of *rpl32* transcripts in the nucleus and translation of the *rpl32* mRNA in the cytoplasm (Vilardell and Warner, 1997). Comparative and structural analyses of protein sequences of certain eukaryotic and archaeal rp genes have led to the identification of an RNA binding multifunctional domain which is conserved in translation suppresser proteins (Koonin et al., 1994). Hence, an extra function of suppression of translation termination is predicted for these ribosomal proteins (S12, L7, L30). Wool, (1996) has reported the possibility of several potential non-translatory roles for ribosomal proteins.

It can be concluded that a RNA-mediated gene transfer mechanism is necessary for the functional mt gene in the nucleus. During the process of gene transfer, extra sequences in addition to the target sequences can be added to the mt gene, which may bring an additional function or deletion of sequences that do not impede its primary function.

6.3 Plant mt DNA rearrangements: impact on mt gene inactivation

Recombination events causing mt DNA rearrangements in the mt genome appear to be the

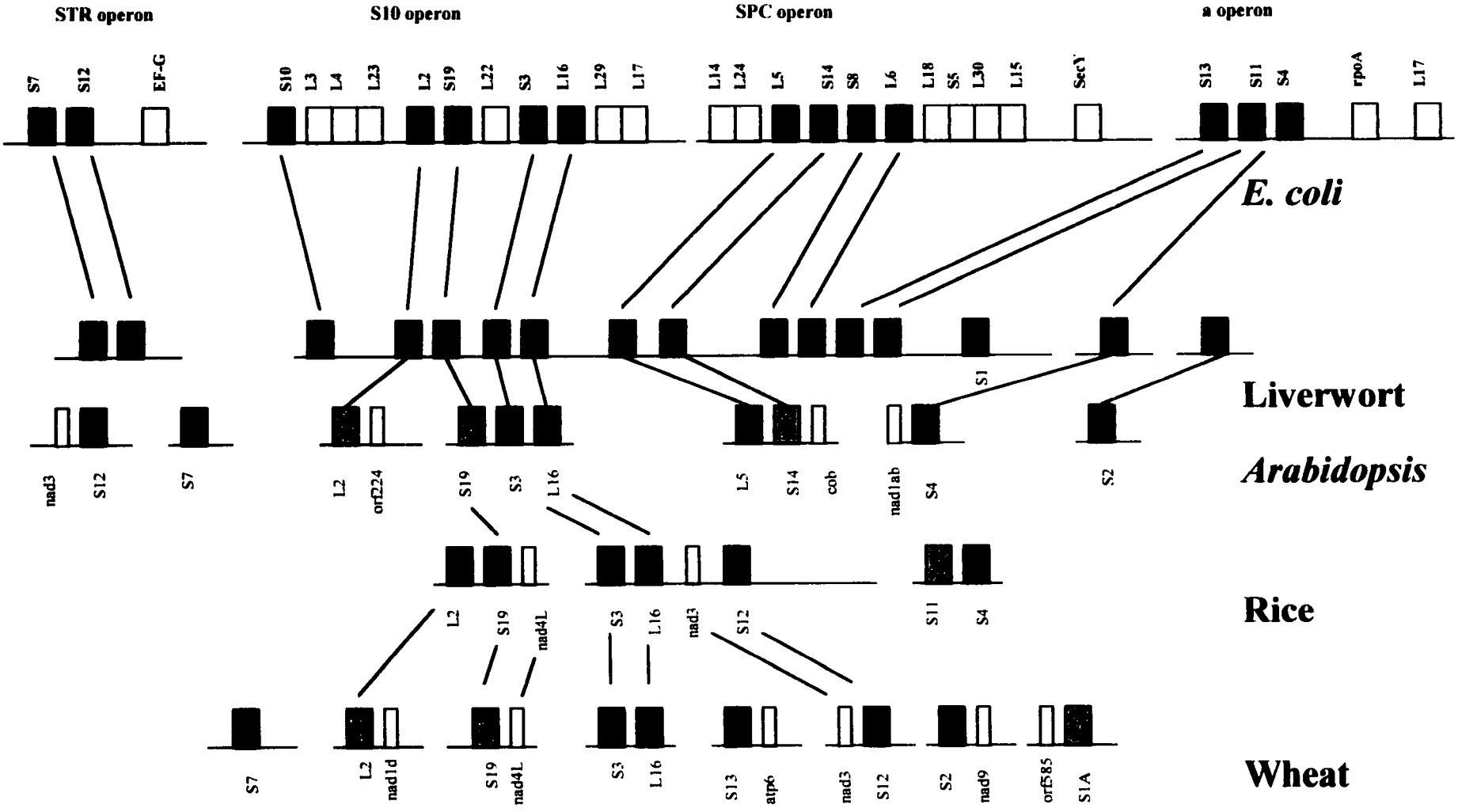
chief mode of gene inactivation. Figure 6.2 portrays the arrangement of some mt rp genes in rice, wheat and *Arabidopsis* in comparison with the liverwort mitochondrial and *E. coli* ribosomal protein gene clusters. The liverwort rp gene cluster clearly indicates that the gene arrangement is same as *E. coli*, which is consistent with the eubacterial origin of mitochondria. As seen from this figure, deletions and pseudogenes are an apparent effect of DNA rearrangements, which also result in the replacement of a respiratory chain gene in the place of a rp gene. For example, the order of five rp genes (*rps10-rpl2-rps19-rps3-rpl16*) are conserved between the S10 operon of *E. coli* and liverwort mitochondria (Fig 6.2 black boxes). When the liverwort arrangement is compared to other flowering plants, several DNA rearrangements are noticed. In wheat, at least four rearrangements are expected in the *rps10-rpl2-rps19-rps3-rpl16* gene cluster. 1.The deletion of the *rps10* and 5' end of *rpl2* gene since no *rps10* sequences have been detected in wheat mt genome (Zanlungo et al., 1995) and only a single copy of *rpl2* 3'segment and not the 5' *rpl2* end is detected in hybridizations. 2. Presence of *nad1d* sequences in the downstream of *rpl2* suggests a replacement event. 3. The detection of a truncated *rps19* gene upstream of *nad4L* suggests a deletion in *rps19* sequences (unpublished results M. Fallahi). 4. The presence of the *nad4L* gene instead of *rps3-rpl16* genes in the *rps19* downstream suggests another replacement event. The prediction of at least four events in one rp gene cluster provides an idea about the magnitude of mt DNA rearrangements recently happening in the lineage leading to wheat. Similarly we have observed rearrangement events which have truncated the *rps7* gene in pea and soybean, and translocated a cytochrome c biogenesis gene (*ccb 248*) into the upstream region (Zhuo et al., 1999). The same type of phenomenon is believed to be responsible for the origin of trans-spliced genes which are functional. Based on my data, the wheat ψ *rpl2* is transcribed as a part of the

nad1d precursor suggesting that DNA rearrangements might have led to the placement of a promoter element upstream of the *ψrpl2-nad1d-nad6* region.

The presence of an enormous amount of promiscuous DNA in the mt genome can have a role in the frequency of DNA rearrangements in the mt genome. For example, the intramitochondrial recombination events involving the integrated chloroplast sequences have led to the duplication and subsequent rearrangement of cp sequences in rice mt DNA (Kanno et al., 1997). Some of the cp sequences located in the mt genomes are found to be conserved between plants while others are consistent with being sequentially inserted. In addition, certain cp genes may be transferred repeatedly as observed by loss and gain of different cp sequences in closely related plants (Nakazono and Hirai, 1993). It is difficult to distinguish between these events as the transferred cp sequences in the mt genome and cp sequences usually share high nucleotide identity (>95%). About, one third of the tRNA genes identified in various plant mt genomes are found to be cp origin (reviewed by Veronica et al., 1996; Joyce and Gray, 1989) and they appear to be the only genes that are functional among the transferred cp DNA fragments. Recently, it has been reported that the cp sequences are being utilized as a source of promoter sequence for the mt *nad9* gene in rice (Nakazono et al., 1996). A similar situation is predicted for the rice mt *rpl2* gene regulatory sequences based on primer extension analysis (Fig. 3.8.A) which indicates the position of the 5' end of the mt *rpl2* transcript within the cp sequences present upstream of the rice *rpl2* gene. This block of 6.8 kb of chloroplast sequences contains *rpoB-rpoC1-ΨrpoC2* and they are present in the mt genomes of all the *Oryza* species examined (Kanno et al., 1997). The cp promoter sequences expected upstream of the cp *rpoB* gene in the opposite orientation may have a role for the transcription of the mt *rpl2* gene if the mt RNA polymerase can recognize the cp promoter

Figure 6.2 Plant mitochondrial ribosomal protein gene organization in comparison to the *E. coli* rp operons

Data were compiled from the complete sequence analysis of *E. coli* (Zurawski and Zurawski, 1985) liverwort mt (Takemura et al., 1992) and *Arabidopsis* mt (Unseld et al., 1997) genomes. For rice and wheat, available information from the literature (Schuster and Brennicke, 1994; Kumar, 1995; Handa et al., 1997 and Bonen, 1998) was included in addition to unpublished results from our lab (my work and that of M. Fallahi). The black boxes indicate the functional rp genes that are common among all organisms. The open boxes of same size represent the genes which are not seen in the liverwort mt rp gene cluster when compared with *E. coli* rp operons. The grey boxes represent the truncated or pseudo genes. The small open boxes correspond to respiratory chain protein genes that are located adjacent to rp genes. Figure adapted from Takemura et al., 1992.


STR operon
S10 operon
SPC operon
a operon
S7 S12 EF-G
S10 L3 L4 L23 L2 S19 L22 S3 L16 L29 L17
L14 L24 L5 S14 S8 L6 L18 S5 L30 L15 SecY
S13 S11 S4 rpoA L17
E. coli
S1
Liverwort
nad3 S12 S7
L2 orf224 S19 S3 L16 L5 S14 cob nad1ab S4 S2
Arabidopsis
L2 S19 nad4L S3 L16 nad3 S12 S11 S4
Rice
S7 L2 nad1d S19 nad4L S3 L16 S13 atp6 nad3 S12 S2 nad9 orf585 S1A
Wheat

sequences.

The occurrence of DNA rearrangements in noncoding sequences is expected to be relatively higher than in coding regions, because of lower functional constraint. The presence of enzymes involved in recombination events and the recombinationally active repeat sequences may play an important role in DNA rearrangements. In yeast, recombination events produce deleted molecules with a high replication rate but bearing little useful information such as "petite" mutants (Albert et al., 1996).

In short, the recombinogenic nature of plant mt DNA and the presence of promiscuous DNA in the mt genome may cause an increase in DNA rearrangements, which result in deletion of mt genes which are not functionally constrained.

6.4 Functional redundancy is an important stage in mt gene transfer

Functional redundancy can be considered as a security stage to maintain the continuous production of the mt protein during any gene transfer process. Although the movement of DNA or RNA is believed to be a random event, data on the location and expression of genes that are transferred to the nucleus in different plants points to a well regulated event as depicted in Figure 6.1. The rice *rpl2* gene located in the mt genome is a functional gene, however detection of related *rpl2* sequences in the rice nuclear genome that are transcribed suggests that it may be a transition stage of gene transfer where both copies are present and they theoretically have an equal possibility of losing its function since only one copy needs to be constrained. Alternatively, the rice mt copy may be maintained for specific reasons such as differential expression of mt and nuclear copies at different developmental stages. Similarly, the transferred *cox2* genes in legumes belonging to several genera (for example Neonotonia, Dumasia and Pseudeminia) were observed

to be transcribed in both compartments (Adams et al., 1998b) whereas, only the mt copy is expressed in pea or only the nuclear copy in legumes such as common bean & soybean (Nugent and Palmer, 1991). In *Neurospora*, two copies of mt *atp9* gene, one copy in each of nucleus and mitochondria are present. The nuclear copy of the mt *atp9* is expressed in all stages in contrast to the mt copy which is found to be expressed only at the spore germination stage (Bittner-Eddy et al., 1994). This observation suggests a potential reason for the maintenance of genes in both compartments. A similar situation may hold for the rice mt *rpl2* because of low steady state levels of mt *rpl2* mRNA compared to other respiratory chain genes in germinating embryo and seedling stages. Protein level analysis is needed to establish whether both genes produce functional proteins which assemble into mt ribosomes.

Besides the possibility of developmentally specific expression of mt and nuclear copies, there may be other reasons why a mt copy is maintained. In the rice mt genome, an intact mt *rps19* gene is located three nucleotides downstream of the mt *rpl2* gene as in *E. coli* (Zurawski and Zurawski, 1985) and they are found to be co-transcribed. The maintenance of *rpl2- rps19* cluster in rice over the period of its evolution may suggest that no DNA rearrangements have occurred because of their close proximity. In addition, the cotranscription of these genes may have some role in the regulation of protein synthesis with regard to proximity or availability for ribosomal assembly. The 50S subunit proteins whose genes are present in large transcriptional units in *E. coli* are often found to be assembled together and they are dependent on each other for ribosomal assembly (Rohl and Nierhaus, 1982). Thus the maintenance of a mt gene when there is a functional nuclear copy, may be either due to differential expression or specific functional requirement.

6.5 Expression of mitochondrial ribosomal protein genes at different developmental stages and in different compartments

The transcript profiles of plant mt genes may vary between species, for example the *cox2* transcript profile is observed to be a simple monocistronic mRNA in wheat (Bonen et al., 1984) compared to maize which has a complex profile (Fox and Leaver, 1981). In rice, the mt *rpl2* gene is transcribed as polycistronic mRNAs with *rps19* and *nad4L* gene sequences. The observation of a complex transcript profile in the mt gene cluster is not an isolated occurrence. Stable large polycistronic mRNAs up to 7.8 kb and several processed mRNAs are observed in other mt rp gene clusters such as *rps3-rpl16-rpl5-rps14-cob* of *B. napus* (Ye et al., 1993). Similarly, the wheat mt *rpl2* 3' segment is also a part of a polycistronic mRNA including *nad1d* and *nad6* genes. Unlike the nuclear genes which have a single abundant transcript corresponding to the mature mRNA, as for the wheat mt *rpl2* (1.2 kb), mt genes often have several stable transcripts corresponding to precursors and processed intermediates. The stability of the precursors in mt RNA population may presumably be due to differences in the efficiency of transcription and splicing or other processing events. However the excised intron levels are higher than mRNA which suggests a more stable form of excised intron than the mRNA.

The excised group II intron RNAs are expected to be lariat structures which are often stable in nature (*B. napus rps3*- Ye et al., 1993; wheat *nad7*- Carrillo and Bonen, 1997). The absence of lariat structures detectable for the excised rice *rpl2* group II intron and the RT- PCR amplification products corresponding to circularized form of intron that lacks nucleotides at the junction suggests that these introns might be excised as linear molecules through hydrolysis. It should be noted that the mt *rpl2* intron has a bulging adenosine in domain 6 which initiates the 2-5'

phosphodiester bond formation in the classical group II splicing. RNase protection analysis using cloned circular intron sequences to distinguish the linear or circular intron molecules would help to resolve the nature of excised *rpl2* intron in vivo.

The expression of genes located in mt genomes also requires C- to U- type RNA editing. RNA editing in mt transcripts changes the amino acids in approximately 1-5% of codons in mt ribosomal protein genes in contrast to 3-13 % in respiratory chain genes (reviewed by Hanson et al., 1996). Consistent with this, a higher percentage of editing (10%) was observed in rice *nad4L* transcripts compared to the rice *rpl2* which has only one editing site (0.002%). In *Oenothera*, the *atp9* and *rps1* genes are cotranscribed but only the *atp9* transcripts are edited (Mundel and Schuster, 1996). The difference in editing between mt transcripts may be correlated with accessibility of the editing machinery to the mRNA or the stability of the mRNA. The possibilities of a reverse transcribed copy integrated into the mt genome with genomically encoded corrected nucleotides may also influence the editing frequency observed for a particular gene. The truncated mt *rpl2* gene in *Oenothera* mt genome is found to be edited at five positions but they are observed to reduce the similarity compared to other proteins (Ac. No.S46947). The truncated nature and editing in the wrong way in *Oenothera* mt *rpl2* may suggest that it is under a reduced or different functional constraint. Shields and Wolfe (1997) have reported that selection to replace the genomic C with a T may account for an accelerated evolution suggesting that mt transcripts that are under functional constraint are edited.

The expression of the transferred mt gene requires an addition of sequences that recognize the nuclear expression machinery. Although integration of transferred mt genes can happen within a nuclear gene, or within untranslated leaders and tails, introns or spacer regions, most of the

transferred genes characterized in the nucleus indicate integration within the 5' untranslated regions of nuclear genes. This is consistent with the acquisition of the target sequences by exon shuffling method in the nuclear-located transferred mt *rps10* gene in *Arabidopsis* (Wischmann and Schuster, 1995) and *cox2* in soybean (Covello and Gray, 1992). Computer searches on nuclear integrated nonfunctional mt DNA also indicate integration within intron and spacer regions (Blanchard and Schmidt, 1996) validating its inability to acquire expression signals. The *rps11* gene transferred to the rice nuclear genome is believed to be activated by the addition of sequences by an exon shuffling method followed by gene duplication and both copies acquired mt target sequences by duplication of the existing target sequences from different genes (Kadowaki et al., 1996). The target sequences and their flanking sequences can also be analyzed to find out the nature of integration sites and to know whether there might have been a single or multiple independent gene transfer events by comparing the corresponding position in other species. The transferred *rps11* gene in pea (Kubo et al., 1998) has different target peptide sequences compared to rice suggesting that they were independent events.

Besides regulation at the levels of transcription, splicing, editing, processing and mRNA stability, the mt *rpl2* mRNA accumulation in rice in germinating embryos and seedling stages is consistent with developmental regulation. There are higher levels of precursors (5.0 & 7.0 kb) and mRNA (3.5 kb) in embryos than in seedlings. This observation may suggest either that transcription rate is higher or the mRNA is more stable in embryos to meet out the increased demand in germinating embryo stage over the seedling stage for the higher mt biogenesis and protein synthesis. Hence the RNA polymerase for transcription and proteins involved in mRNA stability may be regulated accordingly. In contrast, the 2.6 kb transcript levels are higher at

seedling stage than embryo stage. The difference in levels of processing intermediates compared to mRNA and precursors at 24 h and 6 d stages may suggest different processing pathways. In addition, splicing may also be regulated differentially at these stages because some intermediates contain intron sequences. Similarly, stable splicing intermediate accumulation is observed for other mt split genes such as the mt *nad4 in B. campestris* (Gass et al., 1992). This observation suggests that the mRNA processing machinery which is presumably nuclear-encoded is differentially regulated at these stages. The mRNA levels of some bacterial and yeast rp genes are controlled by an autogenous feedback mechanism in which ribosomal proteins bind to their mRNA to control their transcription and /or splicing (reviewed by Mager, 1988; Presutti et al., 1995). It is not known whether plant rp genes are regulated by a similar mechanism.

The steady state level of mRNAs of the wheat nuclear-located mt *rpl2* and *atpβ* genes are about three fold higher in germinating embryos than seedlings when compared to relative cytosolic *rpl2* transcripts. Nuclear run off experiments on plant nuclear genes indicate that variation in mRNA levels arise primarily through the rate of transcription unlike genes located in the mitochondrion where there are important post-transcriptional levels of control (reviewed by Kuhlemeier, 1992). The expression of the maize mt chaperonin HSP60 and *atpβ* genes both at RNA and protein levels are found to be higher at the early stage of seed germination (Prasad and Stewart, 1992). Similarly genes for chloroplast ribosomal proteins located in the nuclear genome are found to be expressed very early after seed imbibition and even precede that of genes encoded in the cp genome (reviewed by Mache et al., 1997). Like organellar rp genes, cytosolic rp genes are also expected to have higher mRNA accumulation during germination. Consistently, tobacco cytosolic *rpl25* (*E. coli* L23 homologue- Gao et al.,1994) and *Arabidopsis rpl16* (*E. coli* L5

homologue-Williams and Sussex, 1995) genes are found to be expressed abundantly, at least two times higher in actively dividing tissues and in the germinating embryo stage compared to the mature tissues or seedlings. In contrast, the transcript accumulation of the cytosolic *rpl2* gene in tobacco is found to be comparable in germinating seeds and in young plant tissues (Marty and Meyer et al., 1992) suggesting a possible regulation at the translation level. In yeast, the ribosomal protein levels are tightly controlled at the level of transcription and they are found to reflect the levels of rp mRNAs which vary depending on the carbon source used for growth (Mager and Planta, 1990; Pan and Mason, 1995). It is not known whether the developmental regulation of the wheat mt *rpl2* transcripts corresponds to protein levels. Western blot analysis on proteins isolated at these stages with the help of L2 specific antibodies would provide information about this.

6.6 Conclusions and future directions

Studies on mitochondrial gene expression and mt gene transfer to the nucleus are exciting areas of research. They not only help in understanding the evolution of mt genes, they also unravel the steps involved in the expression of mt genes orchestrated by mt and nuclear genomes. The available data from this thesis confirm the presence of the mt *rpl2* gene in the nuclear genome of flowering plants consistent with gene transfer event(s) and complete characterization of genes in rice, wheat and soybean nuclear genomes will give even more insight. The N-terminal sequence analysis would help to determine the nature and origin of the mt targeting sequences. The flanking sequences between two nuclear genes may help to resolve whether there has been a single or multiple transfer events and may lead to a better understanding of the mechanisms involved in the process of gene activation. A survey on cp genes from completely sequenced cp genomes of a

wide range of organisms from cyanobacteria to higher plants has revealed that certain cp genes are lost independently and repeatedly during evolution (Martin et al., 1998). Hybridization analyses of several mt rp genes over a wide range of plant species have also indicated multiple transfers and losses of mt genes in flowering plant evolution (Adams et al., 1998a). To date, it is not known whether any organellar gene is re-gained evolutionarilly after its earlier loss from the organelle. However there is evidence for the invasion of mt group II intron as mobile genetic elements and a group I intron by horizontal transfer in plants (reviewed by Bonen, 1998). Additional information on the horizontal interorganellar gene transfers like cp tRNA gene transfer to mitochondria and functional gene transfer from organelles to the nucleus may possibly help us to understand the underlying genetic concepts and early endosymbiotic events. In addition, information on nuclear integration sites would be useful for plant biotechnologists who are trying to develop methods for site specific integration of transgenes.

The expression of the rice mt *rpl2* gene located in the mt genome, particularly the splicing, excision of the intron and forms of excised introns, is interesting to analyze in detail. The study on plant mt group II intron splicing is particularly interesting because of its deviation from classical introns. The excision of rice mt *rpl2* intron suggests a hydrolysis rather than transesterification in the first step of splicing.

The cDNA mediated transfer of mitochondrial genes is interesting to analyze with regard to reverse transcriptase activity in the mt compartment. Plant mt genomes also encode a mat-r gene which is believed to encode a potentially functional protein with reverse transcriptase activity based on its editing status which increases its similarity to the fungal maturases (Thomson et al., 1994). In addition, reverse transcriptase activity has been detected in potato mitochondria by

Moenne et al., (1996). The developmental regulation of mRNA turnover and the factors involved in determining the stability of mt RNA are interesting to analyze. Differential display techniques can be employed to identify proteins involved in mt mRNA splicing and stability.

The DNA rearrangements caused by recombination events in the plant mt genomes are interesting especially when they create functional trans-spliced genes and pseudogenes. Besides, novel chimeric genes which are correlated to the cytoplasmic male sterility are also created during these DNA rearrangement events (reviewed by Bonen and Brown, 1993). The presence of their gene products is correlated with the expression of certain mt genes in a tissue specific manner which affects the functional pollen production. The positioning of promoter sequences that control the transcription of scattered gene pieces suggests a regulation on mt DNA recombination and rearrangement events with regard to the expression of trans-spliced genes. In addition, promiscuous DNA particularly the chloroplast-derived tRNA genes which are functional in the mitochondrion are maintained, expressed and regulated according to the need of mt translation system. The potential involvement of cp DNA sequences in the expression of the adjacent mt genes is quite interesting to analyze in detail especially with regard to mt RNA polymerase recognition. Some of the cp genes are transcribed by the bacterial type cp encoded polymerases while others by the nuclear-encoded phage-type ones which apparently have been recruited by the organellar genomes early in evolution like fungal and animal mt genomes. It may be a possibility that the nuclear-encoded RNA polymerases could be targeted to both compartments and they can recognize both mt and cp promoters. Protein level analysis for the presence of one or more type of RNA polymerases in cp and mt compartments may resolve this issue.

Plant mt rp genes are regulated at many levels including transcription, splicing, editing, RNA processing and translation. In addition, the majority of plant mt rp genes are in the nucleus as a result of transfer events either during early endosymbiosis or more recent evolutionary transfers. The differences in the location of plant mt rp genes and their multilevel regulation is remarkable compared to animal or fungal systems. Why are the plant mt ribosomes different than other systems? Do extra domains in plant mt ribosomal proteins provide additional functions? Is there a need for the maintenance of certain rp genes in plant mt genomes or is it just a random event? Are these differences in gene content related to the dynamic and flexible nature of flowering plant mt DNA? How frequent are successful independent transfers during plant evolution ? These are some of the challenges in understanding the plant mt ribosomal machinery, its function and evolution.

REFERENCES

Adams, K.L., Clements, M.J. and Vaughn, J.C. 1998 a. The Peperomia mitochondrial coxI group I intron: timing of horizontal transfer and subsequent evolution of the intron. J. Mol. Evol. 46: 689-696.

Adams, K.L., Song, K., Qiu, Y., Shirk, A., Cho, Y., Parkinson, L. and Palmer, J. 1998 b. Evolution of flowering plant mitochondrial genomes: Gene content, gene transfer to the nucleus, and highly accelerated mutation rates. In Plant Mitochondria: from gene to function, pp 13-18. Eds. I.M. Moller, P. Gardestrom, K.Glimelius and E. Glaser. Backhuys Publishers, Leiden, The Netherlands.

Albert, B., Godelle, B., Atlan, A., Paepe, R.D. and Gouyon, P.H. 1996. Dynamics of the plant mitochondrial genome: Model of a three-level selection process. Genetics 144: 369-382.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Anderson, S., Bankier, A.T., Barrell, B.G., deBruijin, M.H.L. and Coulson, A.R. 1981. Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.

Andre, C.P. and Walbot, V. 1995. Pulsed-field gel mapping of maize mitochondrial chromosomes. Mol. Gen. Genet. 247: 255-263.

Ayliffe, M.A., Scott, N.S. and Timmis, J.N. 1998. Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. Mol. Biol. Evol. 15: 738-745.

Ayliffe, M.A. and Timmis, J.N. 1992. Plastid DNA sequence homologies in the nuclear genome Mol. Gen. Genet. 236: 105-112.

Bendich, A.J. 1996. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed field gel electrophoresis. J. Mol. Biol. 255: 564-588.

Binder, S. and Brennicke, A. 1993. Transcription initiation sites in mitochondria of *Oenothera berteriana*. J. Biol. Chem. 268: 7849-7855.

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

Bittner-Eddy, P., Monroy, A.F. and Bramb, K. 1994. Expression of mitochondrial genes in the germinating conidia of *Neurospora crassa*. J. Mol. Biol. 235: 881-897.

Blanchard, J.L. and Schmidt, G.W. 1996. Mitochondrial DNA migration events in yeast and humans: Integration by a common end-joining mechanism and alternative perspectives on

nucleotide substitution patterns. Mol. Biol. Evol. 13: 537-548.

Bland, M.M., Levings III, C.S. and Matzinger, D.F. 1986. The tobacco mitochondrial ATP *ase* subunit 9 is closely linked to an open reading frame for a ribosomal protein. Mol. Gen. Genet. 204: 8-16.

Bock, R., Kossel, H. and Maliga, P. 1994. Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. EMBO J. 13: 4623-4628.

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

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

Bonen, L. 1995. The wheat mitochondrial genome. In the molecular biology of plant mitochondria. Eds. S. Leavings and I. K. Vasil. Kluwer Academic Publishers pp 345-364.

Bonen, L. 1998. Mitochondrial genomes: a paradigm of organizational diversity. Adv.Genome Biol. 5B: 415-461.

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

Bonen, L. and Brown, G.G. 1993. Genetic plasticity and its consequences: perspectives on gene organization and expression in plant mitochondria. Can. J. Bot. 71: 645-660.

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

Brandt, P., Unseld, M., Eckert-Ossenkopp, U. and Brennicke, A. 1993. An *rps14* pseudogene is transcribed and edited in *Arabidopsis* mitochondria. Curr. Genet. 24: 330-336.

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. and Schuster, W. 1993. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS lett. 325: 140-145.

Bubenenko, M.G., Schmidt, J. and Subramanian, A.R. 1994. Protein substitution in chloroplast ribosome evolution A eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. J. Mol. Biol. 240: 28-41.

Campbell, C.L. and Thorsness, P.E. 1998. Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. J. Cell. Sci. 111: 2455-2464.

Caoile, A.G. and Stern, D.B. 1997. A conserved core element is functionally important for maize mitochondrial promoter activity in vitro. Nucleic. Acids Res. 25: 4055-4060.

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

Cavalier-Smith, T. 1987. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. Ann. N.Y. Acad. Sci. 503: 55-71.

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

Chapdelaine, Y. 1992. Structure and Expression of the trans-split gene for NADH dehydrogenase subunit 1 in wheat mitochondria. Ph. D thesis. Ottawa-Carleton Institute of Biology. University of Ottawa. Ottawa. Canada.

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

Cho, Y., Qiu, Y.L., Kuhlman, P. and Palmer, J.D. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. USA. 24: 14244-14249.

Clayton, D.A. 1991. Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cytol. 141: 453-478.

Covello, P.S. and Gray, M.W. 1991. Sequence analysis of wheat mitochondrial transcripts capped in vitro: definitive identification of transcription initiation sites. Curr. Genet. 20: 245-251.

Covello, P.S. and Gray, M.W. 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA mediated gene transfer. EMBO J. 11: 3815-3820.

Dean, A.D. and Greenwald, J.E. 1995. Use of filtered pipette tips to elute DNA from agarose gels. Biotechniques, 18: 980.

DePaepe, R., Forchioni, A., Chetrit, P. and Vedel, F. 1993. Specific mitochondrial proteins in pollen: presence of an additional ATP synthetase β subunit. Proc. Natl. Acad. Sci. USA. 90: 5934-5938.

Doolittle, W.F. 1998.a. A paradigm gets shifty. Nature 392: 15-16.

Doolittle, W.F. 1998.b. You are what you eat: a gene transfer rachet could account for bacterial genes in nuclear genomes. Trends Genet. 14: 307-311.

Doyle, J. and Doyle, J. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Systematic. Bot. 20: 272-294.

Drombrowski, S, Brennicke, A. and Binder, S. 1997. 3' inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. EMBO J. 16: 5069-5076.

Dudley, P., Wood, C.K., Pratt, J.R. and Moore, A.L. 1997. Developmental regulation of the plant mitochondrial matrix located HSP 70 chaperone and its role in protein import. FEBS lett. 417: 321-324.

Fauron, C., Casper, M., Gao, Y. and Moore, B. 1995. The maize mitochondrial genome: dynamic, yet functional. Trends Genet. 11: 228-235.

Felsenstein, J.1993. PHYLIP (Phylogeny Inference Package) Version 3.5.c. University of Washington, Seattle

Fennoy, S.L. and Bailey-Serres, J. 1993. Synonymous codon usage in Zea mays L. nuclear genes is varied by levels of C and G- ending codons. Nucleic. Acids Res. 23: 5294-5300.

Fey, J., Dietrich, A., Cosset, A., Desprez, T. and Marechal-Drouard, L. 1997. Evolutionary aspects of "chloroplast-like" trnN and trnH expression in higher-plant mitochondria. Curr. Genet. 32: 358-360.

Flemming, A.J., Mandel, T., Roth, I. and Kuhlemeier, 1993. The patterns of gene expression in tomato shoot apical meristem. Plant Cell 5: 297-309.

Fox, T.D. and Leaver, C.J. 1981. The Zea mays mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. Cell 3: 315-23.

Frohman, M.A., Dosh, M.K. and Martin, G.R. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85: 8998-9002.

Gantt, J.S., Baldauf, S.L., Calie, P.J., Weeden, N.F. and Palmer, J.D. 1991. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 10: 3073-3078.

Gao, J., Kim, S.R., Chung, Y.Y., Lee, J.M. and An, G. 1994. Developmental and environmental regulation of two ribosomal protein genes in tobacco. Plant Mol. Biol. 25: 761-770.

Gass, D.A., Makaroff, C.A. and Palmer, J.D. 1992. Variable intron content of the NADH dehydrogenase subunit 4 gene of plant mitochondria. Curr. Genet. 21: 423-430.

Giege, P., Knoop, V., and Brennicke, A. 1998. Complex II subunit 4 (sdh4) homologous

sequences in plant mitochondrial genomes. Curr. Genet. 34: 313-317.

Goldberg, R.B. 1978. DNA sequence organization in the soybean plant. Biochem. Genet. 16: 45-68.

Gonzalez, D.H. Bonnard, G. and Grienenberger, J.M. 1993. A gene involved in the biogenesis of cytochromes is co-transcribed with a ribosomal protein gene in wheat mitochondria. Curr. Genet. 24: 248-255.

Gornicki, P., Faris, J., King, I., Podkowinski, J., Gill, B. and Haselkorn, R. 1997. Plastid-localized acetyl- coA carboxylase of bread wheat is encoded by a single gene on each of the three ancestral chromosome sets. Proc. Natl. Acad. Sci. USA. 94: 14179-14184.

Graack, H.R., Grohmann, L., Kitakawa, M. and Goldschmidt-Reisin, S. 1995. Gene MRP-L4, encoding mitochondrial ribosomal protein ymL4, is indispensable for proper non-respiratory cell functions in yeast. Gene 152: 107-112.

Graack, H.R. and Whitman-Labeled, B. 1998. Mitochondrial ribosomal proteins (MRPs) of yeast. Biochem. J. 329: 433-448.

Grantham, R., Gautier, C., Gouy, M. Mercier, R. and Pave, A. 1980. Codon catalog usage and the genome hypothesis. Nucleic. Acids Res. 8: r49-r69.

Gray, M.W. 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. 141: 233-357

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

Gray, M.W. and Lang, B.F. 1998. Transcription in chloroplasts and mitochondria: a tale of two polymerases. Trends Microbiol. 6: 1-3.

Gray, M.W., Lang, B.F., Cedergren, R., Golding, G. Lemieux, C., Sanloff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T.G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y. and Burger, G. 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. 26: 865-887.

Green, R. and Noller, H.F. 1997. Ribosomes and translation. Annu. Rev. Biochiem. 66: 679-716.

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

Grohmann, L., Brennicke, A. and Schuster, W. 1992. The mitochondrial gene encoding ribosomal protein S12 has been transferred to the nuclear genome in *Oenothera*. Nucleic. Acids Res. 20: 5641-5646.

Handa, H., Kubo, N. and Kadowaki, K. 1998. Genes for the ribosomal S4 protein encoded in higher plant mitochondria are transcribed, edited and translated. Mol. Gen. Genet. 258: 199-207.

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

Hanekamp, T. and Thorsness, P. 1996. Inactivation of YME2/RNA12, which encodes an integral inner mitochondrial membrane protein, causes increased escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 16: 2764-2771.

Hanson, M.R., Sutton, C.A. and Lu, B. 1996. Plant organellar gene expression: altered by RNA editing. Trends Plant Sci. 2: 57-64.

Hattori, J., Gottlob-Mchugh, S.G. and Johnson, D.A. 1987. The isolation of high molecular weight DNA from plants. Anal. Biochem. 165: 70-74.

Haouazine, N., Takvorian, A., Jubier, M.F., Michel, F. and Lejeune, B. 1993. The nad6 gene and the exon d of nad1 are co-transcribed in wheat mitochondria. Curr. Genet. 24: 533-538.

Hedtke, B. Borner, T. and Weihe, A. 1997. Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science. 277: 809-811.

Heiser, V., Brennicke, A. and Grohmann, L. 1996. The plant mitochondrial 22 KDa (PSST) subunit of respiratory chain complex I is encoded by a nuclear gene with enhanced transcript levels in flowers. Plant Mol. Biol. 31: 1195- 1204.

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

Higgins, D.G.R., Fuchs, R. and Blesby, A. 1992. CLUSTAL: a new multiple sequences alignment program. Comput. Appl. Biosci. 8: 189-191.

Hunt, M.D. and Newton, K.J. 1991. The NCS3 mutation: genetic evidence for the expression of ribosomal protein genes in *Zea mays* mitochondria. EMBO J. 10: 1045-1052.

Itani, K. and Handa, H. 1998. Rapeseed mitochondrial ccb206, a gene involved in cytochrome c biogenesis, is co-transcribed with the nad3 and rps12 genes: organization, transcription, and RNA editing of the nad3/ rps12/ ccb206 locus. Curr. Genet. 34: 318-325.

Joyce, P.B. and Gray, M.W. 1989. Aspartate and asparagine tRNA genes in wheat mitochondrial DNA: a cautionary note on the isolation of tRNA genes from plants. Nuc. Acids. Res. 17: 7865-7878.

Kadowaki, K., Kubo, N., Ozawa, K. and Hirai, A. 1996. Targeting pre-sequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. EMBO J. 15: 6652-6661.

Kanno, A., Nakazono, M., Hirai, A. and Kameya, T. 1997. Maintenance of chloroplast - derived sequences in the mitochondrial DNA of Graminae. Curr. Genet. 32: 413-419.
Kitakawa. M. and Isono, K. 1991. The mitochondrial ribosomes. Biochimie 73: 813-825.

Knoop, V. and Brennicke, A. 1991. A mitochondrial intron sequence in the 5' flanking region of a plant nuclear lectin gene. Curr. Genet. 20: 423-425.

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

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

Knoop,V., Unseld, M., Marienfeld, J., Brandt, P., Sunkel, S. Ulrich, H. and Brennicke, A. 1996. *Copia gypsy* and LINE-like retroposon fragments in the mitochondrial genome of *Arabidopsis thaliana*. Genetics. 142: 579-585.

Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A. and Ohyama, K., 1997. Interorganellar gene transfer in bryophytes: the functional nad7 gene in nuclear-encoded in *Marchantia polymorpha*. Mol. Gen. Genet. 256: 589-592.

Koonin, E.V., Bork, P. and Sander, C. 1994. A novel RNA-binding motif in omnipotent suppressers of translation termination, ribosomal proteins and a ribosomal modification enzyme? Nucleic. Acids Res. 22-2166-2167.

Krug, M.S. and Berger, S.L. 1996. First strand synthesis primed with oligo (dT). Methods. Enzymol. 152: 316-325.

Kubo, N., Harada, K. and Kadowaki, K. 1998. Transfer of an RPS 11gene from mitochondrion to the nucleus in pea. In Plant Mitochondria: from gene to function, pp 25-27. Eds. I.M. Moller, P. Gardestrom, K.Glimelius and E. Glaser. Backhuys Publishers, Leiden, The Netherlands.

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

Kubo, N., Ozawa, K., Hino, T. and Kadowaki, K. 1996. A ribosomal protein L2 gene is transcribed, spliced and edited at one site in rice mitochondria. Plant. Mol. Biol. 31: 853-862.

Kuhlemeier, C. 1992. Transcriptional and post transcriptional regulation of gene expression in

plants. Plant Mol. Biol. 19: 1-14.

Kumar, R. 1995. Mitochondrial ribosomes and their proteins. In The Molecular Biology of Plant Mitochondria Eds. S. Leavings and Indra K. Vasil. Kluwer Academic Publishers.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golden, G.B., Lemieux, C., Sankoff, D., Turmel, M. and Gray, M.W. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387: 493-496.

Laroche, J., Li, P. and Bousquet, J. 1995. Mitochondrial DNA and monocot- dicot divergence time. Mol. Biol. Evol. 12: 115-1156.

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

Li. W-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitutions. J. Mol. Evol. 36: 96-99.

Li, W-H. and Graur, D. 1991. Fundamentals of Molecular Evolution. Seigneur Associates, publishers.

Liang, P., Averboukh, L. and Pardee, A.B. 1993. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinement and optimization. Nuc. Acids. Res.11: 3269-3275.

Lind, C., Halide, C. and Noller, R.M. 1991. Protein synthesis in mitochondria purified from root, leaves and flowers of sugar beet. Physiol. Plant. 83: 7-16.

Lloyd, A.T. and Sharp, P.M. 1992. CODONS: a microcomputer program for codon usage analysis. J. Herd. 83: 239-240.

Mache, R., Zhu, D., Herbs-Mache, S. Hayrack, H., Villain, P. and Gavin, S. 1997. Nuclear control of early plastid differentiation. Plant Physiol. Biochem. 35: 199-203.

Mager, K.H. 1988. Control of ribosomal protein gene expression. Biochem. Biophys. Acta. 949: 1-15.

Mager, K.H. and Planta, R.D. 1990. Muti functional DNA-binding proteins mediate concerted transcription activation of yeast ribosomal protein genes. Biochem. Biophys. Acta. 1050: 351-355.

Maier, R.M., Zeltz, P., Kossel. H., Bonnard, G., Gualberto, M. and Grienenberger, J.M. 1996. RNA editing in plant mitochondria and chloroplasts. Plant Mol. Biol. 32: 343-365.

Margulis, L. 1970. Origin of eukaryotic cells. Yale University Press, New Haven.

Martin, W. and Muller, M. 1998. The hydrogen hypothesis for the first eukaryote. Nature 392: 37-41.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S. Hasegawa, M. and Kowallik, K.V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393: 162-165.

Marty, I. and Meyer, Y. 1992. cDNA nucleotide sequence and expression of a tobacco cytoplasmic ribosomal protein L2 gene. Nucleic Acids Res. 20: 1517-1522.

Mason, T.L., Pan, C., Sanchirico, M.E. and Sirum-Connolly, K. 1996. Molecular genetics of the peptidyl transferase center and the unusual Var1 protein in yeast mitochondrial ribosomes. Experimentia. 52: 1148-1157.

Menassa, R., El-Rouby, N. and Brown, G.G. 1997. An open reading frame for a protein involved in cytochrome C biogenesis is split into two parts in *Brassica* mitochondria. Curr. Genet. 31: 70-79.

Michel, F. and Ferat, J.L. 1995. Structure and activities of group II introns. Annu. Rev. Biochem. 64: 435-461.

Moenne, A. Begu, D. and Jordana, X. 1996. A reverse transcriptase activity in potato mitochondria. Plant Mol. Biol. 31: 365-72.

Moreira, D., Le Guyader,H. and Philippe, H. 1999. Unusually high evolutionary rate of the elongation factor 1 alpha genes from the Ciliophora and its impact on the phylogeny of eukaryotes. Mol. Biol. Evol. 16: 234-245.

Mortan, B.R. (1998) Selection on codon bias of chloroplast and cyanelle genes in different plant and algal lineages. J. Mol. Evol. 46: 449-459.

Muller, E.C. and Wittman-Labeled, B. 1997. Phylogenetic relationship of organisms obtained by ribosomal protein comparison. Cell Mol. Life. Sci. 53: 34-50.

Mundel, C. and Schuster, W. 1996. Loss of RNA editing of *rps1* sequences in *Oenothera* mitochondria. Curr. Genet. 30: 455-460.

Nakagava,A., Nakashima, T., Taniguchi, M. Hosaka, H., Kimura, M. and Tanaka, I. 1999. The three-dimensional structure of the RNA binding domain of ribosomal protein L2; a protein at the peptidyl transferase center of the ribosome. EMBO J. 18: 1459-1467.

Nakazono, M. and Hirai, A. 1993. Identification of the entire set of chloroplast DNA sequences in the mitochondrial genome of rice. Mol. Gen. Genet. 236: 341-346.

Nakazono, M., Itadani, H., Wakazugi, T., Tsutsumi, T., Sugiura, M. and Hirai, A. 1995. The *rps3-rpl16-nad3-rps12* gene cluster in rice mitochondrial DNA is transcribed from alternative prometers. Curr. Genet. 27: 184-189.

Nakazono, M., Nishiwaki, S., Tsutsumi, N. and Hirai, A. 1996. A chloroplast-derived sequence is utilized as a source of promoter of NADH dehydrogenase (*nad9*) in rice mitochondria. Mol. Gen. Genet. 252: 371-378.

Newton, K.J., Winberg, B., Yamato, K., Lupold, S. and Stern, D.B. 1995. Evidence for a novel mitochondrial promoter preceding the cox2 gene of perennial teosintes. EMBO J. 14: 585-593.

Noller, H.F. 1993. Peptidyl transferase: protein, ribonucleoprotein, or RNA? J. Bacteriol. 175: 5297-5300.

Nomura, M., Yates, J.L., Dean, D. and Post, L.E. 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and Ribosomal protein mRNA. Proc. Natl. Acad. Sci. USA. 77: 7084-7088.

Nugent, J.M. and Palmer, J.D. 1991. RNA mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. Cell 66: 473-481.

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

Oliver, J.L., Martin, A. and Martinez-Zapater, J.M. 1990. Chloroplast genes transferred to the nuclear plant genome have adjusted to nuclear base composition and codon usage. Nucleic. Acids Res. 18: 65-73.

Olson, H.M., Nag, B., Etchison, J.R., Trout, R.R. and Glitz, D.G. 1991. Differential localization of the epitopes of *Escherichia coli* ribosomal protein L2 on the large ribosomal subunit by immune electron microscopy using monoclonal antibodies. J. Biol. Chem. 266: 1898-1902.

Palmer, J.D. 1997. The mitochondrion that time forgot. Nature 387: 454-455.

Palmer, J.D. and Delwiche, C.F. 1996. Second-hand chloroplasts and the case of the disappearing nucleus. Proc. Natl. Acad. Sci. USA. 93: 7432-7435.

Palmer, J.D. and Herbon, L.A. 1987. Unicircular structure of the *Brassica hirta* mitochondrial genome. Curr. Genet. 11: 565-570.

Pan, C. and Mason, T. 1995. Identification of yeast nuclear gene for the mitochondrial homologue of bacterial ribosomal protein L16. Nucleic. Acids Res. 23: 3673-3677.

Pfanner, N. 1998. Mitochondrial import: Crossing the aqueous intermembrane space. Curr. Biol. 8: R 262- R 265.

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

Prasad, T.K. and Stewart, C.R. 1992. cDNA clones encoding *Arabidopsis thaliana* and *Zea mays* mitochondrial chaperonin HSP 60 and gene expression during seed germination and heat shock. Plant Mol. Biol. 18: 873-885.

Presutti, C., Villa, T., Hall, D., Petricia, C. and Bozzoni, I. 1995. Identification of the cis-elements mediating the autogenous control of ribosomal protein L2 mRNA stability in yeast. EMBO J. 14: 4022-4030.

Proteome Inc. 1999. http.WWW.proteome.com YPDCYC7.html

Quinones, V., Zanlungo, S., Holuigue, L., Litvak, S. and Jordana, X. 1996. The coxI initiation codon is created by RNA editing in potato mitochondria. Plant Physiol. 108: 1327-1328.

Roger, A.J., Sandblom, O., Doolittle, W.F. and Philippe, H. 1999. An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. Mol. Biol. Evol. 16: 218-233.

Rohl, R. and Nierhaus, H. 1982. Assembly map of the large subunit (50S) of *Escherichia coli* ribosomes. Proc. Natl. Acad. Sci. USA. 79: 729-733.

Sakamoto, W., Tan, S. Murata, M. and Motoyoshi, F. 1997. An unusual *atp9-rpl16* cotranscript found in the maternal distorted leaf mutant of *Arabidopsis thaliana*: Implications of GUG as an initiation codon in plant mitochondria. Plant Cell Physiol. 38: 975-979.

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

Sanchez, H., Fester, T., Kloska, S., Schroder, W. and Schuster, W. 1996. Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in Arabidopsis mitochondria. EMBO J. 15: 2138-2149.

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

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

Schuster, W. and Brennicke, A. 1988. Interorganellar sequence transfer: Plant nuclear DNA is nuclear, is plastid, is mitochondrial. Plant Sci. 54: 1-10.

Schuster, W. and Brennicke, A. 1991. RNA editing makes mistakes in plant mitochondria: editing loses sense in transcripts of a *rps19* pseudogene and in creating stop codons in *coxI* and *rps3* mRNAs of *Oenothera*. Nucleic Acids Res. 19: 6923-6928.

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

Seery, L.T., Schoenberg, D.R., Canning, M.E. and Whitehead, A.S. 1994. Identification and characterization of a cDNA encoding ribosomal protein S12 from *Xenopus laevis*. Gene. 150: 231-233.

Sharp, P.M., Cowe, D., Higgins, D., Shields, K.H., Wolfe K.H. and Wright, F. 1988. Codon usage pattern in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster, and Homo sapiens*: A review of the considerable within-species diversity. Nucleic Acids Res. 16: 8207-8211.

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

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

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

Smith, P.M..C., Mann, A.J., Goggin, D.E. and Atkins, C.A. 1998. AIR Synthetase in cowpea nodules: a single gene product targeted to the organelles. Plant Mol. Biol. 36: 811-820.

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

Subramanian, S. and Bonen, L. 1998. Mitochondrial ribosomal protein L2 genes and pseudogenes in wheat. Proceedings of the 9th International Wheat Genetics Symposium.3: 63-65.

Sunkel, S. Brennicke, A. and Knoop, V. 1994. RNA editing of a conserved reading frame in plant mitochondria increases its similarity to two overlapping reading frames in *Escherichia coli*. Mol. Gen. Genet. 242: 65-72.

Takemura, M., Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Nozato, N., Akashi, K. and Ohyama, K. 1992. Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort *Marchantia polymorpha*. Nucleic. Acids Res. 20: 3199-3205.

Takvorian, A., Coville, J.L., Haouazine, N., Rode, A. and Hartmann, C. 1997. The wheat mitochondrial rps13 gene: RNA editing and co-transcription with the *atp6* gene. Curr. Genet. 31: 497-502.

Thomson, M.C., Macfarlane, J.L., Beagley, C.T., Wolstenholme, D.R. 1994. RNA editing of mat-r transcripts in maize and soybean increases similarity of the encoded protein to fungal and bryophyte group II intron maturases: evidence that mat-r encodes a functional protein. Nucleic. Acids Res. 25: 5745-5752.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic. Acids Res. 22: 4673-4680.

Thorsness, P.E. and Fox, T.D. 1990. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. Nature 346: 376-379.

Tracy, R.L. and Stern, D.B. 1995. Mitochondrial transcription initiation: promoter structure and RNA polymerases. Curr. Genet. 28: 205-216.

Unseld, M., Marienfeld, J.R., Brandt, P. and Brennicke, A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366924 nucleotides. Nature Genet. 15: 57-61.

Vaitilingom, M., Stuper, M., Grienenberger, J.M. and Gualberto, J.M. 1998. A gene coding for an RPS2 protein is present in the mitochondrial genome of several cereals, but not in dicotyledons. Mol. Gen. Genet. 258: 530-537.

VandenBoogaart, P., Samallo, J. and Agsteribbe, E. 1982. Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes *Neurospora crassa.* Nature. 298: 187-189.

Veronico, P. Gallerani, R. and Ceci, L.R. 1996. Compilation and classification of higher plant mitochondrial tRNA genes. Nucleic Acids Res. 24: 2199-2203.

Verwoerd, T.C., Dekker, B. M. and Hoekema, A. 1989. A small scale procedure for the rapid isolation of plant RNAs. Nucleic. Acids Res. 17: 2362.

Vilardell, J. and Warner, J.R.1997. Ribosomal protein L32 of *Saccharomyces cerevisiae* influences both the splicing of its own transcript and processing of rRNA. Mol. Cell Biol. 17: 1959-1965.

Vitart, V.R., De Paepe, C., Mathieu, P., Chetrit, P. and Vedel, F. 1992. Amplification of substochiometric recombinant mitochondrial DNA sequences in nuclear male sterile mutant regenerated from protoplast cultures in *Nicotiana sylvestris*. Mol. Gen. Genet. 233: 193-200.

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

Wahleithner, J.A. and Wolstenholme, D.R. 1988. Ribosomal protein S14 genes in broad bean mitochondrial DNA. Nuc. Acids Res. 16: 6897-6913.

Weglohner, W., Kauschmann, A. and Subramanian, A.R. 1995. Evolution of NH_2- and COOH-terminal extensions of chloroplast ribosomal protein S18. Nucleotide sequences of pea and rye chloroplast *rps18* genes. Biochem. Mol. Biol. Int. 36: 265-273.

Williams, M.A., Kutchner, B.M. and Mulligan, R.M. 1998a. Editing site recognition in plant mitochondria: the importance of 5' flanking sequences. Plant Mol. Biol. 36: 229-237.

Williams, M.E. and Sussex, R.M. 1995. Developmental regulation of ribosomal protein L16 gene in Arabidopsis thaliana. Plant J. 8: 65-76.

Williams, M.A., Tallakson, W.A., Phreaner, C.G. and Mulligan, R.M. 1998b. Editing and translation of ribosomal protein S13 transcripts: unedited translation products are not detectable in maize mitochondria. Curr. Genet. 34: 221-226.

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

Wilson, R.K. and Hanson, M.R. 1996. Preferential RNA editing of specific sites within transcripts of two plant mitochondrial genes does not depend on transcriptional context or nuclear genotype. Curr. Genet. 30: 502-508.

Wischmann, C. and Schuster, W. 1995. Transfer of *rps10* from mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA mediated transfer and exon shuffling at the integration site. FEBS Lett. 374: 152-156.

Wissinger, B., Schuster, W. and Brennicke, A. 1990. Species-specific RNA editing in the mitochondrial *rps13* transcripts of Oenothera and Daucus. Mol Gen Genet. 244: 389-395.

Wittman, H.G.1983. Architecture of prokaryotic ribosomes. Annu. Rev. Biochem. 52: 35-65.

Wolfe, K.H., Sharp, P.M. and Li, K.H. 1989. Rates of synonymous substitution in plant nuclear genes. J. Mol. Evol. 29: 208-211.

Wool, I.G. 1996. Extraribosomal functions of ribosomal proteins. Trends Biochem. Sci. 21: 164-165.

Ye, F., Brenhardt, J. and Abel, W.O. 1993. Genes for ribosomal proteins S3, L16, L5, and S14

are clustered in the mitochondrial genome in *Brassica napus* L. Curr. Genet. 24: 223-329.

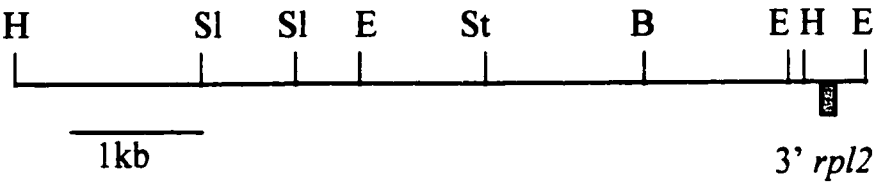
Zanlungo, S., Quinones, V., Moenne, A., Holuigue, L. and Jordana, X. 1994. A ribosomal protein S10 gene is found in the mitochondrial genome in *Solanum tuberosum*. Plant Mol. Biol. 25: 743-749.

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

Zhuo, D., Nguyen-Lowe, H.T., Subramanian, S. and Bonen, L. 1999. The S7 ribosomal protein gene is truncated and overlaps a cytochrome c biogenesis gene in pea mitochondria. Plant Mol. Biol. 40: 91-97.

Zurawski, G. and Zurawski, S.M. 1985. Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acids Res. 13: 4521-4526.

Appendix 1 Sequence analysis of the wheat mt *rpl2* upstream region.
Most sequences presented here have been obtained from only one strand. The restriction sites shown in the figure are given as bold letters and the gaps as series of letter N which are underlined. These sequences are immediately upstream of the EcoR1 fragment shown in figure 3.1.



AAGCTTGTAAAAATGACATCCCCTGCGAAAACCAATAGGATTGTTGGGCCATCGAACACTAGTTTCTGGTTTCTCAAGA-80
CCTTGAGGCAAGGGCAGTTACGCAACATTTTCTCAATATTGAATGTGCATTTTTCCGAATTTTTCGATTCCACTTCTGT-160
ATCAGAAATTAGTGTCAAGGGATGGAGTTCCTGTTTCGTTAACTAAGTCTCGCGTTGTTTGCTTGAGGAAGGGACGGCCG-240
TAGCTCAATTTCCATTGCTAATCTTACATGTTCCACGAACAATAAAGAAAGAGAAAGGCTATTCTTAGACACAGAGGTGG-320
GNNNNNGCATGGGTGGCCCTATCGATCACAAGGGTCTAGATACGGGTCTCGAACGGAAAAGTATGCTACATATTCGTCTG-400
CTCCAGTTCCCTCTGCTTACACCTGGTCAATTCGTTGATCGCCCCTTTTATTTCTGATTTTCAAAAGGGCCCCGCTCTCG-480
TCTCTGCTGACGAAGCCTTTATCTCTGTCTCTGTTATCCTCGTACCTGCTGATCTCTTCGACTGATAGAGTCTATGCTAT-560
TGGTGAAGCTGGTTCTACCTCAACAACAGGATCTGCAAATGGATTTTCAACTAACAGCCTCTGCTTCAATTTGTGATGAA-640
AG**GTCGAC**CAGTTTAGTCCTTGTACTTCGGTGGAACCAAATTCCTGTATATGGATTCGAAACCTTTTCTGCTTTGGGTCT-720
TAAGTGAATGAAAGGGTGGGTTAAAAAACCATGGCCAGCCCCCCTCAAGAAATCTCGCCCGTAGTTATCTGAGGGATCTG-800
ACGTCCGGGTATCAAGGAAAAACGTGGTGCAGAGACAACAAGATGTGCTAACACATATCCCTTATATTTATTGGTTCTAC-880
AGAAGATGTATGATAAAGAGCTGGCATTAAGTCATGCAACCGGCTCAGGCTTTATAAGAGGAAAGACTCTCGCCAAGGTA-960
TAGATTGAACCGTACGGCGCTTGTTCACCAATTCTATCTTTTCCCTCTAAAAAAGGCGTATACTTTTGTTCTATAAGCAA-1040
ATAAAGATACTTCTTTTGCGGATCAAGCAGTCTTCTACCTTGGTGGGTAGCGGAGAATCAAATCAATATCGGAGTATTGC-1120
TCTAGCTGCTGCTGAATTAACCTTGGATTCAATCGTGACTCTCGTGAACTAGGCTGTACTTCGTCTTCCCCTCTGGACCT-1200
CGGAAATGGAAAAAAGTTACCGGTGCGTAAGATTCGGTGAAAACTG**GTCGAC**TCCACCGCAGGTGCGAGATGTTGAGCTT-1280
CCTCGACGCCTACTCTGGGTATCAGCAAGTCTATATGGCTGAGGAAGACAAAGAAAAAACGAGTTTGATCGCCCCTTCCT-1360
TCTGCTACAAGCGGATGCCCTTCGGTTTGATAAATGCCGGCTCCACATTTAAGCGCCTAATGAGCCTCATCCTCAGGTCT-1440
CAGTTGGGTCGTAATGCAGAAGTGTACGTTGACGACGCCGTCATCAAAAGCCGCTTGGCAGGGACTCACCCCAAAAACAC-1520
CTCTTTTTTTGCATCTGCTGAGCAGAACCAATTACCAGATCCTACTTCTAAGCTCTATGCTCTACTTCGGCCTTGCTCTC-1600
ACTTATGGGTAGGAAGGACAATAAGCAG**GAATTC**ATGTTGGCTAACTATTCTATCCAGGCGGAAGTTGTTTTCATAGGAA-1680
AGACGGAAGGTTTTGCGCGTTGGTTGTATCGAATTTCCCTTTCTAATAGCATACACAAATAGATAAGAAAGTTACGACGA-1760
TTGAATAAGGTAAATGAAAACCTTTGATTTACCTTACCATCAGTTCTGTCTTATCTCTGTGCCTCCTCTAGCCATTTCAT-1840
AATTGGATATCGTCATTCAAATCTGTTATGACAACTGTTTGGTAAGAGGGATTCGAATCAAACAATAAATACCAACTCTA-1920
GTGTGATGGATAGGACCTGGTGCCTTAACTACTCAACCTTGAAACCACGACCTACCTACCTCCTGCCTTTTTACCTTTCC-2000
TTCCTTTGTCCCAATTCATATCGAGCATGTCTTTCTCCATCTATAGATAGAGAGATATAGGATAGGGGCGCTCAAAGCAA-2080
ATAGGATTCTGTATCTAATACTAATAGGTACTAACCGCATTCAAGATTGACTTGACCTCTTATTTTGGATAGGCAGAAAT-2160
TTACGCCCTAGGGGTTTAAAGGGCCCACTCCCTTTGGTTGACTCATCAGGCTCACCACCTTTTG**GAGCTC**ATCCAACAGT-2240
CAGTCTACTTAATCATAATCATAATCAATCCGTTGCTATTGCTATCTATGCCATACCCACAAGAGGGTAATTTATCCAAC-2320
TCCCCGTGCGAAGCAAGTCCCGCGTCAGCAGTGATCGAGTTGTTCCTTGAGAGCAGGCTTATGGCATTCCAGGTAAACAT-2400
AACAATATGAGAGAGAACTCCGTTGGCTTATAAGTACTCGATGAACAATTGTTTTTTCATTTCCAGTAGGTCACCTTGGG-2480
CCAGCTGCTGACTTTCTAGAAGAGACAAGGAAGGTCGTCCAACCATAA**NNNNN**TATCCTCTACGAGATAGCGAAGTAGCA-2560
TTAGCGGAAAAGTAAGCTCCTTTTCAAGCCTTCTCTTTCGTAACCTTTGATTAGTTCCTTAGATGCCCTTAGCAAGAATG-2640
ATGCAGGAGCAAGAAAGTCTATTTCCTCCGCCTTTCCAGAATATAGCTATTCAAATAGAACAGTATAAAATGCAATCCAT-2720
AAGTCATGCTGGTCAATCAATGACAATGATAAGCTCTATAAAAAAAAAGAGTCTATTGCTCTTCGACACGTTTTATATCT-2800
ACTTCAGTCATGCTTTATTGAGTTATCTGACGCCAAGGAGGAAGGAATAGAGATTA**GGATCC**TAGCCTTGTTCCGTTTGA-2880
AGTTCCTAAGAGAGGTTACAAAAGAAAGGGATGCTCAAACCAAGGATTTGGAGTCTTATTAGTGCATCAATCTACAGAAA-2960
GGTAGGGGGGCATACCTGTGTGAAAATGGAGTGTAAAGCATATTCTAGATATTTCACTTCGAATTTCTTCCTCTTAAAAG-3040
AGTCAAGTCCGCAGGGTTCCCTAAATTCCCCAGGAATCAAGTGGGACCGGGAGATCGTTGTTCATTCGGGGTTGGCCCTC-3120
CCCCTAATAAGTACCTACCAAGAATTGGGCCAGTTAGTTCAAGCGGTACCGAATCACAAAGAACAGTTGAGTTTTCTAAT-3200
TCATCTATTATCTATTTTTTGCATTTCTAGAATAGCGTTGGAGCCGGGACACACATTTCTGCCCGAGTTCCAACAGCAAA-3280
CCAAATACGTTATCGTGGAAAACGAGTTTCGCCAACACAGAAGAGTTTGTGTGCTTTTTATTTTCACATGAGTTTCACAT-3360
ATGTGATGGTTGGTTGGGAGGGTACGCAATGACTCTAGATCTTTCTGGATACAACCACTGAACCGGAAATGGCCACTAT-3440
CGTCACCGAGAAAATACTATCTTGTGGACTCAGGCTCTACGTTTATGCCTTAAGGTCTTTCTCCTTACCGTGGGGAGCGG-3520
TAACCTTGCGTGATTCTAGAGAAAGAAGGCGGTGTCCCGACCACAAGGCGGTTAAATTCCACTGGTGCTTGCTAGCCAGT-3600
CAAGAACGACTCGAAGACCGTGGGAAGGTCTTCACTGACGCCGAAGACGACGATGGCTCCATACCGGGTGGTAATACACA-3680
GGTATAGGATTTCCTGTAATGATGTGAAACTGTATATTTCAGAGAAATAGCTCAGTTCGAGAGGAAGGCGGAAACAGTTC-3760
CGGTTTTATTAACCTGATTCCTATTCATTTACCGGTAAAGGAAG**GAATTC**-3810

Appendix 2 Sequence analysis of the rice mt *rpl2* and *rps19* genes

Rice mt *cp-rpoB*-mt *rpl2-rps19* gene sequences. These sequences are the ones published by Kubo et al., 1996 Ac. No. D78336. I have also sequenced this region and my data is in agreement. A restriction map is given in figure Figure 3.7.A. The 5.0 kb Bam H1 fragment is given here.

```
   1 gtcgaccaat ccttcctaat tcacattttt gttgaaaaaa tttcttttgt aattcctcac
  61 ataaggattc cgaaaatacc aggtccccgc ctacacaagc aaattgttga taaaactcca
 121 aaatagcttt ttcttttgac tcaatcctct tcttctcctt agcattaggg aaagacaaga
 181 aaatttcagg gtaggaaaca ttatctagaa tttcttttag atttgaaccc atagctgatg
 241 atagaactag aacagatatc ttttgttttc tactcacgcg agcccatatc ctttcttttt
 301 tatcaattgc taattccgat cttcctcccc aatctgatat tatagtcccg gtgtagatag
 361 aaattccttt atggtctaat tccgagcggt agtaaatacc aggacttagc aatatttgat
 421 tgatcacaat tcggtatatt ccatttatta taaaggttcc taaggaattc attataggaa
 481 tgtttccaat ataaatggtt tgcttttgca catcgaaacc aaaaattaat cgcgcggata
 541 cgtataattc ggaagaatag gtgagtgatt catacacagc atccctttct tttatcgaag
 601 gttctagcaa ttgatatcct ttcgcaaata attgaaatga aattccgtga tctggatctt
 661 taattgttgg aaacttctca agttcttctg ccaagccttg attaatgaac ctacaaaatc
 721 cctcgaattg gatctgacta aatccgggta ttgtggacat tccctcattt ccattccgga
 781 gcatcttaat cttaagtttc ctgtttatcg aagaaaaatt ccattatcag ctcactcttc
 841 atcaatccct acagatcgat ctagcaatta tggaatctat attctgttta ctaaatcaca
 901 tgaaattctc gggaactcca catacgtatt tcatatatgt atttcataca tatgaataga
 961 gacaatttca acgaaagttc gaatttgcca ggggtccaaa tggaatgaaa atgggtatat
1021 aaaagaccaa cgaaacgacg agtgagaagg ataaggggga gtagtaggag gagctttcat
1081 tgaagtagag gtttcatcga gtgatgacga gagatagaat gagacaaagc ataaagggga
1141 gagcgcttag acatttcact ttgagtacgg ggaagtccgc agggaggaat tcttccgggc
1201 gtattactgt ttttcaccga gggggtggat cgaagcgatt gcagcgaaaa attgaccttа
1261 aacgaagcac ttcgtctatt ggcattgtgg aaaggatcga atatgaccct aatcgttctt
1321 ctcggatcgc tctagtacga tggatcgaag gggtgctgcc cggccgccag aggaaattca
1381 agacgataga agagttcgct ctgccgcgta agatcctcga atccaccacg gccactatct
1441 tttgcctttt ttcgttctct tccctgtcct ctcccttggc ccagggagag actgcatccc
1501 tttccttcgg cagctcttta ggtttcccaa ggatagcggt agctggggca aagcccgctt
1561 tcttcgcgga gcgaatgaga gagaagaaaa tcggaaaaaa gacgttttct ctttgcgaga
1621 tccgaaagtg gagaacgcat tgcgttctct gggcacatag gatcaaacgt aaagcagcgc
1681 tttcttggca gagtttgagg cagcaaaaaa ctttagaact tgttggagct gctgagcata
1741 acgaatcgaa gctgaaggcg gatcaaggta gcttgctccc aaggcaagtg cttgcttacg
1801 ctttatgtag tggtcggcct tcctaccttc atgcctctag aagcttctac aaagctttgc
1861 ttccggtaga agctagtcgc ttcggtagct tgcccgccaa gccgcctata ggcgaagggc
1921 cgaaggatgg agcgtacaaa gtcgatcgtg cacctgtcgt gtgacccgtt ggtcctaagc
1981 aatgtctttc gcgaagcgac ccacctagaa acagctctcc ttttaggagg caaaaaaatg
2041 gaacttcgat ctgatgcggg tatccaatcc cgccgctggg atgtccagcg gattccgggg
2101 ccttgacaaa tggccgttca ccattgttgt tagaagagca aaaggcccgg ggcatagcag
2161 gatgaaccaa tgtgaatgag tgtaagcttc gttgcccgaa cacgattggt gctgaccaca
2221 ctaggtgcta ccgtggtagc aagagaggcc cggcggtgac aattgagagg ttgtcactga
2281 gcattccgtc tcacacggga agagaggtca aatggcaaag caaaaggcca tacgcccgtg
2341 tggctcctcg cggagtatag ctcacatcca aacatctgat tggggaacgg ggcaacgccc
2401 atgaagctcc ggcaaaggga aggcctgcca ggccgtatgc tcatgggtgc aggattcttc
2461 gaaaaagcgc gggctgactc ggagacctga gaccttggct tagctacgaa tgaaggggag
2521 aaagctcttc gagctttctc cgccagcggc ttatgtagtg gtcggccaat taaagctcgc
2581 taagcttcgc ttccccccct atgaaaaaaa gtagtcggca ttctataagc gacttgtcga
2641 gtttacaaag ctttgcttct tgtagttggc cctccatgcc tcccttcatt tgcttgcctc
2701 cttccagccc agaatgccta ctgacgttaa gctaaccgcc agaagcgaca cccgaagggt
2761 gagcttctgg cgcaggaggc caagcattct gccagacgcc ccgggagccc cgtttgcctt
2821 tactagagtc ttgagttttt ctttcttttt tcattatagc taagtttcaa aagcaaaggt
2881 gaacagcccc ctgtccttta gatttgcata gtcaagggct tataagaaaa gtaaggaagg
2941 aggcatggag atgggaaggc cgaccactac actacacgga gagatttcta taccaagtca
3001 tgagcgatag cgaagccgcc tataggcgaa gggacgaaag cccgttagac ctgatgggaa
```

```
3421 cacaggggct taggtcaaca cttatatatt agccagtcat caattggaag cgggtaatat
3481 ggtgataaat tgcgattgct ccaaaccttc taaaagcggc ttcttgcgac ctgcccagaa
3541 tgctcataca taccttcggt tccaagagct tggacgcaca gtgaataaag gtcgggttga
3601 agggggcagt cagctggcag cttcttggcc acgcccccct gcttatagac acgagatatt
3661 ggatctaaat tcaaaagtag gaaatagcat accattagct gatatacgta tgggaacatg
3721 ggtacatgat attgaatgtc atccaggtca aggcgcaaag ctggctcggg ccgcaggaac
3781 ttatgctaaa ataattaagg agccagcccc acaatgtctt gtgcggctac catcgggtgt
3841 tgaaaaactc atagattccc gatgccgagc tactattggt atagtttcca atcccaacca
3901 tggtgcacgt aagcttagaa aagcaggaca aagccggtgg tcaggcagac gccccattgt
3961 tcgtggtgtt gcaatgaatc cagtggatca tcctcatgga ggaggtgagg ggcgcacgaa
4021 aggaggtaga ccttcggtgt caccttgggg gaagcccacc aaagcaggat ttcgggcagg
4081 agtgggggtg gggaaacgca gaatttagtt catgccacga cgatctatat ggaagggaag
4141 ttttgttgat gctttcctgt ttagaataaa gaagaacaga gaaagtctga tgagcaggaa
4201 aatttggtca cgtagatctt ctatttcgcc ggaattcgtt gattgctccg tactcattta
4261 caatggaaaa actcctgttc gttgtaagat cactgaagga aaggttggtc ataaatttgg
4321 agagtttgct tttacacgga gacgaagacc ctatcaaaca aatagaggaa agggcagaaa
4381 ggggaaaaag taaagtctaa gcgacatatt cggtaatccg aatcccatat aggtcttttc
4441 ttcatactcg accagcccta ggttggggat aatgattttt tcaaatcttg ttgggagatg
4501 cttcgctccg agtccgactg ggacggaatg ggggacccag gcccttcaat gcctccctcc
4561 ggatc
```

Appendix 3 Sequence analysis of the rice mt *nad4L* gene region.

These sequences are immediately downstream of the rice mt *rpl2-rps19* gene sequences Accession No. D78336(Kubo et al., 1996). The sequences presented here are obtained from only one strand. The coding region is shown and the experimentally determined editing sites are given as underlined bold letters with the derived amino acid sequences. The 5'& 3'end of the *nad4L* mRNA are indicated by arrows and the conserved promoter motif by bold underlined sequences. The restriction sites BamH1, EcoR1 and the Hind III are shown as bold letters.

```
GGATCCACTAACTCGGAAGAGCTATTCAAAGGTTTTCTGGCTGACTGTGAGGACACCTCTGCATCCCAGG 70
AGCAGACCGCTCCCCCGGAAGCTGTGGTCAGTCAAGAGCGGCATCCGAATCCCGCTCCAATCCCTGGAGC 140
GGCCGCGAAGACGAACAGAACCCTCCTTCTTTTTCTTGGGAGGGACGACTCTCTCTCTTTGAAAGGAAAG 210
TTCTCTACAAAAGGCCCGCGCCTGTCACGCCCAGATTCTGGACACATTGCGTGATATTACTCTGAAGAAT 280
GGAGAGGAGCTCCATGAAGAAGACTATCGCACCATGCTTGATTTCATGTTGAGCCGACTCCAGTCCGACG 350
TCAATCCCCGGGCGCTGAAGGAGCTCCTGAACAGACTCCAAAACGGTCGGGCACGATCTCAAACTTCTAG 420
                                oligo # 11
GGCGGCAAGGAGATTCCTCGACGAGGACTAGTGAGGCTATAGTACTGACTATGCATAGGACTACAAAGGT 490
TCAAAGACAGGATGTATTCGACCCTTGGATGTGATCGGCAAGAGAAAGATGGAATTCCTTTGGTATAAGG 540
AAGGAAATCAAACGCGTTGTTTGAGAAGGGGTAACTAGAGAAAAAATCCAAATAGGCTCTCCCTGGACTC 630
                                                          M
TACGCAGGACTTCTTTCTAAGCCTCACATAATTTGAATTTCTCTGACATTCCGATGTTTCCGAAACGGAT 700
                                                          T  D
 L                                  F                W
CCTATAAAATATTTCACTTTTTCTATGATCATCTCTATTTTAGGTATTCGGGGAATCCTCCTTAATAGAC 770
 P  I  K  Y  F  T  F  S  M  I  I  S  I  L  G  I  R  G  I  L  L  N  R

                    L     L        L
GAAATATTCTTATTATGTCAATGCCAATTGAATCAATGTTATTAGCTGTCAATTTGAACTTTTTGGTATT 840
R  N  I  L  I  M  S  M  P  I  E  S  M  L  L  A  V  N  L  N  F  L  V  F

                                    L        L        L
TTCCGTTTCTTTGGATGATATGATGGGTCAATCATTTGCTTCATTAGTTCCAACAGTGGCAGCTGCGGAA 910
  S  V  S  L  D  D  M  M  G  Q  S  F  A  S  L  V  P  T  V  A  A  A  E

TCTGCTATTGGATTAGCCATTTTCGTTATTACTTTTCGAGTCCGAGGGACTATTGCTGTCGAATTTATAA 980
 S  A  I  G  L  A  I  F  V  I  T  F  R  V  R  G  T  I  A  V  E  F  I
                                oligo # 12
ATTGCATTCAAGGTTAAACATAACTACAGGGGAGTTACCAAATACAAAGTTCTGTTTTCCTTTCGTTCTC 1050
N  C  I  Q  G  U

TTCTTTCTTTTTGATTTTGTTGTTTTGCCACCGAAATAGCTCTATAGGAGGAATTGTTCAAGACCAGTTC 1120
TCATTCTTTCCTTTTTGCTAATCCATCCGTCCGTCCTCTTTCCATTAGAGCGATTGAAAGCAAGTGAGAA 1190
CCTTTATAGCCAGTGAGCTAGTAAGAATCCGCTTTCGAGACATCATAAGAAAAGGCGGCTAGGGCAGGGG 1260
TGCCATGTCTTAGAGGTTCCAAGCAATCCTCTTCCCTTCAACTCGTACCGAAGGCTATGTCCCAGTAAGC 1330
ATTTCCTCTGTTTCCTTGTTTCCCACCTTCACATGCAAGGAGCAAGCAATCGTACCTCCGCCTATCTGCG 1400
AGATCGACTCTATGGCTGCGTCGGACAAAGCATCAGCTGCTCCATTCTTGGCTCCCTTGCGTCGGTATTA 1470
AGTCCAAAAGCTT 1483
```

Appendix 4 Sequence analysis of the 3'RACE product of the wheat mt *rpl2*

oligo#18

AAGGGATCCAGCATACAAGGC CGAGCGTGCACCTGTCAATTACATAATAGCCAGCCATCAAATGGAACCG 70
R D P A Y K A E R A P V N Y I I A S H Q M E P

GGCAGCATGGTGGTGAACAGCGACTCCTCCAAACCCTCCACGACCGGCTCCTTGATGCGGCCTGCCCACA 140
G S M V V N S D S S K P S T T G S L M R P A H

ATGCCAATTCCATCCTTCGGTTCCAAGAGCTGTTCCGCAAAGCCAGCCAAGAAGGCGAAGAGGGCACTGA 210
N A N S I L R F Q E L F R K A S Q E G E E G T D

TGATCAAGCAAAGGATGCGGCGGTTCCAGCAGC TTGTTGCTCCGAGGGAAAGGTACCGCATCTTTATGTG 280
D Q A K D A A V P A A C C S E G K V P H L Y V

GCTATGGATGACAAAGCTGAGGATTATCATGCTTGTTTTTTTGTTTTCGAAAAAACGGTTTCCGCCGCAG 350
A M D D K A E D Y H A C F F V F E K T V S A A

CACTTTGACTTTATTATTATGTATTAATAATAACAGGTGATGTTTGGACCTGCCTGCGCCCAGTTTGCAT 420
A L * L Y Y Y V L I I T G D V W T C L R P V C I

TCCTTCCTGGCATTTTGACATTAACTTGTTTAGGCATGTTCCTTGTGATTGATAATTTGCTCTGTCTGGT 490
P S W H F D I N L F R H V P C D U U F A L S G

TAATCTAACTGGGAGGGAACAAAATGATGATGTGCATATATTTAGTTCCAGTTTCTGTAGCTGCTGTTCG 560
U S N W E G T K U U C A Y I U F Q F L U L L F

GCCATCACAAGCATGCATGGGATGAGCTGCATACATATACAGTAGTTTATTCAGTTTATCCGAATTAGCT 630
G H H K H A W D E L H T Y T V V Y S V Y P N U L

GATTCAGATCAGTGTAATTTCTGTTGGTGTAAAACCATTACATATGTTGAAAGTTAGCATGATTGATTGA 700
I Q I S V I S V G V K P L H M L K V S M I D U

oligo#24

TTTGCAAAAAAAAAAAAGCTT 720
F A K K K S L

Appendix 5 List of L2 snd S12 sequences used in the phylogenetic analysis

Organism	Gene	Compartment	Accession No.
L2			
Rice	*rpl2* mt	mitochondria	This work & D78336
Wheat	*rpl2* mt	nucleus	This work
Arabidopsis	*rpl2* mt	nucleus	S49579
Liverwort	*rpl2* mt	mitochondria	P26859
Yeast	*rpl2* mt	nucleus	S30827=YEL050c
Rice	*rpl2* cp	chloroplast	P17351
Tobacco	*rpl2* cp	chloroplast	P06379
Yeast	*rpl2* cyt	nucleus	S50243 = L8.e
Arabidopsis	*rpl2* cyt	nucleus	1173018
Wheat	*rpl2* cyt	nucleus	This work
E. coli	*rpl2*		P02387
Oenothera	*rpl2* mt	mitochondria	S46947
Arabidopsis	*rpl2* mt	mitochondria	S46378, S49606
Reclinomonas	*rpl2* mt	mitochondria	O21247
Potato	*rpl2* mt	mitochondria	AF095278
Methanococcus jannaschii	*rpl2*		D64322
S12			
Wheat	*rps12* mt	mitochondria	P10851
E. coli	*rps12*		P02367
Rice	*rps12* mt	mitochondria	P28520
Petunia	*rps12* mt	mitochondria	P49195
Liverwort	*rps12* mt	mitochondria	P26871
Oenothera	*rps12* mt	nucleus	Q01607
Yeast	*rps12* mt	nucleus	P53732
Tobacco	*rps12* cp	chloroplast	R3NT12
Rice	*rps12* cp	chloroplast	P12149
Reclinomonas	*rps12* mt	nucleus	AA11866

Note: The gene referred to as cytosolic *rpl2* in yeast by Presutti et al., (1995) is in fact not the plant cytosolic *rpl2* homologue.

Appendix 6 List of genes used in the codon usage analysis

NUCLEAR GENES	Ac.No.
Wheat mRNA for cytosolic phosphoglycerate kinase	X15232
T.aestivum mRNA for starch branching enzyme I	Y12320
Triticum aestivum COR39 (cor39) mRNA, complete cds	AF058794
Triticum aestivum mRNA for wpk4 protein kinase,	D21204
T. aestivum 1,4-alpha-D-glucanotransferase mRNA,	U66376
T. aestivum L-isoaspartyl methyltransferase mRNA,	L07941
Triticum aestivum manganese superoxide dismutase (SOD3.1) mRNA,nuclear gene encoding mitochondrial protein.	U72212
T.aestivum atp-2 mRNA for ATP synthase beta subunit	X74545
Triticum aestivum ADP-glucose-pyrophosphorylase large subunit (AGP-L) mRNA	AF026539
T.aestivum VDAC3 mRNA for voltage dependent anion channel.	X82148
T.aestivum mRNA for transmembrane proton pump,	Y09815
T.aestivum ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO) gene, exons 1 and 2.	M37328
Triticum aestivum Cu/Zn superoxide dismutase (SOD1.2) mRNA, nuclear gene encoding chloroplast protein.	U69632
Wheat mRNA for plastid omega-3 fatty acid desaturase	D43688
Wheat FBP gene for chloroplast fructose-1,6-bisphosphatase.	X53957
Wheat major chlorophyll a/b-binding protein gene.	M10144
MITOCHONDRIAL GENES	
T.durum mitochondria genes *atpA* and atp9.	X80469
Wheat mitochondrial coxI cytochrome-c oxidase subunit I	X56186
T.aestivum mitochondrial COX2 cytochrome c oxidase subunit II.	X52867
Triticum aestivum mitochondrial cox3 gene.	X15944
Triticum aestivum mitochondrial nad1 gene.	X57966,7,8
Triticum aestivum mitochondrial NADH dehydrogenase subunit 2.	Y14435
Wheat mitochondrion NADH-ubiquinone oxireductase subunit 5 gene	M74157
T.aestivum ND6 gene for NADH-ubiquinone oxidoreductase (subunit 6).S47556	X62100
T.aestivum mitochondrial nad7 gene for NADH dehydrogenase subunit 7	X75036
T.aestivum mitochondrial nad9 gene for 30 kDa subunit of complex I NADH dehydrogenase.	X69720
Wheat mitochondrial ORF25 gene.	X54311
Triticum aestivum mitochondrial matR gene & nad1 gene, exon 5	X57965
T.aestivum mitochondrial rps1A gene and orf589.	X69205
T.aestivum mitochondrial gene for S7 ribosomal protein.	X67242
Triticum aestivum *rps2* gene.	Y13920
Wheat mitochondrial DNA for ribosomal protein S13.	Y00520
Wheat mitochondrial *rps12* gene.	X59153

Appendix 7 Codon usage table for wheat mitochondrial, nuclear and wheat nuclear-located mt *rpl2* genes

7.1 Codon usage table for wheat mitochondrial genes

Wheat mitochondrial ribosomal protein genes (mt) *926 codons*

		RSCU			RSCU			RSCU			RSCU
UUU	24	1.09	UCU	18	1.89	UAU	17	1.62	UGU	12	1.60
UUC	20	0.91	UCC	9	0.95	UAC	4	0.38	UGC	3	0.40
UUA	20	1.38	UCA	10	1.05	UAA	2	0.00	UGA	3	0.00
UUG	19	1.31	UCG	5	0.53	UAG	0	0.00	UGG	8	1.00
CUU	18	1.24	CCU	7	1.00	CAU	19	1.73	CGU	17	1.13
CUC	6	0.41	CCC	5	0.71	CAC	3	0.27	CGC	8	0.53
CUA	10	0.69	CCA	11	1.57	CAA	24	1.55	CGA	18	1.20
CUG	14	0.97	CCG	5	0.71	CAG	7	0.45	CGG	7	0.47
AUU	36	1.32	ACU	13	1.41	AAU	26	1.58	AGU	6	0.63
AUC	23	0.84	ACC	9	0.97	AAC	7	0.42	AGC	9	0.95
AUA	23	0.84	ACA	7	0.76	AAA	50	1.23	AGA	27	1.80
AUG	16	1.00	ACG	8	0.86	AAG	31	0.77	AGG	13	0.87
GUU	11	0.90	GCU	26	1.58	GAU	38	1.52	GGU	16	1.08
GUC	12	0.98	GCC	12	0.73	GAC	12	0.48	GGC	9	0.61
GUA	13	1.06	GCA	14	0.85	GAA	29	1.29	GGA	24	1.63
GUG	13	1.06	GCG	14	0.85	GAG	16	0.71	GGG	10	0.68

Wheat mitochondrial respiratory protein genes (mt) *4929 codons*

		RSCU			RSCU			RSCU			RSCU
UUU	202	1.17	UCU	93	1.39	UAU	127	1.54	UGU	43	1.19
UUC	143	0.83	UCC	63	0.94	UAC	38	0.46	UGC	29	0.81
UUA	168	1.73	UCA	70	1.05	UAA	4	0.00	UGA	3	0.00
UUG	106	1.09	UCG	46	0.69	UAG	5	0.00	UGG	82	1.00
CUU	109	1.13	CCU	76	1.41	CAU	80	1.44	CGU	54	1.21
CUC	59	0.61	CCC	46	0.85	CAC	31	0.56	CGC	25	0.56
CUA	94	0.97	CCA	73	1.35	CAA	109	1.60	CGA	72	1.61
CUG	45	0.46	CCG	21	0.39	CAG	27	0.40	CGG	20	0.45
AUU	174	1.37	ACU	91	1.43	AAU	91	1.35	AGU	83	1.24
AUC	104	0.82	ACC	80	1.26	AAC	44	0.65	AGC	46	0.69
AUA	103	0.81	ACA	56	0.88	AAA	104	1.16	AGA	64	1.43
AUG	149	1.00	ACG	27	0.43	AAG	76	0.84	AGG	33	0.74
GUU	94	1.13	GCU	144	1.53	GAU	117	1.35	GGU	119	1.33
GUC	60	0.72	GCC	94	1.00	GAC	56	0.65	GGC	53	0.59
GUA	105	1.27	GCA	92	0.98	GAA	135	1.34	GGA	125	1.39
GUG	73	0.88	GCG	46	0.49	GAG	66	0.66	GGG	62	0.69

7.2 Codon usage table for wheat nuclear-located mt *rpl2* sequences

Wheat mt rpl2 copy I

TTT	Phe	0	(0.0)	TCT	Ser	0	(0.0)	TAT	Tyr	0	(0.0)	TGT	Cys	0	(0.0)
TTC	Phe	2	(0.9)	TCC	Ser	9	(4.2)	TAC	Tyr	3	(1.4)	TGC	Cys	3	(1.4)
TTA	Leu	0	(0.0)	TCA	Ser	0	(0.0)	TAA	End	0	(0.0)	TGA	End	0	(0.0)
TTG	Leu	2	(0.9)	TCG	Ser	2	(0.9)	TAG	End	0	(0.0)	TGG	Trp	3	(1.4)
CTT	Leu	2	(0.9)	CCT	Pro	4	(1.9)	CAT	His	5	(2.3)	CGT	Arg	3	(1.4)
CTC	Leu	3	(1.4)	CCC	Pro	2	(0.9)	CAC	His	3	(1.4)	CGC	Arg	3	(1.4)
CTA	Leu	2	(0.9)	CCA	Pro	4	(1.9)	CAA	Gln	4	(1.9)	CGA	Arg	2	(0.9)
CTG	Leu	7	(3.3)	CCG	Pro	5	(2.3)	CAG	Gln	3	(1.4)	CGG	Arg	7	(3.3)
ATT	Ile	0	(0.0)	ACT	Thr	1	(0.5)	AAT	Asn	4	(1.9)	AGT	Ser	0	(0.0)
ATC	Ile	7	(3.3)	ACC	Thr	3	(1.4)	AAC	Asn	4	(1.9)	AGC	Ser	6	(2.8)
ATA	Ile	2	(0.9)	ACA	Thr	1	(0.5)	AAA	Lys	2	(0.9)	AGA	Arg	0	(0.0)
ATG	Met	7	(3.3)	ACG	Thr	2	(0.9)	AAG	Lys	10	(4.7)	AGG	Arg	2	(0.9)
GTT	Val	2	(0.9)	GCT	Ala	1	(0.5)	GAT	Asp	6	(2.8)	GGT	Gly	2	(0.9)
GTC	Val	7	(3.3)	GCC	Ala	11	(5.1)	GAC	Asp	4	(1.9)	GGC	Gly	8	(3.7)
GTA	Val	2	(0.9)	GCA	Ala	7	(3.3)	GAA	Glu	5	(2.3)	GGA	Gly	4	(1.9)
GTG	Val	7	(3.3)	GCG	Ala	5	(2.3)	GAG	Glu	5	(2.3)	GGG	Gly	4	(1.9)

Wheat mt rpl2 copy II

TTT	Phe	0	(0.0)	TCT	Ser	1	(0.5)	TAT	Tyr	0	(0.0)	TGT	Cys	0	(0.0)
TTC	Phe	3	(1.6)	TCC	Ser	8	(4.3)	TAC	Tyr	3	(1.6)	TGC	Cys	4	(2.2)
TTA	Leu	0	(0.0)	TCA	Ser	1	(0.5)	TAA	End	0	(0.0)	TGA	End	0	(0.0)
TTG	Leu	2	(1.1)	TCG	Ser	4	(2.2)	TAG	End	0	(0.0)	TGG	Trp	2	(1.1)
CTT	Leu	1	(0.5)	CCT	Pro	4	(2.2)	CAT	His	2	(1.1)	CGT	Arg	4	(2.2)
CTC	Leu	3	(1.6)	CCC	Pro	2	(1.1)	CAC	His	3	(1.6)	CGC	Arg	2	(1.1)
CTA	Leu	2	(1.1)	CCA	Pro	2	(1.1)	CAA	Gln	3	(1.6)	CGA	Arg	2	(1.1)
CTG	Leu	6	(3.2)	CCG	Pro	4	(2.2)	CAG	Gln	3	(1.6)	CGG	Arg	2	(1.1)
ATT	Ile	0	(0.0)	ACT	Thr	2	(1.1)	AAT	Asn	5	(2.7)	AGT	Ser	0	(0.0)
ATC	Ile	4	(2.2)	ACC	Thr	3	(1.6)	AAC	Asn	2	(1.1)	AGC	Ser	6	(3.2)
ATA	Ile	3	(1.6)	ACA	Thr	2	(1.1)	AAA	Lys	1	(0.5)	AGA	Arg	0	(0.0)
ATG	Met	6	(3.2)	ACG	Thr	2	(1.1)	AAG	Lys	9	(4.8)	AGG	Arg	1	(0.5)
GTT	Val	1	(0.5)	GCT	Ala	3	(1.6)	GAT	Asp	8	(4.3)	GGT	Gly	0	(0.0)
GTC	Val	5	(2.7)	GCC	Ala	10	(5.4)	GAC	Asp	3	(1.6)	GGC	Gly	7	(3.8)
GTA	Val	0	(0.0)	GCA	Ala	4	(2.2)	GAA	Glu	4	(2.2)	GGA	Gly	3	(1.6)
GTG	Val	8	(4.3)	GCG	Ala	3	(1.6)	GAG	Glu	6	(3.2)	GGG	Gly	2	(1.1)

Appendix 7.3 Codon usage table for wheat nuclear genes

Wheat nuclear genes coding for mt proteins 1534 codons

		RSCU			RSCU			RSCU			RSCU
UUU	15	0.48	UCU	19	1.11	UAU	14	0.55	UGU	2	0.40
UUC	47	1.52	UCC	18	1.05	UAC	37	1.45	UGC	8	1.60
UUA	4	0.16	UCA	23	1.34	UAA	0	0.00	UGA	4	0.00
UUG	23	0.94	UCG	11	0.64	UAG	1	0.00	UGG	17	1.00
CUU	31	1.27	CCU	20	1.18	CAU	15	0.77	CGU	16	1.35
CUC	51	2.08	CCC	18	1.06	CAC	24	1.23	CGC	16	1.35
CUA	8	0.33	CCA	14	0.82	CAA	15	0.61	CGA	4	0.34
CUG	30	1.22	CCG	16	0.94	CAG	34	1.39	CGG	5	0.42
AUU	34	1.15	ACU	22	0.97	AAU	13	0.43	AGU	11	0.64
AUC	49	1.65	ACC	38	1.67	AAC	48	1.57	AGC	21	1.22
AUA	6	0.20	ACA	20	0.88	AAA	25	0.56	AGA	9	0.76
AUG	23	1.00	ACG	11	0.48	AAG	65	1.44	AGG	21	1.77
GUU	41	1.43	GCU	37	1.01	GAU	27	0.66	GGU	38	1.27
GUC	34	1.18	GCC	59	1.62	GAC	55	1.34	GGC	38	1.27
GUA	3	0.10	GCA	28	0.77	GAA	25	0.53	GGA	27	0.90
GUG	37	1.29	GCG	22	0.60	GAG	70	1.47	GGG	17	0.57

Wheat nuclear genes coding for cp proteins 1434 codons

		RSCU			RSCU			RSCU			RSCU
UUU	14	0.50	UCU	9	0.44	UAU	4	0.16	UGU	3	0.27
UUC	42	1.50	UCC	40	1.95	UAC	46	1.84	UGC	19	1.73
UUA	1	0.05	UCA	5	0.24	UAA	3	0.00	UGA	0	0.00
UUG	12	0.55	UCG	20	0.98	UAG	2	0.00	UGG	28	1.00
CUU	19	0.87	CCU	15	0.71	CAU	15	0.68	CGU	5	0.52
CUC	64	2.93	CCC	29	1.36	CAC	29	1.32	CGC	20	2.07
CUA	2	0.09	CCA	18	0.85	CAA	5	0.22	CGA	5	0.52
CUG	33	1.51	CCG	23	1.08	CAG	40	1.78	CGG	8	0.83
AUU	15	0.76	ACU	10	0.53	AAU	10	0.44	AGU	4	0.20
AUC	37	1.88	ACC	36	1.92	AAC	35	1.56	AGC	45	2.20
AUA	7	0.36	ACA	10	0.53	AAA	7	0.23	AGA	3	0.31
AUG	31	1.00	ACG	19	1.01	AAG	54	1.77	AGG	17	1.76
GUU	14	0.51	GCU	19	0.57	GAU	6	0.17	GGU	26	0.76
GUC	45	1.65	GCC	64	1.92	GAC	66	1.83	GGC	70	2.04
GUA	2	0.07	GCA	14	0.42	GAA	7	0.22	GGA	12	0.35
GUG	48	1.76	GCG	36	1.08	GAG	58	1.78	GGG	29	0.85

Wheat nuclear genes coding for cytosolic proteins 3267 codons

		RSCU			RSCU			RSCU			RSCU
UUU	69	0.97	UCU	39	1.18	UAU	55	0.96	UGU	7	0.64
UUC	74	1.03	UCC	39	1.18	UAC	60	1.04	UGC	15	1.36
UUA	15	0.38	UCA	42	1.27	UAA	2	0.00	UGA	3	0.00
UUG	36	0.90	UCG	16	0.48	UAG	2	0.00	UGG	42	1.00
CUU	54	1.35	CCU	61	1.36	CAU	62	0.96	CGU	20	0.79
CUC	66	1.65	CCC	28	0.63	CAC	67	1.04	CGC	47	1.86
CUA	15	0.38	CCA	55	1.23	CAA	20	0.49	CGA	13	0.51
CUG	54	1.35	CCG	35	0.78	CAG	61	1.51	CGG	25	0.99
AUU	49	1.13	ACU	46	0.96	AAU	45	0.87	AGU	18	0.55
AUC	63	1.45	ACC	79	1.65	AAC	58	1.13	AGC	44	1.33
AUA	18	0.42	ACA	36	0.75	AAA	47	0.46	AGA	22	0.87
AUG	92	1.00	ACG	31	0.65	AAG	158	1.54	AGG	25	0.99
GUU	72	1.32	GCU	69	1.07	GAU	100	0.97	GGU	82	0.95
GUC	57	1.05	GCC	77	1.19	GAC	107	1.03	GGC	125	1.45
GUA	23	0.42	GCA	66	1.02	GAA	94	0.91	GGA	75	0.87
GUG	66	1.21	GCG	47	0.73	GAG	113	1.09	GGG	64	0.74