# The Mitochondrial S7 Ribosomal Protein Gene: Impact of DNA Rearrangements on RNA Expression in Grasses

Evan Byers

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

Ottawa-Carleton Institute of Biology

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

L'Institut de biologie d'Ottawa-Carleton

**Abstract**

Frequent rearrangements, typically through homologous recombination in plant mitochondrial genomes often result in different upstream and downstream sequences for the same gene among a number of species. Transcription and RNA processing signals are therefore different, even among closely related plants. To evaluate the impact of DNA rearrangements on gene expression I conducted a comparative analysis of the S7 ribosomal protein gene (*rps7*) among a number of grasses: wheat, rice, maize, barley, rye, brome, *Lolium* and oats (grasses whose evolutionary divergence times range from about 5 to 60 Mya). Using circularized-RT-PCR to simultaneously map *rps7* transcript termini I found that 3'ends for various RNA species are homogeneous, mapping to conserved sequences among plants. 5' termini are more complex and can be both discrete and heterogeneous for different transcripts, both within and among plants. Genome rearrangements upstream of the *rps7* start codon for some but not all species has led to plant-specific signals for both *rps7* transcription and RNA processing. Termini for *rps7* precursor species in wheat and *Lolium* are very discrete and likely use different upstream tRNAs as processing signals for end-cleavage. A number of potential stem-loop structures have also been identified at or near 5' and 3' termini which may function in maturation of transcript ends or provide transcript stability and protection from degradation by ribonucleases. C-to-U RNA editing of non-coding sequences, a rare event, was observed at multiple sites within the 5' and 3'UTRs among plants. Some sites may even be developmentally regulated as CR-RT-PCR experiments were conducted using mitochondrial RNA isolated from seedlings and germinating embryos. Taken together, my observations demonstrate the frequency of upstream DNA rearrangements and the variety of signals used for expression of *rps7* among grasses, providing new insights into the complexities of mRNA production in plant mitochondria.

**Résumé**

Des réarrangements fréquents, généralement dû à la recombinaison homologue dans les génomes mitochondriaux des plantes, résultent souvent dans des différentes séquences en aval et en amont d'un gène entre un nombre d'espèces. Ainsi, la transcription et la régulation d'ARN post-transcriptionnelle sont différentes, même entre les plantes étroitement apparentées. Pour évaluer l'impact des réarrangements d'ADN sur l'expression des gènes, j'ai mené une analyse comparative du gène codant la protéine ribosomique S7 (*rps7*) parmi des herbes: blé, riz, maïs, orge, seigle, brome, *Lolium* et avoine (des herbes ayant des temps de divergences variant de 5 à 60 millions d'années). En utilisant la technique de la transcription inverse suivie par la réaction en chaîne par polymérase circularisée (TR-RCP) afin de cartographier les extrémités de *rps*7, j'ai trouvé que les extrémités 3' de différentes espèces d'ARN sont homogènes et sont situées dans des séquences conservées parmi les plantes. Les extrémités 5'sont plus complexes et peuvent être soit homogène, soit hétérogène pour différents transcrits et cela, dedans et parmi les plantes. Les réarrangements génomiques en amont du codon d'initiation de *rps*7 pour certaines et non pas tous les espèces a conduit à des signaux spécifiques entre des plantes pour la transcription et pour la régulation d'ARN post-transcriptionnelle. Les extrémités des précurseurs d'ARN pour le gène *rps*7 dans le blé et le *Lolium* sont discrètes et semblent utiliser des ARNt en amont comme signaux de régulation pour le clivage des extrémités. Un nombre de structure potentielles de forme tige-boucle ont été identifié proche des extrémités 5' et 3' et ont possiblement des fonctions dans la maturation des terminaisons des transcrits ou autrement, fournissent de la stabilité et de la protection de la dégradation par des ribonucléases. L'édition C-à-U de l'ARN des séquences non-codantes, un événement rare, a été observé à plusieurs sites dans les régions non-transcrites des extrémités 5 ' et 3' parmi les plantes. Quelques sites peuvent même être régulés de façon développementale étant donné que des expériences TR-RCP circulaire avec de l'ARN mitochondrial isolé de semis développées et des embryons qui germent. Pris ensemble, mes observations démontrent la fréquence des réarrangements d'ADN en amont du gène *rps7* et la variété des signaux utilisés pour l'expression de *rps7* parmi les herbes, fournissant ainsi de nouveaux aperçus dans la complexité de la production d'ARN messager dans les mitochondries des plantes.

**Acknowledgments**

**I would like to take this opportunity thank my supervisor Dr. Linda Bonen for continual support throughout my project. I was fortunate enough to have the opportunity to work with Linda as a summer research assistant and then again the following school term as a 4$^{th}$ year honours student. I was very apprehensive at first about graduate school however having such a wonderful mentor made the experience very rewarding.**

**I would like to also extend my gratitude to my committee members Dr. Doug Johnson and Dr. Jon Vierula for their helpful feedback. Special recognition to Doug, you always had time to swap stories and provide helpful advice.**

**A special thank you to my lab colleagues: Jen Rueger, Jowairyya Ahmed, Boyoung Choi, Thomas Biela, Jinchao Xie, Patrick Williston, Maria Acero, Rayan Dalbani and Stephen Dalby. I don't know how it is that we're always so fortunate to find such amazing people, maybe Linda's just good at picking them but the Bonen lab has always been a really friendly enjoyable place to work. Jen, thank you for starting work on *rps7*, your findings definitely helped me take this project in a very interesting direction. Jo you were the one who taught me the techniques necessary to conduct my experiments, thank you for being a patient tutor. Patrick, Maria and Steve you guys were always there to lend a hand, I really enjoyed working so closely with you. I have to also acknowledge the contribution of former lab members: JL, JC, CLP, TH, DZ, SC. You guys live on in your lab books and your work here has been very helpful. Lastly I'd like to acknowledge Boyoung Choi who started with me as a research assistant and stayed on to do her honours project the following year. I don't think I realized how much I relied on your friendship until you left us to go back to S. Korea. We worked so well together. I appreciate you always pushing me to do better. And I promise someday I'll come to S.Korea and visit, it's just so far!**

**I would like to also thank the support staff and members of the Biology Department. There are a lot of truly awesome people in the department both administrative and academic. Thank you for helping to make the whole process so smooth!**

**Finally I'd like to thank my family and my very close friends, Jill, Josh and Colin. You guys were there with me the entire way and I really appreciated you pretending to be frustrated for me when an experiment didn't work and offering a consoling hand. You guys are the best and I wouldn't have made it if it wasn't for your support.**

# Table of contents

## List of figures

## List of tables

## List of supplementary data

## List of abbreviations

| | |
|---|---|
| ATP | adenosine triposphate |
| *atp* | adenosine triphosphate synthase subunit genes |
| bp | base pairs |
| BSA | bovine serum albumin |
| *ccm* | cytochrome c biogenesis subunit genes |
| cDNA | complementary DNA |
| CDS | coding sequence |
| CMS | cytoplasmic male sterility |
| *cob* | cytochrome bc1 oxidoreductase subunit genes |
| *cox* | cytochrome oxidase genes |
| CR-RT-PCR | circularized reverse transcriptase polymerase chain reaction |
| cv | cultivar |
| $ddH_2O$ | double distilled water |
| ddNTP | dideoxyribonucleotide triphosphate |
| dNTP | deoxyribonucleotide triphosphate |
| EDTA | ethylenediaminetetraacetic acid |
| EtOH | ethanol |
| kb | kilobases |
| mRNA | messenger RNA |
| mt | mitochondria, mitochondrial |
| Mya | million years ago |
| µL | microliter |
| µg | microgram |
| mL | milliliter |
| mg | milligram |
| $MgCl_2$ | magnesium chloride |
| *MMLV* | Moloney murine leukemia virus |
| *nad* | NADH dehydrogenase subunit genes |
| nt(s) | nucleotide(s):<br>R=purine: A=adenosine, G=guanosine<br>Y=pyrimidine: T=thymidine, U=uridine, C=cytidine |
| ORF | open reading frame |
| PNPase | polynucleotide phosphorylase |
| PPR | pentatricopeptide repeat |
| *rpl* | protein of the large ribosomal subunit |
| *rps* | protein of the small ribosomal subunit |
| RNase | ribonuclease |

| | |
|---|---|
| rRNA | ribosomal RNA |
| *sdh* | succinate dehydrogenase subunit genes |
| SDS | sodium dodecyl sulphate |
| str | strain |
| t-element | tRNA-like element |
| TAP | tobacco acid (pyro)phosphatase |
| TBE | Tris-borate-EDTA buffer |
| Tris | tris(hydroxymethyl) aminomethane |
| tRNA | transfer RNA |
| U | units |
| UTR | untranslated region |
| var | variety |
| ψ | pseudogene |

# Chapter 1: Introduction

## 1.1 *Plant mitochondrial genome evolution: origin and gene transfer to the nucleus*

Plants possess some of the largest known mitochondrial genomes and have retained a number of the ancestral genes from the α-proteobacterium that through endosymbiosis with either an archaebacterial or a primitive anerobic eukaryotic host cell became the energy-producing organelle (reviewed in Gray *et al*. 1999; Martin and Koonin 2006). Over evolutionary time many genes were transferred to the nucleus. Dinoflagellates, a large group of flagellate protists are distinguished in having the smallest mitochondrial gene content among functional eukaryotic organelles (Wisecaver and Hackett 2010). *Reclinomonas Americana*, also a flagellate protist has the largest gene content of mitochondrial genomes with 99 genes (Bullerwell and Gray 2004) and it is the case that the protein-coding content among all functional eukaryotic mitochondrial genomes is a subset of that found in *R. Americana* (Gray *et al*. 1999). In plants, mitochondrial gene transfer to the nucleus is still ongoing as evidenced by genome-wide surveys of gene content among a number of plants (Adams *et al*. 2002a). Successful integration of mitochondrial genes into the nucleus is still possible in plants as both compartments use the same genetic code. Using mitochondrial targeting sequences the products of these now nuclear-encoded genes, whose functions remain in the mitochondrion, are translocated back to the organelle through the cytosol.

Mitochondria provide cellular energy in the form of ATP (adenosine triphosphate) by way of the electron transport chain, the final step in the pathway for oxidative phosphorylation of products from both glycolysis and the tricarboxylic acid cycle. The electron transport chain is comprised of five complexes: complex I, NADH dehydrogenase, complex II, succinate dehydrogenase: CoQ, complex III, cytochrome bc1 oxidoreductase, complex IV, cytochrome oxidase and complex V, ATP synthase. Certain subunits for all of these complexes remain encoded within plant mitochondria, with the exception of complex II in some plants (Table 1.1). In plants however some of these complexes are bypassed under environmentally stressful conditions and instead alternative oxidoreductase respiration pathways (Figure 1.1) are used (Eubel *et al*. 2004). Apart from genes for structural subunits of the different complexes other protein-coding

*Table 1.1: Gene content within the mitochondrial genomes of Triticum aestivum (wheat), Oryza sativa (rice), Arabidopsis thaliana and Marchantia polymorpha (liverwort)*

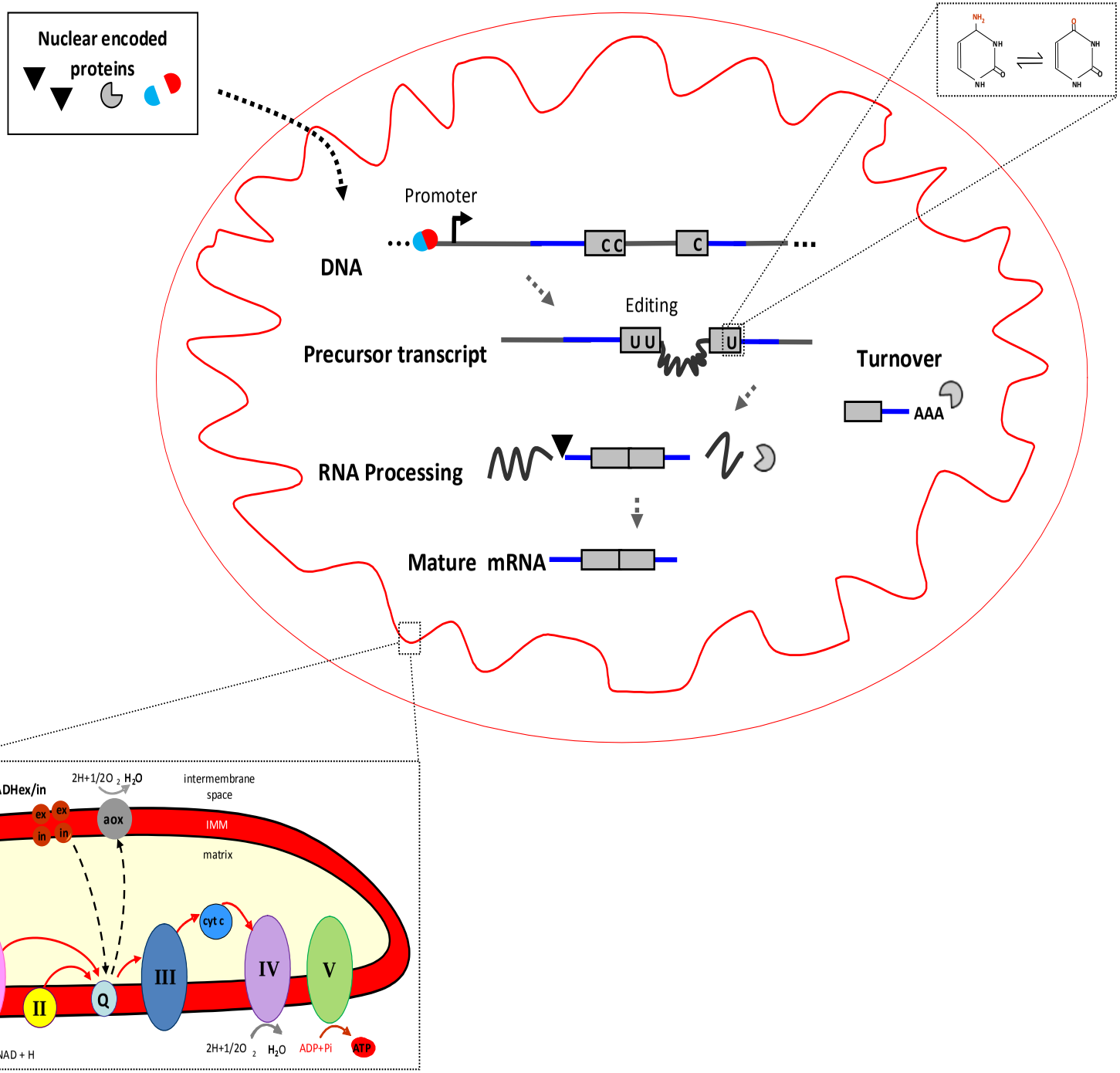
Gene content based on the complete mitochondrial genome sequences for wheat (AP008982, Ogihara *et al*. 2005), rice (BA000029, Notsu *et al*. 2002), *Arabidopsis* (NC_001284, Unseld *et al*. 1997) and *Marchantia* (M68929, Oda *et al*. 1992).

| | Genes[a] | *T. Ae.* | *O. sa* | *A. th.* | *M. po.* |
|---|---|---|---|---|---|
| Respiratory chain genes | | | | | |
| **Complex I**, NADH dehydrogenase | nad1,2,3,4,4L,5,6,7,9 | + | + | + | +[d] |
| **Complex II**, succinate dehydrogenase | sdh3,4 | - | - | - | + |
| **Complex III**, cytochrome bc1 oxidoreductase | cob | + | + | + | +[d] |
| **Complex IV**, cytochrome oxidase | cox1,2,3 | + | + | + | + |
| **Complex V**, ATP synthase | atp1,4,6,8,9 | +[b] | + | +[c] | + |
| **Cytochrome c biogenesis** | ccmB,C,$F_N$,$F_C$ | + | + | +[c] | +[d] |
| Structural RNAs | | | | | |
| **Ribosomal RNAs** | rrn5,18,26 | +[b] | + | + | + |
| **Transfer RNAs** | eg. trnfM,P,S,E,I,K | 15 | 17 | 22 | 29 |
| Ribosomal protein genes | | | | | |
| **Small subunit (SSU)** | rps1 | + | + | - | + |
| | rps2 | + | + | - | + |
| | rps3 | + | + | + | + |
| | rps4 | + | + | + | + |
| | rps7 | + | + | + | + |
| | rps8 | - | - | - | + |
| | rps10 | - | - | - | + |
| | rps11 | - | ψ | - | + |
| | rps12 | + | + | + | + |
| | rps13 | + | + | - | + |
| | rps14 | - | ψ | ψ | + |
| | rps19 | ψ | + | ψ | + |
| **Large subunit (LSU)** | rpl2[e] | ψ | + | 5’ | + |
| | rpl5 | + | + | + | + |
| | rpl6 | - | - | - | + |
| | rpl16 | + | + | + | + |

[a] + = a subset of or all genes are present, - = no genes are present, and ψ = only found as pseudogene.
[b] In wheat there are two copies of *atp6*, *atp8* and 3 copies of *rrn5*, *18*, *26* (1 *rrn26* copy is a pseudocopy).
[c] In *Arabidopsis* there are two copies of *atp6* and *ccmF$_N$* is broken into two pieces, *ccb382* and *ccb203*.
[d] In *M. po. nad7* is a pseudogene, there is an extra pseudo-copy of cob and *ccmF$_C$* is split: *orf169* & *orf332*.
[e] *rpl2* was split into 2 genes in eudicots (eg. *A. th.*) and only the 5’ region is mitochondrial-encoded. The 3’ region is encoded in the nucleus.

*Figure 1.1: Transcription and RNA processing of plant mitochondrial genes*

Mitochondrial DNA and RNA sequences are shown as gray and blue lines respectively. Blue lines represent regions of the DNA that correspond to mRNA 5' and 3'UTRs. A hypothetical mtgene is shown (gray box) with internal C residues that will be converted to U's through RNA C-to-U editing (blow-up shows deamination reaction for conversion of C to U residue). Nuclear-encoded machinery for transcription initiation (half circles), RNA end-cleavage (arrowheads), and exonuclease activity (pacman) are boxed. Transcription start sites (bent arrow) and intron splicing (squiggly line) are also shown. A schematic of the complexes of the electron transport chain (blow-up in bottom left) on the inner mitochondrial membrane (IMM) is also included.


Nuclear encoded
proteins
Promoter
DNA
C C
C
Editing
Precursor transcript
U U
U
Turnover
AAA
RNA Processing
Mature mRNA
NADHex/in
2H+1/2O 2 H2O
intermembrane
space
ex
ex
in
in
aox
IMM
matrix
cyt c
I
II
Q
III
IV
V
NADH NAD + H
2H+1/2O 2 H2O
ADP+Pi
ATP

genes have been retained in plant mitochondrial genomes like cytochrome c biogenesis genes, ribosomal protein genes for both the large and small subunits and structural RNA genes: transfer RNAs and ribosomal RNAs (Table 1.1).

Because of ongoing gene transfer, plant mitochondrial gene content often varies among species. Ribosomal protein genes have been shown to be transferred most often to the nucleus (Adams *et al*. 2002a). A study using Southern hybridization of 280 angiosperms (flowering plants) revealed that many ribosomal protein genes were lost from the mitochondrion independently in a number of different plant lineages (Adams *et al*. 2002a). The *rps7* gene showed the greatest plasticity with up to 42 inferred independent transfer events. A list of mitochondrial-encoded ribosomal protein genes among several plants is included in Table 1.1. This type of comparative analysis allows us to determine the frequency of independent gene loss among lineages. In plants where the gene has been lost from the mitochondrial genome it is likely transferred to the nucleus.

Gene transfer is believed to occur during organelle lysis or 'bursts' and it is usually a cDNA copy that is translocated and incorporated into the nucleus (Adams and Palmer 2003). In order to be successful, the integrated cDNA gene copy must acquire proper expression signals (i.e. promoter) for nuclear transcription and translation in the cytosol as well as protein targeting signals to allow translocation back to the mitochondrion (Adams and Palmer 2003). Such signals can be acquired through duplication of signals from other genes or are derived from an unknown origin (Sandoval *et al*. 2004). In one seemingly unlikely example a transferred mitochondrial gene has been incorporated into the intron of a pre-existing nuclear gene (whose product is targeted to the mitochondrion) and is expressed via alternative splicing, a phenomenon called 'hitchhiking' (Figueroa *et al*. 1999; Kubo *et al*. 1999).

Once gene transfer from the mitochondrion to the nucleus occurs there is a 'transition state' where both compartments have a functional copy. The transfer is successful when the mitochondrial copy is deactivated and the nuclear copy becomes the sole functional copy. In wheat, *rpl5* is currently in a transition state (Sandoval *et al*. 2004) and in some legumes two functional copies of *cox2* are present. However in legumes a number of lineage-specific deactivations of *cox2* have been observed of either the

mitochondrial or nuclear copy, implying recent transfer (Adams *et al*. 1999). In other cases however the mitochondrial copy is not transferred but lost altogether. The protein product of another gene, either a native nuclear gene or a chloroplast copy that has been transferred to either the nucleus or the mitochondrion, can take over or assume the role of the lost mitochondrial gene. In *Arabidopsis*, the mitochondrial genes *rps8* and *rps13* have been replaced by the nuclear gene *rps15A* and the nuclear-encoded chloroplast-origin gene *rps13* respectively (Adams *et al*. 2002b).

Mitochondrial genomes vary in size significantly between kingdoms: plants: 187-2,900 kb (Alverson *et al*. 2011), fungi: 17-176 kb (Allgemeine *et al*. 2005), protists: 6-76 kb (Vaidya and Mather 2009), animals: 15-43 kb (Signorovitch *et al*. 2007). Extreme size differences seen in plants are due to a proliferation of both repetitive and non-repetitive sequences in spacer DNA and expansion of intron sequences. It is apparent than that these organellar genomes do not demonstrate the same genetic economy seen in animal mitochondria. The mitochondrial genome of wheat for example is ~30x larger than that of human. Members of the flowering plant family known as the Cucurbitaceae have the largest mitochondrial DNAs known among plants, with sizes of 1,800 kb for cucumber (*Cucumis sativus*) and 2,900 kb for muskmelon (*Cucumis melo*) (Alverson *et al*. 2011). The increased size of the cucumber mitochondrial genome is due primarily to the proliferation of dispersed repeats (including short 20 to 60 bp repetitive DNA motifs), expansions of existing introns and the acquisition of sequences from diverse sources, including nuclear and chloroplast genomes, viruses and bacteria (Bartoszewski *et al.* 2009; Alverson *et al*. 2011).

Genome sizes for plants including angiosperms (flowering plants) are often represented as ‘master chromosomes or master circles,” a presumptive circular molecule consisting of all of the DNA sequences present in substantial stochiometry in the mitochondrion (Kitazaki and Kubo 2010) but are almost never present in a true circular physical form, unlike for animal mitochondria. Master circles are constructed by chromosome walking which results in very complex structures due to repeated sequences. For example the size of 5 variant master circles in maize range from 536 to 740 kb, but after omitting the repeated sequences the complexity ranges from 507 to 537 kb (Kitazaki and Kubo 2010). Sometimes DNA species are excluded from the master circle; in the

case of muskmelon plasmid-like molecules of either linear or circular form, independent in their replication and mode of inheritance and substochiometric DNA (sublimons) which demonstrate different sequence from the major DNA molecules (Alverson *et al*. 2011).

## 1.2 *Impact of recombination on genome organization and gene expression*

The mitochondrial genomes of angiosperms generally contain fairly large repeat sequences (>1kb) as well as shorter repeat sequences (~100bp-1kb) which can frequently recombine intra- or inter-molecularly. Such events result in subgenomic molecules (across direct repeat sequences) and/or isomers (across indirect repeats) of the mitochondrial genome (Levy *et al*. 1991). In papaya (*Carica papaya*) as few as 36 repeat sequences larger than 50bp have been identified where in rice (*Oryza sativa*) more than 130 repeat sequences have been found (Marechal and Brisson 2009). In the wheat mitochondrial genome there are 16 identified repeat sequences (R1-16) ranging in sizes from 104 bp to 9,882 bp (Ogihara *et al*. 2005). Of the 16 repeats, 9 (R1-9) are involved in intra-molecular recombination. The remaining 7 (R10-16) are also involved in genome rearrangements and duplication events. As a result of many large active repeat sequences, several genes in wheat are in multi-copy. To contrast chloroplast genomes have fewer repeat sequences (often one large inverted repeat >20kb) and are not as prone to recombination as gene order is highly conserved from non-vascular plants such as *M. polymorpha* to flowering plants like wheat (Marechal and Brisson 2009). Microhomologous (6 to 31 bp) repeat sequences have also been found to undergo homologous recombination in plant mitochondria (Knoop 2004). Proliferation of shared upstream sequence motifs for gene expression, gene conversion and even correction of the cytoplasmic male-sterile phenotype known as CMS result from recombination across short repetitive microhomologous sequences (Schardl *et al.* 1985).

Movement of sequences through rearrangements like homologous recombination across both large and small repeat sequences has also resulted in very different transcription patterns in plant mitochondrial genomes. In a study of the rice transcriptome (Fujii *et al*. 2011) not only were transcripts seen for all known functional genes but RNAs resulting from transcription initiating in intergenic regions were also observed. For RNA

isolated from rice calli 48.5% of the mitochondrial genome was actively transcribed, compared to 32.0% seen in etiolated seedlings. This is a huge percentage of the genome considering only about 15% is represented by genic sequences. Differences in the percentage of genomic sequence that was transcribed were due almost exclusively to variability in transcription of spacer DNA as expression levels for previously annotated housekeeping genes were similar between tissue-types. Open reading frames in these spacer sequences were also transcribed however because they are not conserved among flowering plants they likely have no principal functions.

Recombination is also an important mechanism through which DNA repair and thus genome integrity is maintained in plant mitochondria. Deregulation of this process results in genomic instability leading to a variety of phenotypes, therefore particular families of proteins are essential in preventing aberrant recombination (Shedge *et al.* 2007). Plants have co-opted some recombination surveillance proteins of prokaryotic origin as well as evolved wholly novel factors. Homologues of the eubacterial recombinase RecA protein essential for homologous recombination have been found in *Arabidopsis* (Khazi *et al.* 2003). *RecA1* is targeted to the chloroplast, *RecA2* to the mitochondria and *RecA3* to both organelles. This demonstrates the use of organelle-specific nuclear-encoded targeting signals as well as targeting signals recognizable by both cytoplasmic compartments. Mutations in *RecA3* lead to large-scale rearrangements of mtDNA but are recovered by reintroducing a functional copy. The mitochondrial-targeted copy *RecA2* however is essential for organelle viability as insertion mutations were lethal (Shedge *et al.* 2007).

Recombination among repetitive DNAs in plant mitochondria sometimes produces rearrangements which result in phenotypic differences among individuals. In the case of *Arabidopsis thaliana* whose mitochondrial genome (367kb) has been completely sequenced (Unseld *et al.* 1997) 57 “true” genes (known or predictable functions), 28 of which are protein-coding, have been detected and about 156 ORFs of at least 100 codons of unknown function. Because these ORFs are not conserved in other species, it is unlikely that they are functional. This is true of the wild-type (WT) *Arabidopsis*, however variants in nature are sometimes found showing a cytoplasmic male-sterile phenotype (CMS) which is caused by rearranged and transcriptionally active

mitochondrial ORFs. Several plant ecotypes (grown under different environmental conditions) and *in vitro* derived alloplasmic lines (different nuclear/cytoplasmic combinations derived from interspecific crosses) also demonstrate this male-sterile phenotype (reviewed in Pelletier and Budar 2007). The diagnostic feature of CMS is often rearranged mitochondrial-encoded ORFs which leads to an incompatibility between nuclear-mitochondrial interactions. Closer inspection of such ORFs in CMS lines among many different kinds of flowering plants reveals chimeric DNA sequences (Budar and Pelletier 2001). In the case of "polima" CMS in *Brassica napus*, the *orf224* gene contains the upstream expression signals and 5' coding sequence (CDS) of *atp8*, a piece of *rps3* coding sequence and an extended 3'ORF with no known sequence identity (L'Homme and Brown 1993). Interestingly these CMS-causing ORFs are also almost always co-transcribed with 'true' mitochondrial genes and in this case *orf224* is co-transcribed with *trnfM* and *atp6*.

With the accumulation of more and more sequencing data for plant mitochondrial genomes it has become apparent that while the nucleotide substitution rate is very slow, rearrangements occur frequently which leads to variable genome organization even between close relatives (Palmer and Herbon 1988). By comparison plant mitochondrial genomes evolve at slower rates than the genomes of plant chloroplasts which are themselves more slowly evolving than plant nuclear genomes. And it is the case that all three of these cellular compartments' genomes evolve more slowly than animal mitochondrial DNA sequences (Wolfe *et al*. 1987).

Mechanisms for gene expression between animal and plant mitochondria are also very different. Animal mitochondria use two phage-type promoters, one driving expression of the light strand (C-rich) which encodes just 9 genes and the other the heavy strand (G-rich) which encodes 28 genes. Long polycistronic RNAs for both strands are generated from two promoters in the major non-coding region and are then processed, yielding monocistronic transcripts for the 2 rRNAs, 22 tRNAs and 13 protein-coding mRNAs (Lightowlers and Chrzanowska-Lightowlers 2008). In plants, transcription initiates at many different sites throughout the genome and genes are both tightly packed and distal from one another. Therefore transcripts for the ~55 genes in flowering plants are either monocistronic or polycistronic. In *Brassica napus* ribosomal protein genes *rps3,*

*rpl16*, *rpl5*, and *rps14* retain the ancestral bacterial genomic organization and are transcribed as one large ~10kb co-transcript (Ye *et al.* 1993). Often polycistronic RNAs are however processed into mature monocistronic RNAs by endonucleases (reviewed in Gagliardi and Binder 2007).

## 1.3 *Signals for plant mitochondrial transcription initiation*

Because plant mitochondrial DNA is highly recombinogenic, genome organization can be dramatically different even between closely related species. For example the upstream genomic environment for *rps1* differs among closely related legumes; pea, soybean and bean (Hazle and Bonen, 2007a). Each legume must therefore use different promoters for *rps1* expression, resulting in different RNA profiles among plants. Such sequence diversity can only be achieved through multiple lineage-specific DNA rearrangements.

Sometimes homologous regulatory sequences are used for gene expression for several different genes in the same genome (Hazle and Bonen, 2007b). For example wheat *atp4*, *cox2*, and *atp6-2* all possess a very similar upstream sequence, referred to as an expression cassette. This expression cassette contains consensus sequences for transcription initiation and therefore may be very important for expression of these genes in wheat. Signals for RNA stability to protect against 5'UTR degradation or those for translational control to aid in ribosome recognition of mRNAs may also be present within this shared sequence block. The presence of similar upstream signals for gene expression of multiple mitochondrial genes in the same genome is suggestive of some regulatory signal recruitment mechanism.

Expression of a single gene in plant mitochondria can also be controlled by multiple promoter sequences (Kuhn *et al*. 2005). The ATP synthase gene *atp9* in *Arabidopsis* possesses four different promoter motifs which are each responsible for initiating transcription (Kuhn *et al.* 2005). In total, 9 of the 12 genes studied showed transcription initiating from a minimum of two promoter sequences. Some consensus sequences were observed however more than half of the promoters identified did not map to any known motif. No tissue-specific difference in promoter use between leaf and flower was seen either. In another study which characterized all mitochondrial RNA ends

in *Arabidopsis thaliana* tissue-culture only two genes shared consensus motifs of identical sequence and one promoter upstream of *cox3* was shown to be ecotype specific (Forner *et al*. 2007). They caution also that studies carried out in tissue culture may not reflect activities *in planta*.

Signals for transcription initiation of plant mitochondrial genes are different than those used for expression of bacterial gene operons. In bacteria, transcription initiation is signaled by RNA polymerase recognition of a consensus sequence -35 (TTGACA) and -10 (TATAAT) base pairs downstream of the promoter. Plant mitochondria however do not retain these ancestral consensus sequences and instead have adapted an array of sequence motifs that differ even between monocot and eudicots. Plant mitochondria are not 5' capped upon maturation like cytosol RNAs and therefore if unprocessed at their 5' end retain a 5' tri-phosphate group, diagnostic of a de novo primary transcript. Mapping the ends of primary (unprocessed) transcripts which possess either 5' tri- or di-phosphates can therefore determine promoter location. Consensus sequences for transcription initiation like CRTA (where R represents either adenosine or guanosine) can be found upstream of several genes in different monocots (wheat and maize) and some dicots (*Arabidopsis*) and in several cases function as a promoter (Kuhn *et al.* 2005). Among eudicots this tetra-nucleotide consensus sequence can be extended to a nanonucleotide motif CRTAAGAGA (Gagliardi and Binder 2007). In the analysis of *Arabidopsis* RNA ends, variations on this motif including CRTATATAA and CRTATATAG were seen (Forner *et al.* 2007). Other genes however show initiation from promoter sequences that do not correspond to this motif. The abundance of promoters and variation in promoter sequences suggest a relaxed specificity in promoter recognition for plant mitochondrial gene expression.

Despite having different motifs for transcription initiation throughout the genome, plant mitochondrial gene expression is achieved by a nuclear-encoded single subunit bacteriophage-like T3/T7 enzyme, the RNA polymerase RpoTm (reviewed in Ikeda and Gray 1999). In contrast, expression of chloroplast genes is achieved by two different RNA polymerases; a nuclear-encoded eubacterial-type RNA polymerase (Hedtke *et al*. 1997) and a bacteriophage RNA polymerase, RpoTp (which originated from the duplication of RpoTm). In *Arabidopsis* not only are both the RpoTm and RpoTp RNA

polymerases present but another RNA polymerase called RpoTmp, which is dually targeted to the mitochondrion and chloroplast and has diverged more recently from RpoTm than RpoTp (Ikeda and Gray 1999; Hedtke *et al*. 1997; Hedtke *et al*. 2000). This demonstrates that while no sequences have been shown to be successfully integrated into the chloroplast from the mitochondrion the two compartments in fact share some nuclear-encoded machinery to carry out regular housekeeping functions.

## 1.4 *RNA processing of plant mitochondrial transcripts*

For plant mitochondrial gene expression, post-transcriptional modifications such as splicing, C-to-U editing and end-cleavage (RNA processing) are required for transcript maturation. About one third of the protein coding genes contain introns and therefore require splicing in the conversion of precursors to mature messenger RNAs. Virtually all introns in plant mitochondria fall in the group II category and a subset are discontinuous in the genome so that expression requires trans-splicing (reviewed in Bonen and Vogel 2001). Editing is an early event and converts cytidine residues to uridines through a deamination reaction (Figure 1.1) at multiple sites in virtually all protein-coding sequences. Editing usually increases sequence similarity with homologous sequences from other organisms (Shikanai 2006). End-cleavage or 'trimming' of RNA 5' and /or 3' ends removes upstream and downstream regulatory sequences important for gene expression but not for translation initiation. RNA processing steps like splicing and editing are necessary for proper mRNA formation. If translation of productive proteins is efficient there must be some kind of temporal regulation between gene expression and mRNA translation in plant mitochondria. It is difficult however to design models for such regulation as all processes are carried out within the same compartment, much like in bacteria.

Due to the variety of sequence motifs found at transcript ends and prevalence of RNA editing in plant mitochondria, it is assumed that a variety of enzymatic and non-enzymatic proteins as well as motifs play a role in transcript maturation. Virtually all RNA processing machinery is nuclear-encoded and because the vast majority of specific proteins or protein families that carry out these functions have not yet been isolated, universal consensus sequences (*cis*-elements) recognized by such *trans*-factors have also

yet to be determined. In mammalian systems long polycistronic transcripts are processed by RNase P and RNase Z-like endonucleases which recognize and cut tRNA transcript termini (Lightowlers and Chrzanowska-Lightowlers 2008). Bacterial genes are organized into operons and therefore also transcribed as polycistronic transcripts; however transcription and translation occur simultaneously, with no intermediate processing events like editing or end-cleavage. Exonuclease degradation of bacterial 5'UTRs however does occur and is found to be concomitant with RNA translation. Currently, one very large family of nuclear-encoded proteins, the pentatricopeptide repeat (PPR) proteins has been found to be involved in mitochondrial RNA processing. Associated with chloroplast RNA processing (Yamazaki *et al.* 2004), this family of proteins has also been implicated in splicing (Falcon de Longevialle *et al.* 2007) and editing (Zehrmann *et al*. 2009) in the mitochondria of *Arabidopsis* as well as mitochondrial RNA cleavage in various other plant species (Saha *et al.* 2007).

### 1.4.1 *RNA secondary structure, cis-elements for end-cleavage, transcript stability and RNA-protein interactions*

Maturation of plant mitochondrial transcript termini through end-cleavage is necessary for production of mature mRNAs for many mitochondrial genes. For some of these transcripts secondary structures formed from RNA folding are thought to serve as recognition sites for nuclear-encoded proteins involved in end-cleavage or transcript stability. Degenerate tRNAs (i.e. tRNA-like structures), referred to as t-elements for example are thought to be recognized by endonucleases and help generate transcript termini (Forner *et al. 2007*). Examples include RNase Z which recognizes tRNA 3' termini to generate the 5' transcript termini of downstream genes, and RNase P which cleaves tRNA at their 5' ends generating the 3'UTRs of upstream genes (Forner *et al.* 2007). Endonuclease activity has been proposed for the generation of transcript ends for *rpl5, nad7* and *atp6-2* in *Arabidopsis* (Forner *et al.* 2007) and $ccmF_N$ in wheat (Calixte and Bonen 2008) since immediately upstream of these genes, structures referred to as stem-loops or double stem-loops (similar to tRNA acceptor-stems) have been discovered using computational RNA-folding programs. In *Arabidopsis thaliana* the nuclear-encoded RPF1 is required for efficient generation of a 5'end 228 nt upstream of the

mitochondrial *nad4* gene (Holzle *et al*. 2011). RPF1 belongs to a special subclass of PPR proteins which include the RESTORER OF FERTILITY (*RF*) gene products, which reverse cytoplasmic male sterility (CMS) in various plant species. RNA processing factor (RPF) 3 (of the *RF* subclass) in *Arabidopsis* leads to a severe reduction in CcmC proteins (Jonietz *et al*. 2011). Two additional RPF proteins, RPF 1 and 2 are also involved in 5'procesing of different *Arabidopsis* transcripts, demonstrating an emerging role for *RF* genes in post-transcriptional maturation of mitochondrial RNAs. Additional enzymes have been proposed for the generation of transcript ends: The exonuclease RNaseII, which is dual targeted to the mitochondrion and chloroplast in *Arabidopsis* (Perrin *et al*. 2004) and eukaryotic endonucleases containing RNaseIII domains, a prokaryotic enzyme which cleaves double stranded RNA (Susi *et al*. 2004.).

Transcript 5'ends however do not always require end-cleavage for mRNA production. For genes like *cox2* in wheat, transcripts do not require 5' processing (Covello and Gray 1991). Instead the *de novo* primary transcript also functions as the mature messenger RNA and is translated in the mitochondria still with its 5' triphosphate. Both primary transcripts and processed RNA species with discrete homogeneous 5' and/or 3' termini are spared from exonucleolytic degradation at their ends. In plant mitochondrial transcripts, stability conferring secondary structures and/or interactions between RNA *cis*-elements and auxiliary proteins within UTRs is believed to be the mechanism for end protection, as is seen for 'higher' plant chloroplast mRNAs (reviewed in Herrin and Nickelson 2004). In the chloroplast of *Chlamydomonas reinhardtii*, a green alga, transcript stability is achieved through interaction of proteins with *cis*-elements within the 5' and 3'UTRs while it is stem-loop structures at the 3' end that block attack from exonucleases. In both *Arabidopsis* (Forner *et al*. 2007) and wheat (our unpublished data) plant mitochondrial 3' ends are very homogeneous, often exhibiting a single discrete terminus in contrast to heterogeneous 5' ends generated from multiple transcription initiation signals and/or *cis*-elements for end-cleavage. Transcription is thought to simply continue until the polymerase falls off the DNA template. Because 3'ends appear more discrete, processing (either end-cleavage or exo-activity) of the 3'UTR is believed to occur very early during gene expression (reviewed in Gagliardi and Binder 2007).

Non-productive transcripts in plant mitochondria however are usually unstable and as such are often tagged for degradation by the incorporation of tracts of non-encoded nucleotides (often A's) at their 3' ends (Holec *et al*. 2006; Kuhn *et al*. 2001), the same mechanism used for RNA turnover in bacteria (Hoffmann *et al.* 2001) and in plant chloroplasts (Herrin and Nickelson 2004). Addition of a polyA tract at transcript 3'UTR termini (mRNA and/or pre-mRNA) has been found to occur in sunflower *atp1-orf522* as well as in maize *cox2* (Lupold *et al.* 1999), and just like for other plant mitochondrial RNAs, signals transcript degradation by ribonucleases (Gagliardi and Leaver 1999). Although signals for polyadenylation are still unknown, polynucleotide tails of various lengths can occur (1 to 25 is common and on more rare occasions 50 to 100 nt) and at times incorporate nucleotides other than A's (Forner *et al*. 2007), reminiscent of the nucleotidyltransferase activity seen during maturation of tRNA acceptor arms (addition of -CCA motif).

Proteins involved in the actual degradation of RNA molecules in eukaryotes and archaea include the exosome and in bacteria a simpler protein complex, the degradosome. Polynucleotide phosphorylase or PNPase is a functionally similar enzyme common to plants, animals and bacteria (Schilders *et al*. 2006). Not only does this enzyme have phosphorolytic 3' to 5' exonuclease activity but a 3'-terminal nucleotide polymerase activity. It is involved in the processing and degradation of cytosol mRNA and using *Arabidopsis* knock-out mutants has been linked to turnover of plant mitochondrial RNA (Perrin *et al*. 2004a, b; Holec *et al*. 2006). Polyadenylation has also been shown to occur at the ends of truncated mRNAs (Schuster *et al*. 1999); this means that after cuts are made by endonucleases or exo-activity at 3'ends, non-encoded nucleotides are then added, signaling RNA turnover.

### 1.4.2 *Plant mitochondrial C-to-U type RNA editing*

RNA editing in plant mitochondria targets 400-500 cytidine nucleotides, mostly in mRNA coding sequences to be altered to uridine nucleotides. Within coding sequence, C-to-U type RNA editing of plant mitochondrial genes occurs most often in the first or second codon position and therefore usually causes non-synonymous amino acid substitutions. The altered amino acid usually increases sequence identity with

homologues in other organisms (Shikanai 2006). How specific sites are recognized and edited is not well understood as no universal motifs have been found within coding sequences or for those few sites within UTRs and/or intron sequences. Editing factors such as PPR proteins are however thought to primarily recognize upstream sequence elements which sometimes are shared among affected RNA molecules allowing a single PPR protein to be involved in the editing of multiple C residues (Zehrmann *et al*. 2011). Studies in *Arabidopsis thaliana* support the involvement of a nuclear-encoded PPR protein in the editing of several mitochondrial transcripts (Zehrmann *et al.* 2009). The protein MEF1 is involved in editing of three specific sites in three different mitochondrial mRNAs (*rps4*, *nad7*, and *nad2*). Editing activity at such sites is lowered in *Arabidopsis* MEF1 mutants, suggesting the protein confers site-specificity for recognition by editing machinery and is not the only factor involved. More recently another PPR protein in *Arabidopsis*, MEF11 has been shown to edit position 422 of *cox3*, 124 of *nad4* and 344 of *ccb203* (Zehrmann *et al.* 2011). For both *mef1* and *mef11* mutants, loss of editing occurs only at two sites while editing at the third site continues but to a lesser extent. Therefore while some sites require specific PPR proteins to be edited, others potentially containing similar or identical consensus motifs can at least partially undergo editing by alternative proteins within the family. Such minimal specificity allows for a network of PPR *trans*-factors which bind RNA sequence elements to ensure that editing occurs at all sites required and possibly as a side effect at sites not harmful when accidentally edited (Zehrmann *et al.* 2011).

Differences in the degree of editing of mitochondrial protein-coding genes can be seen between members of different groups (ie. *nad* vs, *rp* genes). On average ribosomal protein genes have fewer editing sites within exon sequences than do other protein coding genes like the NADH dehydrogenase subunit genes (Giege and Brennicke 1999). For example *rpl2* (1047nt) in *Arabidopsis* is edited at only one site, while *nad1* (975nt) is edited at 24. Differences have also been observed for the same genes in different plants.

RNA editing is considered an early event in RNA processing as precursor RNAs are also found to be fully or partially edited (Gualberto *et al*. 1991). Editing near intron/exon junctions however has been found to occur later than editing at more distant sites (Li-Pook-Than *et al.* 2007). It is hypothesized that intron secondary structure

sterically impedes the binding of RNA editing machinery, making it harder for editing to occur and/or that splicing may be necessary for generation of the *cis*-element(s) recognized for binding of the editing machinery. Interestingly, in certain plants, mitochondrial RNA editing to some extent is developmentally regulated: In maize *nad3*, editing increased from 50 to 75% in 3 to 7 day seedlings (Grosskopf and Mulligan 1996). More recently, differences in the degree of editing of sites within coding sequence has also been reported under stress conditions such as when plants are grown in the cold. In intron-containing precursors for wheat *cox2* decreases of up to 60% were seen at some sites (Kurihara-Yonemoto and Handa 2001) while in intron-containing genes in rice not only did the degree of editing decrease but increases in the relative abundance of intron-containing precursors were seen as well (Kurihara-Yonemoto and Kubo 2010). Comparatively the spliced mRNAs in rice exhibited virtually 100% editing.

### 1.4.3 *Machinery involved in plant mitochondrial RNA processing*

Preliminary work characterizing the mitochondrial proteome of *Arabidopsis thaliana* identified roughly 500 proteins. Targeting prediction software predicts 1500 proteins are translocated to the organelle (Klodmann *et al.* 2011). Mass spectrometry revealed upwards of 35 different protein complexes within plant mitochondria, several of which were plant-specific PPR protein complexes. 27 PPR proteins were discovered however several of which form part of protein "super complexes" as they ran higher than explainable by their monomeric molecular mass. Two previously unidentified protease complexes were also detected. One of these complexes contained 14 Clp proteases, a family of nuclear and chloroplast-encoded hydrolytic enzymes involved in protein maturation and proteolytic degradation.

In *A. thaliana* 450 genes for PPR proteins are encoded in the nuclear genome with the majority of them predicted to target organelles and bind to RNA (Shikanai 2006). Interestingly in animals 6 nuclear-encoded proteins were identified as potential candidates (Lightowlers and Chrzanowska-Lightowlers 2008). Of these proteins only one has RNA-binding activity; POLRMT (mitochondrial RNA polymerase), another is involved in ribosome assembly; MRPS27 and the remaining 3 analyzed; PTCD1 (pentatricopeptide repeat domain protein 1), PTCD3 and LRPPRC (leucine-rich

pentatricopeptide repeat cassette) are involved in assembly of respiratory chain complexes. The sixth protein candidate PTCD2 appears to also have a close association with the respiratory chain. As it turns out this large family of PPR proteins is almost exclusive to plants, with only 20 in protists and a handful in other non-plant eukaryotes (Lightowlers and Chrzanowska-Lightowlers 2008). In plants as in animals the PPR protein genes are characterized by a structural motif of a degenerate 35-amino-acid sequence which appears as tandem repeats. All PPR proteins contain at their C-terminus an E domain with a subset being further extended by an additional domain which terminates in a DYW triplet (Shikanai 2006). About 140 of the 450 PPR proteins encoded in *A. thaliana* contain C-terminal DYW domains and because these domains contain signature amino acids characteristic of Zn-containing cytidine deaminases they are believed to be exclusively involved in C-to-U RNA editing (Zehrmann *et al.* 2011). Furthermore because all PPR proteins have RNA-binding activity and demonstrate sequence specificity and because RNA-editing is common in flowering-plant mitochondria (400-500 sites) and to a lesser extent in plastids (35-40 sites) but almost never in mammalian mitochondria it is believed that the majority of these proteins, in plants, are primarily involved in mitochondrial transcript C-to-U type RNA editing (Shikanai 2006). Certainly many examples support this (Hammani *et al*. 2011; reviewed in Schmitz-Linneweber and Small 2008) however other PPR proteins have been implicated in other levels of RNA processing (mentioned previously).

### 1.4.4 *Developmental differences in RNA-level events during transcript maturation*

During plant seedling development, the expression of some mitochondrial ribosomal protein genes specifically, appears to be differentially regulated (Li-Pook-Than *et al.* 2004). When compared to the respiratory chain genes (eg. *cox1*, *cox2*, and *atp6*), which show similar relative steady state levels of mRNAs (compared to the 18S and 26S rRNA) over the course of plant development, several ribosomal protein genes like *rps7* show a decrease in relative mRNA abundance, in seedlings compared to germinating embryos. In another example, relative levels of *rps1* mRNAs in wheat mitochondria were found to be higher in not only embryos but also in dormant seeds (Calixte and Bonen, 2008) when compared to developing seedlings. Higher steady-state levels of *rps1*

mRNAs in dormant seeds may simply represent stored messengers that may or may not be translatable as many individual species were found to be truncated and had no translation start codon (Calixte and Bonen 2008). Transcription and translation of ribosomal proteins for the purpose of ribosome synthesis is very important throughout development. During embryonic germination, a period of rapid mitochondrial biogenesis and high oxygen consumption, transcription and RNA processing are likely not as tightly coupled as they are during seedling development and therefore we see higher relative levels of intron-containing precursor RNAs (Li-Pook-Than *et al.* 2004). More efficient RNA processing would explain why steady-state levels of precursor RNAs are relatively lower in seedling stages for all protein coding genes but does not account for the drop in relative abundance of ribosomal protein gene mRNAs specifically (Li-Pook-Than *et al.* 2004). Fewer ribosomal protein mRNAs must result either from an increase in RNA turnover and/or decrease in the level of transcription (and therefore production of mature mRNA).

### 1.4.5 *The ribosomal protein gene rps7 as a model for RNA processing studies in plant mitochondria*

Plant mitochondrial genomes have the distinction of being highly recombinogenic. To this end we are interested in the impact of DNA rearrangements on gene expression over evolutionary time. Transcriptional units for the same gene can vary between different plants, even closely related species. Because duplication followed by homologous recombination is responsible for the shuffling of genetic material it is often a mechanism through which sequence motifs important for gene expression are incorporated upstream of several different genes.

The mitochondrial genomes of agronomically important monocots wheat (*Triticum aestivum*), rice (*Oryza sativa*) and maize (*Zea mays*) have been completely sequenced; National Center for Biotechnology Information (NCBI) accession numbers are given in Table 1.2. In order to study the impact of DNA rearrangements on mitochondrial gene expression in these plants I chose a gene that was distal from any upstream or downstream coding sequences for functional genes and therefore likely independently transcribed. Lineage-specific rearrangements close to the *rps7* coding

sequence would therefore impact expression of only this gene. For example, functional genes are located ~2.9 kb and ~6.3 kb upstream of *rps7* in wheat and rice respectively (Figure 1.2), unlike the non-vascular plants, *P. patens* (moss) and *M. polymorpha* (liverwort) which have retained the ancestral *rps12-rps7* linkage.

The product of this gene, ribosomal protein S7 is universally present in the small subunit of prokaryotic and eukaryotic ribosomes (Lecompte *et al*. 2002). It functions as a primary rRNA binding protein, important during assembly of the ribosome and also helps form the mRNA exit channel at the interface of the large and small subunits (Yusupov *et al*. 2001).

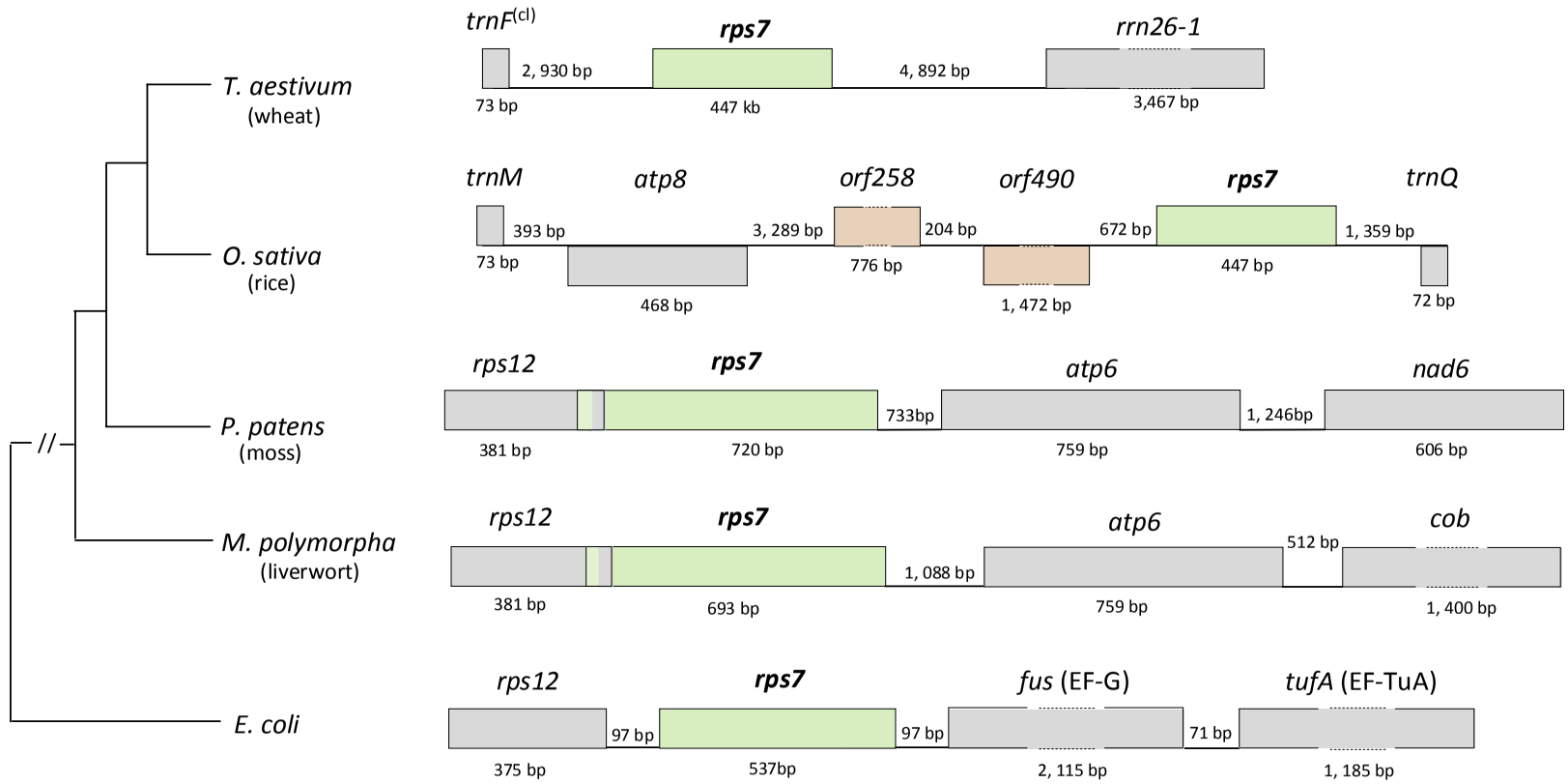
Furthermore from previous work (Li-Pook-Than *et al*. 2004) ribosomal protein gene expression in general was observed to be alternatively regulated during embryo-to-seedling development in wheat. This is in contrast to the expression of respiratory chain genes which showed no variation in mRNA levels (standardized relative to mitochondrial rRNA levels). Because relative levels of mitochondrial ribosomal protein mRNAs appear to differ during plant development, this suggests that gene expression (promoter use) and/or transcript turnover/stability is alternatively regulated during these times. Therefore analysis of *rps7* transcripts from RNA isolated from different stages of seed development might show differences in RNA processing. Two developmental stages were used; germinating embryos (24-36 hours) and developing seedlings (6 to 10 days) since differences in the relative abundance (compared to 18S rRNA) of *rps7* mRNA was observed during these times (Li-Pook-Than *et al*. 2004).

To account for differences in steady state levels of *rps7* mRNAs between stages of development RNA processing must be regulated differently during these times. End-cleavage and RNA editing could be regulated differently throughout development. In a previous study of wheat *rps7* (Zhuo and Bonen 1993) editing was observed at 2 non-silent sites within *rps7* coding sequences and an additional 2 sites within 3'UTRs.

Using close relatives from the family of grasses the Poaceae; rye (*Secale cereale*), barley (*Hordeum vulgare*), brome grass (*Bromus inermis*), annual ryegrass (*Lolium multiflorum*) as well as those previously mentioned; wheat and rice, I conducted a comparative analysis of *rps7* among plants to assess shared upstream signals for transcription initiation as well as evaluate differences in RNA processing for the same

*Figure 1.2: Mitochondrial genome organization of rps7 among angiosperms wheat (Triticum aestivum) and rice (Oryza sativa), non-vascular plants liverwort (Marchantia polymorpha) and moss (Physcomitrella patens) and the bacterium Escherichia coli*

Protein coding sequences, structural RNA genes and open reading frames (*orfs*) are shown by coloured boxes (drawn to scale unless otherwise indicated by dotted lines). Spacer sequences are not to scale. Position relative to black line denotes gene orientation. The cladogram on the left (not to scale) shows the evolutionary relationship of the above species (Knoop *et al.* 2008). Sizes of gene coding sequences, structural RNA genes and length of spacer DNA are indicated in base pairs (bp).




[Streptomycin operon *(str)*]

Scale for genes only: 100bp

gene among closely related species. Divergence times for the grasses studied range from 5-7 million years ago (Mya) for rye and wheat, 15 Mya for wheat and barley, 25 Mya for wheat and *Lolium*, 45-50 Mya for wheat and rice and 60 Mya for wheat and maize (Kellogg and Bennetzen 2004).

### 1.4.6 *Research Goals: Analysis of grass rps7*

1. Are common upstream sequence elements shared among grass mitochondrial *rps7* or have multiple lineage-specific rearrangements swapped ancestral flanking sequences for novel regions that also provide signals for regulation of both transcription and translation? If so what roles might these *cis*-elements play in *rps7* expression and mRNA maturation?

2. What cleavage sites are recognized for *rps7* transcript termini maturation among grasses? Such events are needed for end maturation and the production of mature mRNAs in all stages of seed development. However it may be the case that some *rps7* processing sites are only recognized and cleaved during particular stages of development. A survey of *rps7* transcript profiles in germinating embryos and developing seedlings will help answer this question.

3. What function does C-to-U type editing of *rps7* UTR's serve and might the frequency of *rps7* editing increase or decrease during development? Because non-coding edits are rare it will be interesting to analyze the editing status of *rps7* flanking sequences across a number of grasses and determine if UTR editing helps increase sequence identity among plants or has some structural function.

4. How is *rps7* editing affected by cold-stress? Might there be a difference in the degree of editing of non-silent sites in *rps7* precursors compared to mRNAs under cold-stress conditions? Might non-coding edits like those in *rps7* 3'UTRs also be affected by growth in the cold?

## Chapter 2 Materials and Methods

### 2.1 Plant material

Wheat (*Triticum* aestivum var. Frederick), barley (*Hordeum vulgare* var. OAC Kippen) and oat (*Avena sativa* var. AC Goslin) seeds were kindly provided by Dr. R. Pandeya and Dr. T. M. Choo (Agriculture and Agri-food Canada). Rice (*Oryza sativa* var. Drew), rye (*Secale cereale* var. Gazelle), brome grass (*Bromus inermis*) and annual ryegrass (*Lolium multiflorum*) seeds were commercially purchased. Maize (*Zea mays* var. D39) seeds were provided by Direct Seeds Inc. (Chatham , ON. Canada).

### 2.2 Mitochondrial RNA and DNA isolation

50g of seeds were surface-sterilized in a 1:6 dilution of Javex in distilled water and in 10mN HCl then rinsed in autoclaved water. Whole seeds were either imbibed in vermiculite in an autoclaved tray for seedling growth or dissected to remove embryos for growth on petri dishes over saturated filter papers for embryo germination. Both seedlings and embryos were placed in the dark (etiolated) at room temperature for either 6, 7, 9 or 10 days or 24 or 36 hours respectively. Etiolated wheat seeds were also planted at room temperature for 3 or 4.5 days and then moved into a refrigerated cabinet at 4$^{o}$C for an additional 6 or 4.5 days respectively.

Mitochondrial RNA and DNA were isolated using procedures previously described (Subramanian *et al*. 2001). Working quickly and on ice plant material was placed under a cold mortar and pestle and homogenized with buffer I (0.44M sucrose, 50mM Tris pH 8.0, 3mM EDTA, 1mM β-mercaptoethanol, 0.1% BSA). Homogenate was filtered through cheesecloth and Miracloth (Calbiochem) and crude mitochondria were isolated through differential centrifugation (twice for 5 min. at 1950 rpm and once for 25 min. at 9100 rpm).

For DNA isolation crude mitochondrial pellets were resuspended in Buffer II (50mM Tris pH 8.0, 20mM EDTA) and mitochondria were lysed using Buffer III (20mM Tris pH8.0, 100μM EDTA, 200mM NaCl, 2% SDS, 200 mM β-mercaptoethanol) at 65$^{o}$C for 20 minutes. DNA was then precipitated with KOAc, isopropanol with $NH_4OAc$, ethanol and isopropanol with NaOAc. After collection of the pellet by centrifugation and

vaccum drying, mtDNA was resuspended in TE buffer (10mM Tris pH 7.5, 1mM EDTA) and stored at -20$^{o}$C.

For RNA isolation crude mitochondrial pellets were resuspended in buffer IV (10mM Tris pH 7.5, 50mM KCl, 10mM $MgCl_2$) and with Buffer IV with 8% Triton-X 100. Mitochondria were lysed in a 2X detergent mix (0.17g Sarkosyl, 1.2g Sodium P-aminosalicylate, 0.06g NaCl, 200µl 1M Tris pH 7.4, 9.6ml $dH_2O$) over ice for 5 min. and followed by 2 phenol extractions (1.5 vol. of phenol saturated in TE). Precipitation of nucleic acids was done using 0.1 vol. 5M NaCl and 2 vol. 95% EtOH. Sarkosyl is used here instead of 0.2g of Tri-isopopylnaphtalene sulfanate.

Yields for both DNA and RNA isolations were 1µg and 10µg per gram of wet weight tissue respectively.

NB. During isolation of crude mitochondrial RNA one alteration to the standard procedure was made. In the case of oat RNA additional low speed spins (1950 rpm) were done to remove cellular debris and other contaminants to further purify the extract. Additional phenol extractions were also carried out to remove contaminating polysaccharides and other particulates from the aqueous phase to further purify RNA (primarily for oat RNA extraction as well).

## 2.3 Mitochondrial RNA analysis

### 2.3.1 *RNA blot preparation and northern hybridization*

For northern blot analysis, mitochondrial RNA samples (approximately 5 µg per lane) were electrophoresed until bromophenol blue dye ran 6 cm, about 4hrs at 60V (~24mA) on a 1.2% agarose/formaldehyde gel with a 0.5-10 kb RNA size marker ladder (3µg, Invitrogen). After RNA was transferred to a nylon membrane overnight by capillary action, air-dried and UV cross-linked for 1min. it was hybridized with $^{32}$P-5'-end-labelled oligomer probes using standard procedures (Sambrook *et al*. 1989).

Oligomer probes were prepared using 100 ng of 20-22 nt oligomers incubated with 40 µCi $\gamma$-$^{32}$P-ATP (3000 Ci/mmol, Amersham), 5 units of T4 polynucleotide kinase (Invitrogen) and 1X kinase buffer (50mM Tris pH 9.5, 10mM $MgCl_2$, 5mM DTT) for 45 minutes at 37$^{o}$C. 37.5µl of TE was added to stop the reaction and probes were added to a

Sephadex G-50 column (equilibrated with TE) to remove unincorporated radioactive label. Three 50µl eluants of TE were collected from the column and the 2nd eluant was used for hybridization experiments (probe has not completely passed through the column in 1st eluant and free label comes out in 3rd eluant). Blots were hybridized overnight in 1-2ml hybridization buffer (deionized formamide, 20X SSC, 10% SDS, 10mg/ml yeast tRNA) at 42°C.

Membranes were then washed twice in 20ml of 20X SSC and 0.1% SDS for 20min. at the same temperature as that used for hybridization in order to remove free-label and non-hybridized probes. Exposure for ½ hr for hybridizations using *rrn18*-specific probes (LB211) and either 1 day or 4 days when using *rps7*-specific probes (LB29, LB37, LB524, LB578) was done on phosphoimaging screens (Bio-rad) at room temperature. Phosphoimaging screens were scanned in the phosphoimager scanner (Bio-rad Molecular imager FX) at a resolution of 50µm. Oligomer sequences are given in Table 2.1 and primer maps are provided in Figure 2.1.

NB. Before hybridization if blots had not been used previously they were pre-hybridized overnight bathing in an appropriate hybridization buffer in a shaking water bath at 42°C as opposed to pre-hybridization under the same conditions but in falcon tubes in large glass *Robbins* tubes in the hybridization ovens.

### 2.3.2 *RNA ligation, reverse-transcription and polymerase chain reaction of mtRNA (CR-RT-PCR)*

To simultaneously map the 5' and 3' termini of *rps7* mRNAs and precursor RNAs, the CR-RT-PCR strategy (cf. Kuhn and Binder 2002; Calixte and Bonen 2008) was used. To circularize transcripts, approximately 5-10 µg of mitochondrial RNA was incubated overnight at 14°C or for ½hr at 37°C with 18 units of T4 RNA ligase 1 (10units/µl, New England Biolabs) in the presence of 10mM ATP and 10X RNA ligase buffer (New England Biolabs), 5 µg of BSA and 50 units of the RNase inhibitor RNAsin (40units/µl, Promega). After two, 1:1 phenol extractions (first with buffer saturated phenol then Phenol:Chloroform:Isoamyl alcohol, 25:24:1) and ethanol precipitation in 2.5 vol. of 95% EtOH and 0.1 vol. of 5M NaCl, the self-ligated RNA was heated at 65°C for 5 min

*Table 2.1: Oligomers used in this study*

Oligomer name, sequence, orientation (direction in Figure 2.1) and gene region for *rps7* among grasses. Plant-specific primers are indicated.

| Plant | Name | Seqeunce 5'-3' | S/A[a] | Gene Region |
|---|---|---|---|---|
| Wheat[b] | | | | |
| | LB27 | ATCAATTTATCGGCCTCGTC | S | *rps7* mRNA 5'UTR |
| | LB28 | GCAGGCCTCTTGTGGATTCC | A | *rps7* 3' flanking |
| | Lb29 | ACTGAATGAGGAAGAGCTCC | A | *rps7* transcript 3'UTR |
| | Lb37 | GTTCAGTTCGAGCTAGGCGGTG | A | *rps7 5'* coding sequence |
| | LB211[c] | GTGATCATTGGTCCGATGCT | A | upstream *rrn18* |
| | Lb521 | TCACGTTACATGCTAAATCAGGC | S | *rps7* precursor 5'UTR |
| | Lb522 | AACCCCTATCTCAGTCTCC | S | *rps7* mRNA 5'UTR |
| | Lb523 | CGCATTTCAGATGGTGGTAAAGTG | S | *rps7* 3' coding seqeunce |
| | Lb524 | TTCAGTTCGAGCTAGGCGGTG | A | *rps7 5'* coding sequence |
| | LB550 | TTGCTCACCATCAAAGTCCC | A | *rps7 5'* coding sequence |
| | LB551 | GAGCTCTTCCTCATTCAGTC | S | *rps7* transcript 3'UTR |
| | LB553 | GCCAGGGATCGTCAACAAAC | S | *rps7* coding sequence |
| | LB577 | CCGTGAAACACATAGGCTCC | A | *rps7* precursor 5'UTR |
| | LB578 | GGAGGTGCGTAGTGTCTTAC | A | *rps7* precursor 5'UTR |
| | LB582 | GAGATACTGGATGCTTACCG | S | *rps7* 3' coding seqeunce |
| | LB602 | AGGAAGGCCGATTTTCTTTC | A | *rps7* transcript 3'UTR |
| | LB743 | TTTGCGGAAACCACTACTGG | A | *rps7* precursor 5'UTR |
| rice | | | | |
| | LB703 | GAGCACACTGTGAACTATCC | A | *rps7* precursor 5'UTR |
| | LB737 | CAAGGGTATGATGACCACTC | A | *rps7* precursor 5'UTR |
| rye | | | | |
| | LB704 | AGCATTTCGTCGCTTGCTAC | A | *rps7* precursor 5'UTR |
| | LB717 | GCTGGTCCTTGTTGACTCGC | A | *rps7* precursor 5'UTR |
| *Lolium* | | | | |
| | LB705 | ATTAAGGTCGTCACCCTCCG | A | *rps7* precursor 5'UTR |

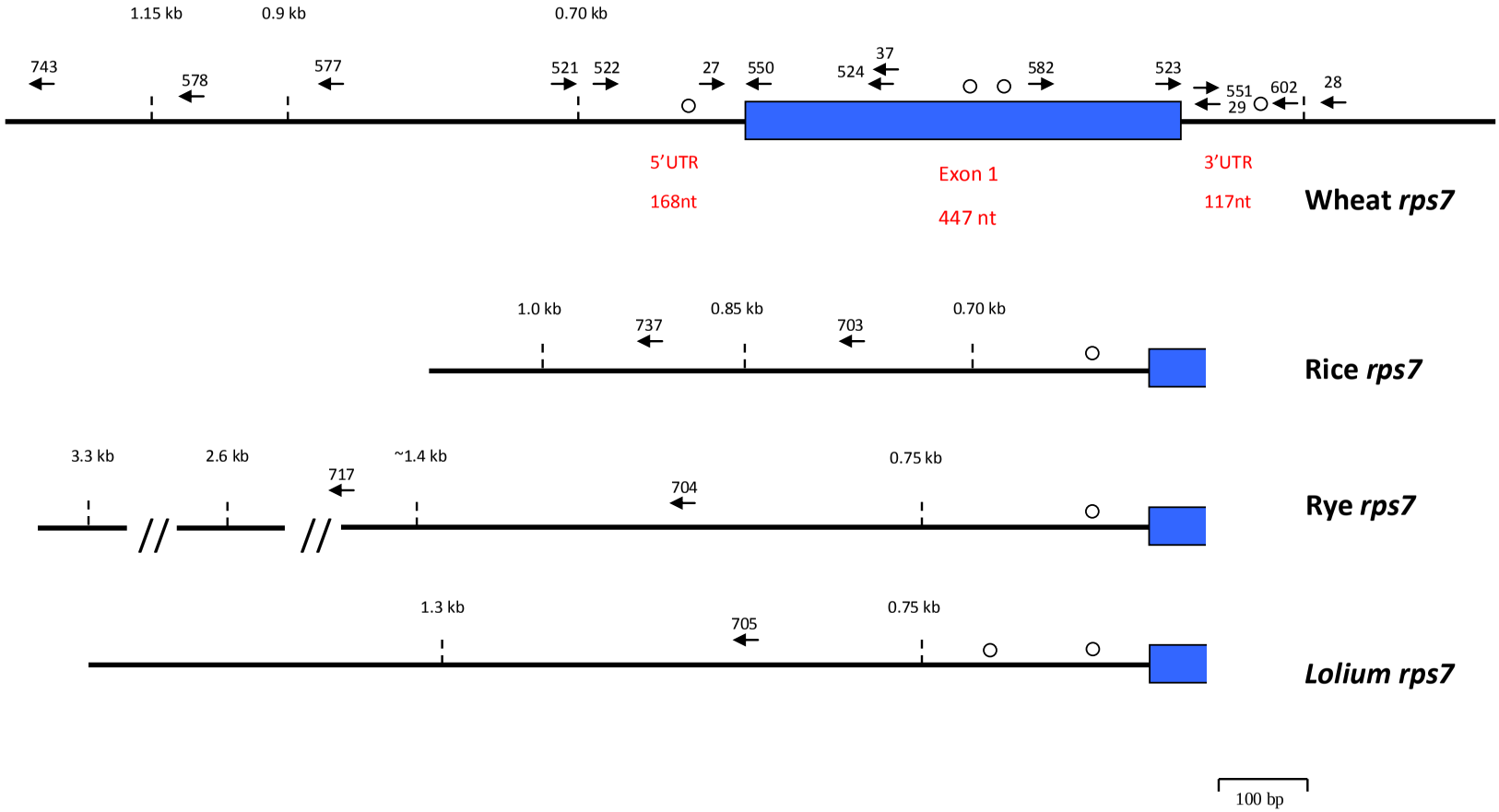
[a] Sense or Antisense orientation

[b] Primers designed on wheat coding sequence and within the region for the 3'UTR (+LB27) were also used for *rps7* in other plants.

[c] Primer used for *rrn18* northern hybridization experiments.

*Figure 2.1: Schematic showing the positions of oligomers used in this study for rps7 among grasses*

Oligomers used for either RT-PCR, CR-RT-PCR, inverse PCR, PCR and northern hybridization are shown as black arrows relative to *rps7* coding (blue box) and flanking sequences (black lines) (drawn to scale). RNA editing sites are given as open circles. Lengths of identified RNA species among plants (dotted lines) and length of coding sequence (exon 1) and UTRs for wheat are given. Oligomers designed on wheat *rps7* coding and 3'UTR sequences and LB27 were used on all grasses studied due to the 98% nucleotide identity among plants within these regions.


1.15 kb
0.9 kb
0.70 kb
743
578
577
521
522
27
550
524
37
582
523
551
29
602
28
5'UTR
168nt
Exon 1
447 nt
3'UTR
117nt
Wheat rps7
1.0 kb
0.85 kb
0.70 kb
737
703
Rice rps7
3.3 kb
2.6 kb
~1.4 kb
717
704
0.75 kb
Rye rps7
1.3 kb
705
0.75 kb
Lolium rps7
100 bp

with *rps7*-specific RT primers. Reverse transcription using M-MLV reverse transcriptase (Invitrogen) for 3hrs at 37$^{o}$C was then carried out for cDNA synthesis. The gene-specific primer used for reverse transcription of all grass mRNAs was LB524. Primer LB523 was then used in combination with LB524 for CR-RT-PCR amplification of single-stranded cDNA molecules corresponding to 5'-to-3' ligated *rps7* transcripts. Double nested checks on gel-purified CR-RT-PCR products or isolated plasmid DNA from cloned CR-RT-PCR products for mRNA amplicons were done using LB550 and LB551. For reverse transcription of precursor RNA species the RT primers used were as followed; LB577, LB578 and LB743 for wheat, LB703 and LB737 for rice, LB704 and LB717 for rye and LB705 for *Lolium*. Single nested checks for higher molecular weight species were done using the original RT primer and LB550. Size marker ladders used for agarose gels were from NBI Fermentas.

Synthetic oligomers (Invitrogen) were designed based on the wheat mitochondrial *rps7* gene region and DNA flanking sequences for additional grasses. All wheat *rps7* primers that anneal within the region transcribed for the 0.7 kb mRNA excluding LB522 were used on other grasses because of 100% sequence identity (Figure 2.1). Grass-specific primers were also designed for rice, rye and *Lolium* (Table 2.1, Figure 2.1). A list of accession numbers for plants used in this study is given in Table 2.2.

### 2.3.3 *RT-PCR analysis of rps7 coding region edits among grasses*

To assess RNA editing status within the *rps7* coding sequences of rye, barley, brome and *Lolium* embryo RNA (24hr or 36hr) clones of direct RT-PCR products were generated from mitochondrial RNAs which had been DNase-treated (Promega) one or two times, and the primers used were LB27 and LB29 (Figure 2.1) which map to the 3' UTR and 5'UTR respectively. DNase treatments were done according to the Promega RQ1 RNase-Free DNase protocol. 1 unit of enzyme (1unit/µl DNase) per µg of RNA with 8 vol. of Low TE (10mM Tris pH 7.5, 0.1mM EDTA pH 8.0) and 1 vol. of 10X RQ1 DNase Reaction Buffer were incubated for ½hr at 37$^{o}$C.

NB. Reaction was not stopped with RQ1 DNase Stop Solution as recommended by the manufacturer's protocol. Instead two 1:1 phenol extractions, first using buffer saturated

*Table 2.2:* Genome sequences used for *rps7* coding and flanking sequence comparisons

| Organism | Genome | Accession # | Publication | Info |
|---|---|---|---|---|
| *Megaceros aenigmaticus* (hornwort) | mito. | EU660574 | Li et al. 2009 | |
| *Isoetes engelmannii* (quillwort)[a] | mito. | FJ010859 | Grewe et al. 2009 | |
| *Marchantia polymorpha* (liverwort) | mito. | M68929 | Oda et al. 1992 | |
| *Physcomitrella patens* (moss) | mito. | NC_007945 | Terasawa et al. 2007 | |
| *Zea mays* (maize) | mito. | AY506529 | Clifton et al. 2004 | strain NB |
| *Triticum aestivum* (wheat) | mito. | AP008982 | Ogihara et al. 2005 | cv. Chinese Spring |
| *Triticum aestivum* (wheat)[b] | mito. | EU534409 | Cui et al. 2009 | cv. Chinese Yumai |
| *Triticum aestivum* (wheat)[c] | mito. | GU985444 | Liu et al. 2011 | cv. K-type Ch. Yu. CMS line |
| *Triticum aestivum* (wheat) | chl. | NC_002762 | Ogihara et al. 2002 | |
| *Oryza sativa* (rice) | mito. | BA000029 | Notsu et al. 2002 | Japonica |
| Bambusa oldhamii (bamboo) | mito. | EU365401 | unpublished | Green bamboo |
| *Hordeum vulgare* (barley) | mito. | HM581684 | Byers et al. 2010 | |
| *Secale cereale* (rye) | mito. | HM581685 | Byers et al. 2010 | |
| *Bromus inermis* (brome grass) | mito. | HM581686 | Byers et al. 2010 | |
| *Lolium multiflorum* (ryegrass) | mito. | HM581683 | Byers et al. 2010 | |
| *Avena sativa* (oats)[d] | mito. | | unpublished | var. AC Goslin |
| *Arabidopsis thaliana* (thale cress) | mito. | NC_001284 | Unseld et al.1997 | |
| *Cycas taitungensis* (cycad) | mito. | NC_010303 | Chaw et al. 2008 | |
| *Escherichia. coli* | bac. | NP_417800 | Riley et al 2006 | str. K-12, substr. MG1655 |

[a] First of 5 fosmid insert sequences; FJ010859, FJ536259, FJ390841, FJ176330 and FJ628360
[b] Wheat K-type maintainer line (km3)
[c] *Aegilops kotschyi* cytoplasm in *T. ae.* nuclear background (ks3)
[d] Oat *rps7* coding and flanking sequences are in appendix 2

phenol and then Phenol:Chloroform:Isoamyl Alcohol (25:24:1), followed by an ethanol precipitation in 2.5 vol. of 95% EtOH and 0.1 vol. of 3M NaOAc were used to stop the reaction and recover mtRNA respectively.

Degree of editing at two sites within *rps7* coding sequence and multiple sites within both 5' and 3' UTRs was analyzed using direct sequencing of RT-PCR products for 9 day cold-grown (3+6d) and room temperature wheat (9d). Products were amplified using primers LB522 and LB602 (mixed population of mRNA and precursor species) from two separate cDNA syntheses and primers LB521 and LB602 (population of precursor RNAs only) from only one cDNA synthesis reaction. 9d and 3+6d wheat mtRNA preps were 3x DNase treated according to the RQ1 DNase I protocol from Promega (as described above).

## 2.4 Cloning and sequencing of DNA and cDNA

PCR, RT-PCR, CR-RT-PCR and inverse PCR products were gel-purified using Ultra-Clean15 (MoBio Laboratories) prior to ligation into the pGEM-T Easy vector (Promega) and then cloned. The resulting recombinant plasmid DNAs were isolated using the QIAprep spin Miniprep kit (Qiagen) and automated sequencing was performed by the Ottawa Health Research Institute DNA sequencing facility (OHRI, StemCore laboratories). Universal primers M13 reverse and M13 forward were used for the sequencing of isolated recombinant plasmid DNA. Gel-purified RT-PCR products for analysis of *rps7* coding region edits for RNA isolated from cold-grown wheat were not cloned, instead they were sent for direct sequencing using custom oligomers; LB521, LB522, LB524, LB582 and LB602 (Figure 2.1).

DNA sequence information for the upstream flanking regions for rye, barley, brome and *Lolium* was obtained using an inverse PCR strategy. After restriction of mitochondrial DNA with either *Bam*HI or *Hin*dIII (since sites are located within the *rps7* coding region) and ligation with T4 DNA ligase (Invitrogen), the *rps7* coding-region specific primers LB553 and LB524 were used for PCR amplification. To obtain DNA sequence downstream of *rps7* for rye, brome, *Lolium* and oats, direct PCR products using

the oligomers LB27 and LB28 (Figure 2.1) which map within the 5' UTR and downstream of the 3' terminus respectively, were used.

For each of the plants under study, CR-RT-PCR clones were obtained from several independent experiments using different RNA preparations. Three clones which corresponded to mRNAs lacking intact coding sequences were omitted from the analysis, namely two brome clones and one rye clone, which had full 3'UTR sequences but 5' coding truncation.

## 2.5 Bioinformatics analysis

Sequences were obtained from the NCBI (National Institute for Biotechnology Information) databank for primer design (http://www.ncbi.nlm.nih.gov) and sequence analysis. The *rps7* gene sequences and flanking regions published in Byers *et al.* 2010 have been deposited in the NCBI GenBank under accession numbers HM581683 for *Lolium*, HM581684 for barley, HM581685 for rye and HM581686 for brome. Additional mitochondrial DNA sequence for the oat *rps7* coding region and flanking sequences (amplified using LB27 and LB28) can be found in Appendix 2. Included also are cDNA sequences upstream of LB27 in oat *rps7* and flanking sequences upstream of the 5' *HindIII* site in rye *rps7*, obtained from CR-RT-PCR experiments. CLUSTALW (European Bioinformatics Institute website; Chenna *et al*. 2003; http://www.ebi.ac.uk/Tools/clustalw/index.html) and BLAST searches (Altschul *et al*. 1990; http://www.ncbi.nlm.nih.gov/blast) were used for mitochondrial *rps7* sequence comparison with the following genomes; *Megaceros* (mt), *Isoetes* (mt), *Marchantia* (mt), *Physcomitrella* (mt), maize (mt), spring wheat (mt), yumai wheat (mt), K-type yumai wheat (mt), wheat (chl), rice (mt), bamboo (mt), *Arabidopsis* (mt), cycad (mt) and *E. coli* (bacterial). For a list of accession numbers (with publications) used in this study see Table 2.2. The search for possible RNA secondary structures in the *rps7* UTRs was conducted using the mfold program version 3.2 (Zuker 2003; http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html).

# Chapter 3 Results: Impact of genomic environment on mitochondrial *rps7* mRNA features in grasses

This chapter has been published as:

Byers E, Rueger J, Bonen L (2010) Impact of genomic environment on mitochondrial *rps7* mRNA features in grasses. Mol Genet Genomics 284:207-216

## 3.1 Research contribution from other students

Jen Rueger, a fourth year honours student began work on *rps7* 5'ends using the CR-RT-PCR method. Her work was integral in identifying a shared mRNA of 0.7 kb among grasses. Her northerns are shown in Figure 3.1A.

## 3.2 Additional data to accompany manuscript (Byers *et al*. 2010).

After the time of publication we became interested in comparing and contrasting transcript termini for the same RNA species between different stages of plant development. Continuing to work with *rps7* I mapped seedling transcript ends for all of the plants surveyed in the manuscript. Information for *rps7* in additional plants oats and maize was also collected. Presented in an addendum to chapter 3 (pg.58) is CR-RT-PCR data for seedling and/or embryo clones for *rps7* mRNA transcripts among grasses and northern hybridization data for oat *rps7*.

## 3.3 Abstract

The mitochondrial genomes of flowering plants are highly recombinogenic and this can lead to altered transcriptional units, even between closely-related species. We are interested in the effects that DNA rearrangements have on the generation of mature mRNAs, and to this end we have determined the termini of mitochondrial S7 ribosomal protein (*rps7*) mRNAs from selected grasses, using circularized-RT-PCR. Although the *rps7* mRNAs show a similar size of about 750 nt by northern hybridization analysis and have virtually identical 3' UTRs, their 5' terminal extremities differ among plant species, and this is attributable to genome rearrangements in some but not all cases. In wheat, rice

and barley, the 5’ ends are homogeneous for each plant but map to non-homologous sites among the three species. In contrast, the rye, brome and *Lolium* 5’ ends are quite heterogeneous in length even though they are located within conserved genomic regions. Comparative sequence analysis suggests that certain grass lineages have retained an ancestral organization upstream of *rps7* that includes a 170-bp block homologous to sequences preceding several other mitochondrial genes, whereas others have undergone independent rearrangements at a recombination-prone site. Our analysis of mature *rps7* transcripts revealed two non-silent RNA edits within the coding sequences, and also editing at several sites within the conserved 5’ and 3’ UTR regions in these plants, raising the possibility of their having a role in *rps7* expression at the post-transcriptional level. Taken together, our observations illustrate the dynamic nature of upstream regulatory cis-elements for mitochondrial *rps7* mRNA production in contrast to conservative 3’ end-formation signals, during evolution in grasses.

**Keywords:** Mitochondria, Ribosomal protein, RNA processing, Editing, Grasses

## 3.4 Introduction

The S7 ribosomal protein is universally present in the small subunit of prokaryotic and eukaryotic ribosomes (Lecompte et al. 2002) where it is an important component of the translational machinery. It is a primary rRNA binding protein which assists in rRNA folding and the binding of other proteins during small subunit assembly, and it also helps form the mRNA exit channel at the interface of the large and small subunits (Yusupov et al. 2001). The S7 proteins in mitochondrial ribosomes have retained strong sequence similarity with their bacterial counterparts (reflecting their endosymbiotic ancestry) although in various eukaryotic lineages the gene has been relocated into the nucleus during evolution. Based on a large-scale Southern hybridization survey for flowering plants (Adams et al. 2002), the *rps7* gene is still located in the mitochondrion in many of the grasses examined, although it appears to have been lost numerous independent times in eudicot lineages. This suggests that functional copies have been transferred to the nucleus, however it is somewhat surprising that a nuclear-located *rps7* gene has been identified in only a few plants (cf. Liu et al. 2009). In representatives of earlier-diverging

plant lineages, such as the gymnosperm *Cycas taitungensis* (Chaw et al. 2008), liverwort *Marchantia polymorpha* (Oda et al. 1992) and moss *Physcomitrella patens* (Terasawa et al. 2007), the mitochondrial *rps7* gene retains a bacterial-type linkage with the *rps12* gene (or *rps12* pseudogene in the case of *Cycas*), however in the hornwort *Megaceros aenigmaticus* mitochondrion, there is only an *rps7* pseudogene (Li et al. 2009) and in the quillwort *Isoetes engelmannii*, the *rps7* gene is absent from the mitochondrial genome (Grewe et al. 2009).

As yet, relatively little is known about the expression of the mitochondrion-located *rps7* gene in flowering plants. It encodes a highly-conserved protein of 148 amino acids, and from mitochondrial genomic sequencing studies, it is clear that the *rps7* gene has not retained a bacterial-type operon organization and it is typically located distant from other genes. The disruption of transcriptional units and shuffling of gene order is common in flowering plants because of the highly recombinogenic nature of their mitochondrial genomes (reviewed in Kubo and Newton 2008). The *rps7* mRNAs have been determined to be monocistronic and their termini mapped in wheat (Zhuo and Bonen 1993) and *Arabidopsis* (Forner et al. 2007). However their 5' ends are not located near the tetra-nucleotide CRTA which has been identified as a loose consensus promoter motif in plant mitochondria (reviewed in Binder et al. 1996), and they likely undergo endonucleolytic cleavage to generate the mature termini. The presence of several high molecular weight RNAs in northern blot analysis for wheat (Zhuo and Bonen 1993) is consistent with this view. Although the cis-regulatory signals for end-cleavage and RNA stability are not yet well understood, it appears that in some cases secondary structures such as tRNA-like elements or stem-loops may act as docking sites for the machinery (Forner et al. 2007). RNA editing is another processing step that is needed for the maturation of virtually all plant mitochondrial protein-coding transcripts (reviewed in Takenaka et al. 2008) and specific cytidines in precursor RNAs are converted to uridines. Editing occurs most frequently at positions which increase amino acid similarity with homologues from other organisms, and in the case of wheat and rice *rps7*, two non-silent codon conversions have been observed (Zhuo and Bonen 1993; Notsu et al. 2002).

To gain more insight into the expression of plant mitochondrial *rps7* genes, we have characterized their transcripts in selected grasses which have diverged from a

common ancestor ranging from about 5 to 50 million years ago (Kellogg and Bennetzen 2004). Because the *rps7* gene is located far away from other genes, and hence perhaps more prone to undergoing DNA rearrangements in its flanking regions, it was a good candidate to learn more about the nature and volatility of gene regulatory signals. In addition, this gene has the distinction of being predicted to have been lost from the mitochondrion the greatest number of times during plant evolution (Adams et al. 2002). In our analysis, we have used RNA isolated from germinating embryos because the abundance of wheat *rps7* mRNA was seen to be relatively higher in early stages of development than in seedlings (Li-Pook-Than et al. 2004). We find that although the 3' UTRs and part of the 5'UTRs are very similar among all these grasses, their extreme 5' termini differ markedly in length and complexity. Our CR-RT-PCR sequencing strategy also enabled us to examine untranslated regions for RNA editing, and rather unexpectedly (given the rarity of non-coding edits reported in plant mitochondria, cf. Giegé and Brennicke 1999) we observed editing within both the 5' and 3' UTRs at sites conserved among these plants.

## 3.5 Materials and Methods

### 3.5.1 *Mitochondrial RNA and DNA isolation*

Mitochondrial RNA was isolated from germinating embryos of wheat (*Triticum* aestivum var. Frederick), rice (*Oryza sativa* var. Drew), barley (*Hordeum vulgare* var. OAC Kippen), rye (*Secale cereale* var. Gazelle), brome grass (*Bromus inermis*) and annual ryegrass (*Lolium multiflorum*), using previously described procedures (Subramanian et al. 2001). Surface-sterilized embryos were dissected prior to germination in the dark for 36 hrs before RNA extraction. Mitochondrial DNA was isolated from 6-day etiolated seedlings grown in vermiculite at room temperature.

### 3.5.2 *Mitochondrial RNA analysis*

For northern blot analysis, mitochondrial RNA samples (approximately 5 μg per lane) were electrophoresed on 1.2% agarose/formaldehyde gels and after membrane transfer, hybridized with a $^{32}$P-end-labelled *rps7*-specific oligomer probe

5’GTTCAGTTCGAGCTAGGCGGTG 3’, using standard procedures (Sambrook et al. 1989).

To simultaneously map the 5’ and 3’ termini of *rps7* mRNAs, the CR-RT-PCR strategy (cf. Kuhn and Binder 2002; Calixte and Bonen 2008) was used. To circularize transcripts, approximately 5 μg of mitochondrial RNA was incubated overnight with T4 RNA ligase (New England Biolabs) in the presence of RNAsin (Promega). After phenol extraction and ethanol precipitation, the self-ligated RNA was heated at 65$^{o}$C for 5 min with the *rps7*-specific RT primer 5’ TTCAGTTCGAGCTAGGCGGTG 3’ prior to cDNA synthesis with M-MLV reverse transcriptase (Invitrogen) for 3 h at 37$^{o}$C. For subsequent PCR amplification, the above primer and one located in the 3’ coding region, namely 5’ CGCATTTCAGATGGTGGTAAAGTG 3’ were used, and amplicons were checked using various nested primers. Synthetic oligomers (Invitrogen) were designed based on the wheat mitochondrial *rps7* gene region [AP008982] and size marker ladders were from NBI Fermentas.

To assess RNA editing status within the *rps7* coding sequences of rye, barley, brome and *Lolium*, direct RT-PCR products were generated from mitochondrial RNAs which had been DNAse-treated (Promega) one or two times, and the primers used were 5’ ACTGAATGAGGAAGAGCTCC 3’ and 5’ ATCAATTTATCGGCCTCGTC 3’, which map to the 3’ UTR and 5’UTR respectively.

### 3.5.3 *Cloning and sequencing of DNA and cDNA*

PCR and RT-PCR products were gel-purified using Ultra-Clean15 (MoBio Laboratories) prior to ligation into the pGEM-T Easy vector (Promega) and then cloned. The resulting recombinant plasmid DNAs were isolated using the QIAprep spin Miniprep kit (Qiagen) and automated sequencing was performed by the Ottawa Health Research Institute DNA sequencing facility.

DNA sequence information for the upstream flanking regions for rye, barley, brome and *Lolium* was obtained using an inverse PCR strategy. After restriction of mitochondrial DNA with either *Bam*HI or *Hin*dIII (since sites are located within the *rps7* coding region) and ligation with T4 ligase (Invitrogen), the *rps7* coding-region specific primers 5’ GCCAGGGATCGTCAACAAAC 3’ and 5’

TTCAGTTCGAGCTAGGCGGTG 3' were used for PCR amplification. To obtain DNA sequence downstream of *rps7*, direct PCR products using the oligomers 5' ATCAATTTATCGGCCTCGTC 3' and 5' GCAGGCCTCTTGTGGATTCC 3' which map within the 5' UTR and downstream of the 3' terminus, respectively were used.

For each of the plants under study, CR-RT-PCR clones were obtained from several independent experiments using different RNA preparations. Three clones which corresponded to mRNAs lacking intact coding sequences were omitted from the analysis, namely two brome clones and one rye clone, which had full 3'UTR sequences but 5' coding truncation.

The *rps7* gene sequences and flanking regions obtained in this study have been deposited in the NCBI GenBank under accession numbers HM581683 for *Lolium*, HM581684 for barley, HM581685 for rye and HM581686 for brome. CLUSTALW was used for comparisons with mitochondrial *rps7* genes from the following plants (and their GenBank accession numbers): *Arabidopsis* (NC_001284) (Unseld et al.1997), rice (BA000029) (Notsu et al. 2002), maize (AY506529) (Clifton et al. 2004) and wheat (AP008982) (Ogihara et al. 2005). The search for possible RNA secondary structures in the *rps7* UTRs was conducted using the mfold program version 3.2 (Zuker 2003).

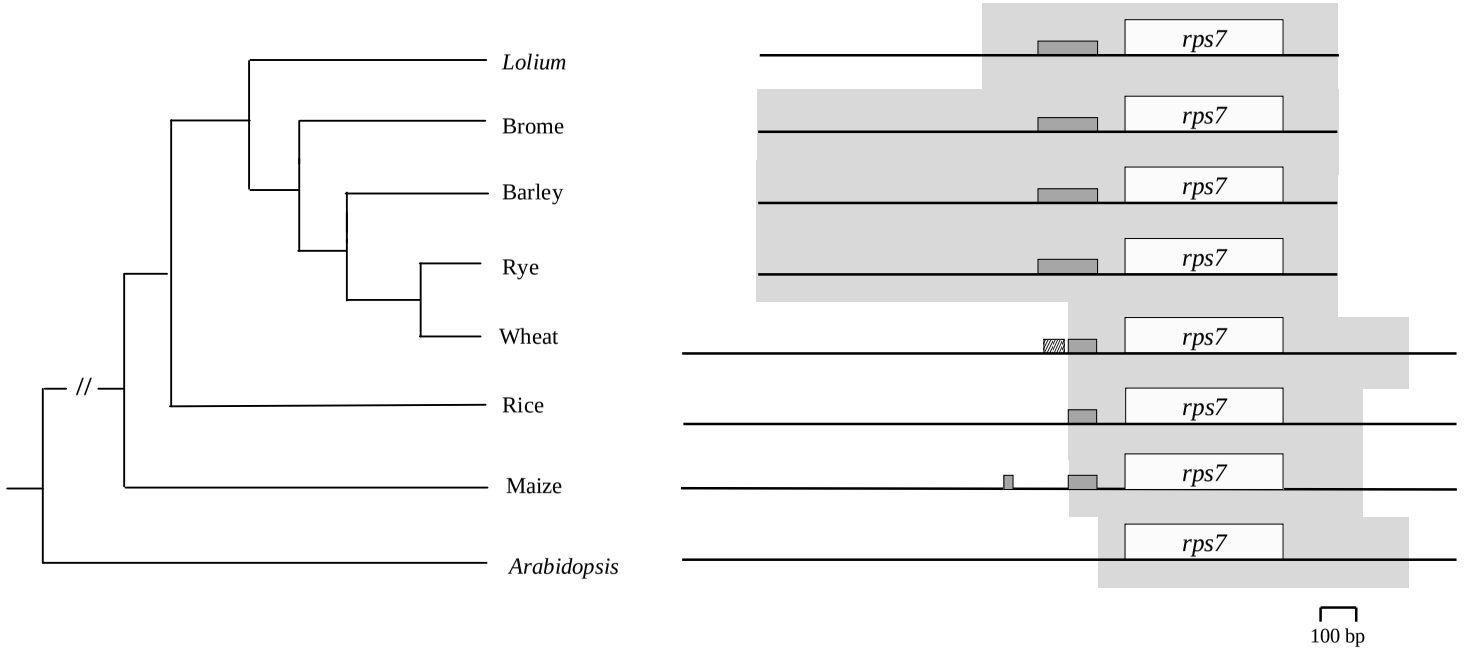
## 3.6 Results

### 3.6.1 *Genomic environment of the mitochondrial rps7 gene in selected grasses*

We have determined *rps7* and flanking sequences for rye, barley, brome, and *Lolium* and compared them with those previously-determined for wheat (Zhuo and Bonen 1993; Ogihara et al. 2005), rice (Notsu et al. 2002), maize (Clifton et al. 2004) and *Arabidopsis* (Unseld et al. 1997), as shown schematically in Figure 3.1 The *rps7* coding sequences (of 444 bp) for the four newly-investigated grasses are all virtually identical to that of wheat, having at most one nucleotide substitution in any pairwise comparison. Interestingly however, two independent non-synonymous substitutions were observed among the six grasses, one in each of rye and rice, and such behaviour is unexpected for proteins under strong functional constraint. The immediate flanking sequences of *rps7* (Figure 3.1 schematic, shading) are also very highly conserved among all these grasses, exhibiting about 90-99% nucleotide identity. The regions upstream and downstream of

*Figure 3.1: Genomic environment of the mitochondrial rps7 gene in selected grasses*

Schematic showing breakpoints in sequence homology flanking *rps7* (white block) among the grasses maize, rice, wheat, rye, barley, brome and *Lolium*, as well as the eudicot *Arabidopsis*. Background shading denotes regions conserved among these plants and lines are extended for those plants whose mitochondrial genomes have been completely sequenced, namely, *Arabidopsis* (Unseld et al. 1997), rice (Notsu et al. 2002), maize (Clifton et al. 2004) and wheat (Ogihara et al. 2005). The dark grey blocks denote sequences homologous to ones also found upstream of several other mitochondrial genes in grasses (Hazle and Bonen 2007) and hatched block is homologous to sequences preceding the 26S rRNA gene in wheat (Zhuo and Bonen 1993). In the cladogram on left (not to scale) the wheat-rye, wheat-*Lolium*, and wheat-rice divergence times are estimated to be approximately 5-7 Mya, 25 Mya and 50 Mya, respectively (cf. Kellogg and Bennetzen 2004).


Lolium
Brome
Barley
Rye
Wheat
//
Rice
Maize
Arabidopsis
rps7
rps7
rps7
rps7
rps7
rps7
rps7
rps7
100 bp

*rps7* show 96% and 80% nucleotide identity, respectively, when compared with the homologous regions from the eudicot *Arabidopsis*.

Preceding the *rps7* start codon, there is an 80-bp stretch which is held in common among all the plants shown in Figure 3.1 and it is preceded by a grass-specific conserved block (Figure 3.1, dark grey) which is found in either full-length or half-length, so either about 170 bp or 80 bp long. Notably it is very similar to ones located immediately upstream of several other protein-coding genes in grasses (Hazle and Bonen 2007). For example, full-length elements precede wheat *cox2, atp4*, and *atp6-2*, as well as a half-length copy in front of *atp6-1*. Moreover, this site appears to be recombination-prone as the sequences upstream of the *rps7* half-copy block are unrelated in wheat, rice and maize (Figure 3.1). In contrast, rye, barley, brome and *Lolium* appear to have retained an ancestral organization with a full-length copy.

The DNA sequences downstream of *rps7* are virtually identical among wheat, rye, barley, brome, and *Lolium* for at least ~160 bp (which was the limit of sequence information obtained by our PCR strategy) and based on the fully-sequenced genomes for rice and maize, a breakpoint in homology occurs about ~220 bp downstream of the stop codon, even though homology between wheat and *Arabidopsis* extends until ~350 bp (Figure 3.1).

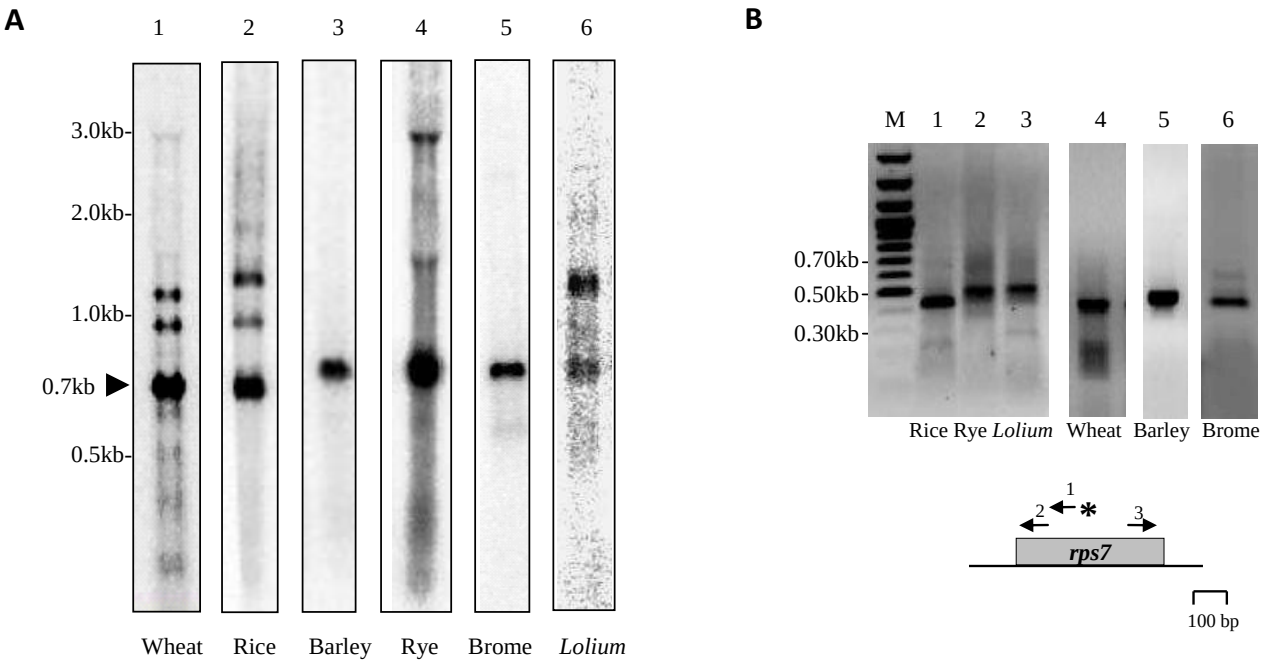
### 3.6.2 *Analysis of mitochondrial rps7 mRNAs and their termini in selected grasses*

From northern blot analysis, it can be seen that the *rps7* mRNAs in wheat, rice, barley, rye, brome and *Lolium* germinating embryos have a rather similar length of about 750 nt (Figure 3.2a), but the sizes and relative abundance of higher molecular weight transcripts vary among the plants. In barley and brome no larger precursors were detected (Figure 3.2a, lanes 3 and 5), whereas in *Lolium*, a precursor of approximately 1.3 kb is as abundant as the mature mRNA. The wheat *rps7* profile (Figure 3.2a, lane 1) is in keeping with our earlier observations (Zhuo and Bonen 1993).

To simultaneously map the 5’ and 3’ termini of the *rps7* mRNAs from these plants, we used the CR-RT-PCR strategy (cf. Forner et al. 2007) on ligase-treated RNA with coding-region specific primers as shown in Figure 3.2b (arrows 2 and 3). The products (as illustrated in Fig. 3.2b) were cloned and sequenced, and resulting data are

*Figure 3.2: Analysis of mitochondrial rps7 transcripts in selected grasses*

**a**) Northern hybridization in which germinating embryo RNAs from wheat (lane 1), rice (lane 2), barley (lane 3), rye (lane 4), brome (lane 5) and *Lolium* (lane 6) were probed with an *rps7* specific oligomer (arrow 1 with asterisk in panel (**b**)). The mature *rps7* mRNA is denoted by a black arrowhead and size markers are on left side. Less RNA was loaded in the *Lolium* lane. **b**) CR-RT-PCR products for rice (lane 1), rye (lane 2), *Lolium* (lane 3), wheat (lane 4), barley (lane 5), and brome (lane 6) obtained using primers 2 and 3, shown by arrows in schematic. Size markers are on left.


A
1
2
3
4
5
6
3.0kb-
2.0kb-
1.0kb-
0.7kb
0.5kb-
Wheat
Rice
Barley
Rye
Brome
Lolium
B
M
1
2
3
4
5
6
0.70kb-
0.50kb-
0.30kb-
Rice
Rye
Lolium
Wheat
Barley
Brome
1
2
3
rps7
100 bp

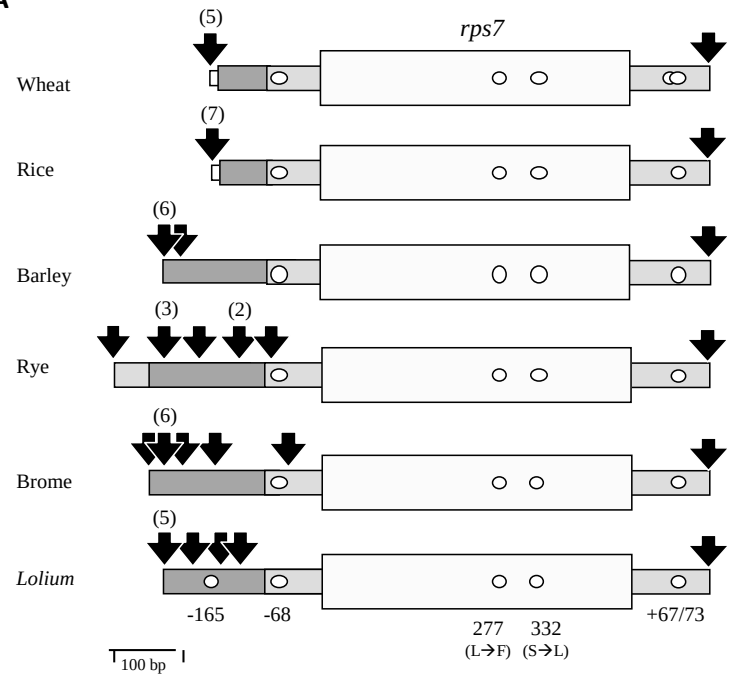
shown in Figure 3.3. The black arrows denote the positions of termini (Figure 3.3a) and the coordinates are given in Figure 3.3b. Notably, all these grasses share virtually identical 3' termini, so that the *rps7* 3'UTRs are typically 116-118 nt long, and this position is in agreement with our earlier S1 nuclease protection data for wheat (Zhuo and Bonen 1993). The 5' termini of *rps7* mRNAs, in contrast, show lineage-specific variation (Figure 3.3a, arrows). In the case of rice, wheat, and barley, the *rps7* 5' ends appear uniformly simple, whereas in rye, brome and *Lolium* there are multiple termini. In wheat, the sequenced clones all mapped about 165 nt upstream of the initiation codon, consistent with our previous analysis. In rice, the 5' terminus maps to position -156/157 also corroborating earlier observations for that plant (Zhang and Liu 2006). Interestingly, even though the 5' UTRs for wheat and rice *rps7* mRNAs are virtually identical (except for a 6-nt indel close to the initiation codon), their extreme termini map to non-homologous sequences and the first 10 nt of the *rps7* mRNAs (as well as upstream regulatory sequences) are unrelated (Figure 3.3a, narrow white bar and Figure 3.4b, lower case nucleotides).

Unlike the simple (yet plant-specific) 5' termini seen for wheat, rice and barley *rps7* mRNAs, those of rye, brome and *Lolium* are quite heterogeneous and multiple termini were observed (Figure 3.3a,b). Alignments of these 5' terminal regions (as well as those of barley) are shown in Figure 3.3d. Thus even though sequences are virtually identical at the DNA level, the 5' ends map to a wide variety of different positions in these plants, albeit with a bias for the same 5' terminus in the case of barley, brome and *Lolium* (Figure 3.3b). It should be noted that in all clones examined the same 3' terminus (within a few nucleotides) was observed, suggesting that the 5' heterogeneity is unlikely due to the quality of RNA preparations. With respect to the presence of non-encoded nucleotides at the 5'-3' junction, which could reflect tagging for RNA degradation (reviewed in Holec et al. 2008), we observed only five such cases, namely one rye clone (-220) with TCG, and four having a single extra A or T, namely, barley (-228 and -225), *Lolium* (-227), and a coding-truncated rye clone omitted from the analysis.

*Figure 3.3: Positions of the 5' and 3' termini of mitochondrial rps7 mRNAs in grasses*

**a**) Schematic of the *rps7* mRNAs with ends denoted by black arrows and numbers in brackets indicating the number of clones with either identical or similar (within 12 nt) positions. Wide white blocks and light grey blocks denote *rps7* coding regions and homologous non-coding UTRs, respectively. Narrow white blocks represent non-homologous UTR regions. Dark grey blocks denote sequences homologous to ones found upstream of several other genes in grasses (cf. Hazle and Bonen 2007). Editing positions are shown by white circles, and their coordinates as well as resulting amino acid changes are shown at the bottom. **b**) 5' and 3' termini designations based on CR-RT-PCR clone sequencing data. Boxed coordinates for 5' termini correspond to those mapping close to CGTATA (see panel D). **c**) Frequency of editing seen in *rps7* cDNA clones. The rice and wheat coding (CDS) edits, shown by plus symbols, are from Notsu et al. 2002 (Genbank accession number BA000029) and Zhuo et al. 1999, respectively. Negative and positive numbers represent distance from the *rps7* start and stop codons, respectively. The +67 editing site in wheat was also previously observed (Zhuo and Bonen 1993). **d**) Alignment of *rps7* upstream sequences for brome, barley, rye and *Lolium*. Boxed individual nucleotides denote 5' termini (cf. coordinates in panel B) and the large blocked region corresponds to the dark grey blocks in panel A. The CGTATA motif is shown as white on gray, and black dots show the positions of editing sites -68 and -165 (the latter being edited only in *Lolium*). The triangle denotes the position of breakpoint in homology with wheat and rice, which have half-length blocks (dark grey in panel A), and their 5' terminal sequences are shown in Figure 3.4b.

**A**




**B**

| | 5'UTR | 3'UTR |
|---|---|---|
| Wheat | -170 | +117 |
| | -168 | +117 |
| | -166 | +117 |
| | -165 | +116 |
| | -164 | +114 |
| Rice | -157 | +116 |
| | -157 | +116 |
| | -157 | +116 |
| | -157 | +118 |
| | -156 | +117 |
| | -156 | +116 |
| | -156 | +117 |
| Barley | -228 | +115 |
| | -227 | +114 |
| | -227 | +119 |
| | -225 | +108 |
| | -224 | +116 |
| | -224 | +115 |
| | -208 | +117 |
| Rye | -305 | +116 |
| | -227 | +116 |
| | -227 | +117 |
| | -220 | +117 |
| | -186 | +112 |
| | -123 | +118 |
| | -113 | +114 |
| | -76 | +116 |
| Brome | -249 | +117 |
| | -225 | +117 |
| | -224 | +111 |
| | -220 | +114 |
| | -218 | +117 |
| | -216 | +116 |
| | -214 | +116 |
| | -203 | +118 |
| | -159 | +117 |
| | -49 | +116 |
| *Lolium* | -234 | +115 |
| | -227 | +103 |
| | -226 | +117 |
| | -226 | +118 |
| | -223 | +109 |
| | -188 | +117 |
| | -144 | +115 |
| | -124 | +118 |

**C**

RNA Editing Status

| | 5'UTR | | 3'UTR | | CDS | |
|---|---|---|---|---|---|---|
| | -165 | -68 | +67 | +73 | 277 | 332 |
| Wheat | 0/4 | 5/5 | 1/5 | 2/5 | + | + |
| Rice | 0/0 | 7/7 | 0/7 | 2/7 | + | + |
| Barley | 0/7 | 7/7 | 0/7 | 4/7 | 5/5 | 5/5 |
| Rye | 0/5 | 8/8 | 0/8 | 7/8 | 5/5 | 5/5 |
| Brome | 0/8 | 8/9 | 0/10 | 3/10 | 4/4 | 4/4 |
| *Lolium* | 4/6 | 8/8 | 0/8 | 5/8 | 3/4 | 4/4 |

**D**

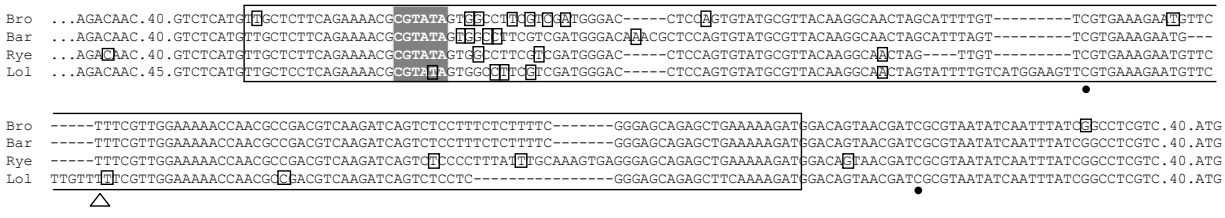
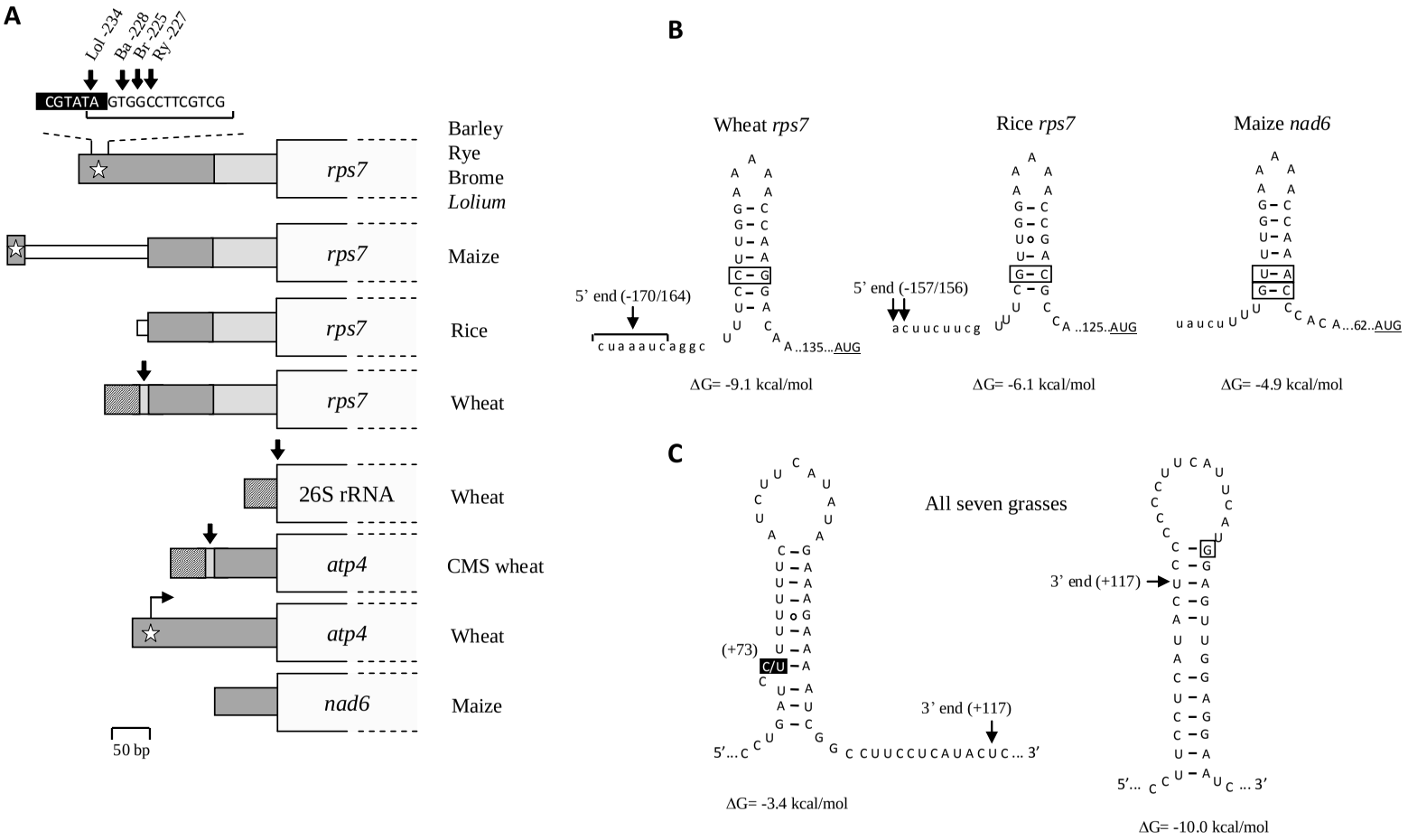


*Figure 3.4: Potential sequence motifs and RNA structures near rps7 mRNA termini*

**a**) Conserved blocks located upstream of *rps7* and several other genes in grasses are designated as in Figures 1 and 3. A 50-nt block (hatched) is located upstream of the 5' termini (arrows) of wheat *rps7*, wheat 26S rRNA, and *atp4* in wheat CMS (*Ae. crassa* mitochondrion) (Ogihara et al. 1999). Stars represent the position of a CGTATA motif, identified as a transcriptional initiation site for wheat *atp4* (bent arrow), and in the case of barley, rye, brome, and *Lolium* the arrows and bracket indicate the most distal 5' termini and range, respectively, of those boxed in Figure 3b. **b**) Potential stem-loop structures near the 5' terminus (arrows) of wheat and rice *rps7*, as well as within the homologous region upstream of maize *nad6* (cf. Hazle and Bonen 2007). Nucleotides within the conserved block are in upper case (and correspond to the 5' end of dark grey block in panel A). The barley, rye, brome and *Lolium* sequences are identical to rice except for U-A rather than U-G base pair, and similar structures are seen for the other genomic copies of this element. The positions of 5'termini are indicated by arrows, compensatory base changes are boxed, and initiation codons are underlined. **c**) Potential stem-loop structures either upstream of the *rps7* 3' terminus (on left) or overlapping it (on right). The C/U with black background (on left) represents a partial edit at position +73 and the boxed nucleotide (on right) denotes the only site (+133) which varies in sequence among these grasses. Structures in (B) and (C) were determined using the 'mfold' software (Zuker 2003).


A
Lol -234
Ba -228
Br -225
Ry -227
CGTATA GTGGCCTTCGTCG
rps7
Barley
Rye
Brome
Lolium
rps7
Maize
rps7
Rice
rps7
Wheat
26S rRNA
Wheat
atp4
CMS wheat
atp4
Wheat
nad6
Maize
50 bp
B
Wheat rps7
Rice rps7
Maize nad6
5' end (-170/164)
c u a a a u c a g g c
A..135..AUG
5' end (-157/156)
a c u u c u u c g
A ..125..AUG
u a u c u U U C C A C A ...62..AUG
ΔG= -9.1 kcal/mol
ΔG= -6.1 kcal/mol
ΔG= -4.9 kcal/mol
C
All seven grasses
(+73)
C/U
3' end (+117)
5'... C C G G C C U U C C U C A U A C U C ... 3'
ΔG= -3.4 kcal/mol
3' end (+117)
5'... C C U A U C ... 3'
ΔG= -10.0 kcal/mol

### 3.6.3 *RNA editing within mitochondrial rps7 transcripts*

To assess the status of C-to-U editing of *rps7* mRNAs from rye, barley, brome and *Lolium*, we combined our CR-RT-PCR sequence information with that obtained from direct RT-PCR data for the coding regions. From earlier studies of wheat and rice *rps7* coding sequences (Zhuo and Bonen 1993; Notsu et al. 2002), editing was expected in the other grasses at only two sites (Figure 3.3a, white circles), namely at positions 277 and 332 relative to the start codon. Indeed this was seen to be the case (Figure 3.3c). They create leucine-to-phenylalanine and serine-to-leucine conversions, respectively. The latter increases similarity with homologous proteins from other organisms as expected, whereas the former actually decreases amino acid similarity.

Within the *rps7* leader, we identified a site which is fully edited in all these grasses and it is located at position -68 (Figure 3.3d) or the homologous -62 site in rice. In *Lolium* an additional upstream site (at position -165) was seen to be partially edited. Interestingly, although editing was not seen at this latter site in the other grasses, the homologous position in the wheat *cox2* leader is edited (Covello and Gray 1989). Moreover the *Lolium rps7* and wheat *cox2* are identical immediately upstream of this editing site, whereas there is a 9-nt deletion in the *rps7* region of barley, rye and brome (Figure 3.3d), consistent with the view that editing recognition signals are primarily located within 15 nt upstream of the site (Takenaka et al. 2008). Within the *rps7* 3' UTR, a site 73 nt downstream of the stop codon was observed to be partially edited in all the plants and in wheat, one clone also showed an edit at +67, a site which we had previously observed (Zhuo and Bonen 1993).

### 3.6.4 *Sequence motifs or potential secondary structures near the ends of rps7 mRNAs*

Because the regions upstream of the *rps7* gene differ among grasses in the vicinity of the 5' termini, it is expected that there are lineage-specific regulatory elements. As mentioned above, these *rps7* genes are preceded by either a full- or half-length copy of a block (Figure 3.3a and Figure 3.4a, dark grey; Figure 3.3d, triangle to denote breakpoint) that is observed upstream of several other protein-coding genes (Hazle and Bonen 2007). The copies preceding wheat *cox2* and *atp4* have been demonstrated by guanylyl transferase capping experiments to contain the transcription initiation site (Covello and

Gray 1991; Ogihara et al. 1999) mapping just downstream of a CRTATA motif that characterizes some plant mitochondrial promoters (Binder et al. 1996). The majority of the 5' ends of the barley, brome and *Lolium rps7* mRNAs map slightly downstream of such a motif (Figure 3.3d, white on black shading), whereas the rye 5' termini are more heterogeneous. It is worth noting that our northern analysis revealed larger precursors for all of the grasses except barley and brome (Figure 3.2a) and that in the case of rye and brome, slightly longer transcripts were detected by CR-RT-PCR, so that this cannot be the sole promoter. Incidentally, almost immediately downstream of the CGTATA motif the *rps7* copies have a 4 nt-deletion relative to those preceding wheat *cox2* and *atp4*. Also, in maize a 25-bp homologous stretch containing this CGTATA motif is located further upstream of the *rps7* gene (Figure 3.1, short grey block and Figure 3.4, star).

In the present study, CR-RT-PCR products were obtained from RNAs which had not been treated with tobacco acid pyrophosphatase, so are expected to be processed transcripts. That said, it has been noted that in *Arabidopsis* mitochondria some primary and processed transcripts map to the same positions and that some genes have multiple promoters (Kühn et al. 2005). Very little is known as yet about cis-elements that direct specificity of 5'end-cleavage, but it is worth noting that the 50-nt stretch which precedes the uniform 5' terminus of the wheat *rps7* mRNA (Figure 3.4, hatched block) is very closely related to sequences which precede the mature 26S rRNA in wheat, and also the *atp4* mRNA in an alloplasmic line of cytoplasmic male sterile wheat in which the *rps7* upstream sequences have recombined in front of the *atp4* gene (Ogihara et al. 1999).

In wheat and rice, where rearrangements have resulted in only a half-length conserved block being located upstream of *rps7* (Figure 3.4a, dark grey blocks) so that their extreme 5' termini differ (Figure 3.4b, lower case), we identified a possible stem-loop structure using the 'mfold' program (Zuker 2003). Notably there are compensatory base changes (Figure 3.4b, blocked positions) that would conserve folding in the homologous half-length copy in wheat and rice, as well as in those which precede several other genes in grasses (Hazle and Bonen 2007), such as the *nad6* gene in maize (Figure 3.4b), where these sequences are within the *atp6-nad6* dicistronic mRNA (Haouazine-Takvorian et al. 1997). These observations thus provide phylogenetic support for such a structure which might be involved in RNA stability or translational control.

The 3’ terminal region of the *rps7* mRNA in grasses can be folded into a weak stem-loop structure whose stability would be slightly improved by editing at position +73 (Figure 3.4c, black shaded nucleotides) and a more thermodynamically-stable structure would span the 3’ cleavage site (Figure 3.4c, right side). The latter would require the activity of an endoribonuclease with specificity for double-stranded RNA. The corresponding region of the *Arabidopsis rps7* mRNA can be folded into the latter structure although it has a 3’UTR of 351 nt (Forner et al. 2007), and interestingly its 3’ terminus maps to the position of the breakpoint in sequence homology between *Arabidopsis* and wheat (Figure 3.1, shaded). Figure 3.4c also illustrates that pyrimidine stretches are present on both sides of the 3’ end-cleavage site in the grasses.

## 3.7 Discussion

Our analysis of mitochondrial *rps7* mRNA termini in grasses provides a rather striking example of the contrast in behaviour between 5’ ends and 3’ ends. One might have expected that the 5’ regions would be more conservative during evolution, given the importance of upstream regulatory elements for gene expression, but the opposite was seen. Each of the six grasses has its own distinctive 5’ UTR features, whereas they all share the same simple discrete 3’ terminus. Interestingly, although wheat and rye are the most closely-related plants in this study, their 5’ termini exhibit the most pronounced differences in complexity. In wheat and rice, differences can be attributed to DNA rearrangements, but it is notable that among barley, rye, brome and *Lolium*, even though sequences are virtually identical at the genomic level, their 5’ ends exhibit considerable difference. In a previous study mapping the termini for the complete set of *Arabidopsis* mitochondrial protein-coding genes (Forner et al. 2007), the pattern of simple 3’ ends but more complex 5’ termini was also observed. In addition, different accessions of *Arabidopsis* have been seen to exhibit 5’ end polymorphisms for certain mitochondrial genes, a subset of which were attributed to mitochondrial DNA differences and the others to nuclear genetic variation (Forner et al. 2008).

Based on the variation seen in the 5’ termini of *rps7* mRNAs among grasses, as well as the differing complexities of their precursor RNA profiles, it is clear that there are lineage-specific post-transcriptional pathways. Very little is known about cis-elements

which signal end-cleavage, although potential stem-loop and tRNA-like structures have been implicated in certain cases (cf. Forner et al. 2007). The maturation of 3' termini appears to require the concerted action of both endo- and exoribonucleases (Perrin et al. 2004, Forner et al. 2007) and tRNA enzymes may in some cases be co-opted. However for *rps7* in grasses, we did not detect a convincing stem-loop structure immediately upstream of its 3' terminus. The 5' terminal heterogeneity seen in certain grasses may reflect a combination of the use of multiple promoters, discrete endonucleolytic cleavage events, and random exonucleolytic attack, the latter being mitigated if there is protection by secondary structure and/or RNA stability proteins. It is worth noting that the 5' heterogeneity seen for *rps7* mRNAs is less than we observed for *rps1* monocistronic transcripts in wheat embryos, where many molecules lacked the expected initiation codon, even though they all possessed full-length 3' UTRs (Calixte and Bonen 2008). Although little is known about the nature of factors which specify end-cleavage, an RNA processing factor RPF2 has recently been shown to be required for 5' end maturation of *nad9* and *cox3* mRNAs in *Arabidopsis* (Jonietz et al. 2010). It is a member of the PPR (pentatricopeptide repeat) family of proteins (reviewed in Schmitz-Linneweber and Small 2008), which have been implicated in various plant organellar RNA processing events such as splicing (de Longevialle et al. 2007) and editing (Zehrmann et al. 2009). Interestingly, a PPR protein has been found to play an important role in defining and stabilizing the 5' and 3' termini of mRNAs in maize chloroplasts, perhaps by acting as a barrier to RNA degradation (Pfalz et al. 2009).

It was somewhat unexpected to find multiple RNA editing sites within non-coding regions of *rps7* mRNAs as it has rarely been observed in UTRs. For example, in a large-scale analysis of *Arabidopsis* mitochondrial editing, a total of only 7 edits were observed in leaders and tails (Giegé and Brennicke 1999; Forner et al. 2007). Interestingly, position -68 within the *rps7* mRNA leader is fully edited in all the grasses we examined, and in *Arabidopsis* there is a genomically-encoded T at this site. This raises the possibility that it might have some biological function in RNA folding/stability or translation. The possibility that the editing could be fortuitous if the specificity cis-element happened to resemble that of an important coding edit cannot be excluded, however this region did not show detectable sequence similarity to anywhere else in the genomes.

Our observations for mitochondrial *rps7* gene expression in grasses illustrate the plasticity of mRNA maturation pathways operating in plant mitochondria and they point to a complex and continually-evolving assortment of cis-regulatory signals (accompanied by compatible trans-acting machinery). This study provides examples of two such cis-elements, namely an "expression cassette" present in either full- or half-length (Figure 3.4a, dark grey blocks) and a 50-nt element (Figure 3.4a, hatched blocks) which precedes the processed 5' termini for several genes in wheat. The acquisition of new regulatory signals for *rps7* during evolution is sometimes mediated by DNA rearrangements, but in other cases, differences appear due either to more subtle (or more distal) mitochondrial genome changes or to variation in nuclear-encoded specificity factors which recognize the cis-elements. This study has focused on mitochondrial RNA events in germinating embryos, and because mitochondrial ribosomal protein transcripts appear to be particularly sensitive to developmental regulation (cf. Li-Pook-Than et al. 2004; Calixte and Bonen 2008; Naydenov et al. 2008), it will also be of interest to examine the behaviour of mitochondrial RNA maturation/stability pathways during other stages of plant development. Our preliminary examination of *rps7* transcript profiles for other grasses in the seedling stage suggests that they too have relatively lower steady-state levels of precursors and mature mRNAs than in germinating embryos (unpublished observations). Thus it could be argued that this phenomenon is likely controlled at the post-transcriptional level, since promoters for the high molecular weight *rps7* precursors seen in wheat, rice, rye and *Lolium* (Figure 3.2) are predicted to be in non-homologous genomic regions. Elucidation of the trans-factors which interact with the transcripts will also be needed to further our understanding of mitochondrial *rps7* mRNA regulation during development.

**Acknowledgements**

We thank J. Ahmed and B. Choi for expert technical assistance, as well as Dr R. Pandeya and Dr. T. M. Choo (Agriculture and Agri-food Canada) for kindly providing seeds. Financial support from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

## 3.8 Chapter 3 addendum: Expression of mitochondrial *rps7* in oats and maize.

Upon comparing upstream DNA sequences between wheat, rice and maize I determined that maize also possessed the half expression cassette. I became interested in mRNA ends and their positions relative to this sequence. In rice and wheat mRNA 5'ends map immediately upstream of this sequence element which happens to also be the breakpoint in DNA homology among wheat, rice and maize. Might maize *rps7* mRNA 5'termini also map to this 'hotspot' for recombination among the 3 plants? Preliminary CR-RT-PCR data is presented below.

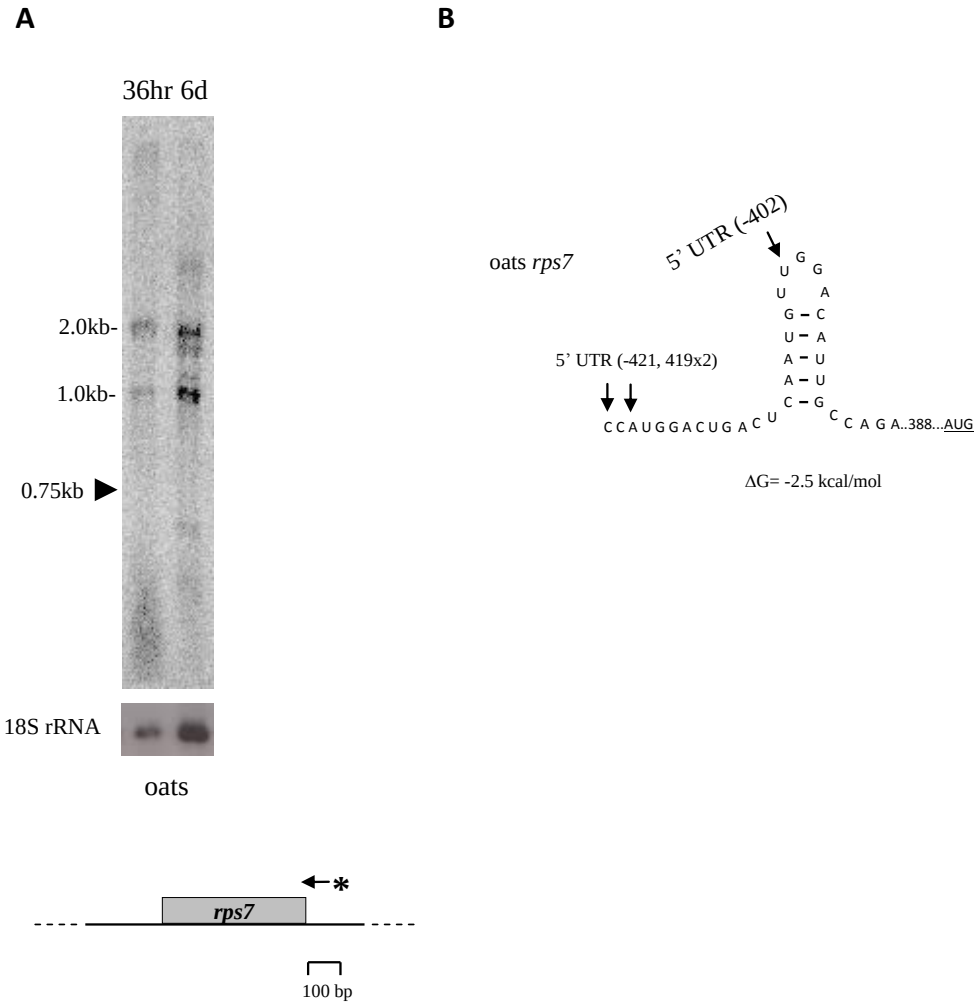
Oat *rps7* northern hybridization experiments (Jen Rueger 4$^{th}$ year honours project) revealed mRNAs larger than the conserved 0.7 kb species observed among other grasses (Byers *et al*. 2010). Preliminary CR-RT-PCR data for oat *rps7* as well as data for seedling mRNA termini among the grasses studied in Byers *et al*. 2010 is also presented here. No differences in positions of mRNA termini between stages of development (embryo vs. seedling) with the exception of rye were seen among plants. Results for rye seedling *rps7* mRNA however are derived from a single RNA prep.

### 3.8.1 *Northern hybridization profiles for oat rps7 and CR-RT-PCR end analysis of rps7 transcripts for maize and oats.*

Hybridization using an *rps7*-specific oligomer probe shows two major oat *rps7* transcripts of 1.0 kb and 2.0 kb (Figure 3.5a). Interestingly the 1.0 kb species which likely represents the oat *rps7* mRNA is of a higher molecular weight than the *rps7* mRNAs seen for other grasses such as wheat, rye, barley, brome, and *Lolium* (Byers *et al*. 2010). In both stages of plant development used for northern hybridization (36 hour germinating embryos and 6 day developing seedlings) the relative stochiometries of the 2.0 kb precursor and 1.0 kb mRNA are very similar. Analogous observations were made for northern hybridization profiles comparing relative levels of *rps7* transcripts in embryo RNA from *Lolium* (Figure 3.1). Using the CR-RT-PCR method, data for oat *rps7* mRNA

*Figure 3.5: Analysis of oat rps7 transcripts*

**a**) Northern analysis of oat mitochondrial *rps7*. Blots were hybridized with a $^{32}$P-end-labelled *rps7*-specific oligomer probe LB29 (arrow with asterisk in schematic). Lanes contain RNA isolated from germinating embryos (36hr) and developing seedlings (6d). 18S rRNA loading controls are indicated below and size markers in kilobases are shown on the left. Black arrowhead denotes position of *rps7* mRNA in grasses; wheat, rice, barley, rye, brome and *Lolium*. **b**) RNA secondary structure found upstream of oats *rps7*. Position of *rps7* 5'UTR termini -421, -419 and -402 (numbers relative to start codon) are indicated by black arrows. Distance to *rps7* start codon (AUG) given in brackets. Structure was determined using 'mfold' software (Zuker 2003) and corresponding value for Gibbs free energy is given below.


A
36hr 6d
2.0kb-
1.0kb-
0.75kb
18S rRNA
oats
rps7
100 bp
B
oats rps7
5' UTR (-402)
5' UTR (-421, 419x2)
CCAUGGACUGA
CAGA..388...AUG
ΔG= -2.5 kcal/mol

termini from 6 day seedlings show very heterogeneous 5'ends versus very conservative 3' ends, the pattern seen for grasses like rye, brome and *Lolium* (for a summary of mRNA termini among grasses see Table 3.1). Oat *rps7* mRNA with 5'ends mapping between 403 and 421 nt upstream of the translational start codon however were seen in 4 out of 9 clones. The 5'ends for the 5 other clones showed no bias for any upstream position and were shorter (54 to 165 nt upstream of *rps7* start codon). Shorter than expected 5'UTRs are not believed to result from RNA degradation as virtually all 3'UTR termini mapped to similar positions downstream of the *rps7* stop codon.

Sequence analysis reveals that the oat *rps7* gene is preceded by the 3'half of the expression cassette found in either full or half length upstream of all mitochondrial *rps7* genes in the grasses surveyed thus far (Byers *et al*. 2010). Upstream of the half cassette in front of oat *rps7* there is a breakpoint in DNA homology compared to other grasses, including the close relative *Lolium* (Kellogg and Bennetzen 2004) (Appendix 1). This is also the position of the breakpoint in DNA homology between wheat, rice and maize (Figure 3.1). For plants; wheat, rice, maize and now oats the breakpoint in *rps7* DNA homology occurs immediately upstream of the 3' half of the expression cassette making this position very convincing as a true 'hotspot' for recombination.

Preliminary data for transcript 5'ends for maize mRNA isolated from germinating embryos shows short 5'UTRs mapping around 50 nt in front of the *rps7* start codon, within a region of 80 bp specific to *rps7* in all grasses studied (Figure 3.1). One species 5'UTR terminus however maps within the half cassette, 103 nt upstream of the translational start codon (Table 3.2). An additional embryo transcript was found missing the first nucleotide of the start codon and also had a shorter than expected 3'end. 5'ends for CR-RT-PCR clones generated from seedling RNA for maize *rps7* map both to the region 50 nt in front of the *rps7* start codon and further upstream around 100 nt and 260 nt (Table 3.1). The apparent 'hotspot' for DNA recombination, roughly 160 bp upstream of the *rps7* start codon for grasses wheat, rice, maize and oats to which homogeneous 5'termini map in both wheat and rice is therefore not the site of 5'end processing in either maize or oats. For a summary of mRNA 5'ends and their positions relative to upstream flanking sequences for *rps7* among grasses see Table 3.2.

*Table 3.1: Transcript termini for mitochondrial rps7 mRNAs from various grasses*

CR-RT-PCR data for *rps7* mRNA termini is summarized here for all grasses surveyed. For some but not all plants, mRNA 5' and 3'ends were analyzed for both seedling and germinating embryo RNA.

**Transcript termini**

| Plant and stage of development for *rps7* mRNA | Total RNA lengths are given in nucleotides (nt). Positions of 5'UTR termini given relative to *rps7* start codon (-ve), unless otherwise indicated for 5' truncated mRNAs. Positions for 3'UTR termini given relative to *rps7* stop codon (+ve). RNA species with non-encoded nucleotides indicated, with length of particular transcripts in brackets. | | | |
|---|---|---|---|---|
| | **RNA length (nt)** | **(-) 5'UTR** | **(+) 3'UTR** | **Non-encoded NTs** |
| **Wheat** | | | | |
| Embryo mRNA | 734, 732, 730, 728, 725, | 170, 168, 166, 165, 164, | 117, 117, 117, 116, 114 | |
| Seedling mRNA | 731, 728, 727, 727, 726, 726, 720 | 168, 165, 165, 165, 165, 165, 159 | 116, 116, 115, 115, 114, 114, 114 | |
| **Rice** | | | | |
| Embryo mRNA | 722, 720, 720, 720, 720, 720, 719 | 157, 156, 157, 157, 157, 156, 156 | 118, 117, 116, 116, 116, 117, 116 | |
| Seedling mRNA | 724, 721, 721, 720, 720 | 159, 155, 157, 156, 157 | 118, 119, 117, 117, 116 | |
| **Barley** | | | | |
| Embryo mRNA | 793, 790, 788, 787, 786, 780, 772, 556 | 227, 228, 227, 224, 224, 225, 208, 37 | 119, 115, 114, 116, 115, 108, 117, 72 | T (780), A (790), TT (556) |
| Seedling mRNA | 866, 791, 786, 784, 778, 691, 631 | 170, 227, 225, 223, 218, 128, 67 | 249, 117, 114, 114, 113, 116, 117 | A (866), A (778) |
| **Rye** | | | | |
| Embryo mRNA | 868, 821, 790, 784, 745, 688, 676, 639, 520 | 305, 227, 227, 220, 186, 123, 113, 76, (+)37 | 116, 147, 116, 117, 112, 118, 116, 116, 110 | 62bp E. coli (639), TCG (784), A (520) |
| Seedling mRNA | 679, 630, 617, 617, 572, 571 | 115, 66, 55, 52, 10, 11 | 117, 117, 115, 118, 115, 113 | |
| **Brome** | | | | |
| Embryo mRNA | 813, 789, 782, 782, 781, 779, 777, 768, 723, 612, 510, 488 | 249, 225, 224, 218, 220, 216, 214, 203, 159, 49, (+)54, (+)76 | 117, 117, 111, 117, 114, 116, 116, 118, 117, 116, 116, 116 | |
| Seedling mRNA | 776, 767, 737, 685 | 212, 202, 175, 120 | 117, 118, 115, 118 | T (767) |
| ***Lolium*** | | | | |
| Embryo mRNA | 796, 791, 790, 779, 777, 752, 706, 689 | 234, 226, 226, 223, 227, 188, 144, 124 | 115, 118, 117, 109, 103, 117, 115, 118 | AA (689), A (777) |
| Seedling mRNA | 846, 792, 792, 790, 789, 789, 775, 763, 746, 718, 716, 688 | 282, 229, 229, 228, 228, 229, 211, 200, 182, 154, 156, 124 | 117, 116, 116, 115, 114, 113, 117, 116, 117, 117, 113, 117 | CC (716), A (792), AAAAACA (789) |
| **Maize** | | | | |
| Embryo mRNA | 660, 623, 619, 619, 485 | 103, 59, 56, 55, (+1) | 110, 117, 116, 117, 38 | AAAA (660) |
| Seedling mRNA | 834, 824, 675, 660, 623, 622, 619, 617, 572 | 272, 266, 111, 103, 59, 61, 55, 57, 9 | 115, 111, 117, 110, 117, 114, 117, 113, 116 | A (824), AAAA (660) |
| **Oats** | | | | |
| Seedling mRNA | 985, 984, 984, 966, 730, 700, 677, 619, 618 | 419, 421, 419, 403, 165, 136, 99, 56, 54 | 119, 116, 118, 116, 118, 117, 131, 116, 117 | A (966) |

*Table 3.2: Positions of 5' ends for mitochondrial rps7 mRNAs from various grasses*

A schematic showing sequences flanking mRNA termini among grasses. Potential stem-loop structures, consensus sequences for transcription initiation and upstream sequences homologous to those of other mitochondrial and chloroplast genes are illustrated here.

**Transcript termini**

| Plant and stage of development for *rps7* mRNA | 5' Termini and DNA[a] flanking sequence (5' and 3'). RNA ends are in bold letters with potential promoter motifs (CRTA(TA), where R=A or G) underlined. Possible stem-loop (SL) structures upstream or downstream of RNA ends are double underlined. Transcript termini near SL structures likely generated by RNase P or Z cleavage at either the structures' 5' or 3' end respectively. | Remarks |
|---|---|---|
| **Wheat** | | |
| Embryo mRNA | TTGTGTGGGTGTCCGCTCATGTTCACGTTACATG**C**T**A**A**ATC**AGGCTTTCCTTGGAAAAACCAAGGACAACCCCTATCTCAGTCTC.....119……ATG | SL, RNase P. Structure virtually identical in the 3' half of the purple cassette (purple letters) found in either full or half length in all grass *rps7* mRNAs. Pink sequence corresponds to a 35 bp stretch homologous to the rrn26S pre-RNA. |
| Seedling mRNA | TTGTGTGGGTGTCCGCTCATGTTCACGTTACATGCT**A**AA**T**CAGGC**T**TTCCTTGGAAAAACCAAGGACAACCCCTATCTCAGTCTCC…..119…..ATG | |
| **Rice** | | |
| Embryo mRNA | TGTGCTCATTCTCAAAAAAAGAAAAAA**AC**TTCTTCGTTTCGTTGGAAAAACCGACGCCAACGTTAAGATCAGTCTCCTTTCTCTTTT.……97.....ATG | Sequence homologous (blue letters) to a portion of the 5'UTR of wheat *rpl5* |
| Seedling mRNA | GTGCTCATTCTCAAAAAAAGAAAA**A**A**ACT**TCTTCGTTTCGTTGGAAAAACCGACGCCAACGTTAAGATCAGTCTCCTTTCTCTTTT.....…97.....ATG | SL, Rnase P |
| **Barley** | | |
| Embryo mRNA | TCAGAAAACGCGTATAG**TG**G**CC**TTCGTCGATGGGACA**A**ACGCTCCAGTGTATGCGTTACAAGGCAACTAGCATTTAGTTCGTGAA.….160.…ATG | |
| Seedling mRNA | TCAGAAAACGCGTATAGT**G**G**C**C**T**TCGT**C**GATGGGACAAACGCTCCAGTGTATGCGTTACAAGGCAACTAGCATTT**A**GTTCGTGAA.….160.…ATG | |
| **Rye** | | |
| Embryo mRNA | GGAAAAGAAAGTCTCATGTTGCTCTTCAGAAAACGCGTATAGTG**G**CCTTCG**T**CGATGGGACCTCCAGTGTATGCGTTACAAGGCA**A**C…184…ATG | ~180bp sequence (green letters) homologous to 5' piece of plastid *rrn23* found at virtually identical positions in brome, barley and *Lolium rps7* (only 65bp in *Lolium*). |
| Seedling mRNA | CAGTCTCCCCTTT**A**TTTGCAAAGTGAGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCG**C**GTAATATCAA**T**TT**A**TCGGCCTCG.….42....ATG | |
| **Brome** | | |
| Embryo mRNA | AAGTCTCATGT**T**GCTCTTCAGAAAACGCGTATAGT**GG**CCT**T**C**G**T**C**G**A**TGGGACCTCC**A**GTGTATGCGTTACAAGGCAACTAGCAT….175…ATG | |
| Seedling mRNA | AAGTCTCATGTTGCTCTTCAGAAAACGCGTATAGTGGCCTTCGTCGAT**G**GGACCTCCA**G**TGTATGCGTTACAAGGCAACTAGCAT**T**TT…172…ATG | |
| ***Lolium*** | | |
| Embryo mRNA | GCCAGAGGAAAAGAAAGTCTCATGTTGCTCCTCAGAAAACGCGTA**T**AGTGGC**CT**TC**G**TCGATGGGACCTCCAGTGTATGCGTTAC.....194.….ATG | Promoter motif in 5' half of full purple cassette known to initiate transcription of other genes in wheat mitochondria like *cox2* and *atp4*. |
| Seedling mRNA | GCCAGAGGAAAAGAAAGTCTCATGTTGCTCCTCAGAAAACGCGTATAGTG**GC**CTTCGTCGATGGGACC**T**CCAGTGTATG**C**GTTAC….194.….ATG | |
| **Maize** | | |
| Embryo mRNA | CGCCCACATATTGATCTTTAAGTCTCCTTTCTC**T**TTTGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCGCGTA**A**TA**TC**AATTTA....48...ATG | SL, possible stability conferring structure for longer maize *rps7* mRNAs. |
| Seedling mRNA | CGCCCACATATTGATCTTTAAGTCTCCTTTCTC**T**TTTGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCGCG**T**A**A**T**A**T**C**AATTTA....48...ATG | |
| **Oats** | | |
| Seedling mRNA[a] | **C**C**A**TGGACTGACTCAATG**T**TGGACATTGCCCAATTTTCCATTTGTTTAGCTAATATTCTGATGACCGGGGCCGCCAAGCCTCAAGGA….334….ATG | SL, RNase Z |

[a] Sequence information corresponds to cDNA data from oats CR-RT-PCR clones

### 3.8.2 *Potential secondary structures upstream of oat rps7 may act as cis-elements for RNA processing*

Secondary structures corresponding to sequences in the 3'half of the expression cassette (Hazle and Bonen 2007b) found within the 5'UTR for all grasses studied and additional genes in wheat (Figure 3.4 and Table 3.2) have been proposed as signals for RNA stability or end-cleavage in wheat and rice, plants that demonstrate homogeneous 5' mRNA termini immediately upstream of the structures' 5' end. Preliminary data for maize and oats *rps7*, two monocots that also have the homologous upstream 3'half of the expression cassette (maize mitochondrial genome sequence from NCBI database, oats RNA-level data from CR-RT-PCR method), demonstrate heterogeneous 5' mRNA termini (Table 3.1) that do not map upstream of the conserved stem-loop structure. Oat *rps7* northern data shows mRNAs longer than the 0.7 kb species seen in wheat and rice therefore for at least oat *rps7* we do not expect mRNA termini to map to this position. Sequences upstream of similar 5'termini (-419, -421 and -419) for 4 CR-RT-PCR clones (Table 3.1) from oat *rps7* mRNA were able to be folded into a stem-loop (Figure 3.5b) with a $\Delta G$ of -2.5 kcal/mol. Messenger RNAs mapped using the CR-RT-PCR method for oat *rps7* that are shorter than the predicted 1.0 kb from northern analysis are believed to result from degradation to 5'ends as 3' termini for all clones sequenced were homogeneous, mapping to the conserved region within the downstream flanking sequence for *rps7* among grasses. Maize *rps7* transcripts were also variable in length, both shorter and longer (572 to 834) than the predicted ~700 nt from northern hybridizations for *rps7* among grasses. Like oats *rps7* the variation in transcript size was also due to different length 5'UTRs. These observations suggest the presence of strong stability-conferring factors for 3'UTRs and an apparent lack of stability of 5'termini.

### 3.8.3 *Editing within rps7 leaders and tails at predicted sites is observed for both maize and oats.*

Non-coding edits at positions -68 and +73 (numbers relative to *rps7* start and stop codons respectively) were observed in 6/7 and 4/9 clones and 5/5 and 9/13 clones for oats and maize respectively. Because CR-RT-PCR products for oats and maize were generated using oligomers LB524 and LB523 (Figure 2.1) sequences for coding region

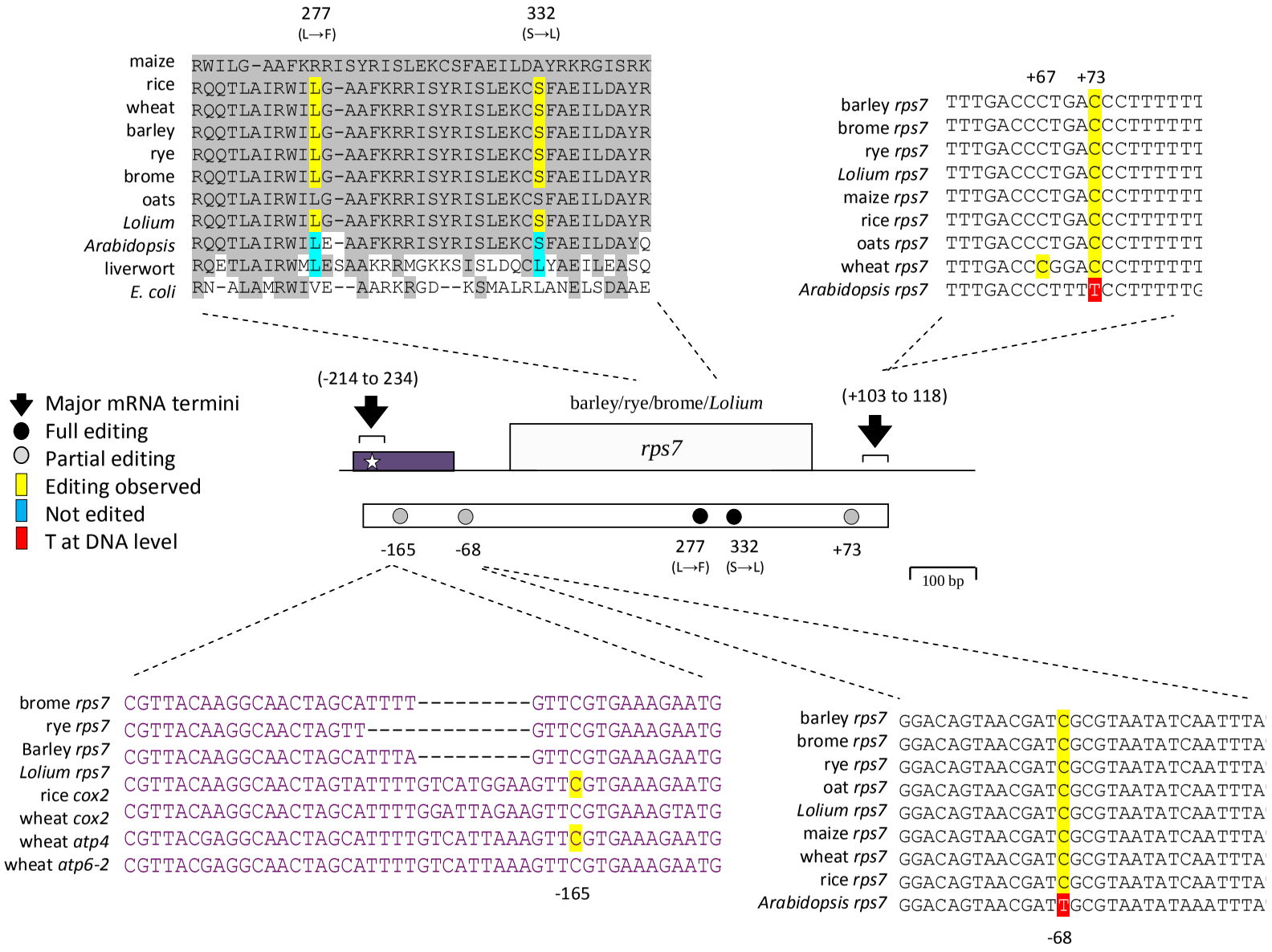
edits were not included in this analysis. RT-PCR data for the coding region edits in *rps7* for grasses rye, brome, barley and *Lolium* however was and is shown in Figure 3.6 (yellow-highlighted genomically encoded amino acids). Editing for wheat and rice *rps7* coding sequence was previously known (Zhuo and Bonen 1993; Notsu *et al*. 2002). No other editing sites for maize *rps7* mRNA apart from -68 and +73 were observed in leaders and tails when sequences from CR-RT-PCR products were compared with the complete mitochondrial genome sequence for *Zea Mays* (NCBI database). In *Arabidopsis rps7* the homologous site for position -68 among grasses (-62 in rice), is already a T and therefore does not require editing (red highlighted T Figure 3.6). All other grasses analyzed, including both maize and oats exhibit virtually complete editing at this position, regardless of stage of development.

While it is not clear if editing within UTRs serves any biological function, editing of at least position +73 in the 3'UTR of all grasses studied (shown by genomically encoded T highlighted in yellow in Figure 3.6) increases the thermodynamic stability of the stem-loop structures predicted in this region (Figure 3.4c).

Interestingly while 1 clone for wheat *rps7* mRNA from germinating embryos showed editing at position +67 (Byers *et al*. 2010), a site previously observed in wheat *rps7* (Zhuo and Bonen 1993), 4 out of 7 seedling clones for wheat *rps7* mRNA did show editing. An additional site in wheat seedling RNA at position -108 (relative to *rps7* start codon) was edited in 4/7 CR-RT-PCR clones but in no clones for embryo mRNAs. Positions +67 and -108 therefore demonstrate developmental differences in RNA editing potentially due to use of alternative machinery in later stages of development. Editing of these positions within *rps7* UTRs is not likely to happen fortuitously as flanking sequences do not demonstrate any sequence identity with homologous regions in coding sequences of other genes also known to be edited.

*Figure 3.6: Mitochondrial rps7 coding and non-coding RNA C-to-U editing among plants: wheat, rice, maize, barley, brome, rye, Lolium, oats, Arabidopsis, liverwort and the bacterium E. coli*

Coding region for *rps7* (white box) with flanking sequences (black line) and expression cassette (purple box) (Hazle and Bonen 2007b) are drawn to scale. Major mRNA termini for barley, rye, brome and *Lolium* are denoted by black arrows. Position of coding and non-coding edits (numbers relative to *rps7* start and stop codons) and corresponding amino acid conversion are given below. The position of the star indicates the promoter motif known to initiate transcription for other mitochondrial genes in wheat (Covello and Gray 1991; Ogihara *et al*. 1999). Unedited amino acid and nucleotide sequences are shown above and below respectively. Highlighted yellow letters in *rps7* sequences show homologous sites that undergo editing in the different species. Other genes with the full expression cassette are also aligned and of them only wheat *cox2* reports 5'UTR editing (yellow highlighted nucleotide) (Covello and Gray 1989) at a site homologous to position -165 upstream of *Lolium rps7* which is also edited. At position -68 upstream of *Arabidopsis rps7* a thymidine (red highlighted) is already genomically encoded. Previous analyses of wheat and rice revealed editing at the 2 coding sites and position +67 in wheat (Zhuo and Bonen 1993) and the 2 coding edits for rice (Notsu *et al*. 2002). All other sites shown in yellow were identified in this study.


277
(L→F)
332
(S→L)
maize RWILG-AAFKRRISYRISLEKCSFAEILDAYRKRGISRK
rice RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
wheat RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
barley RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
rye RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
brome RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
oats RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
Lolium RQQTLAIRWILG-AAFKRRISYRISLEKCSFAEILDAYR
Arabidopsis RQQTLAIRWILE-AAFKRRISYRISLEKCSFAEILDAYQ
liverwort RQETLAIRWMLESAAKRRMGKKSISLDQCLYAEILEASQ
E. coli RN-ALAMRWIVE--AARKRGD--KSMALRLANELSDAAE
+67 +73
barley rps7 TTTGACCCTGACCCTTTTTT
brome rps7 TTTGACCCTGACCCTTTTTT
rye rps7 TTTGACCCTGACCCTTTTTT
Lolium rps7 TTTGACCCTGACCCTTTTTT
maize rps7 TTTGACCCTGACCCTTTTTT
rice rps7 TTTGACCCTGACCCTTTTTT
oats rps7 TTTGACCCTGACCCTTTTTT
wheat rps7 TTTGACCCGGACCCTTTTTT
Arabidopsis rps7 TTTGACCCTTTTCCTTTTTG
Major mRNA termini
Full editing
Partial editing
Editing observed
Not edited
T at DNA level
(-214 to 234)
barley/rye/brome/Lolium
rps7
(+103 to 118)
-165
-68
277
(L→F)
332
(S→L)
+73
100 bp
brome rps7 CGTTACAAGGCAACTAGCATTTT---------GTTCGTGAAAGAATG
rye rps7 CGTTACAAGGCAACTAGTT----------GTTCGTGAAAGAATG
Barley rps7 CGTTACAAGGCAACTAGCATTTA---------GTTCGTGAAAGAATG
Lolium rps7 CGTTACAAGGCAACTAGTATTTTGTCATGGAAGTTCGTGAAAGAATG
rice cox2 CGTTACAAGGCAACTAGCATTTTGGATTAGAAGTTCGTGAAAGTATG
wheat cox2 CGTTACGAGGCAACTAGCATTTTGTCATTAAAGTTCGTGAAAGAATG
wheat atp4 CGTTACGAGGCAACTAGCATTTTGTCATTAAAGTTCGTGAAAGAATG
wheat atp6-2 CGTTACGAGGCAACTAGCATTTTGTCATTAAAGTTCGTGAAAGAATG
-165
barley rps7 GGACAGTAACGATCGCGTAATATCAATTTA
brome rps7 GGACAGTAACGATCGCGTAATATCAATTTA
rye rps7 GGACAGTAACGATCGCGTAATATCAATTTA
oat rps7 GGACAGTAACGATCGCGTAATATCAATTTA
Lolium rps7 GGACAGTAACGATCGCGTAATATCAATTTA
maize rps7 GGACAGTAACGATCGCGTAATATCAATTTA
wheat rps7 GGACAGTAACGATCGCGTAATATCAATTTA
rice rps7 GGACAGTAACGATCGCGTAATATCAATTTA
Arabidopsis rps7 GGACAGTAACGATTGCGTAATATAAATTTA
-68

# Chapter 4: RNA processing of mitochondrial *rps7* transcripts during development of grasses from embryo to seedling stages

## 4.1 Comments:

This chapter has been written up in a manuscript format for future publication. After looking at mRNA ends for *rps7* among grasses I became interested in processing signals for precursor species. This gene is particularly interesting because it allows us to evaluate end maturation quite extensively as there are a number of *rps7* precursor RNAs among grasses. This work demonstrates that a variety of different signals are used for conversion of higher molecular weight *rps7* precursors to mature mRNAs.

## 4.2 Data for RNA processing of room temperature vs. cold-grown wheat seedlings

Since wheat mitochondrial gene expression is regulated differently during embryo-to-seedling development our lab also became interested in how gene expression might be handled under stress conditions. Following experiments conducted by Kurihara-Yonemoto and Handa 2001 for *cox2* in wheat I exposed etiolated wheat seedlings to cold temperatures of $4^{o}$C in the dark. For room temperature growth, wheat seedlings were grown for 36 hours and 3 or 9 days. Growth in the cold was carried out for 4.5 and 6 days after allowing 4.5 and 3 days respectively for initial development at room temperature. RNA was then isolated and used for RT-PCR experiments. Direct sequencing results were analyzed to assess differences in the degree of editing at various positions within wheat *rps7* transcripts in response to cold growth. I also looked at northern hybridization profiles for mitochondrial RNA isolated from cold-grown wheat compared to RNA from room temperature wheat. My observations are explained in the addendum to chapter 4 (pg.101).

## 4.3 Abstract

Plant mitochondrial RNAs require various RNA processing events for transcript maturation, including endonucleolytic cleavage. Northern data suggests processing of ribosomal protein gene S7 precursor transcripts for grasses such as wheat, rice, rye and *Lolium* is regulated differently during plant development and analysis of *rps7* pre-mRNA

termini among these plants shows that end-cleavage occurs at the 5'-end. The relative abundance of *rps7* precursor transcripts and mRNAs among grasses was observed to be lower in seedling stages compared to germinating embryos suggesting decreased transcriptional activity and/or increased RNA turnover in later stages of development. Using circularized RT-PCR to map RNA termini we have determined that multiple discrete 5'end-cleavage events occur during maturation of *rps7* transcripts from rice, wheat and rye while upstream processing of precursor RNAs for grasses such as *Lolium* appears to be carried out by a single endonucleolytic cut. Interestingly it appears that the expression of wheat *rps7* transcripts is dependent on signals from the upstream chloroplast-origin tRNA gene $trnF^{(cl)}$. The 5'UTR terminus of a 3.4 kb precursor RNA maps exactly to the mature 3'end of the tRNA gene. In *Lolium*, the single discrete higher molecular weight species for *rps7* also maps just downstream of a transfer RNA gene, *trnS*. Precursor 5' termini for wheat, rice and *Lolium* are homogenous in length but map to non-homologous sequences between plants. Precursor RNA 3'ends are virtually identical mapping within homologous sequences to the same position as mRNA 3' termini. Furthermore from northern as well as RNA-end data it appears that some plants, like rye show different length precursor RNAs between developmental stages, in contrast to mRNAs. Our findings demonstrate the plastic nature of plant mitochondrial transcriptional units and suggest that chloroplast sequences are co-opted for transcription and processing of native mitochondrial genes.

**Keywords:** Grasses, Ribosomal protein, Precursor RNA, Mitochondria, Development, Seedlings, Germinating Embryos.

## 4.4 Introduction

When seeds leave dormancy and begin to germinate there is a period of high energy demand and rapid mitochondrial biogenesis (reviewed in Bewley and Black 1994). Specifically during the first 12hrs post-imbibition there is a sharp increase in oxygen consumption as components of the electron transport chain are activated for energy production during oxidative phosphorylation. This is followed by a lag in respiration and then at the time of radicle emergence (germination, typically ca. 18hrs post-imbibition)

there is a second respiratory burst when the mitochondrial machinery is fully functional. Studies conducted by Li-Pook-Than *et al*. 2004 using northern hybridization with wheat RNA isolated from different stages of development show that transcript profiles for respiratory chain genes (eg. *nad7*, *cox1*, *cox2*, *atp6*) parallel those of mitochondrial ribosomal rRNAs (eg. 5S, 18S, 26S) showing little to no difference in the relative abundance of steady-state levels of mRNAs. Profiles for ribosomal protein genes (eg. *rps2*, *rps3*, *rps7*) however showed a decrease in later stages (2-6 days) of seedling development. The relative abundance of precursor RNAs for intron-containing genes is also higher in earlier stages of development, as are excised introns (Li-Pook-Than *et al.* 2004), evidence supporting an inefficient coupling between transcription and splicing.

Loose consensus motifs like the tetranucleotide motif CRTA (where R is either an A or G) have been proposed for transcription initiation of monocot mitochondrial genes however are not used universally for gene expression (Kuhn *et al.* 2005). Among eudicots this tetra-nucleotide consensus sequence can be extended to a nanonucleotide motif (CRTA)AGAGA (Gagliardi and Binder 2007). In the analysis of *Arabidopsis* RNA ends, variations on this motif including (CRTA)TATAA and (CRTA)TATAG were seen (Forner *et al.* 2007). Furthermore the majority of higher plant mitochondrial promoters exhibit an A/T-rich sequence element immediately upstream of the promoter core, which has been proven essential for the full function of several dicot and monocot mitochondrial promoters *in vitro* (Dombrowski *et al*. 1999).

Plant mitochondrial RNA processing requires multiple steps including end-cleavage, intron splicing and RNA C-to-U editing. Pentatricopeptide repeat (PPR) proteins involved in organellar RNA processing are targeted to the mitochondrion and provide machinery for transcript maturation (reviewed in Zehrmann *et al*. 2011). PPR proteins involved in end-cleavage and RNA editing have been primarily identified in *Arabidopsis*. Recently RPF1, a special subclass of PPR proteins belonging to the *RF* (RESTORER OF FERTILITY) gene products has been implicated for efficient generation of a 5'end 228 nt upstream of the mitochondrial *nad4* gene (Holzle *et al*. 2011). It is estimated that there are more than 450 nuclear-encoded PPR proteins in *Arabidopsis* with the majority believed to be involved in RNA editing (Shikanai 2006). PPR proteins have binding activity for single-stranded RNA and using a somewhat

relaxed specificity for consensus-motifs are able to recognize and edit multiple C residues in several mitochondrial transcripts (Hammani *et al*. 2011).

Plant mitochondrial precursor RNAs with discrete homogeneous 5' and/or 3' termini are believed to be protected from exonucleolytic degradation at their ends (Forner *et al*. 2007). Interactions between stability conferring secondary structures and/or RNA *cis*-elements within transcript UTRs, with auxiliary proteins like PPR proteins are believed to be the mechanism for end protection, as is seen for 'higher' plant chloroplast mRNAs (reviewed in Barkan 2011). In *Arabidopsis* mitochondria 3'ends in general are very homogeneous, often exhibiting a single discrete terminus in contrast to heterogeneous 5'ends generated from multiple transcription initiation signals and/or *cis*-elements for end-cleavage (Forner *et al*. 2007). Endonucleases such as RNase P and RNase Z-like enzymes are believed to generate transcript ends at RNA secondary structures formed by tRNAs and degenerate tRNAs called t-elements (Forner *et al*. 2007). Stem-loops and double stem-loops are also thought to be involved in both endonucleolytic generation of transcript termini and RNA stability.

Mitochondrial genomes possess a mosaic of DNA sequences, derived from their endosymbiont ancestor, plastid genomes, nuclear-origin sequences like retro-elements and sequences of unknown origin (Adams and Palmer 2003). They are also highly recombinogenic and as such flanking sequences can be very different even for genes from closely-related species (Hazle and Bonen 2007a). Stretches of chloroplast genome sequences are incorporated into plant mitochondrial DNA, sometimes with copies of chloroplast tRNA genes or pieces of protein-coding sequence (Joyce *et* al. 1988). In wheat there are chloroplast-origin genes for *trnF* and *trnS*. Both tRNAs are expressed and likely involved in translation of mitochondrial RNAs (Joyce and Gray 1989). Both tRNAs in wheat are derived from the same piece of chloroplast DNA and are likely transcribed together. In rice a longer piece of the same chloroplast-origin sequence can be found upstream of *nad9* and it is thought that the mitochondrial respiratory chain gene derives expression signals from these chloroplast sequences (Nakazono *et al*. 1996).

To gain more insight into the regulation of gene expression and specifically the processing required for maturation of messenger RNAs we have determined the 5' and 3'ends of precursor RNAs for the ribosomal protein gene *rps7* from RNA isolated from

both germinating embryos and developing seedlings of several grasses. We find that mitochondrial encoded tRNA genes of both mitochondrial-origin and chloroplast-origin are recruited and act as *cis*-regulatory signals for end-cleavage. Therefore for *rps7* in some plants promoter sequences for cotranscription with upstream genes have been co-opted for gene expression. Additional promoter motifs downstream of pre-mRNA termini in cases such as *Lolium* suggests the possible use of multiple promoters for *rps7* expression, like has been seen for various genes in *Arabidopsis* mitochondria (Kuhn *et al*. 2005). Analysis of UTR sequences using RNA-folding programs has allowed us to also identify potential secondary structures for end-stability or site-specificity for end-cleavage. Precursor RNA profiles in general are complex with different populations of higher molecular weight species demonstrating either discrete or heterogeneous 5'ends. Some plants even exhibit differences in northern hybridization signals for *rps7* between stages of plant development. RNA editing is also regulated differently during development as some sites appear to be recognized more frequently in seedling stages.

## 4.5 Materials and Methods

### 4.5.1 *Mitochondrial RNA and DNA isolation*

Mitochondrial RNA was isolated from dormant seeds, germinating embryos and etiolated seedlings of wheat (*Triticum* aestivum var. Frederick), rice (*Oryza sativa* var. Drew), rye (*Secale cereale* var. Gazelle) and annual ryegrass (*Lolium multiflorum*), using previously described procedures (Subramanian *et al*. 2001). Surface-sterilized embryos were dissected prior to germination in the dark for 36 hrs before RNA extraction. Seedlings were derived from sterilized un-dissected seeds grown in vermiculite at room temperature in the dark for 6-10 days. Wheat seeds were kindly provided by Dr. R. Pandeya and Dr. T. M. Choo (Agriculture and Agri-food Canada). Seeds for other grasses were commercially purchased.

### 4.5.2 *Mitochondrial RNA analysis*

For northern blot analysis, mitochondrial RNA samples (approximately 5 μg per lane) were electrophoresed on 1.2% agarose/formaldehyde gels and after membrane

transfer, hybridized with $^{32}$P-end-labelled *rps7*-specific oligomer probes 5'ACTGAATGAGGAAGAGCTCC 3' (LB29) and 5'GTTCAGTTCGAGCTAGGCGGTG 3' (LB37) using standard procedures (Sambrook *et al*. 1989).

To simultaneously map the 5' and 3' termini of *rps7* precursor transcripts, the CR-RT-PCR strategy (cf. Kuhn and Binder 2002; Calixte and Bonen 2008) was used. To circularize transcripts, approximately 5 µg of mitochondrial RNA was incubated for ½ hour at 37$^{o}$C with T4 RNA ligase (New England Biolabs) in the presence of RNAsin (Promega). After phenol extraction and ethanol precipitation, the self-ligated RNA was heated at 65$^{o}$C for 5 min with various *rps7*-specific RT primers; 5' TTCAGTTCGAGCTAGGCGGTG 3' (LB524) for rice, 5'CCGTGAAACACATAGGCTCC 3' (LB577), 5'GGAGGTGCGTAGTGTCTTAC 3' (LB578), and 5'TTTGCGGAAACCACTACTGG 3' (LB743) for wheat, 5'AGCATTTCGTCGCTTGCTAC 3' (LB704), and 5' GCTGGTCCTTGTTGACTCGC 3' (LB717) for rye and 5'ATTAAGGTCGTCACCCTCCG 3' (LB705) for *Lolium*, prior to cDNA synthesis with M-MLV reverse transcriptase (Invitrogen) for 1½ hours at 37$^{o}$C. For subsequent PCR amplification, the above primers and one located in the 3' coding region, namely 5' CGCATTTCAGATGGTGGTAAAGTG 3' (LB523) were used, and amplicons were checked using various nested primers. Synthetic oligomers (Invitrogen) were designed based on the wheat mitochondrial *rps7* gene region [AP008982], rye mitochondrial *rps7* gene region [HM581685], *Lolium* mitochondrial *rps7* gene region [HM581683] and size marker ladders were from NBI Fermentas.

### 4.5.3 *Cloning and sequencing of cDNA*

PCR and RT-PCR products were gel-purified using Ultra-Clean15 (MoBio Laboratories) prior to ligation into the pGEM-T Easy vector (Promega) and then cloned. The resulting recombinant plasmid DNAs were isolated using the QIAprep spin Miniprep kit (Qiagen) and automated sequencing was performed by the Ottawa Health Research Institute DNA sequencing facility.

### 4.5.4 *Sequence analysis and modeling of RNA secondary structures*

BLAST searches (Altschul *et al*. 1990; http://www.ncbi.nlm.nih.gov/blast) were used for mitochondrial *rps7* sequence comparison with the following genomes; maize (mt), spring wheat (mt), yumai wheat (mt), K-type yumai wheat (mt), wheat (chl), rice (mt), bamboo (mt) and *Arabidopsis* (mt). For a list of accession numbers (with publication) used see Table 2.2. The search for possible RNA secondary structures in the *rps7* UTRs was conducted using the mfold program version 3.2 (Zuker 2003; http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html).

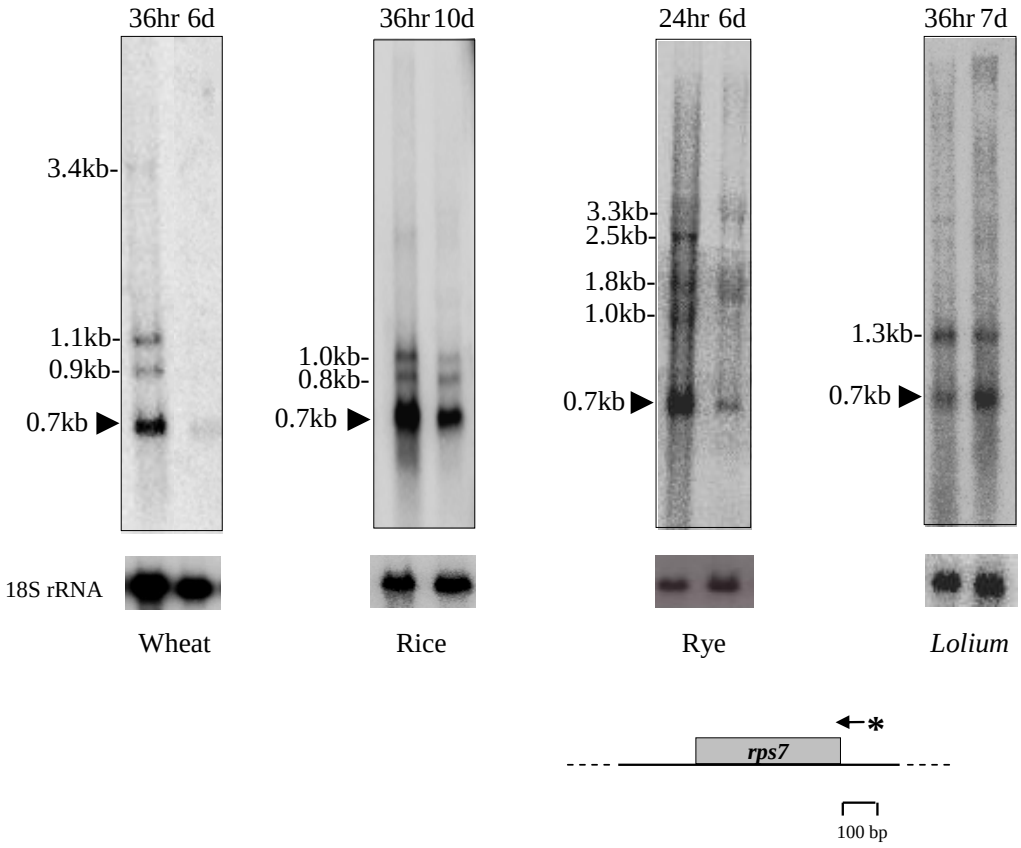
## 4.6 Results

### 4.6.1 *Developmental differences in gene expression among wheat, rice, rye and Lolium rps7*

Profiles for northern hybridization using an oligomer probe specific to a region of the *rps7* locus identical among wheat, rice, rye and *Lolium* show multiple higher molecular weight species in addition to mature mRNAs (Figure 4.1). In wheat a minor RNA species of ~3.4 kb can be seen only for northern hybridizations using RNA isolated from germinating embryos (36hr, Figure 4.1) as the relative abundance of precursors in the seedling stage (6d) is reduced. Species for wheat *rps7* of ~1.1 kb and ~0.9 kb were also more prominent in embryo RNA. In rice, higher molecular weight species of ~0.8 kb and ~1.0 kb were seen using northern hybridization for both embryo and seedling RNA (36hr and 10d respectively) while in rye, differences between stages were observed (Figure 4.1). In rye, species of roughly 1.0 kb, 1.8 kb and 2.5 kb are visible for embryo RNA (24hr) while minor species of 1.8 kb and 3.3 kb are visible for RNA isolated from 6d seedlings (Figure 4.1). Northern hybridizations for wheat, rice and rye *rps7* transcripts show a decrease in the relative abundance of all *rps7* RNA species between stages. In *Lolium*, *rps7* precursors (~1.3 kb) as well as mRNAs (~0.7 kb) are present in relatively similar stochiometries between stages (Figure 4.1). Hybridization experiments using different oligomer probes and independent mitochondrial RNA preps show very similar RNA profiles for all these grasses.

*Figure 4.1: Northern analysis of mitochondrial rps7 among wheat, rice, rye and Lolium.*

Blots were hybridized with a $^{32}$P-end-labelled *rps7*-specific oligomer probe LB29 (arrow with asterisk in schematic). Lanes contain RNA isolated from germinating embryos (24hr, 36hr) and developing seedlings (6d, 7d, 10d). 18S rRNA loading controls are indicated below and size markers in kilobases are shown on the left.


36hr 6d
36hr 10d
24hr 6d
36hr 7d
3.4kb-
1.1kb-
0.9kb-
0.7kb
1.0kb-
0.8kb-
0.7kb
3.3kb-
2.5kb-
1.8kb-
1.0kb-
0.7kb
1.3kb-
0.7kb
18S rRNA
Wheat
Rice
Rye
Lolium
rps7
100 bp

Using CR-RT-PCR we have simultaneously mapped 5' and 3' UTR termini for pre-mRNAs for wheat, rice, rye and *Lolium rps7* (Table 4.1). While *rps7* precursor RNA 3'ends were homogeneous (90% of termini mapping within +114 to +119, numbers relative to *rps7* stop codon), with few exceptions among clones, 5'ends were not. Wheat *rps7* transcript ends for the 2 major and 1 minor higher molecular weight species seen in northern analysis (Figure 4.1) have homogeneous 5'UTR termini at positions roughly 0.4 kb, 0.6 kb, and 2,9 kb upstream of the *rps7* start codon (Table 4.1). In rice, *rps7* 5'ends were identified for 2 precursor species and mapped to similar positions (Table 4.1). Rice transcript termini however did not demonstrate the same degree of homogeneity as 5'ends in wheat and 3'termini among plants. *Lolium* northern hybridization shows a single HMW species for *rps7* of about ~1.3 kb (Figure 4.1). CR-RT-PCR data supports this with homogeneous transcript 5'ends mapping 765 nt (5'ends within 10 nt among clones, with few exceptions) upstream of the *rps7* start codon.

Unlike 5'ends among grasses wheat, rice and *Lolium*, pre-mRNA species for rye *rps7* are heterogeneous (Table 4.1). Furthermore data for RNA isolated from both germinating embryos and seedlings for rye also shows that transcript 5'ends are heterogeneous between these stages of development (Table 4.1). Two different size classes for rye *rps7* transcripts isolated from embryo RNA have been identified; 2.4 to 2.6 kb and 1 to 1.8 kb, while species ranging in sizes from 2 to 2.1 kb have been identified for seedling RNA. We believe that transcript ends are not derived from degraded RNA as 3'UTR termini for all plants, regardless of developmental stage, are homogenous, mapping to homologous sequences. The possibility of heterogeneous transcript 5'ends resulting from *in vitro* 5'exo-activity however cannot be excluded.

Individual transcripts having non-encoded nucleotides are designated by total length of RNA species in Table 4.1. Of the 5 clones with non-encoded nucleotides at the position of the 5-to-3' UTR junction, extensions consisted primarily of A tails but other nucleotides were seen as well. One clone among rye precursor transcripts had a tail of 15 nt with a fairly even distribution of all As, Cs, Ts and Gs. All clones were derived from RNA not treated with tobacco acid pyrophosphatase and likely do not represent primary transcripts, unless the triphosphate of a primary transcript was removed by slight damage to the RNA prior to RNA ligation (Forner *et al*. 2007).

*Table 4.1: Transcript termini for mitochondrial rps7 precursor transcripts from wheat, rice, Lolium and rye*

CR-RT-PCR data for *rps7* precursor transcript termini is summarized here for all grasses surveyed. For some but not all plants, mRNA 5' and 3'ends were analyzed for both seedling and germinating embryo RNA.

**Table 4.1.** Positions of 5' ends for mitochondrial *rps7* precursor RNAs from various grasses

| **Transcript termini** | | | | |
|---|---|---|---|---|
| Plant and stage of development for particular *rps7* RNA species | Total RNA lengths are given in nucleotides (nt). Positions of 5'UTR termini given relative to *rps7* start codon (-ve), unless otherwise indicated for 5' truncated mRNAs. Positions for 3'UTR termini given relative to *rps7* stop codon (+ve). RNA species with non-encoded nucleotides indicated, with length of particular RNA species in brackets. | | | |
| | **RNA length (nt)** | **(-) 5'UTR** | **(+) 3'UTR** | **Non-encoded NTs** |
| **Wheat** | | | | |
| Embryo Precursor Species 3 | 3494, 3493, 3492, 3491, 3490, 3413 | 2928, 2928, 2927, 2927, 2926, 2928 | 119, 118, 118, 117, 117, 38 | |
| Embryo Precursor Species 2 | 1210, 1150, 1150, 1149, 1147, 1146 | 647, 587, 587, 586, 584, 586 | 116, 116, 116, 116, 116, 113 | AAA (1146) |
| Seedling Precursor Species 2 | 1151, 1151, 1150, 1150, 1150, 1150, 1150, 1150, 1149, 1147, 1146, 1059 | 586, 586, 586, 586, 586, 586, 586, 586, 586, 586, 583, 499 | 118, 118, 117, 117, 117, 117, 117, 117, 116, 114, 116, 113 | AA (1059) |
| Embryo Precursor Species 1 | 989, 983 | 424, 421 | 118, 115 | A (989) |
| **Rice** | | | | |
| Embryo Precursor Species 2 | 1048, 1047, 1028 | 484, 482, 463, 484 | 117, 118, 118 | |
| Embryo Precursor Species 1 | 886, 851, 836 | 325, 290, 270 | 114, 114, 119 | |
| ***Lolium*** | | | | |
| Embryo Precursor Species | 1329, 1329, 1324, 1280, 1252 | 765, 764, 760, 760, 765 | 117, 118, 117, 73, 40 | T (1329) CGAGG (1280) |
| Seedling Precursor Species | 1400, 1329, 1328, 1327, 1326, 1318, 1263, 1252 | 765, 764, 765, 764, 765, 764, 765, 695 | 188, 118, 116, 116, 114, 107, 51, 110 | AAAACAAAA (1252) |
| **Rye** | | | | |
| Embryo Precursor Species 2[a] | 2621, 2613, 2604, 2604, 2468, 2466, 2466, 2465, 2465 | 2052, 2048, 2041, 2041, 1902, 1902, 1902, 1900, 1900 | 122, 118, 116, 116, 119, 117, 117, 118, 118 | |
| Seedling Precursor Species 1[a] | 2161, 2092 | 1469, 1527 | 245, 118 | |
| Embryo Precursor Species 1 | 1884, 1625, 1217, 1056 | 1321, 1062, 655, 493 | 116, 116, 115, 115 | GACGACTATGATGGT (1625) |

[a]Sequences correspond to data for cDNA 5'/3' ligated junction clones.

### 4.6.2 *Relative amounts of wheat rps7 precursors differ over the course of plant development*

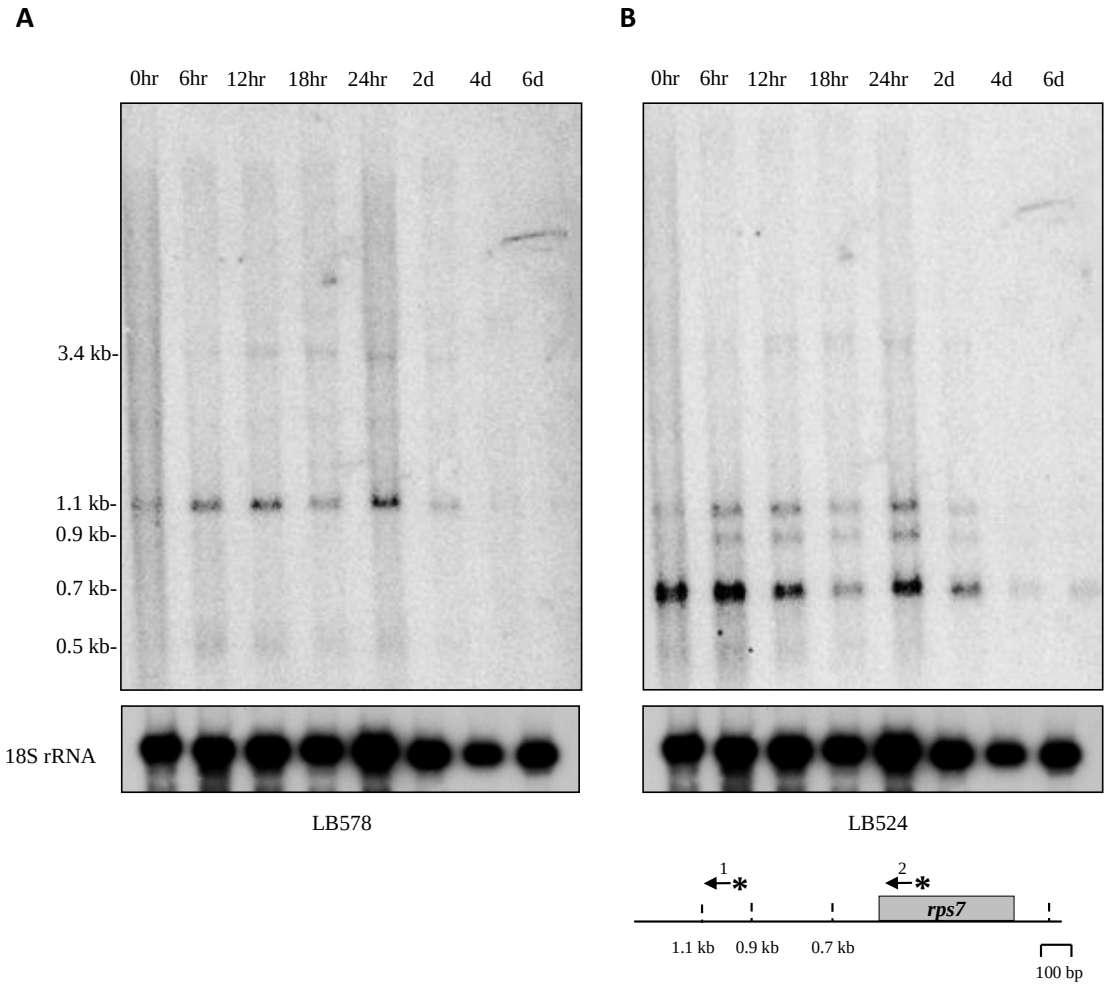
In a northern hybridization of wheat RNA isolated from various stages of seed development (Figure 4.2a,b) from dormant seeds to developing seedlings (i.e. 0hr, 6hr, 12hr, 18hr, 24hr, 2d, 4d, 6d) we see differences in not only relative abundance of mature *rps7* mRNAs (as seen previously in Li-Pook-Than *et al*. 2004) but also in the amount of precursor species. Using probes specific to only wheat *rps7* precursors (LB578, Figure 4.2a) and the coding region for wheat *rps7* (LB524, Figure 4.2b) we see distinct profiles showing processing events at the 5'end of the higher molecular weight transcripts. Only larger precursors; 3.4 kb and 1.1 kb, are visible (in virtually all embryo RNA isolates) when a probe within the 5'UTR is used in northern hybridization (Figure 4.2a). Two additional species (0.9 kb precursor and 0.7 kb mRNA) are visible when a coding region oligomer (LB524) is used. Interestingly when loading controls (hybridization for 18S rRNA) are compared, the decrease in relative abundance of both precursor species and *rps7* mRNAs is apparent as seeds germinate (post-imbibition, 6hr) and leave dormancy. Stored mRNAs in dormant seeds (0hr) are abundant however signals for HMW species are of low-intensity. Abundant higher molecular weight species (1.1 kb, 0.9 kb) as well as mRNAs (0.7 kb) at 24 hours (Figure 4.2b) compared to lower relative levels (18S rRNA loading controls) of transcripts for 6d seedlings suggests that in later stages of development RNA turnover increases and/or transcription is reduced. An additional signal seen at approximately 0.5 kb is visible when using the 5'UTR oligomer probe and is consistent with the upstream byproduct generated from processing of the 1.1 kb species to the 0.7 kb species and is likely tagged for degradation. Interestingly this band is less intense in later stages (Figure 4.2) as might be expected if during this time RNA turnover is occurring at a higher rate.

### 4.6.3 *Upstream sequence analysis of rps7 pre-mRNA*

Upstream of *rps7* in wheat (fully sequenced mtgenome, Ogihara *et al*. 2005) there are 2 discrete stretches of chloroplast-origin sequences (Gray shading Figure 4.3a,b). Roughly 700 bp upstream of the wheat *rps7* start codon, about 100 bp in front of the

*Figure 4.2: Northern analysis of wheat rps7 during seed development*

**a**) Northern blot hybridized first with the $^{32}$P-end-labelled wheat *rps7*-specific oligomer probe LB578 (arrow 1 with asterisk in b). **b**) The same blot was then hybridized with an *rps7* coding region-specific oligomer probe LB524 (arrow 2 with asterisk in b). Lanes contain RNA from wheat embryos and seedlings (hr = hour, d = day). 18S rRNA loading controls are indicated below and size markers in kilobases are shown on the left. Wheat *rps7* 5'/3'UTR termini (dotted lines) for mRNA and precursor transcripts (lengths given in kb) are given in b. Blots prepared by J. Li-Pook-Than.

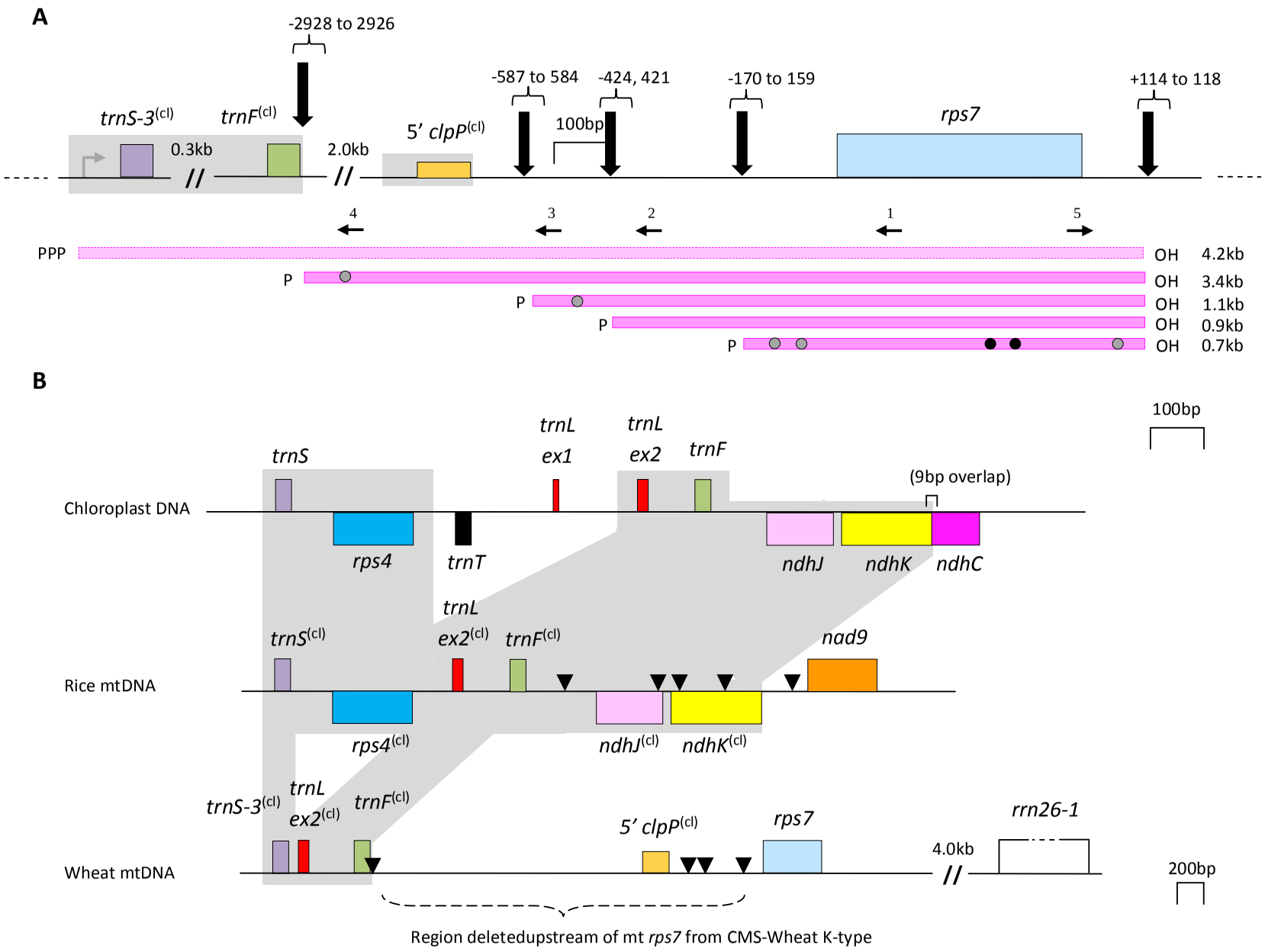

A
B
0hr 6hr 12hr 18hr 24hr 2d 4d 6d
3.4 kb-
1.1 kb-
0.9 kb-
0.7 kb-
0.5 kb-
18S rRNA
LB578
LB524
1
2
rps7
1.1 kb
0.9 kb
0.7 kb
100 bp

5'UTR terminus for the 1.1 kb precursor (black arrows Figure 4.3a), there is a stretch of chloroplast sequence 104 bp long corresponding to the 5'end of the *clpP* gene (Figure 4.3a). In front of the piece of *clpP* coding sequence there is an additional 80 bp homologous to *clpP* upstream flanking sequence. 2.9 kb upstream of the wheat *rps7* start codon there is an additional stretch of chloroplast-origin sequence of ~680 bp encoding a full copy *trnS-3*$^{(cl)}$, *trnF*$^{(cl)}$ and exon 2 of *trnL*$^{(cl)}$. An internal deletion in the ancestral wheat mitochondrial genome has removed exon 1 of *trnL*$^{(cl)}$, *trnT* and *rps4* (Figure 4.3b), sequences present at the corresponding locus in the chloroplast genome among flowering plants (Figure 4.3b). The 5'UTR terminus for the 3.4 kb wheat *rps7* precursor transcript maps within 1 to 2 nucleotides of the mature 3'end of the *trnF*$^{(cl)}$ (Figure 4.4b). Both chloroplast-origin tRNAs are expressed in the wheat mitochondrial genome (Joyce and Gray 1989) and may provide initiation signals (bent gray arrow Figure 4.3a) for transcription of *rps7*. A 4.2 kb polycisronic transcript containing the full structural RNA sequence for both chloroplast-origin tRNAs would be subsequently processed by RNase P and Z-like enzymes at the 5' and 3'ends of the tRNA genes respectively. The mature 5' end of the wheat *rps7* mRNA can then be generated by endonucleolytic cleavage.

Interestingly ~360 bp upstream of the rice *nad9* translation start codon there is a longer stretch of the same chloroplast-origin sequence (Nakazono *et al*. 1996) found upstream of wheat *rps7* (Gray shading Figure 4.3b). The genes for chloroplast *ndhK*, *ndhJ* and *rps4* are included in this upstream sequence in rice and are in their corresponding orientation and organization as seen in the chloroplast genome (Figure 4.3a,b). However, like in wheat the upstream exon for *trnL* and the gene for *trnT*, sequences that would have been present upon acquisition of this region of chloroplast DNA are missing upstream of rice *nad9*. Termini for *nad9* transcripts as described in Nakazono *et al*. 1996 as well as RNA ends for wheat *rps7* are shown by black arrowheads in Figure 4.3b. The deletion of *trnT* and exon 1 of *trnL* upstream of rice *nad9* likely occurred in the common ancestor to wheat and rice. In wheat a lineage-specific rearrangement placed the chloroplast sequence corresponding to the region from the *trnS* to the *trnF*, upstream of *rps7*. A subsequent deletion then likely occurred removing the *rps4* coding sequence. The chloroplast-origin sequence upstream of rice *nad9* is therefore more ancestral-like than that found upstream of *rps7* in wheat. Sequences for deleted

*Figure 4.3: Sequences flanking transcript termini in wheat rps7*

**a**) Coding sequences are represented by different coloured boxes with gene names and are full-length unless otherwise indicated. Vertical arrows represent transcript termini (values relative to *rps7* start codon) and horizontal arrows represent primers used for RT (1, LB524) and CR-RT-PCR (1; LB524, 2; LB577, 3; LB578, 4; LB743, 5; LB523). Precursor transcripts and mRNAs (lengths indicated) are shown as pink boxes. A potential consensus motif (CATA) for transcription of the hypothetical 4.2 kb precursor (hatched pink rectangle) is shown by the bent gray arrow. Gray and black circles represent partially and fully edited sites within RNAs respectively (editing sites in precursor 5'/3'UTRs not shown since very low-level). **b**) Origin of chloroplast sequence in front of rice *nad9* and wheat *rps7*. Sequences upstream of rice *nad9* and wheat *rps7* homologous to chloroplast sequences are shown by gray shading. Genes homologous to those found in the chloroplast genome are indicated with a 'cl' and gene orientation is depicted as above or below the line. Transcript termini as previously seen for rice *nad9* (Nakazono *et al*. 1996) are shown as black arrowheads. The upstream sequence missing in front of *rps7* in the mitochondrial genome of a K-type CMS line (GU985444, Liu *et al*. 2011) is indicated by a dotted bracket. Schematics are to scale unless otherwise indicated (hatched bars or dotted lines).


A
-2928 to 2926
-587 to 584
-424, 421
-170 to 159
+114 to 118
trnS-3(cl)
trnF(cl)
0.3kb
2.0kb
5' clpP(cl)
100bp
rps7
4
3
2
1
5
PPP
P
OH 4.2kb
OH 3.4kb
OH 1.1kb
OH 0.9kb
OH 0.7kb
B
100bp
trnS
trnL ex1
trnL ex2
trnF
(9bp overlap)
Chloroplast DNA
rps4
trnT
ndhJ
ndhK
ndhC
trnL ex2(cl)
trnS(cl)
trnF(cl)
nad9
Rice mtDNA
rps4(cl)
ndhJ(cl)
ndhK(cl)
trnS-3(cl)
trnL ex2(cl)
trnF(cl)
5' clpP(cl)
rps7
rrn26-1
4.0kb
Wheat mtDNA
200bp
Region deletedupstream of mt rps7 from CMS-Wheat K-type

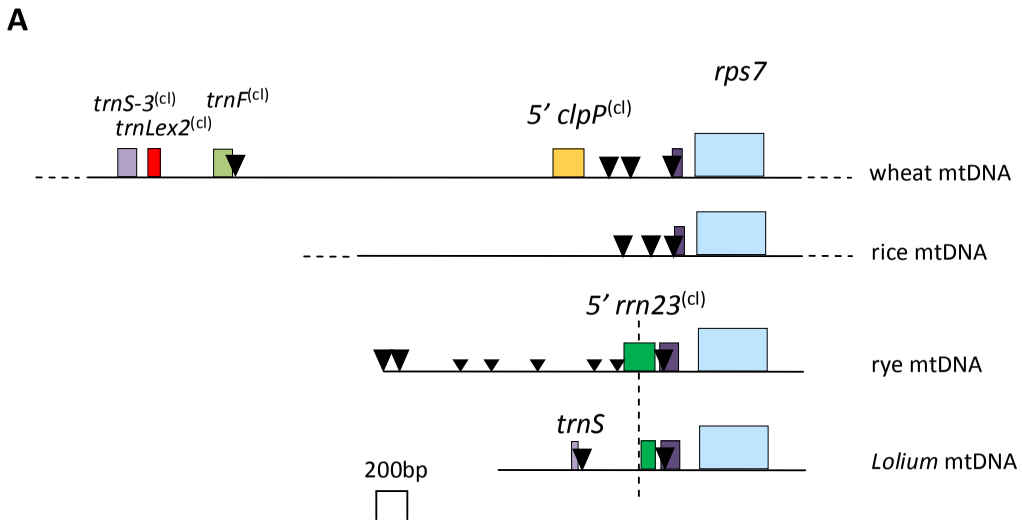
regions upstream of *rps7* in wheat and *nad9* in rice as well as *ndhJ* and *ndhK* in wheat are not present elsewhere within the mitochondrial genomes for these plants.

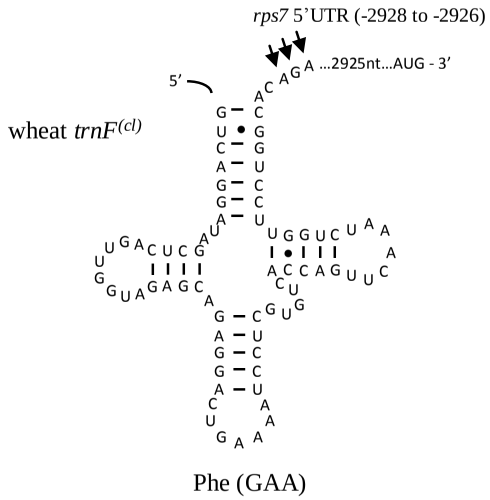
Upstream of *Lolium*, barley, rye and brome *rps7* (GenBank accession numbers HM581683, HM581684, HM581685, HM581686 respectively, Byers *et al*. 2010) there is a sequence corresponding to an expression cassette found in front of several cereal mitochondrial genomes in either full or half copy (Hazle and Bonen 2007b). Immediately upstream of the expression cassette in *Lolium*, barley, rye and brome is a sequence corresponding to the 5'end of the mature *rrn23*$^{(cl)}$ gene. In barley, rye and brome the sequence is 175 bp long while in *Lolium* only 65 bp at the 3'side are homologous (dotted line and green box in Figure 4.4a). Transcript termini for the most predominant mRNAs (Byers *et al*. 2010) and precursor species for rye and *Lolium* are shown by black arrowheads in Figure 4.4a. Because barley and brome *rps7* show no higher molecular weight species from northern analysis of embryo RNA (Byers *et al*. 2010) they were excluded from precursor transcript analysis by CR-RT-PCR. In *Lolium* at the position of the breakpoint between *Lolium* and barley, rye and brome *rps7*, there is 1.2 kb (total sequence available from GenBank accession) of upstream sequence which is homologous to sequences 810 bp upstream of the wheat *nad7* start codon (wheat genome sequence, accession AP008982, Ogihara *et al*. 2005) and also contains the gene for tRNA serine. This tRNA serine is homologous to the second copy *trnS-2* in the wheat mitochondrial genome and is of mitochondrial origin, not chloroplast. The 5'UTR terminus for the single precursor species seen for *Lolium rps7* maps just 12 nt downstream of the amino acid acceptor stem (Figure 4.4c) of the *trnS* gene.

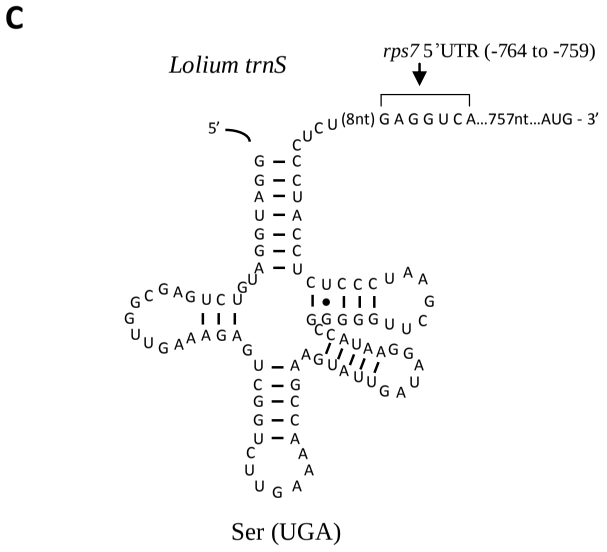
Upstream of the breakpoint between *Lolium*, barley, rye and brome for about 800 bp (limit of sequence data from GenBank accessions) barley, rye and brome are virtually identical (98%) (Byers *et al*. 2010). 100 bp upstream of the extended piece of 5' *rrn23*$^{(cl)}$ sequence in barley, rye and brome (green box in Figure 4.4a) there are ~40 bp homologous to spacer sequence from the wheat mitochondrial genome (accession AP008982). Immediately upstream an additional 100 bp homologous to intergenic DNA from CMS wheat (accession GU985444), and ~450 bp homologous to spacer sequences in both wheat (accession AP008982) and bamboo (accession EU365401) are also present in front of barley, rye and brome *rps7*. CR-RT-PCR data for *rps7* precursor transcripts in

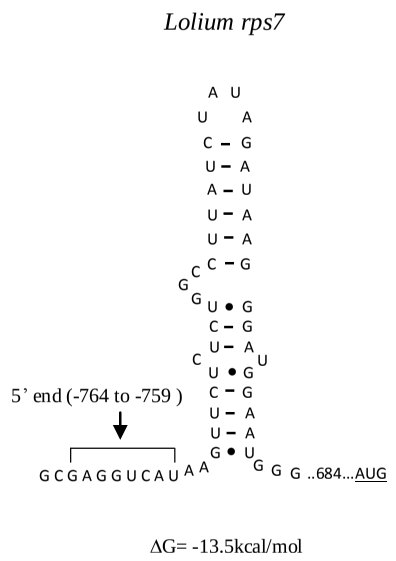
*Figure 4.4: Upstream flanking sequences for mitochondrial rps7 among wheat, rice, rye and Lolium and potential secondary structures important for RNA processing*

**a**) *rps7* coding and flanking sequences (pale-blue box) as well as upstream chloroplast-origin (cl) sequences (coloured boxes) are drawn to scale and transcript ends are denoted by large and small (major vs. minor) arrowheads. A piece of chloroplast sequence corresponding to the 5' region of *rrn23* is shown by the green box. The breakpoint in DNA homology between rye and *Lolium* is indicated by the doted line. **b**) RNA secondary structure of the chloroplast-origin *trnF* gene upstream of *rps7* in the wheat mitochondrial genome. **c**) RNA secondary structure of mitochondrial *trnS* upstream of *rps7* in *Lolium*. Amino acid encoded and anti-codon triplets are given for tRNA genes. **d**) Stem-loop structure downstream of *Lolium rps7* precursor transcript termini with value for Gibbs free energy indicated below. Transcript termini are shown as black arrows with positions relative to the *rps7* start codon. Stem-loop structures were determined using the 'mfold' software (Zuker 2003).


A
rps7
trnS-3(cl)
trnF(cl)
trnLex2(cl)
5' clpP(cl)
wheat mtDNA
rice mtDNA
5' rrn23(cl)
rye mtDNA
trnS
Lolium mtDNA
200bp


B
rps7 5'UTR (-2928 to -2926)
...2925nt...AUG - 3'
5'
wheat trnF(cl)
Phe (GAA)


C
Lolium trnS
rps7 5'UTR (-764 to -759)
5'
(8nt) G A G G U C A...757nt...AUG - 3'
Ser (UGA)


D
Lolium rps7
5' end (-764 to -759 )
G C G A G G U C A U A A G G ..684...AUG
ΔG= -13.5kcal/mol

rye provide an additional ~900 nt of sequence data in front of the upstream *HindIII* site (1,157 bp upstream of rye *rps7* start codon, accession HM581686) used for generation of DNA sequence in the inverse PCR method (Byers *et al*. 2010). Roughly 380 and 460 bp of this sequence is homologous to sequences found in the bamboo mitochondrial genome. Entire chunks or shorter segments of these sequences can be found at single or multiple places in maize mitochondria (accession AY506529). Upstream of the homogeneous mRNA 5'end for rice *rps7* (Byers *et al*. 2010) there is a stretch of sequence about 70 bp long that corresponds to the 5'UTR (299 nt in length from CR-RT-PCR end-mapping experiments; unpublished data) of wheat *rpl5*. These observations illustrate the frequency of lineage-specific rearrangements in plant mitochondrial genomes and demonstrate that bits and pieces of spacer DNA that correspond to regulatory regions of other genes might be recruited for RNA processing.

### 4.6.4 *C-to-U RNA editing in rps7 precursors among grasses*

Using sequence data from CR-RT-PCR experiments it is possible to assess RNA editing within precursor transcript coding sequence and UTRs depending on primer location for PCR amplification (Figure 4.3a horizontal black arrows). We observed low-level editing of wheat *rps7* precursor RNAs at positions -439, -490 and -577 in the 1.1 kb transcript and -2503 in the 3.4 kb transcript (numbers relative to the *rps7* start codon): 1/12 seedling, 2/12 seedling, 1/6 embryo and 2/6 embryo clones were edited at these sites respectively. Editing was also observed at positions –836 and –855 (1/2 and 2/2 embryo clones respectively) in rye and positions -233, -259 and –403 (2/3, 2/6 and 1/6 embryo clones respectively) in rice. No 5'UTR edits were observed for precursor RNAs for *Lolium rps7*. As predicted from previous work on editing in grass *rps7* mRNA 3'UTRs (Byers *et al*. 2010), position +67 in wheat and position +73 (numbers relative to *rps7* stop codon) among grasses were partially edited in precursor RNAs as well. In wheat however editing at position +67 was only observed in 1 embryo clone. Editing of this position was not observed in precursor transcripts from seedling RNA, unexpected as unpublished data (chapter 3 addendum) shows an increase in the degree of editing at this position in wheat *rps7* mRNAs isolated from seedlings. CR-RT-PCR analysis showed no differences in mRNA ends between embryo and seedling stages (Figure 3.2). Because clone sequences

correspond to CR-RT-PCR amplicons there is no risk of amplification of contaminating DNA. Among grasses whose precursor transcript termini were mapped using CR-RT-PCR, 75%, 90% and 73% of the 5'UTRs were sequenced for rice, rye and *Lolium* but for the 3.4 kb transcript in wheat only 37% was examined. These numbers consider flanking sequences previously observed for *rps7* mRNA UTRs among grasses (Byers *et al*. 2010).

### 4.6.5 *Possible Secondary structures as models for RNA processing of precursor RNAs*

RNA secondary structures such as stem loops, double stem-loops, t-elements and 'true' tRNA genes can act as *cis*-elements for processing of plant mitochondrial transcripts (Forner *et al*. 2007; Kuhn and Binder 2002). Upstream of the wheat 3.4 kb and the *Lolium* 1.3 kb *rps7* pre-mRNAs there are the *trnF*$^{(cl)}$ and *trnS* genes respectively (Figure 4.4b,c). Both RNA secondary structures likely serve as recognition sites for an RNase Z-like activity. In wheat this cleavage event would generate the 3.4 kb transcript 5'UTR termini however in *Lolium* the cut would be made 12 nt upstream of the 1.3 kb precursor transcript 5'end (ends shown as black arrows with coordinates for termini in Figure 4.4b,c) implying that some exonuclease activity may occur before the transcripts stable 5'end is generated. Sequences just downstream of the stable 5'end for the 1.3 kb precursor for *Lolium rps7* can be folded into a stem-loop structure (Figure 4.4d) using the 'mfold' software (Zuker 2003). With a convincing Gibbs free energy of -10.0 kcal/mol compared to -3.4 kcal/mol for the stability conferring structure proposed for the 3'UTR (Figure 3.4), this structure may provide stability for the *rps7* precursor transcript in *Lolium* after endonucleolytic generation of its 5' end. Sequences highlighted in green and red letters in Table 4.2 correspond to the *trnF*$^{(cl)}$ and *trnS* genes with downstream flanking sequence in black letters. The transcript termini are given in large bold font with distance to start codon indicated for individual plants (only those termini that fit within the window shown for individual *rps7* genes and flanking sequences are shown).

The lower molecular weight precursors for wheat *rps7* also possess potential RNA secondary structures near their 5'ends. The 1.1 kb species (-584 to -587, position of 5'UTR termini relative to *rps7* start codon) has a stem-loop structure upstream of the positions mapped for 5'ends and if generated by an endonuclease would require RNase Z-like activity. The 0.9 kb species has a stem-loop structure downstream of its mapped 5'

*Table 4.2: Positions of 5' ends for mitochondrial rps7 precursor RNAs among wheat, rice, Lolium and rye.*

A schematic showing sequences flanking precursor transcript termini among grasses. Potential stem-loop structures, consensus sequences for transcription initiation and upstream sequences homologous to those of other mitochondrial and chloroplast genes are illustrated here.

**Table 4.2** Sequences surrounding 5' ends of mitochondrial *rps7* precursor transcripts from various grasses

| **Transcript termini** | | |
|---|---|---|
| Plant and stage of development for particular *rps7* RNA species | 5' Termini and DNA[a] flanking sequence (5' and 3'). RNA ends are in bold letters with potential promoter motifs (CRTA(TA), where R=A or G) underlined. Possible stem-loop (SL) structures upstream or downstream of RNA ends are double underlined. Transcript termini near tRNA or SL structures likely generated by RNase P or Z cleavage at either the structures' 5' or 3' end respectively. | Remarks |
| **Wheat** | | |
| Embryo Precursor Species 3 | CGTGTCACCAGTTCAAATCTGGTTCCTGGCAC**AGA**ACGAACAATGAATGCCTTTTCGGGAAAGAAAAGGGCCACATATTTTTTT…..2876…..ATG | Upstream *trnF*$^{cl}$ gene in green letters, Rnase Z |
| Embryo Precursor Species 2 | AGAAAGA**A**GTGCTTTTCCACCGGTATCAACATTCTATCTGGAATCAGTAGCGATAATTGTTGTATTG**TG**A**G**CCAGCCCCGTAAG…..570…..ATG | SL, RNase Z |
| Seedling Precursor Species 2 | TAGAAAGAAGTGCTTTTCCACCGGTATCAACATTCTATCTGGAATCAGTAGCGATAATTGTTGTATTG**T**G**AG****C**AGCCCCGTAAG…..570…..ATG | |
| Embryo Precursor Species 1 | CTTGCTTCTTCGAATCTCGAAATAACATATAG**A**AA**G**TGTTTCTGTGATGAGACCATTCTCGATCGATAAATGCGATAGGAGCCTAT…...370....ATG | SL, RNase P |
| **Rice** | | |
| Embryo Precursor Species 2 | AATCGGAGGGGCCGGCCGGATTCCCTACAAATAAATTCTGTATTAA**G**A**A**ACGAAGGAAGAGTGCGCG**T**TCGGACCTCAACTAGTC….445....ATG | Rice, wheat, maize and oats *rps7* possess half purple cassettes while *Lolium*, rye, barley and brome have full-length sequences. |
| Embryo Precursor Species 1 | TGTCTGTCTTATG**T**CATCCTGGTGATTTATCATCGGAATAGCTCAGTT**C**GAGGGAGGGGGGGGTGGTA**A**GCTGAAGCGTCGAATAGT…251..ATG | |
| ***Lolium*** | | |
| Embryo Precursor Species | CGAAGTATTGATAGGAATACCGGGGGTTCGAATCCCTCTCCATCCG**CG**AGG**T**CATAAGTTCTCTCTTGCCTTATCTATAGATAAGAA….724....ATG | Upstream *trnS* gene in red letters, Rnase Z |
| Seedling Precursor Species | CGAAGTATTGATAGGAATACCGGGGGTTCGAATCCCTCTCCATCCG**CG**AGGTCATAAGTTCTCTCTTGCCTTATCTATAGATAAGAA….724....ATG | |
| **Rye** | | |
| Embryo Precursor Species 2[a] | AAC**A**A**A**CAAAGTACGTAATGGCCTCAGGCCTGCTCTATTACATCGACCACCTTCGAGTTCTCTATATGCGAAGAGCCTGGAATGGTG..1825…ATG | SL, Rnase P |
| Seedling Precursor Species 1[a] | TTGATCTTACTCCAACCTACAGAA**T**ACACTGTCTTGCTCAATGTGCAAGGGATCATTCTATAAACTTGGTCTTGGAAAGCCT**A**TCA…..1465....ATG | |
| Embryo Precursor Species 1 | GAGTCGAGGAATAGGAGGAGAATCCCCAGGCAAA**C**AACTAGGGTTCCAGATTCACTGGGCAGAGAAGGGAGAATAGTATCCTCTC…1270…ATG | |

[a]Sequences correspond to cDNA data for CR-RT-PCR clones.

termini (-424, -421, relative to *rps7* start codon) and would require RNase P-like activity for end-cleavage (Figure 4.5a).

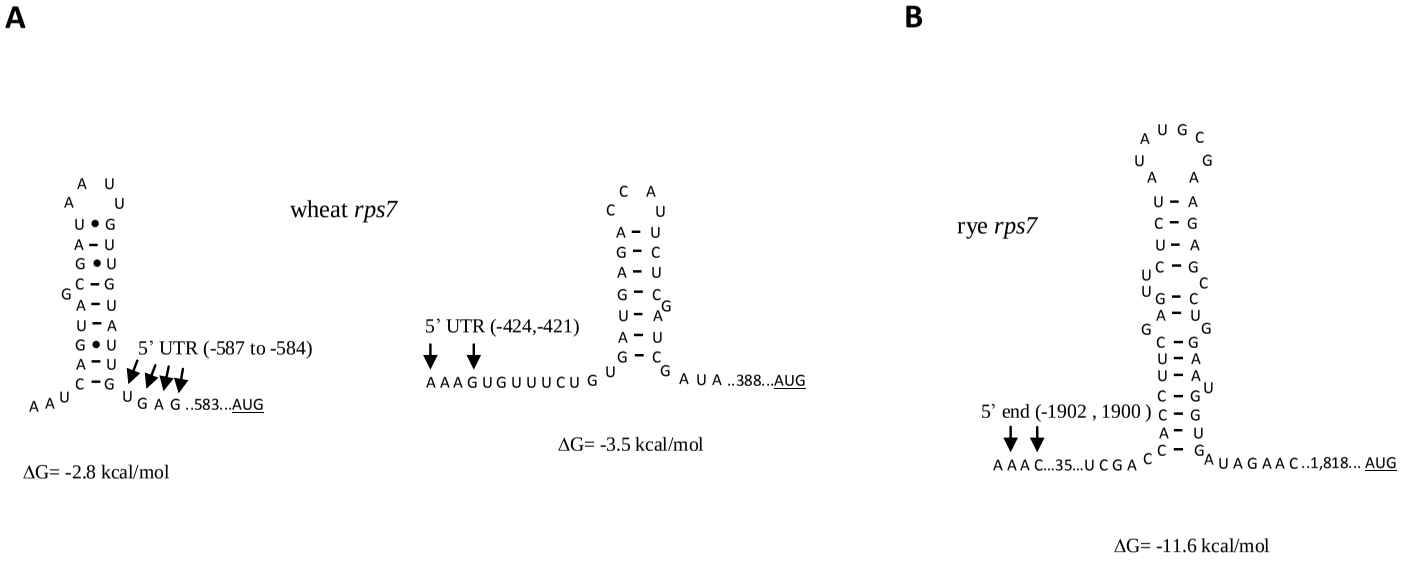
Unlike both wheat and *Lolium rps7* precursor transcripts, pre-mRNAs for rye *rps7* showed significant heterogeneity of 5'ends (Table 4.1). 5/9 CR-RT-PCR clones for precursor transcripts in embryo RNA however shared 5'ends at positions -1902 and -1900 (numbers relative to *rps7* start codon). Sequences just downstream of this homogeneous 5'end can be folded into a stem-loop structure with a Gibbs free energy of -11.6 kcal/mol (Figure 4.5b). This potential stem-loop however is ~40 nt downstream of the conserved 5'end and therefore is unlikely to serve as an end-protection conferring structure. If such a structure were to provide site-specificity for a PPR protein which recruited endonucleases like PPR proteins of the *RF* group that are believed to recognize and cut upstream flanking sequences (Holzle *et al*. 2011) it may serve some biological function. The RNA binding sites (*cis*-elements) and stretch of upstream sequence recognized by these proteins however has yet to be determined.

## 4.7 Discussion

To date, few analyses have looked at RNA processing of precursor transcript termini in plant mitochondria. Our findings demonstrate that maturation of mitochondrial *rps7* 3'ends requires fewer steps than upstream 5' termini. Similar observations were seen for *cox2* in pea (Kuhn and Binder 2002) and *ccmFN-rps1* in wheat (Calixte and Bonen 2008). In these studies it was also noted that 3' ends are held in common between both precursor forms and mature transcripts, which is also the case for *rps7* among wheat, rice, rye and *Lolium* (Byers *et al*. 2010). This suggests that like RNA editing, 3' processing is an early event (reviewed in Takenaka *et al*. 2008). RNA processing of 3'UTRs may be more efficient because it requires fewer steps or involves less machinery. Upstream of the 3' terminus (+69 to +102, relative to *rps7* stop) in all full-length *rps7* precursor and messenger RNAs there is a stem-loop structure that may function as a barrier to exo-activity (Byers *et al*. 2010). If this structure does provide stability 3'ends may be generated from exonuclease degradation in the 3' to 5' direction of the *de novo* primary transcript. Also interesting is that *rps7* precursor 5' ends for plants wheat, rice and *Lolium* are homogenous among grasses, even when mRNA termini for *Lolium* show

*Figure 4.5: Potential secondary structures near 5'ends of wheat and rye pre-mRNAs*

**a**) Stem-loop structures upstream (left) and downstream (right) of wheat *rps7* precursor transcript termini (black arrows). **b**) Stem-loop structure downstream of rye *rps7* precursor transcript termini. Positions of transcript ends are relative to *rps7* start codon. Distances from the *rps7* start codon and values for Gibbs free energy are given. Stem-loop structures were determined using the 'mfold' software (Zuker 2003).


A
wheat rps7
5’ UTR (-587 to -584)
A A U
U G A G ..583...AUG
ΔG= -2.8 kcal/mol
5’ UTR (-424,-421)
A A A G U G U U U C U G
A U A ..388...AUG
ΔG= -3.5 kcal/mol
B
rye rps7
5’ end (-1902 , 1900 )
A A A C...35...U C G A
A U A G A A C ..1,818... AUG
ΔG= -11.6 kcal/mol

several different heterogeneous ends (Byers *et al*. 2010). Rye *rps7* precursor 5'ends however are the exception as significant heterogeneity can be seen among transcripts. These differences in 5'end processing likely correlate with variations in the number of signals for transcription initiation and/or RNA processing embedded in the different DNA sequences that precede *rps7* among grasses.

5' termini for wheat *rps7* transcripts map to the 3'end of the $trnF^{(cl)}$, and through RNase Z-like activity likely generates a conserved 5'UTR terminus at positions -2926 to -2928 (relative to *rps7* start codon). Flanking wheat *rps7* 5'UTR termini (-584-587 and -421, -424) for other higher molecular weight precursors are two potential stem-loop structures (Figure 4.4d) which may function as elements for transcript stability or recognition sites for machinery involved in end-cleavage. Because RNA profiles using *rps7*-specific oligomer probes consistently show 2 relatively abundant and 1 minor precursor species, such RNAs must utilize mechanisms for protection from RNA turnover and/or *cis*-elements for generation of transcript ends. Termini upstream of the 0.9 kb precursor for wheat *rps7* are 10 and 13 nt in front of the stem-loop. If endonucleolytic cuts are made at the base of the stem of the stem-loop structure or exo-activity proceeds to this point the structure will not fit either model. However similar structures have been proposed for the generation of 3'ends for *ccmC* mRNA (Forner *et al*. 2007). Precursor 5' termini for *Lolium rps7* map to the 3'end of *trnS* (Figure 4.4c). Homogeneous termini (-759-764) are 12 nt from the acceptor stem for the mature tRNA and not immediately downstream making classical processing by an RNase Z-like enzyme somewhat unlikely. If it is endonucleolytic activity that generates this RNAs (1.3 kb) 5'end, exo-activity might be responsible for shorter than expected 5' termini.

From northern data rice *rps7* has 2 abundant (relative to 18S rRNA), very discrete transcripts for precursor RNAs (Figure 4.1). End-mapping data shows these species (1.05 and 0.83 kb) to be homogeneous (Table 4.1), like precursors for both *Lolium* and wheat *rps7*. No convincing secondary structures however could be identified in their 5'UTRs. Since certain PPR proteins like those of the *RF* subclass possess endonucleolytic activity for single stranded RNA (Holzle *et al*. 2011) perhaps there are consensus sequences within these upstream regions that signal end-cleavage from such a protein. However since no canonical motifs have been proposed, candidate sequences are difficult to

identify. Studies in *Arabidopsis* using different ecotypes with distinct upstream sequences for *ccmC* and knock-out mutants for the PPR protein RNA processing factor 3 (RPF3) demonstrate that for either ecotype different trans-factors are capable of generating the mature 5'end of the *ccmC* mRNA (Jonietz *et al*. 2011). It is also hypothesized that PPR proteins function as multi-protein complexes with individual proteins providing site-specificity while others carry out the actual enzymatic reaction. For editing of mitochondrial transcripts some PPR proteins are responsible for editing at various positions within different genes however when missing, a subset of those sites are still edited (Zehrmann *et al*. 2011). This means that additional editing factors, likely other PPR proteins can compensate for the loss of function. These observations suggest that there are a variety of signals for recognition by PPR proteins and that multiple PPR proteins have similar if not identical RNA processing functions.

Between stages of development *rps7* transcript profiles for individual grasses appear to be similar with the exception of *rye rps7* (Figure 4.1). End-mapping data for rye *rps7* precursor RNAs (Table 4.1) shows not only that transcript 5'termini vary between RNA isolated from germinating embryos and seedlings but that they are also heterogeneous in comparison to precursor species for *rps7* in wheat, rice and *Lolium*. While messenger RNAs for rye *rps7* were shown to favour 5'ends close to the known upstream promoter motif (Byers *et al*. 2010) they also mapped to various other positions, demonstrating similar heterogeneity (Byers *et al*. 2010). This non-specific 5'end location for rye *rps7* precursors may simply be due to random exo-activity. However somewhat discrete RNA species seen for northern hybridization with *rps7*-specific oligomer probes suggests some species must have somewhat protected ends (Figure 4.1).

C-to-U editing of plant mitochondrial *rps7* precursor transcripts demonstrates a temporal relationship between editing and end-cleavage as precursors in the RNA population showed less complete editing of sites in the UTRs compared to mRNAs. Editing is known to be an early processing event so it is not surprising that positions within UTRs are edited before end-cleavage. It is however curious as to why such positions upstream of the terminus for the mature *rps7* mRNAs among plants would be edited, to then be removed through end processing and eventual RNA turnover. Sites are perhaps fortuitously edited like is seen for regions within UTRs resembling the coding

sequence for protein-coding genes (Forner *et al*. 2007) or are in fact necessary for proper folding of RNA secondary structures. Consensus motifs for site-specificity for end-cleavage by processing machinery may also be generated through editing. The editing site at position -577 in the 1.1 kb precursor transcript for wheat *rps7* may be one such site. Developmental differences in site recognition for RNA editing, as is seen for position -108 in seedling *rps7* mRNA in wheat (chapter 3 addendum) may be due to a difference in the editing machinery activated during different stages of plant development. Later in seedling stages expression levels for the nuclear-encoded PPR proteins involved in mitochondrial editing may become altered, allowing newly expressed proteins access to different sites within *rps7* transcripts. Since editing is typically an early event it is difficult to determine why editing at position -108 was not seen for *rps7* precursor RNAs from wheat. One possibility is steric hindrance of editing machinery by secondary structures formed through RNA folding of longer 5'UTRs in wheat *rps7* precursors compared to mRNAs.

When comparing RNA profiles for embryo and seedling wheat, rice, *Lolium* and rye (Figure 4.1) it is apparent that the relative abundance (compared to 18S rRNA) of *rps7* precursor transcripts decreases in later stages of development. This same observation was made for intron-containing species for genes in wheat (Li-Pook-Than *et al*. 2004). Interestingly, unlike respiratory-chain genes, the ribosomal protein genes also show a decrease in the relative abundance of mRNAs (Li-Pook-Than *et al*. 2004). Stored messengers contribute to higher levels of *rps7* mRNAs in dormant seeds (Figure 4.2) however as seeds are rehydrated and begin to germinate (6-24hr, Figure 4.2) transcriptional and RNA processing machinery is activated and it is during these times that we begin to see a higher relative abundance of precursor species. In seedling stages (Figure 4.1, 4.2) lower levels of both *rps7* precursor transcripts and mRNAs for plants like wheat, rice and rye likely result from a decrease in the level of transcription and/or an increase in RNA turnover.

While no convincing transcription initiation consensus motifs were identified upstream of wheat, rice, rye or *Lolium rps7* precursor transcript termini many potential motifs perhaps active as minor transcription start sites were found throughout (underlined, Table 4.2). Sequences corresponding to the loose CRTA (where R is either A or G) and

extended variations like CATA(TA) found both upstream and downstream of major and minor RNA ends may contribute to some of the slight (wheat, rice and *Lolium*) and more pronounced (rye) heterogeneity seen among precursor transcript termini for the surveyed grasses. Multiple initiation sites were seen for several genes in *Arabidopsis* mitochondria and functioned as both transcription initiation sites and sites for RNA processing, either through site-directed end-cleavage or random 5'end degradation (Kuhn *et al*. 2005).

Sequences derived from the transfer of chloroplast DNA are located upstream of native mitochondrial genes in wheat and rice. The transfer of sequences upstream of wheat *rps7* and rice *nad9* is likely to have occurred in the common ancestor shared by the two plants ~50 Mya (Kellogg and Bennetzen 2004). It is likely that prior to speciation of wheat and rice an internal deletion in the upstream region of the chloroplast-origin sequence occurred. A subsequent independent lineage rearrangement would then have occurred in wheat to move a portion of the chloroplast-derived sequence upstream of *rps7*. Shuffling of the mtDNA was then followed by an additional internal duplication of the chloroplast-origin sequence in wheat (Figure 4.3b). Sequences corresponding to the deleted regions in rice and wheat cannot be found elsewhere in the mitochondrial genomes of these plants. Sequences homologous to *ndhJ* and *ndhK* in the wheat mitochondrial genome are also not present; likely due to rearrangement to the point of non-homology or deletion of mtDNA. Regardless of the mechanisms used for incorporation of chloroplast DNA, transferred sequences likely provide expression signals for mitochondrial genes like wheat *rps7* (this study) and rice *nad9* (as proposed in Nakazono *et al*. 1996) demonstrating a close association of RNA processing machinery for both mitochondria and chloroplasts.

## 4.8 Chapter 4 addendum: Affect of cold-growth on RNA processing of wheat mitochondrial *rps7* transcripts

### 4.8.1 *Differences in RNA processing and C-to-U editing of rps7 transcripts in cold-grown wheat seedlings*

Differences in RNA processing in response to cold-stress have been observed for wheat *cox2* (Kurihara-Yonemoto and Handa 2001) and a subset of the intron-containing

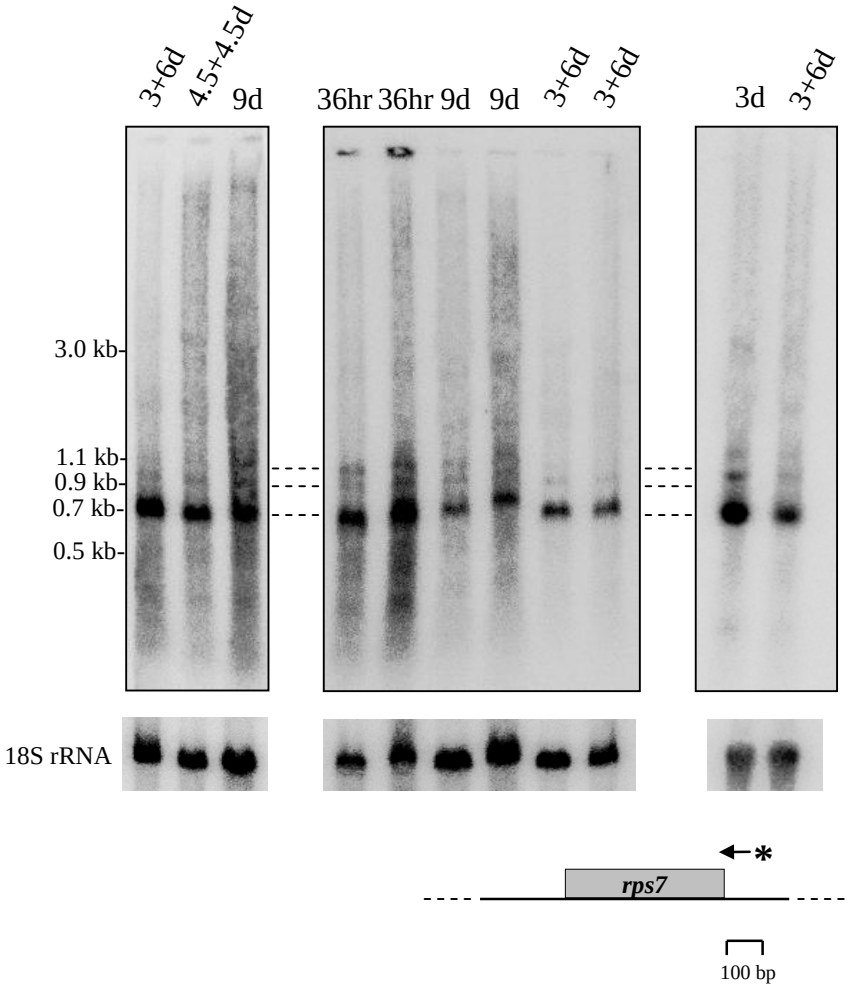
genes in rice (Kurihara-Yonemoto and Kubo 2010). Of the 17 editing sites within the coding region for wheat *cox2*, 12 showed a decrease in the degree of editing in the cold while the other 5 sites either increased or showed virtually no response. In rice, the relative abundance (compared to ethidium bromide staining of ribosomal RNA) of 14 out of 23 intron-containing RNA species was found to increase under cold-growth. Activity of RNA processing machinery for splicing and RNA editing is therefore likely affected by cold stress. An amino acid altering editing site in exon 1 of *cox2* in rice close to the exon-intron boundary in precursors showed no editing in the cold while the homologous site in wheat behaved similarly, demonstrating less than 10% editing. Sites close to exon-intron boundaries which typically require intron excision for efficient editing (Li-Pook-Than *et al*. 2007) appear to be the most greatly affected by growth in the cold. From preliminary work on editing in response to cold-stress for respiratory chain genes in wheat it also appears that the effect on silent sites is greater than for non-silent ones (P. Williston, B. Choi in Bonen lab).

Northern hybridization profiles for mitochondrial *rps7* in wheat grown at room temperature and in the cold for varying lengths of time show differences in the relative abundance (compared to 18S rRNA) of precursor transcripts, in particular the higher molecular weight species of 1.1 kb (Figure 4.6). All wheat RNA isolated from seedlings grown in the cold show faint or absent northern hybridization signals for the 1.1 kb precursor transcript compared to wheat grown at room temperature. Work with wheat *nad3*, another intronless gene, also shows that under cold-stress precursor RNAs decrease compared to RNA isolated from wheat grown at room temperature (Patrick Williston, 4$^{th}$ year honours project), different from previous observations for intron-containing species in rice (Kurihara-Yonemoto and Kubo 2010). The relative abundance of mRNAs between 3d room temperature and 3+6d cold-grown wheat also appears to vary (Figure 4.6), however because the relative abundance of mRNAs for *rps7* in wheat decrease in later stages of development (Li-Pook-Than *et al*. 2004) it is difficult to say whether the difference is solely a response to growth in the cold.

Lower levels of only certain *rps7* RNA species in cold-grown wheat seedlings suggests that processing machinery may be affected by growth in the cold. If production

*Figure 4.6: Northern analysis of wheat mitochondrial RNA from germinating embryos and seedlings grown under various conditions*

Blots were hybridized with a $^{32}$P-end-labelled *rps7*-specific oligomer probe LB29 (arrow with asterisk in schematic). 18S rRNA loading controls are indicated below and size markers on left. Seedlings were grown for 3 and 9 days at room temperature and for 4.5 and 6 days in the cold after initial development at room temperature for 4.5 and 3 days respectively.


3+6d
4.5+4.5d
9d
36hr
36hr
9d
9d
3+6d
3+6d
3d
3+6d
3.0 kb
1.1 kb
0.9 kb
0.7 kb
0.5 kb
18S rRNA
rps7
100 bp

or activity of endonucleases responsible for generation of the 1.1 kb precursor RNA from the 3.4 kb higher molecular weight transcript was decreased in the cold this would explain a decrease in the relative abundance of only this species. Alternatively if what we are seeing is actually an accumulation of the 0.9 kb transcript as opposed to inefficient processing of the 3.4 kb species we would invoke that factors involved in the maturation of the 0.7 kb mRNA are retarded in the cold.

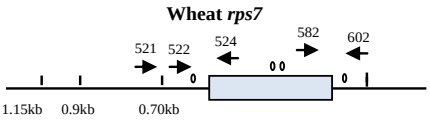
Direct sequencing of RT-PCR products derived from a mixed population of mRNA and precursor transcripts revealed that editing of the two non-silent edits within the coding region for wheat *rps7* was not greatly affected by cold growth (Figure 4.7 and 4.8a,b). The degree of editing at both sites 277 and 332 (numbers relative to *rps7* start codon) was virtually 100%. Responses to growth in the cold compared to room temperature were seen however for editing at positions 277 and 332 in RNA populations of precursor transcripts only: decreases of 25% and 35% respectively. Because 100% editing was seen at multiple sites I am confident that the degree of editing observed reflects the status of each RNA population and is not confounded by contaminating DNA. Since editing is typically an early event, populations of precursor RNAs are virtually fully edited at all non-silent positions (Figure 4.7). The decrease in editing for precursor populations at positions 277 and 332 therefore is an effect of growth in the cold, perhaps due to a decrease in efficiency of enzymatic activity. Editing is carried out by a family of highly similar proteins the PPR proteins and therefore it is difficult to imagine why some sites are affected more than others as enzymatic activity should be lowered uniformly.

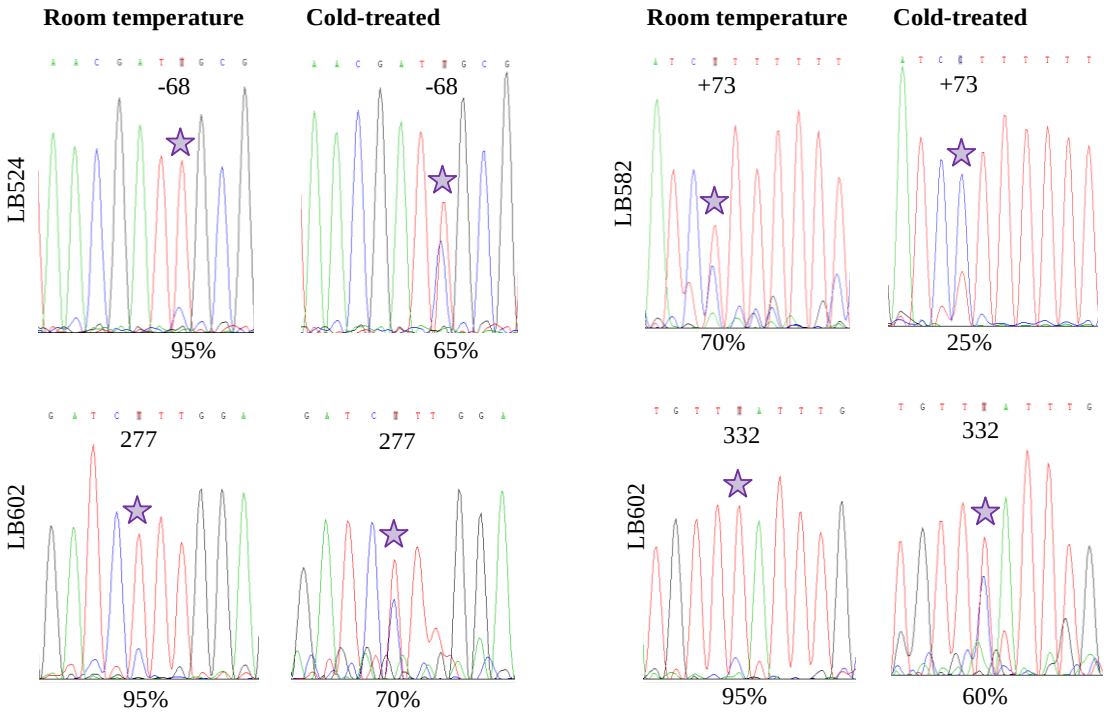
Affect of cold-stress on non-coding edits has never before been studied. The degree of editing of wheat *rps7* transcripts at positions -68 and +73 (numbers relative to *rps7* start and stop codons), in the 5' and 3'UTR respectively, decreases in the cold for both pre-mRNA and RNA for mixed populations (Figure 4.8a,b). Position -68 was affected less (30% and 20% for pre-mRNA and mixed populations respectively) than position +73 which showed the greatest influence under cold-stress with a decrease of ~45% and 40% in pre-mRNA and mixed RNA populations respectively. Two or more replicates of independent cDNA syntheses were generated for mixed populations and only one as yet for pre-mRNA (Figure 4.8b). Differences in the degree of editing of sites;

*Figure 4.7: Chromatogram results for direct sequencing of wheat rps7 RT-PCR products from 9 day room temperature and 3+6 day cold-grown seedlings*

RT-PCR products using LB521 and LB602 (black arrows in schematic) correspond to precursor transcripts while products generated from LB522 and LB602 correspond to a mixed population of precursor transcripts and mRNA. Transcript 5' and 3'UTR termini are indicated by dotted lines with RNA species' lengths given in kilobases. Editing sites are indicated by open circles. Percent editing is given below picture of chromotogram read for particular *rps7* editing sites (values relative to *rps7* start codon), for either RNA isolated from room temperature or cold-treated wheat. Custom oligomers for sequencing (LB524, LB602, LB582, black arrows in schematic) are shown vertically next to each chromatogram read.

Direct sequencing results for *rps7* editing Precursors (LB521/602)




mRNA+ (mixed population) (LB522/602)

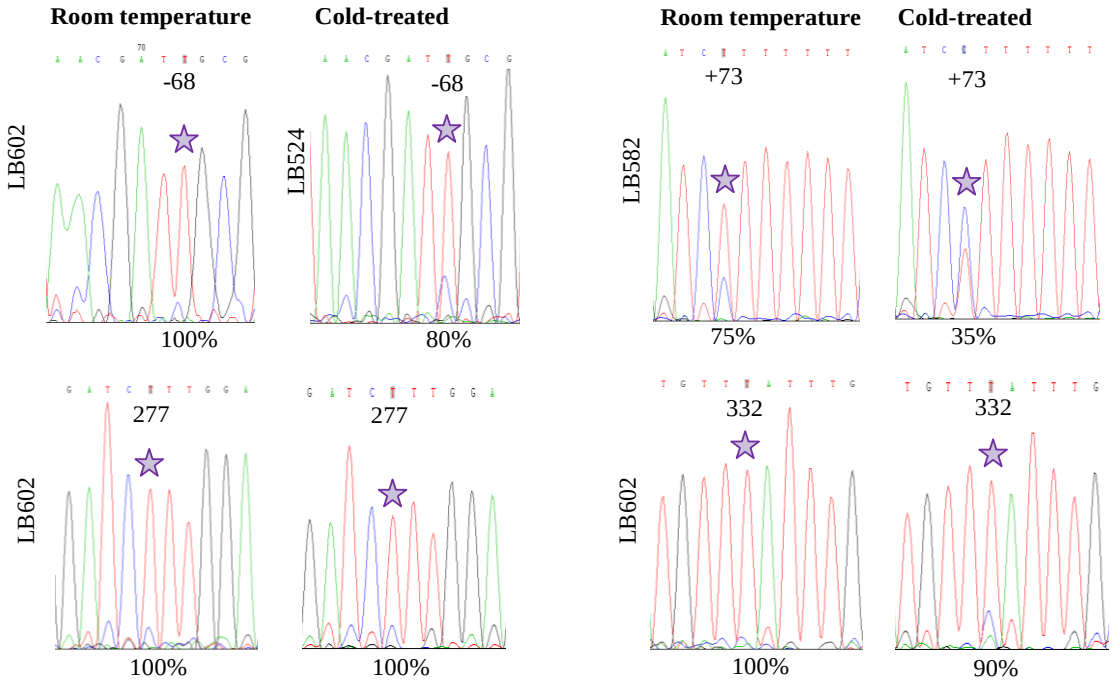
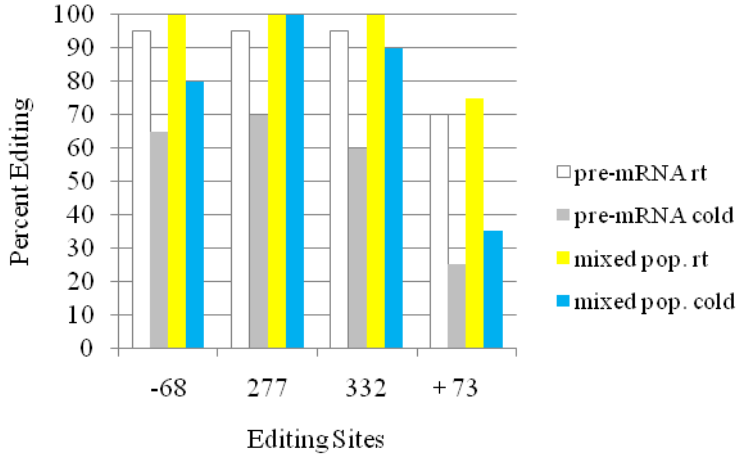


*Figure 4.8: Proportion of editing at various sites within 9 day room temperature and 3+6 day cold-grown wheat rps7 mRNAs and precursor transcripts.*

**a**) Bar graph showing the percentage of editing (y-axis) at a given site (x-axis, numbers relative to *rps7* start codon) according to direct sequencing results from trial 1 (b). Editing sites were analyzed for precursor transcripts from room temperature (rt) and cold-grown wheat (white vs. gray bars). A mixed population comprised of precursor species and mRNAs from both room temperature and cold-grown wheat (yellow and blue bars respectively) were also surveyed for degree of editing. **b**) Editing results for 2 trials for direct sequencing of *rps7* transcripts from rt and cold-grown wheat. **c**) CR-RT-PCR data for non-coding edits in wheat *rps7* mRNA 5'/3'UTRs (left). Number of clones edited at particular sites from room temperature embryo (E) and seedling (S) RNA are indicated. A schematic showing the location of all wheat *rps7* mRNA edits is given as well (right).

**A**



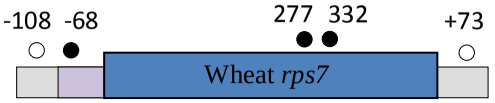

**B**

% Editing

| Editing Sites | Trial 1 | | | | Trial 2 | |
|---|---|---|---|---|---|---|
| | precursor | | mixed pop. | | mixed pop. | |
| | rt | cold | rt | cold | rt | cold |
| -68 | 95 | 65 | 100 | 80 | 85 | 75 |
| 277 | 95 | 70 | 100 | 100 | 95 | 95 |
| 332 | 95 | 60 | 100 | 90 | 95 | 90 |
| +73 | 70 | 25 | 75 | 35 | 85 | 50 |

**C**

| | 5'UTR | | 3'UTR |
|---|---|---|---|
| | -108 | -68 | +73 |
| wheat (E) | 0/5 | 5/5 | 2/5 |
| wheat (S) | 4/7 | 7/7 | 7/7 |



-68 and +73 in wheat *rps7* UTRs shows that particular sites within non-coding sequences are affected differently during cold-stress.

While editing of coding region non-silent sites is essential for proper protein formation, silent sites and sites within non-coding sequences like those of wheat *rps7* in the 5' and 3'UTRs may be important for formation of signals for translation, stability of RNA secondary structures or motifs for end-cleavage. Non-coding sites sometimes are fortuitously edited if UTRs contain sequence blocks similar to those of biologically significant edits within protein-coding regions for other genes (Forner *et al*. 2007). Flanking sequences for position -68 and +73 however, do not share sequence identity with coding sequences for other mitochondrial genes and are therefore not likely edited by chance. If template specificity of editing machinery however is compromised during exposure to cold temperatures we might expect to see an increase in the degree of fortuitous editing.

Editing was observed at position -108 and +67 in wheat seedling mRNAs generated from CR-RT-PCR experiments in 4/7 and 4/6 clones respectively (chapter 3 addendum). This is in contrast to the single clone observed for position +67 and no clones at position -108 in *rps7* mRNA clones from germinating embryos (Byers *et al*. 2010). Rather surprising, direct sequencing data for wheat 9d room temperature seedlings and 3+6d cold-grown seedlings did not show editing of *rps7* mRNAs at position -108 or +67. CR-RT-PCR suggests that editing of these particular sites may be developmentally regulated however due to conflicting results for direct sequencing this observation remains unclear.

Work on cold-grown wheat is still very preliminary and more research is required to elucidate some of the interesting differences in relative abundance of intron-containing species compared to precursors RNAs that do not contain introns. Furthermore it will be interesting to look at the degree of editing of additional non-coding sites in other genes to delineate their more dramatic behavior when wheat is grown in the cold.

## Chapter 5: General Discussion

The highly recombinogenic nature of plant mitochondrial genomes makes these systems very dynamic and interesting to study at the molecular level. Mitochondrial gene

content among plants is variable particularly for ribosomal protein genes and pathways for gene expression differ, sometimes even for the same genes between closely related species. Focusing on the structure, expression and evolution of mitochondrial genes we are able to identify genome rearrangements and their impact on RNA processing among plants, allowing us to elucidate how plants are able to tolerate what appear to be very disorganized, impracticable genomes. Analysis of genes such as the ribosomal protein gene *rps7* surveyed among a number of related species of grasses demonstrates the variety of upstream sequences used for transcription initiation, assortment of potential signals for end-cleavage among plants, plant-specific non-coding edits and possible mechanisms for transcript stability.

## 5.1 *Impact of DNA rearrangements on mitochondrial rps7 expression among grasses*

In a survey of ribosomal protein genes, *rps7* was predicted to be transferred from the mitochondrion to the nucleus the largest number of times during angiosperm evolution (Adams *et al*. 2002a). This however appears to be primarily from orders outside the Poales, a group of angiosperms that includes the grass family Poaceae. In a survey of 10 species from the order of Poales *rps7* was missing from the mitochondrial genomes of 2 lineages; *Lachnocaulon* and *Guzmania* (Adams *et al*. 2002a). *Lachnocaulon* however is unique in that the protein-coding content of its mitochondrial genome is predicted to be quite reduced compared to other plants (Adams *et al*. 2002a). Successful transfer of mitochondrial *rps7* to the nucleus, however has only been confirmed for *Rhododendron* and *Lactuca* and surprisingly no mito-copy of *rps7* can be found in the NCBI databases for four angiosperms that have lost the mitochondrial gene (Liu *et al*. 2009).

Among the grasses analyzed in this study, *rps7* is mitochondrialy-encoded however lineage-specific rearrangements have placed the gene in different genomic environments in almost all these plants. Rye and wheat for example, the most closely related grasses studied (5-7 Mya) have different upstream flanking sequences (Figure 3.1). Rye, barley and brome, separated by roughly 18-25 million years (Kelloggs and Bennetzen 2004) are the only plants of those surveyed that have retained homologous sequences upstream of the conserved *rps7* coding region. Upstream flanking sequences were obtained from PCR-amplified circular restriction fragments of 1.5 kb for barley, rye

and brome (accession numbers; HM581684, HM581685, HM581686) and therefore it is not known how far upstream sequence homology extends. Barley and brome however likely share the same pathway for *rps7* expression as neither show northern hybridization signals for species larger than the ~0.7 kb mRNA common among all grasses surveyed except oats (Figure 3.2, 3.5a). Other grasses which do show breakpoints in DNA homology close to the *rps7* start codon have different pathways for *rps7* expression. An alignment of grasses used in this study is given Appendix 1; shown here are the breakpoints in DNA homology upstream and downstream of *rps7* among plants. Sequence homology for rice, wheat, maize and oats extends ~160 bp upstream, mapping to the end of the 3' half of the expression cassette (Hazle and Bonen 2007b), what looks like a 'hotspot' for recombination among these plants. Downstream sequences immediately flanking homogeneous 3'termini among grasses are conserved for at least ~160 bp for plants whose complete mitochondrial genomes have not been sequenced (i.e. barley, brome, rye, *Lolium*, and oats). Of plants whose mitochondrial genomes have been completely sequenced like wheat, rice and maize, downstream sequence homology extends for an additional 70 bp. Sequences close to the *rps7* start codon are important for translational control while those downstream of the translational stop are likely crucial for transcript stability. Messenger RNA flanking sequences therefore must be conserved for efficient translation however it appears that upstream flanking sequences can be highly variable.

The mitochondrial genomes of angiosperms are the largest mitochondrial genomes reported so far and are highly variable in size among plant species (Mikami and Kubo 2007). This large genome size is attributable to a proliferation of intergenic DNA. The majority of spacer DNA appears to be of unknown origin however it is more likely that it has been scrambled to the point of non-homology due to extensive rearrangements and duplication of pre-existing sequences. It is in these spacer regions where signals for gene expression, like those for *rps7* precursor transcripts among wheat, rice, *Lolium* and rye are found (Table 4.2). Duplicate copies of expression signals shared among several genes like those identified for wheat *cox2* and *atp4* can also be found in spacer regions distal from functional genes (Hazle and Bonen 2007b). Through DNA rearrangements in plant mitochondria, expression signals in spacer sequences like these and/or foreign DNA

such as the chloroplast sequence upstream of wheat *rps7* can be brought into context for expression of native mitochondrial genes. In plant mitochondria, chloroplast-origin sequences likely provide expression signals for transcription of some mitochondrial genes. The respiratory chain gene *nad9* (Nakazono *et al*. 1996) and ribosomal protein gene *rpl2* (Subramanian *et al*. 2001) in rice are believed to be expressed this way. It will be interesting to see if primary transcripts for wheat *rps7* also use acquired chloroplast-origin *trnS*, *trnF* and flanking sequences for transcription initiation (Figure 4.3). Chloroplast sequences for 3 tRNA genes; *trns*, *trnF* and *trnP* were integrated as a unit in potato mitochondria and are co-transcribed (Remacle and Marechal-Drouard 1996). It is therefore not unlikely that transcription for *rps7* initiates upstream of the $trnS^{(cl)}$ creating one long polycistronc primary transcript (position of gray arrow in Figure 4.3). S1 nuclease protection assays or *in vitro* capping experiments with guanylyl transferase could be used to identify such primary transcripts.

A bias for *rps7* 5'mRNA ends near the consensus motif for transcription initiation in the expression cassette, for grasses *Lolium*, rye, brome and barley (about 230 nt upstream of *rps7* start codon among plants) suggests this *cis*-element acts as a promoter for gene expression (Figure 3.3). Transcript termini for *Lolium*, rye and brome with 5'ends mapping upstream of this position from either embryo or seedling CR-RT-PCR data (Table 3.1) suggest that there are additional sites upstream of the conserved CGTATA motif for transcription initiation. Precursor species for both rye and *Lolium rps7*, identified by northern hybridization (Figure 4.1) and CR-RT-PCR (Table 4.1) confirms that transcription initiates upstream of the above mentioned promoter motif for *rps7* in these plants. Northern analysis of brome *rps7* however shows a single discrete species of ~0.7 kb and no higher molecular weight species. Examples of consensus motifs like CATA and CGTA can be found 375 nt and 416 nt respectively upstream of the conserved promoter motif for brome *rps7*. If these *cis*-elements also serve as sites for transcription initiation of brome *rps7* RNAs, slightly longer transcripts may not have appeared as discrete species in our northern analysis. No transcript 5'ends however were seen at these positions (Table 3.1). Since many different sequence motifs are believed to initiate transcription in plant mitochondria (Kuhn *et al*. 2005) it cannot be excluded that other upstream sequences may have a role to play in *rps7* gene expression for brome. In

pea *cox2* for example *in vitro* transcription assays identified transcription start sites at non-canonical motifs 317 nt upstream of the ATG (Kuhn and Binder 2002).

## 5.2 *Potential signals for RNA processing*

Mapping transcript termini for mitochondrial *rps7* among grasses has demonstrated the diversity in 5'end processing signals contrasted with simpler mechanisms for 3'end maturation and to my knowledge this study is the most extensive survey of upstream and downstream mitochondrial regulatory sequences for a given mitochondrial gene from multiple plant species. Comparative analysis of the same gene among a number of different plants allows us to determine whether signals for gene expression are shared among species or are species-specific. Of the 8 grasses studied only barley and brome have the potential to share the same pathway for expression of mitochondrial *rps7*; both appear to use the same duplicated sequences from regulatory regions of other mitochondrial genes. Others use upstream signals from a native-mitochondrial tRNA gene (*Lolium*), genic regions of chloroplast DNA (wheat) and potential *cis*-elements such as stem-loop structures (rye, oats).

DNA rearrangements in plant mitochondria are somewhat rampant as sequences preceding the same mitochondrial protein-coding gene, like *rps7* are highly variable among plants, even closely related species. Some sequences like those upstream of wheat, rice, maize and oat *rps7* might even be prone to rearrangements since breakpoints in DNA homology among these plants are so similar (Appendix 1). In an analysis of sequences upstream of *rps1* among legumes variability was seen among 4 different species (Hazle and Bonen 2007a). In the mitochondrial genomes for wheat, rice and maize upstream flanking sequences for *rps1* are homologous, presumably because of the close association with *ccmFN*. In comparison *rps7* upstream flanking sequences are different among wheat, rice and maize fairly close to the translational start codon (~160 bp). Very different evolutionary events have occurred upstream of *rps1* and *rps7* in these plants; In my analysis of mitochondrial *rps7* I see a diversity of signals for 5'end formation for the same gene among plants while for *rps1* very conservative expression signals have been retained for plants such as wheat, rice and maize. This demonstrates that although nucleotide substitution rates in plant mitochondria are known to be slow

(Wolfe *et al*. 1987) and we expect flanking sequences to be highly conserved, DNA rearrangements are numerous and from our analysis, occur close in to *rps7* 5'ends in upstream regulatory regions among grasses (Appendix 1).

The majority of 5'ends for messenger RNAs in plants like *Arabidopsis*, are believed to be processed and are not primary transcripts (Forner *et al*. 2007). Furthermore these processed ends are believed to result from end-cleavage and not random degradation by exonuclease activity. Multiple, heterogeneous 5'ends for *rps7* mRNAs in rye, brome, *Lolium*, maize and oats (Table 3.1) suggests a lack of 5'signals for discrete end protection, in contrast to uniform 3' termini, leaving ends vulnerable to exo-attack. Transcript ends however are not entirely ragged as might be expected from exo-activity as there is a slight bias for *rps7* mRNA 5'ends close to a predicted promoter motif in some of these plants (eg. rye, brome, *Lolium*, Figure 3.4). Plant-specific ribonucleases with 5'exo-activity could explain ragged termini generated *in vitro* for plants like maize and oats. In a broad survey of mRNA termini for mitochondrial protein-coding genes in wheat virtually all transcripts have homogeneous 3'ends while only a subset of genes have discrete 5'termini (B. Choi, M. Acero, Bonen unpublished data). While it appears there is a general phenomenon of discrete 3'ends among mitochondrial mRNAs for wheat protein-coding genes and *rps7* among grasses, homogeneous *rps7* mRNA 5'termini for plants like wheat, rice and barley may be the exception.

Higher molecular weight precursor RNAs were also surveyed for *rps7* among a subset of the grasses: wheat, rice, rye and *Lolium*. Homogeneous 3'ends and 5'ends, with the exception of rye 5'termini were observed for *rps7* precursors among plants (Table 4.1). Taken together, our observations for precursor and mature RNAs show that all RNA species for wheat and rice have discrete ends. Similar observations were reported for *cox2* in pea (Kuhn and Binder 2002) and *rps1-ccmF*$_N$ in wheat (Calixte and Bonen 2008). *Lolium rps7* was different in that very conservative 5' and 3'termini were observed for precursor transcripts while mRNA 5'ends showed significant heterogeneity. Rye *rps7* also showed novel features; 5'ends for both mRNAs and precursor transcripts mapped to a variety of different upstream positions, however like all *rps7* RNAs among grasses also showed very homogeneous 3'ends. Furthermore these observations held for RNAs isolated from different stages of development. No obvious developmental differences

were seen for particular species however more work is needed for rye *rps7* to determine whether or not different 5'processing sites are recognized in embryo versus seedling RNA for end-cleavage.

Signals for RNA turnover such as incorporation of polyA tails at the 3'ends of aberrant transcripts can be analyzed using the CR-RT-PCR method. We observed very few non-encoded nucleotides at the 5'/3'UTR junction for circularized *rps7* transcripts, supporting the idea that transcripts were not ones tagged for degradation and do contribute to the pool of productive RNAs. Of the transcripts having non-encoded nucleotides only one was truncated at the 5'end and virtually all had 'full-length' 5'and 3'UTRs (Table 3.1, 4.1). Editing was also observed at positions within the 5' and 3'UTRs for these transcripts. The transcript non-encoded nucleotide additions can be grouped into two categories; single nucleotide extensions of A or T and short nucleotide extensions of primarily A's (observed by Kuhn *et al*. 2001). No 3'extensions like the –CCA motif determined for turnover of truncated *rps12* mRNAs in maize were seen (Williams *et al*. 2000). Single nucleotide additions may not reflect activity *in planta* as they can result from *in vitro* activity of the RNA ligase enzyme used in the CR-RT-PCR method (Kuhn and Binder 2002).

Homogeneous 3'termini for *rps7* precursor transcripts and mRNAs map to the same position downstream of the *rps7* stop codon suggesting 3'end processing is an early event. Plant mitochondrial transcription is not believed to have any defined termination signals and therefore RNA secondary structures and/or sequence motifs must be recognized efficiently for either stability conferring proteins or endonucleases respectively to generate discrete ends during the early stages of RNA processing. The mode and structures involved in transcription termination are still however completely unknown. Furthermore double stem-loop and stem-loop structures are unable to terminate transcription *in vitro* and are instead believed to be *cis*-elements associated with end protection (Dombrowski *et al*. 1994). Like for *rps7* transcript 3'ends there are two potential stem-loop structures upstream of the 3'ends in pea *cox2* believed to act as barriers to exonucleolytic degradation in postranscriptional 3'trimming (Kuhn and Binder 2002). Pea *cox2* 3'ends were also located in a polythymidine stretch of 9 nt proposed as a potential signal for transcription termination (Kuhn and Binder 2002). Grass *rps7* 3'ends

also map to a homopolymeric stretch of nucleotides, a sequence of 6 cytidines. This is similar to prokaryotic transcription terminators and therefore if there are indeed elements of bacterial-type termination of transcription in plant mitochondria further *in vitro* studies are required to learn more about 3'end formation.

Among the different 5'termini for *rps7* precursor RNAs and mRNAs, potential transcription initiation signals and stem-loops were located upstream and downstream of transcript termini (Table 3.2, 4.2). Short motifs of similar sequence however were not identified as potential cis-elements for end-cleavage by protein machinery like PPR proteins with endonuclease activity for single-stranded RNA. Genetic analysis of mutants defective in organellar functions has revealed many nuclear-encoded post-transcriptional regulators of mitochondrial gene expression (reviewed in Woodson and Chory 2008). Most of the cytoplasmic male sterility (CMS) RESTORER OF FERTILITY (*Rf*) genes encode PPR proteins and therefore in plants with nuclear encoded *Rf* genes additional PPR proteins are available for organellar RNA processing. Normally for plants with CMS, fertility is often restored by crossing the infertile line with a 'restorer' line which has the proper nuclear background (i.e. necessary *Rf* genes) to process and eliminate aberrant chimeric transcripts (Pelletier and Budar 2007). End processing of mitochondrial *rps7* RNAs may therefore be regulated differently in different plant lineages. Variance among nuclear backgrounds and therefore machinery for transcript maturation or differences in post-transcriptional or post-translational modification of mRNAs for RNA processing machinery or PPR proteins respectively could explain lineage-specific characteristics of *rps7* transcript ends.

## 5.3 *C-to-U Editing of mitochondrial rps7 transcripts*

Plant mitochondrial editing has been identified as C-to-U and results from a deamination reaction that converts cytidines to uridines primarily in coding sequences for protein-coding genes. Such editing is also seen in chloroplast transcripts however in chloroplasts between 4 and 25 RNA editing events occur in the entire transcribed genome (~110 kb) whereas in the mitochondria of flowering plants several 100 such changes are estimated. In the *Arabidopsis* genome which is 367 kb, 441 sites are edited (Giegé and Brennicke 1999). Some mitochondrial protein-coding sequences are edited very

frequently and others not at all; *cox1* in *Arabidopsis* has no editing sites while *ccmB* is edited at 39 sites. Ribosomal protein genes are by comparison less edited than respiratory chain genes and most cytochrome c biogenesis genes too. Among grasses, *rps7* is edited at two sites within mRNA coding sequences and one site within the 5'UTR and the 3'UTR. Both coding sites are non-silent and demonstrated full editing from RT-PCR experiments using RNA isolated from germinating embryos. Position 277 is a leucine to phenylalanine conversion and takes *rps7* away from sequence identity with other plants like *Arabidopsis* and liverwort (Figure 3.6), an example of 'wrong-way' editing. Position 332 is an example of 'right-way' editing as serine residues are converted to leucine residues, increasing sequence identity with more distantly related species like liverwort and *E. coli*. Interestingly in some plant chloroplasts the *rps7* homologue is also edited at the homologous site to position 332 creating the same serine to leucine conversion (Yura K *et al*. 2009). Editing at homologous sites in chloroplast and mito *rps7* is explained by dual targeting of PPR proteins to both organelles. Studies analyzing chloroplast editing have shown that machinery is shared between compartments (review in Woodson and Chory 2008).

Differences in the degree of editing of multiple positions within UTRs for *rps7* mRNAs indicates that different sites are likely recognized by different editing machinery and/or that some sites are more easily edited than others. Difference in the degree of editing of particular sites in plant mitochondrial transcripts between stages of development may result from differential expression of editing machinery throughout development. Developmental differences in RNA editing have only been previously documented for intron-containing transcripts for wheat *cox2* (Li-Phook-Than *et al*. 2007). If editing machinery in later stages of seed development is more active compared to earlier stages this might explain the increased editing observed at novel sites in *rps7* UTRs.

In a large-scale survey of editing in *Arabidopsis* only 15 edits were seen in non-coding sequences, 8 were in introns and 7 were in 5' and 3'UTRs (Giegé and Brennicke 1999) with additional sites found in the 5'UTR of *ccmC* (Forner *et al*. 2007). Sequences in the 5'leader for *ccmC* however are homologous to coding sequences for the *nad6* reading frame and are therefore fortuitously edited. Of the editing sites observed in

leaders and tails for *Arabidopsis* protein-coding genes 6 were in 5'UTRs and only 1 in 3'UTRs. Of the 6 sites in 5'leaders 2 were within *rps7* 5'UTRs however not at positions homologous to *rps7* among grasses. No coding edits were identified for *rps7* in *Arabidopsis* with the exception of one site in experiments using RNA from rosette leaves (Bentolila *et al*. 2008). The *rps7* transcripts among grasses therefore demonstrate a very high degree of editing within non-coding sequences, both for RNAs isolated from germinating embryos and seedlings. Editing at multiple sites within UTRs may increase stability of potential secondary structures like the stem-loop in the 3'UTR of all *rps7* transcripts among grasses (Figure 3.4) or help with regulation of protein synthesis by facilitating recognition by translation machinery. The generation of *cis*-elements for RNA processing events like end-cleavage is also possible through editing.

Non-silent coding edits and editing positions within 5' and 3'UTRs for wheat *rps7* transcripts from 9 day seedlings showed marked responses to growth under cold-stress compared to room temperature. In RNA populations including both precursors and mRNA at position +73 within the 3'UTR a drop of 60% in degree of editing was seen. This is the first time a non-coding UTR edit has been studied under cold-stress conditions. While these results are still preliminary the change in the degree of editing can be attributed entirely to the editing status of reverse transcribed cDNA as additional sites within the coding regions were observed to be fully edited (i.e. no contaminating mtDNA). While it is arguable that non-coding edits are not biologically important and therefore show the greatest affect under cold-stress it seems unlikely that RNA editing machinery for non-coding sequences would be affected by growth in the cold and not machinery for editing within coding regions. Other possibilities for differences in editing between coding and non-coding sequences include secondary structure and therefore increased steric hindrance of particular sites under different environmental conditions. Position +73 lies within the stem of a potential stem-loop structure. If under cold-conditions the RNA secondary structure is not able to unfold for PPR protein recognition of single-stranded RNA than perhaps editing at this position is compromised. This is also the first time editing within the coding region for an intronless gene has been reported for seedlings grown in the cold. Interestingly, while the relative abundance of intron-containing precursor species for rice mitochondrial RNAs increase with increasing time

in the cold (Kurihara-Yonemoto and Kubo 2010), it appears precursors for intronless genes like *rps7* may decrease, at least for some higher molecular weight species (Figure 4.6).

## 5.4 Future directions

More and more data for primary transcripts in plant mitochondria demonstrates the variability in signals used for transcription initiation (Kuhn and Binder 2002; Kuhn *et al*. 2005; Calixte and Bonen 2008). Shared upstream expression signals for multiple genes (Hazle and Bonen 2007b) are not a common phenomenon among plant mitochondrial coding sequences. This observation appears to be the exception and not the rule. End mapping data for wheat and *Lolium rps7* precursors suggests *rps7* expression likely initiates upstream of the $trnF^{(cl)}$ and *trnS* genes respectively. To identify transcription initiation signals upstream of wheat and *Lolium rps7* using the CR-RT-PCR method it is necessary to treat the RNA with tobacco acid pyrophosphatase (TAP), an enzyme required to remove the 5'triphosphate from primary transcripts, prior to RNA ligation. Because precursor RNAs could be 5' to 3'-end-ligated in the absence of TAP in our study, they did not contain a 5'triphosphate, characteristic of primary transcripts. RNA isolated from germinating embryos should be used for such experiments since embryo RNA shows higher relative levels of precursor transcripts (Li-Pook-Than *et al.* 2004).

For plants like barley and brome whose northern profiles for *rps7* showed no higher molecular weight species (see Figure 3.1) it would be interesting to determine whether or not transcription initiates at the known consensus motif within the full-length upstream expression cassette (Byers *et al*. 2010) using techniques like *in vitro* capping experiments. This type of experiment uses guanylyl transferase, an *in vitro* 5'capping enzyme which adds guanosine monophosphate groups to the 5'end of primary transcripts, radiolabelled GMP is used to identify species with a 5' triphosphate. Additional experiments like primer extension which requires a radiolabelled primer for gene-specific reverse transcription of upstream sequences can also be used to analyze positions of 5'ends. After using an *rps7*-specific primer in the antisense orientation for transcription run-off experiments, products can be run on a polyacrylamide gel giving the positions of

5’ends among RNA species within the population. Primer extension however does not identify primary transcripts.

Northern hybridization data for rye *rps7* shows different sized precursors in germinating embryos compared to seedlings implying there might be developmental differences in RNA processing. To date only a few species for rye *rps7* embryo and seedling RNA have been analyzed using the CR-RT-PCR method. Alternative methods like primer extension would be informative in identifying multiple higher molecular weight species for either embryo or seedling mitochondrial RNA. RNA ligation is more efficient for smaller more easily ligated RNA molecules and PCR preferentially amplifies smaller molecules of cDNA, making it both more difficult to go after higher molecular weight transcripts and potentially biasing the RNA pool. Since different size transcripts are seen between stages of development different signals for transcription initiation and/or end-cleavage are being used. Sequences flanking these termini might resemble *cis*-elements for site-recognition by protein machinery previously identified for RNA processing of other genes.

CR-RT-PCR data for oat *rps7* mRNA is interesting because the 5’UTRs for some clones are so much longer than those in other grasses, including its close relative *Lolium* (Kellogg and Bennetzen 2004). Northern profiles for oats (Figure 3.5a) show species larger than 0.7 kb, the length of the *rps7* mRNA in all other grasses studied. Preliminary end mapping data for oats (Table 3.1) confirms that longer stable messengers are generated. Shorter transcripts were also amplified using the CR-RT-PCR method. Sequence analysis shows that oat *rps7* contains the 3’half of the upstream expression cassette, just like for wheat, rice and maize *rps7* (Table 3.2) and therefore does not share upstream sequence homology with *Lolium* past this point. Mapping of RNA ends for precursor transcripts in oats will help determine what type of rearrangement has occurred upstream of oat *rps7* and how exactly expression differs from that of *rps7* in *Lolium*. Resolving questions such as these helps demonstrate the variety of signals for gene expression and the highly rearranging nature of plant mitochondrial genomes.

Many sites within *rps7* UTRs undergo C-to-U conversions compared to other protein-coding genes in plants examined (eg. *Arabidopsis*). In general very few non-coding edits have been reported and of those several are fortuitously edited due to

homology with genic sequences. Evaluating non-coding edits that show a high degree of editing in the RNA population will help elucidate potential consensus or loose consensus motifs for the machinery involved. Non-coding edits were seen at multiple sites in *rps7* UTRs. Several sites like -165 in *Lolium* and -108 and +67 in wheat (numbers relative to *rps7* start and stop codons) were plant-specific while others; -68 (-62 in rice) and +73 were edited in all grasses surveyed. Because CR-RT-PCR oligomers for reverse transcription and amplification of precursor species in wheat, rice, rye and *Lolium* were designed close to 5'UTR termini the entire non-coding regions for these species were not surveyed for C-to-U editing. Because the sense primer LB523 (Figure 2.1) was used for CR-RT-PCR only 4 nt of the 3'UTR were not analyzed for editing among grasses. Upstream sequences among grasses however are variable and therefore a number of plant-specific primers were used to amplify the 5'UTRs of precursor species. Amplification of mRNAs however was done with the same antisense coding-region primer, LB524 (Table 2.1 and Figure 2.1), therefore for the mature *rps7* mRNAs among plants, editing data was collected for all sites within 5'UTRs (eg. barley, brome, maize and oats).

Editing analysis of both *rps7* precursor and mature transcripts from cold-grown wheat and wheat grown at room temperature has shown that while editing is an early event but is reduced in populations of precursor RNAs, both at coding and non-coding sites; the largest effect is seen at positions for UTR edits in populations of precursor RNA from cold-grown wheat. The same was seen for populations of mRNA and precursors but to a lesser extent. It will be interesting to determine if non-coding edits behave similarly in other transcripts for genes like *rps4*, where at least 2 non-coding edits are confirmed (Jinchao Xie, 4$^{th}$ year honours project). It will also be interesting to analyze the degree of editing of sites in coding (silent and non-silent) and non-coding sequence for wheat *rps7* RNA isolated from different stages of development. From CR-RT-PCR clone data we see that certain editing sites may be recognized in different stages of development so perhaps degree of editing is also regulated developmentally. This more rigorous analysis will also help determine how much the change in the propensity of editing at non-coding sites in the cold is a result of cold-stress versus developmental regulation.

## 5.5 Concluding remarks

My analysis of mitochondrial *rps7* among grasses has demonstrated that many different pathways are used for production of mature messenger RNAs even for closely-related species. In some cases chloroplast-origin sequences for transfer RNAs appear to provide signals for transcription initiation and RNA-processing as do native-mitochondrial tRNA genes. Analysis of upstream *rps7* flanking sequences has highlighted the degree of lineage-specific rearrangements among grasses and demonstrates how plant mitochondrial genomes can tolerate frequent shuffling of genetic material.

An immerging trend for homogeneous 3'termini among plant mitochondrial transcripts provides insight into 'early' and 'simple' 3'end processing. Complexity of processing at the 5'end of mitochondrial transcripts may result from a larger array of machinery for both maturation of 5'ends and regulation of translation. Preliminary data for regulation of RNA processing in the cold demonstrates that specificity and/or efficiency of machinery for end-cleavage and C-to-U editing may be compromised under stress conditions.

Much is still to be learned about gene expression in plant mitochondria. A significant amount of data continues to be unearthed concerning the regulation of post-transcriptional events such as C-to-U editing, end-cleavage and transcript stability (reviewed in Woodson and Chory 2008). More and more we are also being presented with protein-level data which helps to elucidate the machinery involved in the regulation of RNA processing. Current research specifically is providing insights into assembly of PPR protein super complexes (Klodmann *et al.* 2011), an important factor for providing site-specificity for either RNA editing or endonucleolytic generation of transcript ends.

Unfortunately a lack of transformation system in plant mitochondria makes it difficult to conduct site-directed mutagenesis experiments to evaluate potential RNA processing signals. *In organello* molecular techniques involving introduction of DNA constructs into electroporated mitochondria may be integral in elucidating some of the outstanding questions about the regulation of gene expression in plant mitochondria.

Furthermore since genetic engineering of crop plants through the use of techniques like protoplast fusion to create cytoplasmic male sterile (CMS) hybrids, is

important for agricultural purposes it is more important than ever to understand the nuclear-mitochondrial interactions required for highly regulated production of productive mRNAs. If we can demonstrate how PPR proteins belonging to the sub-class of RESTORER OF FERTILITY genes recognize and process aberrant transcripts (reverting the sterile phenotype in CMS plants) it will be easier to generate hybrid lines which produce large quantities of seeds, improving both quantity and quality of available food stocks (hybrid vigour).

Plant mitochondrial genomes are very dynamic compelling systems to study scientifically. Their highly rearranging nature begets complex and unique gene expression pathways. With more and more information emerging about protein machinery and RNA-level signals for transcript maturation we are beginning to understand the complex control mechanisms involved in post-transcriptional regulation of plant mitochondrial RNAs. Expanding our knowledge of gene-specific processes like those for *rps7* is necessary for understanding plant mitochondrial gene expression as a whole.

**References**

Adams K, Song K, Roessler P, Nugent J, Doyle JL, Doyle JJ, Palmer J (1999) Intracellular gene transfer in action: dual transcription and multiple silencing of nuclear and mitochondrial *cox2* genes in legumes. Proc Natl Acad Sci USA 96:13863-13868

Adams K, Qui Y, Stoutemyer M, Palmer J (2002a) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA 99(15):9905-9912

Adams K, Daley D, Whelan J, Palmer J (2002b) Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast and cytosol counterparts. Plant Cell 14:931-943

Adams K and Palmer J (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 29:380-395

Allgemeine L, Botanik M, Bochum R-U, Universitätsstr (2005) Fungal intervening sequences. Genes and Genomics 5:71-92

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

Alverson A, Rice D, Dickinson S, Barry K, Palmer J (2011) Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber. The Plant Cell 23:2499–2513

Barkan A (2011) Studying the structure and processing of chloroplast transcripts. Methods Mol Biol. 774:183-97

Bartoszewski G, Gawronski P, Szklarczyk M, Verbakel H, Havey MJ (2009) A one-megabase physical map provides insights on gene organization in the enormous mitochondrial genome of cucumber. Genome 52:299-307

Bentolila S, Elliott L, Hanson M (2008) Genetic Architecture of Mitochondrial Editing in *Arabidopsis thaliana*. Genetics 178:1693-1708

Bewley J, Black M (1994) Seeds: physiology of development and germination, 2nd edn. Plenum, New York

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

Budar F, Pelletier G (2001) Male sterility in plants: occurrence, determinism, significance and use. C.R. Acad. Sci. Paris, Sciences de la vie/Life Sciences 324:543–550

Bullerwell C, Gray M (2004) Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. Current Opinion in Microbiology 7:528–534

Byers E, Rueger J, Bonen L (2010) Impact of genomic environment on mitochondrial *rps7* mRNA features in grasses. Mol Genet Genomics 284:207-216

Calixte S, Bonen L (2008) Developmentally-specific transcripts from the *ccmFN-rps1* locus in wheat mitochondria. Mol Genet Genomics 280:419-426

Chaw SM, Shih AC, Wang D, Wu YW, Liu SM, Chou TY (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol 25:603-615

Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, Meyer L, Wilson RK, Newton KJ (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol 136:3486-3503

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

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

Cui P, Liu HT, Lin Q, Ding F, Zhuo GY, Hu SN, Liu DC, Yang WL, Zhan KH, Zhang AM, Yu J (2009) A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants**.** Journal of Genetics 88:299-307

de Longevialle AF, Meyer EH, Andres C, Taylor NL, Lurin C, Millar AH, Small ID (2007) The pentatricopeptide repeat gene OTP43 is required for *trans*-splicing of the mitochondrial *nad1* intron 1 in *Arabidopsis thaliana*. Plant Cell 19:3256-3265

Dombrowski S, Brennicke A, Binder S (1997) 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. EMBO J. 16:5063-5076

Dombrowski S, Hoffmann M, Guha C, Binder S (1999) Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria. J. Biol. Chem. 274:10094–10099

Eubel H, Heinemeyer J, Sunderhaus S, Braun H (2004) Respiratory chain supercomplexes in plant mitochondria. Plant Physiol Biochem 42:937-942

Figueroa P, Gómez I, Carmona R, Holuigue L, Araya A, Jordana X (1999) The gene for mitochondrial ribosomal protein S14 has been transferred to the nucleus in Arabidopsis thaliana. Mol Gen Genet. 262:139-144

Forner J, Weber B, Thuss S, Wildum S, Binder S (2007) Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5' and 3' end formation. Nucl Acids Res 35:3676-3692

Forner J, Holzle A, Jonietz C, Thuss S, Schwazlander M, Weber B, Meyer RC, Binder S (2008) Mitochondrial mRNA polymorphisms in different Arabidopsis accessions. Plant Physiol 148:1106-1116

Fujii S, Toda T, Kikuchi S, Suzuki R, Yokoyama K, Tsuchida H, Yano K, Toriyama K (2011) Transcriptome map of plant mitochondria reveals islands of unexpected transcribed regions. BMC Genomics 2011 12:279-287

Gagliardi D, Binder S (2007) Expression of the plant mitochondrial genome, In Logan D (ed.) Plant mitochondria. Blackwell Publishing, Ames, IA, USA, pp50-96

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

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

Gray M, Burger G, Lang B (1999) Mitochondrial evolution. Science 283:1476-1481

Grewe F, Viehoever P, Weisshaar B, KnoopV (2009) A *trans*-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. Nucl Acids Res 37:5093-5104

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

Gulaberto J, Bonnard G, Lamattina L, Grienenberger J (1991) Expression of the Wheat Mitochondrial *nad3-rpsl2* Transcription Unit: Correlation between Editing and mRNA Maturation. The Plant Cell 3:1109-1120

Hammani K, Colas des Francs-Small C, Takenaka M, Tanz S, Okuda K, Shikanai T, Brennicke A, Small I (2011) The Pentatricopeptide Repeat Protein OTP87 Is Essential for RNA Editing of *nad7* and *atp1* Transcripts in *Arabidopsis* Mitochondria. Journal of Biological Chemistry 286: 21361–21371

Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus L.*): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Res. 31:5907-5916

Haouazine-Takvorian N, Takvorian A, Jubier MF, Lejeune B (1997) Genes encoding subunit 6 of NADH dehydrogenase and subunit 6 of ATP synthase are co-transcribed in maize mitochondria. Curr Genet 31:63-69

Hazle T, Bonen L (2007a) Status of genes encoding the mitochondrial S1 ribosomal protein in closely-related legumes. Gene 405:108-116.

Hazle T, Bonen L (2007b) Comparative Analysis of Sequences Preceding Protein-Coding Mitochondrial Genes in Flowering Plants. Mol. Bio. Evol. 24:1101-1112.

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

Hedtke B, Borner T, Weihe A (2000) One RNA polymerase serving two genomes. EMBO Rep. 1:435-440

Herrin D, Nickelsen J (2004) Chloroplast RNA processing and stability. Photosynth Res. 82:301-314

Hoffmann M, Kuhn J, Daschner K, Binder S (2001) The RNA world of plant mitochondria. Prog Nucleic Acid Res Mol Biol. 70:119-154

Holec S, Lange H, Canaday J, Gagliardi D (2008) Coping with cryptic and defective transcripts in plant mitochondria. Biochim Biophys Acta 1779:566-573

Holec S, Lange H, Kuhn K, Alioua M, Borner T, Gagliardi D (2006) Relaxed transcription in *Arabidopsis* mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase, Mol Cell Biol. 26:2869-2876

Holzle A, Jonietz C, Torjek O, Altmann T, Binder S, Forner J (2011) A RESTORER OF FERTILITY-like PPR gene is required for 5'-end processing of the *nad4* mRNA in mitochondria of *Arabidopsis thaliana*. The Plant Journal 65:737-744

Ikeda T, Gray M (1999) Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. Plant Mol Biol. 40:567-578

Jonietz C, Forner J, Hölzle A, Thuss S, Binder S (2010) RNA PROCESSING FACTOR2 is required for 5' end processing of *nad9* and *cox3* mRNAs in mitochondria of Arabidopsis thaliana. Plant Cell 22:443-453

Jonietz C, Forner J, Hildebrandt T, Binder S (2011) RNA PROCESSINGFACTOR 3 is crucial for the accumulation of mature ccmC transcripts in mitochondria of *Arabidopsis thaliana* accession Columbia. Plant Physiol. (Epub ahead of print)

Joyce P, Michael G (1989) Chloroplast-like transfer RNA genes expressed in wheat mitochondria. Nucl. Acid Research 17:No.14

Joyce P, Spencer D, Bonen L, Gray M (1988) Genes for $tRNA^{Asp}$, $tRNA^{Tyr}$, and two $tRNAs^{Ser}$ in wheat mitochondrial DNA. Plant Mol. Biol. 10:251-262

Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. Am J Bot 91:1709-1725

Khazi FR, Edmondson AC, Nielsen BL (2003) An *Arabidopsis* homologue of bacterial RecA that complements an *E. coli* RecA deletion is targeted to plant mitochondria. Mol Genet Genomics 269:454-463

Kikuchi S, Takaiwa F, Oono K (1987) Variable copy number DNA sequences in rice. Mol Gen Genet 210:373-380

Kitazaki K, Kubo T (2010) Cost of Having the Largest Mitochondrial Genome: Evolutionary Mechanism of Plant Mitochondrial Genome. Journal of Botany 1-12

Klodmann J, Senkler M, Rode C, Braun H (2011) Defining the "protein complex proteome" of plant mitochondria. Amer. Soc. Of Plant Biology DOI:10.1104/pp.111.182352

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46: 123–139

Rüdinger M, Polsakiewicz M, Knoop V (2008) Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. Mol Biol Evol. 25:1405-1414

Krásný L, Vacík T, Fučík V, Jonák J (2000) Cloning and Characterization of the *str* Operon and Elongation Factor Tu Expression in *Bacillus stearothermophilus*. J Bacteriol. 182:6114–6122

Kubo N, Harada K, Hirai A, Kadowaki K (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. Proc Natl Acad Sci USA. 96:9207-9211

Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. Mitochondrion 8: 5-14

Kuhn J, Binder S (2002) RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. Nucl Acids Res 30:439-446

Kuhn J, Tengler U, Binder S (2001) Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. Mol Cell Biol. 21:731-742

Kühn K, Weihe A, Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. Nucl Acids Res 33:337-346

Kurihara-Yonemoto S, Kubo T (2010) Increased accumulation of intron-containing transcripts in rice mitochondria caused by low temperature: is cold-sensitive RNA editing implicated? Curr Genet 56:529–541

Kurihara-Yonemoto S, Handa H (2001) Low temperature affects the processing pattern and RNA editing status of the mitochondrial *cox2* transcripts in wheat. Curr Genet 40:203-208

Lecompte O, Ripp R, Thierry J, Moras D, Poch O (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucl Acids Res 30: 5382-5390

Levy A, Andrh C, Walbot V (1991) Analysis of a 120-Kilobase Mitochondrial Chromosome in Maize. Genetics 128:417-424

L'Homme Y, Brown GG (1993) Organizational differences between cytoplasmic male sterile and male fertile *Brassica* mitochondrial genomes are confined to a single transposed locus, Nucleic Acids Res. 21:1903–1909.

Li L, Wang B, Liu Y, Qiu YL (2009) The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. J Mol Evol 68:665-678

Lightowlers R, Chrzanowska-Lightowlers Z (2008) PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression. 416:e5-e6

Lin N.-S, Hu C.-C, Chien L.-F, Hsu Y.-H (2007) A Mitochondrial Genome Sequence of Bamboo and Comparison to Those of Cereals (unpublished)

Li-Pook-Than J, Carrillo C, Bonen L (2004) Variation in mitochondrial transcript profiles of protein-coding genes during early germination and seedling development in wheat. Curr Genet 46: 374-380

Li-Pook-Than J, Carrillo C, Niknejad N, Calixte S, Crosthwait J, Bonen L (2007) Relationship between RNA splicing and exon editing near intron junctions in wheat mitochondria. Physiol. Plant. 129:23-33

Liu H, Cui P, Zhan K, Lin Q, Zhuo G, Guo X, Ding F, Yang W, Liu,D, Hu,S, Yu J, Zhang A (2011) Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. BMC Genomics 12:163-177

Liu S, Zhuang Y, Zhang P, Adams KL (2009) Comparative analysis of structural diversity and sequence evolution in plant mitochondrial genes transferred to the nucleus. Mol Biol Evol 26:875-891

Lupold D, Caoile A, Stern D (1999) Polyadenylation Occurs at Multiple Sites in Maize Mitochondrial *cox2* mRNA and Is Independent of Editing Status. The Plant Cell 11:1565-1577

Marechal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. New Phytologist 186:299-317

Martin W and Kooin E (2006) Introns and the origin of nucleus-cytosol compartmentalization. Nature 440:41-45

Nakazono M, Nishiwaki S, Tsutsumi N, Hirai A (1996) A chloroplast-derived sequence is utilized as a source of promoter sequences for the subunit 9 of NADH dehydrogenase (*nad9*) in rice mitochondria. Mol Gen Genet 252:371-378

Naydenov NG, Khanam SM, Atanassov A, Nakamura C (2008) Expression profiles of respiratory components associated with mitochondrial biogenesis during germination and seedling growth under normal and restricted conditions in wheat. Genes Genet Syst 83:31-41

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics 268:434-445

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

Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N, Takumi S, Ikeo K, Gojobori T, Murai R, Murai K, Matsuoka Y, Ohnishi Y, Tajiri H, Tsunewaki K (2002) Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol. Genet. Genomics 266:740-746

Ogihara Y, Kurihara Y, Futami K, Tsuji K, Murai K (1999) Photoperiod-sensitive cytoplasmic male sterility in wheat: nuclear-mitochondrial incompatibility results in differential processing of the mitochondrial *orf25* gene. Curr Genet 36:354-362

Ogihara Y, Terachi T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proceedings of the National Academy of Sciences 85:8573–8577

OgiharaY, YamazakiY, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, Takumi S, Murata M, Futo S, Tsunewaki K (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucl Acids Res 33:6235-6250

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

Pelletier G, Budar F (2007) The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. Curr. Op. In Biotech. 18:121–125

Perrin R, Lange H, Grienenberger JM, Gagliardi D (2004a) AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria. Nucleic Acids Res. 32:5174-5182

Perrin R, Meyer E, Zaepfel M, Kim Y, Mache R, Grienenberger J, Gualberto J, Gagliardi D (2004b) Two exoribonucleases act sequentially to process mature 3'-ends of *atp9* mRNAs in *Arabidopsis* mitochondria. J Biol Chem. 279:25440-25446

Pflaz J, Bayraktar OA, Prikryl J, Barkan A (2009) Site-specific binding of a PPR protein defines and stabilizes 50 and 30 mRNA termini in chloroplasts. EMBO J 28:2042-2052

Remacle C, Marechal-Drouard L (1996) Characterization of the potato mitochondrial transcription unit containing 'native' *trnS* (GCU), *trnF* (GAA), and *trnP* (UGG). Plant Mol. Bio. 30:533-563

Riley M, Abe T , Arnaud M, Berlyn M, Blattner F, Chaudhuri R, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G III, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D and Wanner BL (2006) Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. Nucleic Acids Res. 34:1-9

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

Sandoval P, Leon G, Gomez I, Carmona R, Figueroa P, Holuigue L, Araya A, Jordana X (2004) Transfer of RPS14 and RPL5 from the mitochondrion to the nucleus in grasses. Gene 324:139-147

Schardl CL, Pring DR, Lonsdale DM (1985) Mitochondrial DNA rearrangements associated with fertile revertants of s-type male-sterile maize. Cell 43:361–368

Schilders G, van Dijk E, Raijmakers R, Pruijn G (2006) Cell and molecular biology of the exosome: how to make or break an RNA. Int Rev Cytol. 251:159–208

Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 13:663-670

Schuster G, Lisitsky I, Klaff P (1999) Polyadenylation and degradation of mRNA in the chloroplast. Plant Physiol. 120:937-944

Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. Plant Cell 19: 1251-1264

Shikanai T (2006) RNA editing in plant organelles: machinery, physiological function and evolution. Cell Mol Life Sci. 63:698-708

Signorovitch A, Buss L, Dellaporta S (2007) Comparative Genomics of Large Mitochondria in Placozoans. PLos Genetics 3:44-50

Subramanian S, Fallahi M, Bonen L (2001) Truncated and dispersed *rpl2* and *rps19* pseudogenes are cotranscribed with neighboring downstream genes in wheat mitochondria. Curr Genet 39:264-27

Susi P, Hohkuri M, Wahlroos T, Kilby NJ (2004) Characteristics of RNA silencing in plants: similarities and differences across kingdoms. Plant Mol Biol. 54:157-174

Takenaka M, Verbitskiy D, van der Merwe JA, Zehrmann A, Brennicke A (2008) The process of RNA editing in plant mitochondria. Mitochondrion 8:35-46

Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N (2007) The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. Mol Biol Evol 24:699-709

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

Vaidya A, Mather M (2009) Mitochondrial Evolution and Functions in Malaria Parasites. Annu. Rev. Microbiol. 63:249–67

Ward B.L., Anderson R.S., Bendich A.J (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell 25:793–803

Wisecaver J, Hackett J (2010) Dinoflagellate Genome Evolution. Annu Rev Microbiol. Sep 28. [Epub ahead of print].

Williams M, Johzuka Y, Mulligan R, (2000) Addition of non-genomically encoded nucleotides to the 3'-terminus of maize mitochondrial mRNAs: truncated *rps12* mRNAs frequently terminate with CCA. Nucleic Acids Res 28: 4444–4451

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

Woodson J, Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. Nature Reviews Genetics 9:383-395

Yura K, Sulaiman S, Hatta Y, Shionyu M, Go M (2009) RESOPS: a database for analyzing the correspondence of RNA editing sites to protein three-dimensional structures. Plant Cell Physiol. 50:1865-1873

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD, Noller HF (2001) Crystal structure of the ribosome at 5.5 Å resolution. Science 292:883-896

Zehrmann A, Verbitskiy D, Härtel B, Brennicke A, Takenaka M (2011) PPR proteins network as site-specific RNA editing factors in plant organelles. RNA Biology 8:1, 67-70

Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M (2009) A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. Plant Cell 21: 558-567

Zhang Q-Y and Liu Y-G (2006) Rice mitochondrial genes are transcribed by multiple promoters that are highly diverged. J Integr Plant Biol 48:1473-1477

Zhuo D, Bonen L (1993) Characterization of the S7 ribosomal protein gene in wheat mitochondria. Mol Gen Genet 236:395-401

Zhuo D, Nguyen-Lowe HT, Subramanian S, 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

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucl Acids Res 31:3406-3415

**Appendix 1:**

Gray: homologous sequences (asterisks in C show identical sequences so shading not used)
Blue: *trnS* with acceptor stem in red letters
Green: sequence homologous to the 5'*rrn23*
Underline: possible promoter motifs (CRTA)
Red and green underline: BamHI and HindIII sites respectively
Double underline: potential stem-loop structures
~~Strikethrough~~: sequence homologous to pre-RNA of *rrn26*
Red: sequence homologous to sequences within the 5'UTR of *rpl5*
Purple: full and half expression cassette with end of cassette denoted by T
Blue: *rps7*-specific sequence
Green: *rps7* start codon
Red: *rps7* stop codon
Pink: nucleotide differences in rps7 coding sequence
Blue: sites exhibiting editing.

NB. Upstream sequence for oat rps7 corresponds to cDNA-level data and therefore editing sites are shows in yellow in the edited form.

A) Upstream sequence alignment for *rps7* among grasses surveyed in this study.

```
brome           GGGAGATTCTTTCACCCT----TTAGTAGTGCGAGGAGGATAGAGCTCCCAGCAACGAGT 56
barley          GGGAGATTCTTTCACCCTGTTTTTAGTAGTGCGAGGAGGATAGAGCTCCCAGCAACGAGT 60
rye             GGGAGATTCTTTCACCCT----TTAGTAGTGCGAGGAGGATAGAGCTCCCAGCAACGAGT 56
Lolium          CCCGCAGGGAGCGCAGCAACCGACGGTGCATATTATCATATAGAAAGAAGCGAAGCGTAG 60

brome           GTGAAGTTGACATTCACATTGTTCGTACGAGCGAAGGAGCGAGCATGCTTTACCCTTAGA 116
barley          GATTGTTCGTACGAGCGAAGGAGCGAGCATGCTTTACCCTTAGAGTGAAGTTGACATTCA 120
rye             GATTGTTCGTACGAGCGAAGGAGCGAGCATGCTTTACCCTTAGAGTGAAGTTGACATTCA 116
Lolium          CAGCGATTCGTACTACCGGAAAAGTGTCGCTAACCGCTAGGCAATAGAGTCAGCTTACGG 120

brome           GAGGTCTTGTGGGAACTCCTTCAGTTCAGGTAGCCATAGTCACTCAGGGCTATAGTGGTA 176
barley          CAGGTCTTGTGGGAACTCCTTCAGTTCAGGTAGCCATAGTCACTCAGGGCTATAGTGGTA 180
rye             CAGGTCTTGTGGGAACTCCTTCAGTTCAGGTAGCCATAGTCACTCAGGGCTATAGTGGTA 176
Lolium          GGTTGGGGCGCAAGCAAGCGTAGCGAATCACATAGAGTTGCCCCTACCTGCCAGCGCGCA 180

brome           GTAGCAAGATCGAAATTGTGAGATTGGCTCGATCACAAACTATCAAGAAGAAAGATCAAG 236
barley          GTAGCAAGATCGAAATTGTGAGATTGGCTCGATCACAAACTATCAAGAAGAAAGATCAAG 240
rye             GTAGCAAGATCGAAATTGTGAGATTGGCTCGATCACAAACTATCAAGAAGAAAGATCAAG 236
Lolium          GATAGATAGGGCGGGCGAGCGCAGGGATGGATGTCTGAGCGGTTGAAAGAGTCGGTCTTG 240

brome           ATTCACTTTTGTCGGGCAAAGAACTCTATTGAGGAAATGACTTCGCTCACTGAAGACCAA 296
barley          ATTCACTTTTGTCGGGCAAAGAACTCTATTGAGGAAATGACTTCGCTCACTGAAGACCAA 300
rye             ATTCACTTTTGTCGGGCAAAGAACTCTATTGAGGAAATGACTTCGCTCACTGAAGACCAA 296
Lolium          AAAACCGAAGTATTGATAGGAATACCGGGGGTTCGAATCGCTCTCCATCCGCGAGGTCAT 300

brome           GCAGCAGCCATTCTAGATGAAAGGACACAAGCCTAAAACTTGAATGCGGATCTATTTATC 356
barley          GCAGCAGCCATTCTAGATGAAAGGACACAAGCCTAAAACTTGAATGCGGATCTATTTATC 360
rye             GCAGCAGCCATTCTAGATGAAAGGACACAAGCCTAAAACTTGAATGCGGATCTATTTATC 356
```

```
Lolium          AAGTTCTCTCTTGCCTTATCTATAGATAAGAACGAATCTCCTCGACTCGACTGATATGAT 360

brome           TGGAATTCCTATCTTCATAAATAAAGAGTTGCCTATCTTCAGAAATGCTGGTATAGTTGA 416
barley          TTGAATTCCTATCTTCAGAAATAAAGAGTTGCCTATCTTCAGAAATGCTGGTATGGTTGA 420
rye             TGGAATTCTTATCTTCATAAATAAAGAGTTGCCTATCTTCAGAAATGCTGGTATAGTTGA 416
Lolium          GGATGGAATGGGTAGAAGGTTGAGGTTATTGTGTGTTGATTTAGTTAGGACTTTGTCTCC 420

brome           CGACTTGCCTCTGTTCTCGCACCGGTGTCTGCCAGGGTCTGAACCATTAGGTAGAGGATC 476
barley          CGACTTGCCTCTGTTCTCGCACCGGTGTCTGCCAGGGTCTGAACCATTAGGTAGAGGATC 480
rye             CGACTTGCCTCTGTTCACGCACCGGTGTCTGCCAGGGTCTGAACCATTAGGTAGAGGATC 476
Lolium          CTTTCGTTATCTTCCGCCCCGGGTGGATGACCTGTGGGAGCTAAGTCG-AAGATCTCGGT 480

brome           AGCTGAAAGAAAAGGCGGACAGAGATCTTGTGCCAGATCCGTATACATTCAGCACCCTCA 536
barley          AGCTGAAAGAAAAGGCAGACAGAGATCTTGTGCCAGATCCGTATACATTCAGCACCCTCA 540
rye             AGCTGAAAGAAAAGGCGGACAGAGATCTTGTGCCAGATCCGTATACATTCAGCACCCTCA 536
Lolium          GTCGTCGTGAGTGGTAGCAGCGATTGGAGTTTGATTTAGGCCCTGTGAAGCTTAACAGAC 540

brome           TTACAGCATTTGCTAAGCATGAAAAGTGGTTGGAGCTATGCTTATGTGCACGAAAATGCA 596
barley          TTACAGCATTTGCTAAGCATGAAAAGTGGTTGGAGCTATGCTTATGTGCACGAAAATGCA 600
rye             TTACAGCATTTGCTAG-CATGAAAAGTGGTTGGAGCTATGCTTATGTGCACGAAAATGCA 595
Lolium          AAACAAAGAAAACGATAGAGACTTCCGGAGGGTGACGACCTTAATGAATCAAGTATTCAG 600

brome           G-AGAAAGGATGGCACCCAGAGACGAGGATGG--CGTAGCAAGCGACGAAATGCTTCGGG 653
barley          GGAGAAAGGATGGCACCCAGAGACGAGGAAGG--CGTAGCAAGCGACGAAATGCTTCGGG 658
rye             GGAGAAGG-ATGGCACCCAGAGACGAGGAAGGGCCGTAGCAAGCGACGAAATGCTTCGGG 654
Lolium          GTGGCAGAGTTATTAGCAACAGCTCAAATTCATGAATACTTATTGCCCTTGCTTCTATGA 655
Oats            ------------------------------------------------------CCATGGA 7

brome           G-AGTTGAAAATAAGCATAGATCCGGAGATTCCCAAATAGGTCAACCTTTTGAAC----- 707
barley          G-AGTTGAAAATAAGCATAGATCCGGAGATT-CCAAATAGGTCAACCTTTTGAAC----- 711
rye             GGAGTTGCAAATAAGCATAGATCCGGAGATTCCCAAATAGGTCAACCTTTTGAAC----- 709
Lolium          TCCTGGCACATTTAGGTTTTGATCTGAGGCTATCCTGGTCTGCAACCTTTTGAACTGTCT 715
Oats            CTGACTCAATGTTGGACATTGCCCAATTTTCCATTTGTTTAGCTAATATTCTGATGACCG 67

brome           TGCCTGCTGAATCCATGAGCAGGCAAGAGACAACCTGGCGAACTGAAACATCTTAGTAGC 767
barley          TGCCTGCTGAATCCATGAGCAGGCAAGAGACAACCTGGCGAACTGAAACATCTTAGTAGC 771
rye             TGCCTGCTGAATCCATGAGCAGGCAAGAGACAACCTGGCGAACTGAAACATCTTAGTAGC 769
Lolium          GCTGAATCCATGAGCAGGCAAGAGACAACCTGGCGAACTGAAACATCTTTCTTTTCTAGC 775
oats            GGGCCGCCAAGCCTCAAGGACTTATCAATTTTTCCTAGGGGGGAATCCAATCCATTTCTT 127
wheat           GGCAGCCGGCCAATAGGGGAAGGTTGTGAATCCGGCGAACCGACCGCTTTAAGAACGTGA 60
rice            CGTAGGTTTTGTGTCTGTCTTATGTCATCCTGGTGATTTATCATCGGAATAGCTCAGTTC 60
maize           AATCAAGTCTCATGTTGCTCCTCAGAAAACGCGTATAGTATATAGTATATAGTATATAGT 60

brome           CAGAGGAAAAGAAAGTCTCATGTTGCTCTTCAGAAAACGCGTATAGTGGCCTTCGTCGAT 827
barley          CAGAGGAAAAGAAAGTCTCATGTTGCTCCTCAGAAAACGCGTATAGTGGCCTTCGTCGAT 831
rye             CAGAGGAAAAGAAAGTCTCATGTTGCTCTTCAGAAAACGCGTATAGTGGCCTTCGTCGAT 829
Lolium          CAGAGGAAAAGAAAGTCTCATGTTGCTCCTCAGAAAACGCGTATAGTGGCCTTCGTCGAT 835
oats            TTATGAAATCAATGTTAATGTAGTTACAGATGTGAAAAAGGTAAATATCTTTCTGTCTGT 187
wheat           TTGTCTTCTCTCACTCAGTTATCAATTGACAAAGATGACCCATTCTTTTTCGCCCTAAAA 120
rice            GAGGGAGGGGGGGGTGGTAAGCTGAAGCGTCGAATAGTCTTTTTAGTGTACAGTGTCTTG 120
maize           ATATAGTATATAGTATATAGTATATAGTAGTAGTAGTATCATTGGCCAAAGTAAAGTCGA 120

brome           GGGAC-----CTCCAGTGTATGCGTTACAAGGCAACTAGCATTTTG----------TTCGT 873
barley          GGGACAAACGCTCCAGTGTATGCGTTACAAGGCAACTAGCATTTAG----------TTCGT 882
rye             GGGAC-----CTCCAGTGTATGCGTTACAAGGCAACTAGTTG-------------TTCGT 871
Lolium          GGGAC-----CTCCAGTGTATGCGTTACAAGGCAACTAGTATTTTGTCATGGAAGTTCGT 890
oats            GTAGGTTCGGTTTGTTCTTATCGATCCATTTTATCGAGTGTGGGTGTCCTTCCGTTCATG 247
wheat           ACGACAATGGTATGGTACTTTTTCTTCAAATCGAGATTGTGTGGGTGTCCGCTCATGTTC 180
rice            GTTTTGGTTCGATCAACTATCCGCTTCAAAAAGGATAGTTCACAGTGTGCTCATTCTCAA 180
maize           TGGGACAAACGCTGTTCGTTAGAACTAGCATTTTGTTTTGTCATGGAATCAAGTCTATTT 180

brome           GAAAGAATGTTC---------TTTCGTTGGAAAAACCAACGCCGACGTCAAGATC----A 920
barley          GAAAGAATG------------TTTCGTTGGAAAAACCAACGCCGACGTCAAGATC----A 926
rye             GAAAGAATGTTC---------TTTCGTTGGAAAAACCAACGCCGACGTCAAGATC----A 918
Lolium          GAAAGAATGTTCTT----GTTTTTCGTTGGAAAAACCAACGCCGACGTCAAGATC----A 942
oats            TTTACGTTCTCAAAATCAGGCTTTCATTGGAAAAACCAACGCCGACCCCTATCTC----A 303
wheat           ACGTTACATGCTAAATCAGGCTTTCCTTGGAAAAACCAAGGACAACCCCTATCTC----A 236
rice            AAAAAGAAAAAAACTTCTTCGTTTCGTTGGAAAAACCGACGCCAACGTTAAGATC----A 236
maize           GTTCGAATGTTCTTTTTTTCGTTTCGTTGGAAAAACCTACGCCCACATATTGATCTTTAA 240
```

```
brome           GTCTCCTTTCTCTTTT-------CGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCG 973
barley          GTCTCCTTTCTCTTTT-------CGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCG 979
rye             GTCTCCCCTTTATTTGCAAAGTGAGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCG 978
Lolium          GTCTCCT----------------CGGGAGCAGAGCTTCAAAAGATGGACAGTAACGATCG 986
oats            GTCTCCTTT--------TCTTTTCGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATTG 355
wheat           GTCTCCTTTCC--TTTCTCTTTTCGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCG 294
rice            GTCTCCTTT-------CTCTTTTCGGGAGCAGAGCTTAAAAAGATGGACAGTAACGATCG 289
maize           GTCTCCTTT--------CTCTTTTGGGAGCAGAGCTGAAAAAGATGGACAGTAACGATCG 292

brome           CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 1033
barley          CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 1039
rye             CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 1038
Lolium          CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 1046
oats            CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 415
wheat           CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 354
rice            CGTAATATCAATTTATCGGCCTCGTCATCGATT------TCCAATTGCTCGGAAATTCTC 343
maize           CGTAATATCAATTTATCGGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTC 352

brome           AGCTATATG 1042
barley          AGCTATATG 1048
rye             AGCTATATG 1047
Lolium          AGCTATATG 1055
oats            AGCTATATG 424
wheat           AGCTATATG 363
rice            AGCTATATG 352
maize           AGCTATATG 361
```

B) Downstream sequence alignment for *rps7* among grasses surveyed in this study.

```
rice            TAAAGTGATACCATATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60
maize           TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTCAGAAA 60
wheat           TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60
oats            TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60
rye             TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60
brome           TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60
barley          TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60
Lolium          TAAAGTGATACCACATAAGGAGCTCTTCCTCATTCAGTCATACTCAACAAAAGTAAGAAA 60

rice            TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
maize           TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
wheat           TGTTTGACCCGGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
oats            TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
rye             TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
brome           TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
barley          TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120
Lolium          TGTTTGACCCTGATCCTTTTTTCATCTTCATATAGAAAGAAAATCGGCCTTCCTCATACT 120


rice            CCCCCCTTCATTCATGGAGTTGGAGGAATCCACAAGAGGCCTGCCCGTTCATAATTGCAT 180
maize           CCCCCCTTCATTCATGGAGTTGGAGGAATCCACAAGAGGCCTGCCCGTTCATAATTGCAT 180
wheat           CCCCCCTTCATTCATAGAGTTGGAGGAATCCACAAGAGGCCTGCCCGTTCATAATTGCAT 180
oats            CCCCCCTTCATTCATAGAGTTGGA---------------------------------- 144
rye             CCCCCCTTCATTCATAGAGTTGGA---------------------------------- 144
brome           CCCCCCTTCATTCATAGAGTTGGA---------------------------------- 144
barley          CCCCCCTTCATTCATAGAGTTGGA---------------------------------- 144
Lolium          CCCCCCTTCATTCATAGAGTTGGA---------------------------------- 144

rice            AAAAGAACCAT----TTTATGAAAACTCTTGTTCCAAACAAGCAACGGATTGAGCGCACT 236
maize           AAAAGAACCATTCTTTTTATGAAAAC---------AAACAAGCAAGGGATTGAGC-AACT 230
wheat           AAAAGAACCATTCTTTTTATGAAAACTCTTGTTCCAACCTCACCTCAGGTCGAATGAATA 239

rice            AGCGCGAAAGCGTTAGCACGCGCATCCGTTTTCTTGCTTTGAATCCAATGGGTTTCGTCA 296
maize           AGCGC--------TAG---GCGCATCC----TCTTGCT--GGG----ATGGTTTGAGACA 269
wheat           CGAAAGGGGGATCAATCAAATCAATAAGCCATGAATGAAG-------------------- 280
```

C) Alignment of *rps7* coding sequence among grasses surveyed in this study.

```
brome           ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
oats            ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
barley          ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
rye             ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
wheat           ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
maize           ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
Lolium          ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
rice            ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACTTTCGCATG 60
                ************************************************************

brome           ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTCCACCGCCTAGCT 120
oats            ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
barley          ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
rye             ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
wheat           ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
maize           ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
Lolium          ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
rice            ATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACTTTTCACCGCCTAGCT 120
                ********************************************** ************

brome           CGAACTGAACGCGATGTAATAAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
oats            CGAACTGAACGCGATGTAATAAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
barley          CGAACTGAACGCGATGTAATAAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
rye             CGAACTGAACGCGATGTAATAAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
wheat           CGAACTGAACGCGATGTAATAAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
maize           CGAACTGAACGCGATGTAATCAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
Lolium          CGAACTGAACGCGATGTAATCAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
rice            CGAACTGAACGCGATGTAATCAAACTTATGGTTGACGCCGTAGATAATATAAAGCCAATA 180
                ******************* ****************************************

brome           TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
oats            TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
barley          TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
rye             TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
wheat           TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
maize           TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
Lolium          TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
rice            TGCGAAGTGGTCAAAGTAGGAGTCGCAGGTACTATTTATGATGTTCCTGGGATTGTAGCC 240
                ************************************************************

brome           AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
oats            AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCTGCTTTCAAACGACGT 300
barley          AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
rye             GGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
wheat           AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
maize           AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
Lolium          AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
rice            AGGGATCGTCAACAAACCTTAGCTATTCGTTGGATCCTTGGAGCAGCTTTCAAACGACGT 300
                .****************************************** ***************

brome           ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
oats            ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
barley          ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
rye             ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
wheat           ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
maize           ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
Lolium          ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
rice            ATAAGCTACAGGATAAGCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGA 360
                ************************************************************

brome           AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG 420
oats            AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG 420
barley          AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCCTCCACCAATCGG 420
rye             AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG 420
wheat           AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG 420
maize           AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG 420
Lolium          AAGAGGGGAATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG 420
rice            AAGAGGGGAATTTCACGTAAGAGAAGGGCGAATCTTCATGGACTGGCTTCCACCAATCGG 420
```

```
                ************************** ***************** ************

brome           AGTTTCGCGCATTTCAGATGGTGGTAA 447
oats            AGTTTCGCGCATTTCAGATGGTGGTAA 447
barley          AGTTTCGCGCATTTCAGATGGTGGTAA 447
rye             AGTTTCGCGCATTTCAGATGGTGGTAA 447
wheat           AGTTTCGCGCATTTCAGATGGTGGTAA 447
maize           AGTTTCGCGCATTTCAGATGGTGGTAA 447
Lolium          AGTTTCGCGCATTTCAGATGGTGGTAA 447
rice            AGTTTCGCGCATTTCAGATGGTGGTAA 447
                **************************
```

**Appendix 2:**

Unpublished mtDNA sequence for the oat *rps7* coding region and upstream cDNA-level data from CR-RT-PCR clones for mitochondrial oat and rye *rps7* flanking sequences. Upstream oat and rye sequence correspond to the longest 5'UTRs among clones in table 3.1 and 4.1.

>oatrps7
ATGGGGGACTTTGATGGTGAGCAAAAAGAATTGATCAAGAAATTGGTAAACT
TTCGCATGATCGATGGTAAAAGAACGAGAGTTCGTGCTATTGTTTATAAAACT
TTTCACCGCCTAGCTCGAACTGAACGCGATGTAATAAAACTTATGGTTGACGC
CGTAGATAATATAAAGCCAATATGCGAAGTGGTCAAAGTAGGAGTCGCAGGT
ACTATTTATGATGTTCCTGGGATTGTAGCCAGGGATCGTCAACAAACCTTAGC
TATTCGTTGGATCCTTGGAGCTGCTTTCAAACGACGTATAAGCTACAGGATAA
GCTTAGAGAAATGTTCATTTGCTGAGATACTGGATGCTTACCGAAAGAGGGG
AATTTCACGTAAGAGAAGGGAGAATCTTCATGGACTGGCTTCCACCAATCGG
AGTTTCGCGCATTTCAGATGGTGGTAA

>oatrps7up
CCATGGACTGACTCAATGTTGGACATTGCCCAATTTTCCATTTGTTTAGCTAA
TATTCTGATGACCGGGGCCGCCAAGCCTCAAGGACTTATCAATTTTTCCTAGG
GGGGAATCCAATCCATTTCTTTTATGAAATCAATGTTAATGTAGTTACAGATG
TGAAAAAGGTAAATATCTTTCTGTCTGTGTAGGTTCGGTTTGTTCTTATCGAT
CCATTTTATCGAGTGTGGGTGTCCTTCCGTTCATGTTTACGTTCTCAAAATCAG
GCTTTCATTGGAAAAACCAACGCCGACCCCTATCTCAGTCTCCTTTTCTTTTCG
GGAGCAGAGCTGAAAAAGATGGACAGTAACGATTGCGTAATATCAATTTATC
GGCCTCGTCATCGAAAGCGGCTTCCAATTGCTCGGAAATTCTCAGCTAT(ATG)

>ryerps7up
TAGACCTATTCCATGTCTATATCTCTCTGTTTTGTCCTTTTTACTTTTTCCTATC
CTCTCTTGCATGGCATTTCATCTCGCATGTAGAAAAACAGCGCTCATTGTGTA
TAGCTATCATTGAAAGACCCTCTCAATAATAACTTAAACAAACAAAGTACGT
AATGGCCTCAGGCCTGCTCTATTACATCGACCACCTTCGAGTTCTCTATATGC
GAAGAGCCTGGAATGGTGATAGAACTATCGGAAACTACCCAACGACGAGCG
AGCCTGAGATCGCCTACCAGTCATTCCATTGCTTGAAAAAGTGAGCCTGGAT
CTTAGGAGGAAGGATTGAAGAGCTCTAGAGAAGGTCTGGAAAGGATGGTCG
GCCACTAGAAAGCAACAATTCAGCCGGAAGTATGGACACATTGGCTCTCGTC

TTAAAGTTCCCATTGACGACCAAAGGATTAATTCCTCATGGAGGAAAGCCTC
ACTACCGATGCTTTACTTTCGGCTCCATTGATCTTACTCCAACCTACAGAATA
CACTGTCTTGCTCAATGTGCAAGGGATCATTCTATAAACTTGGTCTTGGAAAG
CCTATCAAGAGTAGCTCGGCATCGGGAATCTATGAAGGAAGAATGCCCCTTT
GCTTTCTTAGGGGAATAGCCGCTCTTTTATCAGATGCGAGATTACCGGTCTGT
CTCCTTAATAAAGAGTCGAGGAATAGGAGGAGAATCCCCAGGCAAACAACT
AGGGTTCCAGATTCACTGGGCAGAGAAGGGAGAATAGTATCCTCTCACAGTA
GGTGACGTTTTGGGCGAACTCACTAGTCTCCTAATTAATTCCCCACGATAGGA
AAATAGAGAGAGCACAAGACTTTTAAGCACCTTAGGGGTTGATGTAGAAAGC
TT