Evolutionary status of mitochondrial ribosomal protein genes *rps19* and *rpl2* and their transfer to the nucleus in grasses

Sruthi Atluri

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

Massive mitochondrial gene transfer to the nucleus occurred very early during eukaryotic evolution following endosymbiosis, and is still ongoing in plants. Tracking recent gene transfer events can give us insight into the evolutionary processes by which a transferred gene becomes functional in the nucleus and how its protein gets targeted back into the mitochondrion, where it is needed. *Rps19* and *rpl2* are two such ribosomal protein genes that are known to have been transferred to the nucleus, many times independently during flowering plant evolution. My research project focusses on determining the status and expression of *rps19* and *rpl2* in the mitochondrion and nucleus of selected grasses and in particular brome (close relative to agronomically important crops such as wheat, rye and barley). My results at the level of DNA and RNA (PCR and RT-PCR, respectively) show that the mitochondrial brome *rpl2* copy is a pseudogene while its functional gene is in the nucleus. The brome mitochondrial genome has a copy of *rps19* which is transcribed and C-U edited. Surprisingly, the brome nuclear genome also has functional copies of *rps19.*The targeting sequence for the nuclear *rps19* gene was acquired from duplication of mitochondrial targeting heat shock protein (hsp70) presequence. Comparative analysis strongly suggests that a functional *rps19* gene was transferred to the nucleus before rice and maize lineages split and now that brome *rps19* has been found to be present in both compartments, this implies a transition stage of about 60 million years. Oats was found to have a functional *rps19* copy in the nucleus and has a novel presequence due to lineage specific rearrangements and exon shuffling. Functional paralogous copies were found in wheat, and maize while barley lost one of the copy. Thus, following transfer, duplication of *rps19* gene must have occurred in the ancestor of barley and wheat clade. Maize might have had a recent duplication or gene conversion events along its lineage as its paralogous copies are very similar to each other. More information is needed to determine if this duplication event extends to wheat-brome, wheat-oats or even before rice and maize split. Barley was also found to have a recent independent DNA mediated transfer in addition to the common transfer, as it possesses an unedited nuc-mt *rps19* in its nuclear genome. This suggests that barley must also have had a transition stage for ~60MY and lost its mitochondrial copy very recently.

Résumé

Le transfert massif d’ADN du génome des mitochondries, organite d’origine eubactérienne, vers celui du noyau est un processus qui a lieu depuis le début de l’évolution eucaryote (après l’endosymbiose) et qui est toujours actif chez les plantes. Détecter les évènements récents de transfert de gènes peut nous donner des informations sur les processus évolutifs selon lequel un gène transféré devient fonctionnel dans le noyau et comment la protéine est ciblée dans la mitochondrie, là où elle est nécessaire. *rps19* et *rp12* sont deux protéines ribosomales connues pour avoir été transférées dans le noyau plusieurs fois et indépendamment pendant l’évolution des plantes à fleurs. Mon projet de recherche s’est focalisé sur la détermination du statut et de l’expression de *rps19* et *rp12* dans la mitochondrie et le noyau de certaines plantes herbacées et en particulier dans le genre *Bromus* (phylogénétiquement proche de cultures importantes pour l’agronomie comme le blé, l’orge et le seigle). Mes résultats au niveau de l’ADN et de l’ARN (PCR et RT-PCR, respectivement) ont montré que la copie du gène mitochondrial *rp12* chez les bromes est un pseudogène, alors qu’il est fonctionnel dans le noyau. Le génome mitochondrial du brome a une copie de *rps19* qui est transcrit et édité C-U, le génome nucléaire a aussi des copies fonctionnelles de *rps19*. La séquence cible pour le gène nucléaire *rps19* fut acquise par duplication de la protéine de choc thermique (hsp70). Une analyse comparative suggère qu’un gène *rps19* fonctionnel fut transféré au noyau avant la séparation des lignées du riz et du maïs, et que maintenant le *rps19* du brome est présent dans les deux compartiments, ce qui implique une étape de transition vers 60 million d’années. Chez l’avoine, il y a une copie fonctionnelle de *rps19* dans le noyau et il y a une nouvelle pré-séquence due à des réarrangements spécifiques de brassage d’exon dans la lignée. Des copies paralogues fonctionnelles ont été trouvées chez le blé et le maïs tandis que l'orge a perdu l'un de la copie . Ainsi, après le transfert, le dédoublement des *rps19* gène doit avoir eu lieu dans l'ancêtre de l'orge et du blé clade. Le maïs peut avoir eu une récente duplication ou les événements de conversion génique le long de son lignage comme ses copies paralogues sont très semblables les uns aux autres. Plus de renseignements sont nécessaires pour déterminer si cet événement de duplication s'étend au blé-brome, blé-avoine ou avant même l’ancêtre du riz et le maïs. On a trouvé que l'orge ont un ADN indépendante récente médiée transfert en plus du transfert commun, car il possède un non éditées nuc-mt *rps19* dans son génome nucléaire. Ceci suggère que l'orge doit aussi avoir eu une étape de transition pour ~59ma et perdu sa copie mitochondriale très récemment.

ACKNOWLEDGMENTS

I would like to express deepest gratitude to my thesis supervisor Dr. Linda Bonen, for providing me an opportunity to do research in her lab. It has been a wonderful learning experience and a great start to my research career. I really appreciate all the helpful discussions and constructive feedback you have provided me with throughout my project. I have been extremely fortunate to have your guidance and I cant thank you enough for giving me the freedom to explore new directions with my project. I would also like to take this opportunity to thank you Linda, for editing many drafts of my manuscript. I would also like to thank my committee members Dr. Nicolas Corradi and Dr. Shelley Hepworth for their encouragement and insightful comments and questions during committee meetings. A special thanks also goes to Sarah Rampersad (an honours project student) whose findings formed a solid foundation for my project and allowed me to extend further as well as Jordan Silke for revising my thesis. Working in the lab has been a very pleasurable experience, thanks to all my friends/colleagues: Jordan Silke, Karen Massel, Mostafa Kamal as well as previous lab members (SD, AN, CM, and AZ) for your support and help when needed. Thanks to Dr. Douglas Johnson and his lab members for sharing interesting conversations and friendship over the past years.

I would like to say thanks to my family members for being there for me: my sisters (Swetha Atluri, Sravya Atluri and my brother Sai Sandeep Atluri as well as my beautiful niece Ishika Chalasani).

Last but not least, I would like to extend special thanks and dedicate this work to my parents Uma Devi and Durga Prasad Atluri for their support and love without which this would not have been possible.

This one is for you Mom and Dad!

Table of Contents

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3 Results: Retention of functional genes for S19 ribosomal protein in both the mitochondrion and nucleus for over 60 million years

3.5 Chapter 3 addendum:

Chapter 4 Results ……………………………………………………………………………

Chapter 5: General Discussion………………………………………………………….….......

List of Figures

List of Tables

Supplementary tables

Appendices

.

List of abbreviations

atp: adenosine triphosphate synthase subunit genes
cDNA: complementary DNA
Chr: chromosome
CMS: cytoplasmic male sterility
cob: cytochrome bc1 oxidoreductase subunit genes
CR-RT-PCR: circularized reverse transcriptase polymerase chain reaction
CTAB: Cetyl trimethylammonium bromide
dI-dVI: group II introns domains I-VI
DIUF: deionized ultrafiltered
DNA: Deoxyribonucleic acid
EtOH: ethanol
kb: kilobases
mg: milligram
ml: milliliter
M-MLV: Moloney murine leukemia virus
mRNA messenger RNA
Mt: mitochondria, mitochondrial
mya: million years ago
Mya: million years ago
nad: NADH dehydrogenase subunit genes
NADH: nicotinamide adenine dinucleotide (reduced form)
NCBI: National Center for Biotechnology Information
nt: nucleotide(s)
Nuc: nuclear
ORF: open reading frame
PPR: pentatricopeptide repeat
RNA: Ribonucleic acid
rpl: protein of the large ribosomal subunit
rps: protein of the small ribosomal subunit
rRNA ribosomal RNA
RT: reverse transcriptase
RT-PCR: reverse transcriptase polymerase chain reaction
TIM: translocase of the inner membrane
TOM: translocase of the outer membrane
UTR: untranslated region
μg: microgram
μL: microliter

CHAPTER 1 Introduction

1.1 Origin of mitochondria

The widely accepted endosymbiotic theory explains the origin of mitochondria as a result of a eukaryotic or an archaebacterial host cell engulfing an aerobic alpha proteobacterial endosymbiont (reviewed by Gray 1999, Martin *et al 2006*). A symbiotic relationship has been established between the two, as the host cells were able to respire in an aerobic environment and the endosymbiont received a place to reside in. Following endosymbiosis, the endosymbiont lost genes that are no longer needed due to this symbiotic relationship and retained some genes that are useful such as respiratory genes. Many of the genes remaining within the protomitochondrion such as ribosomal protein genes were relocated to the nucleus or have been functionally replaced by genes in the nucleus or chloroplast (Timmis *et al* 2004, Gray 1992). The mitochondrial genomes in different lineages such as animals, plants and fungi differ, as they each acquired their own unique features with respect to organization, size and gene expression, following this event. Animal mitochondrial genomes are usually 16-18kb in size, gene dense and do not have introns. With a few exceptions, all animal mitochondrial genomes have the same 37 genes: two for ribosomal RNAs, 13 for proteins and 22 for tRNAs (Li *et al* 2015). Typical animal mitochondrial DNA also does not undergo frequent recombination. Mitochondrial genomes of plants on the other hand undergo frequent recombination with repeated DNA sequences, variable introns and ORFs. Fungal mitochondrial genomes vary between major phyla with respect to gene order although the gene content is fairly conserved. They also are more close to plants with respect to rearrangements as their genomes undergo frequent recombination (Aguielta *et al* 2014).

The presence of genes inside the mitochondrion and their similarity to alpha proteobacterial genes is a very compelling evidence for endosymbiosis. Recently, a pre-endosymbiotic hypothesis has been suggested in addition to the standard endosymbiotic theory. Pre-endosymbiotic theory states that the host cell, which engulfed an alpha-proteobacteria, already has non-alpha proteobacterial content in the form of a membrane bound organelle also known as pre-mitochondrion (Gray *et al* 2014). This view, has been opposed by Zimorski *et al*, as they state that it does not explain how mitochondria originated but instead explains how mitochondrial proteins came to be in the first place (2014).

1.2 Plant mitochondrial genomes and their notable aspects

Plant mitochondrial genomes are remarkable as they have a number of unique features, which help in distinguishing them from animal and fungal counterparts. Mitochondrial genomes of plants are larger than those of animals (and other eukaryotes). They have circular mitochondrial DNAs of 15-17kb in size and are quite different even between very close relatives. In angiosperms the size of mtDNA ranges from 200KB to 600KB (Sloan *et al* 2012). Still, some flowering plant lineages have bigger mitochondrial genomes seen in cucumber (*Cucumis satinus*) and *Silene nocliflora,* which range from 200KB to 10MB (Gualberto *et al* 2014). Plant mitochondrial genomes contain repeated sequences within their genome, which can have direct or indirect orientation. This enables them to pair up, recombine to generate sub genomic circular DNA molecules (Gualberto *et al* 2014). A high frequency of homologous recombination has been observed in plant mitochondrial genomes by some studies (Klein *et al* 1994, Sugiyama *et al* 2005). Duplication or deletion of genomic sequences is also seen due to presence of numerous intermediate size repeats (ISRs), which vary between about 50-700bp in length. ISRs occur due to break induced replication pathways and single strand annealing (Gualberto *et al* 2014).

Mitochondrial RNA metabolism is very complex in eukaryotes such as plants as the machinery is a combination of new traits in the host cell plus the existing bacterial traits. These mechanisms of RNA metabolism include transcription, RNA editing, the splicing of group I and II introns, maturation of transcript ends, translation and RNA degradation. The abundance of the final protein in plant mitochondria is highly dependent on these post-transcriptional mechanisms. Pentatricopeptide repeat (PPR) proteins which are made up of repeated units of 35 amino acid motifs are found to be the eukaryotic specific factors involved in RNA metabolism in plants (Hammani and Giegé 2014). RNA processing or end maturation occurs at 5’ and 3’ ends following transcription. RNA stem loop folds along with RNA binding proteins, such as PPR proteins, have been implicated in 5’ end or 3’ end maturation events (Forner *et al* 2007). Both function to form secondary structure in order to block and protect the ends from exoribonuclease activity. This prevents RNA degradation and forms stable transcript ends (Binder and Brennicke 2003, Dombrowski *et al* 2007, Forner *et al* 2007).

RNA editing is a posttranscriptional modification process where the RNA molecules undergo nucleotide changes making the RNA sequence different from the DNA sequence. Many different RNA editing methods exist among viruses, primitive eukaryotes, vertebrates, fungi and plants. RNA editing in plants was first discovered in 1989 when they found that a portion of the Cs in the DNA changed to U in the RNA (Covello and Gray 1989). C-to-U editing is also present in plastids but not in the nucleus of plants. U-to-C editing also known as reverse editing is known to occur in ferns, mosses and *lycopodiaceae* (Grewe *et al* 1995, Kugita *et al* 2003). When editing occurs, it typically impacts the first or second position of the codon, which means that the mature messenger RNA will encode a different amino acid at the affected codon than the genomic DNA. RNA editing can change any codon of the messenger and this includes the initiation and termination codons (Takenaka *et al* 2013).

Plant mitochondrial ribosomes are made up of three ribosomal RNAs (26S, 18S and 5S) and 70-80 ribosomal proteins (Pinel *et al* 1986 and Maffey *et al* 1997). Gene density is very low in plant mitochondria for example the 360kb mitochondrial genome of *Arabidopsis* encodes 57 genes, which cover only 10% of the genome (Unseld *et al* 1997). These genes encode proteins for rRNAs, tRNAs and for components of respiratory chain and cytochrome c maturation complexes (Hammani and Giege 2014).

1.3 Plant nuclear genome content and expression

The size of nuclear genomes can vary from species to species. *Arabidopsis* has a small nuclear genome as its size is 125Mb and contains 25,498 genes (The *Arabidopsis* Genome Initiative, 2000) while barley (5.3GB) is among the largest within grasses (Bosler *et al* 2015). Variation in the size of nuclear genome between species usually comes from duplication events. It can also most likely occur due to amplification of DNA, which generates high amounts of repetitive DNA. Nuclear genomes of grasses consist of repetitive DNA that has been derived from transposable elements (currently or previously active) (Diez *et al* 2014). Some of the repetitive DNA encodes ribosomal RNA genes while others can have structural roles at the telomeres or centromeres (Mehrotra and Goyal, 2014). The variation in size of nuclear genomes among different species, are due to these repetitive DNA sequences. For instance species with small genome size such as rice and *Arabidopsis thaliana* have 20% repetitive DNA but species with larger genome such as maize, wheat and barley have up to 85% repetitive DNA (Diez *et al* 2014).

Duplication of single genome (autopolyploidy) or multiple genomes (allopolyploidy) is a common ongoing process in plants (Adams *et al* 2004). Rice is a diploid plant consisting of 12 chromosomes and its nuclear genome size is 382 MB. Maize is also a diploid and its genome size is 2,000MB and consists of 10 chromosomes. Wheat is an allohexaploid (AABBDD) with 21 chromosomes and a 17,000MB genome (Zhiguo *et al* 2013). Bromus inermis is a C3 grass that is classified as an octaploid (AAAA BBBB) consisting of A and B genomes (Armstrong 1991). Thus genome duplication has led to extensive variation even among this small clade of grasses.

Rice (Goff et al 2002), maize (Schnable *et al* 2009), and barely (International Consortium 2012) nuclear genomes are sequenced. The draft nuclear genome sequences are available in the Ensembl Plants database (http://plants.ensembl.org) for *Triticum avestivum* (bread wheat) and includes the genomes of two bread wheat's diploid progenitors: *Triticum uratu* (A genome progenitor) and *Aegilops tauschii* (D genome progenitor) (Kersey *et al* 2014).

Nuclear gene expression is different than mitochondrial gene expression as different signals are required. Three different types of RNA polymerases (Pol I, II and III) control transcription of nuclear genes in plants. RNA polymerase II is responsible for transcribing protein coding genes (Reddy 2007). In plant mitochondria, transcription is mediated by RpoTm, a nuclear encoded single subunit T3/T7 bacteriophage type RNA polymerase (reviewed in Liere *et al* 2011). Plant nuclear genes have spliceosomal introns, while mitochondrial genes have group I and group II introns. Group II introns are hypothesized to be progenitors of spliceosomal introns (Fica *et al* 2013). Intron containing precursor mRNAs undergo "pre-mRNA splicing" which occurs co-transcriptionally at the spliceosome to produce functional mRNAs (Fica *et al* 2013).

The PPR (Pentatricopeptide repeats) protein family has been identified in nuclear genomes of plants. In *Arabidopsis* thaliana, the PPR proteins family is made up of 450 members. They are involved in gene expression and more than 15% of mitochondrial proteins in *Arabidopsis* are made up of PPR proteins. They act as RNA binding proteins which can cause interactions between substrates and enzymes that act upon them. PPR proteins thus allow interaction between mitochondria and the nucleus and help regulate plant mitochondrial gene expression (Chase 2006).

1.4 Mitochondrial gene transfer to the nucleus

Mitochondria and chloroplasts, which were derived from prokaryotes, only contain some of their ancestral eubacterial genome. This is due to relocation of many genes to the nucleus (Timmis *et al* 2004). Up to 99% of mitochondrial proteins are nuclear encoded. Angiosperm mitochondrial genomes have experienced remarkable high rates of gene loss and frequent transfer to the nucleus (Liu *et al* 2003). Although many genes were either lost or transferred to the nucleus during endosymbiosis, some of the genes remained in the mitochondria (67 protein coding genes remained in protist *Reclinomonas americana*, 13 in animals and up to 40 somewhat variable protein coding genes in flowering plants) (reviewed in Adams *et al* 2002, Liu et al 2009). Gene transfer is an ongoing evolutionary process in land plants, protists and green algae (reviewed in Adams and Palmer 2003; Bonen and Calixte 2006 and Liu *et al* 2009). Functional gene transfer to the nucleus ceased in animals and fungi, more than 600 million years ago because of changes in the genetic code of their mitochondrial genomes (Adams and Palmer, 2003).

In the mitochondrial genome of liverwort, there are 16 ribosomal protein genes present and at least 14 have been identified in the mitochondrial genomes of angiosperms (Takemura *et al* 1992). A survey of different ribosomal protein genes in the mitochondrial genome of plants such as liverwort (*Marchantia*), *Arabidopsis* (Eudicot), *Phoenix* or date palm, rice, wheat, maize, bamboo and lolium or rye grass (monocot) shows that only liverwort still retains all 16 in the mitochondrion (Table 1). Genes that are missing, are predicted to be transferred to the nucleus (Table 1). This following table gives us an idea of the prevalence of gene transfer between the mitochondrion and nucleus among different plants for ribosomal protein genes.

There are three main reasons that a eukaryotic cell can tolerate mitochondrial gene loss. First is that the gene that is being lost by the mitochondria is no longer essential. For instance, genes that were needed for making the bacterial cell wall were lost in the early endosymbiont; however, this gene loss is rare in modern day eukaryotes with an active mitochondrion. Mitochondrial gene loss can also occur, because the function of the gene has been replaced by another gene in the nucleus. Finally, gene loss occurs in the mitochondria when the gene has been functionally transferred to the nucleus, which is by far the most common present day reason (Adams *et al* 2003). Mitochondrial cDNA or DNA makes its

Table 1.1:- List of mitochondrial ribosomal protein genes identified in mitochondria and nuclear genomes of different plants

+ present based on sequencing information; - absent based on Southern data; ψ pseudogene; +? Truncated and potentially functional, s+based on Southern data; (table has been updated from Subramanian thesis 2001 so now it includes *Phoenix* (JN375330.1); *maize* (AY506529.1); *Bambusa* (EU365401); and *Lolium* (JX999996.1)

Gene	**Liverwort**	***Arabidopsis***	**Phoenix**	**Maize**	**Rice**	**Wheat**	***bambusa***	***Lolium***
Rps1	+	-	+	+	+	+?	+	+
Rps2	+	-	+	+	+	+	+	+
Rps3	+	+	+	+	+	+	+	+
Rps4	+	+	+	+	+s	+	+	+
Rps7	+		+	+			+	+
Rps8	+							
Rps10	+				-	-		
Rps11	+		+		ψ	+		
Rps12	+	+	+	+	+	+	+	+
Rps13	+		+	+	+s		+	+
Rps14	+	ψ	+					+
Rps19	+	ψ	+		+	ψ	+	ψ
Rpl2	+	+?	+		+	ψ		ψ
Rpl5	+	+	+				+	+
Rpl6	+							
Rpl16	+	+	+	+	+	+	+	+

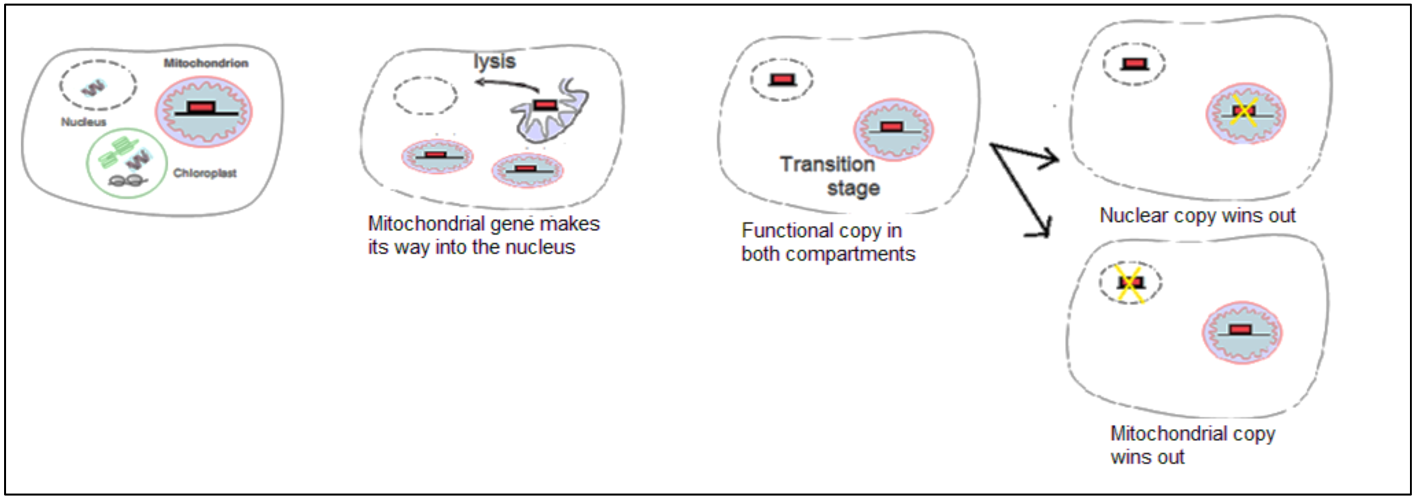
way into the nucleus upon lysis, division, or disruption of mitochondrial membrane during digestion of organelles by lysosomes or vacuoles during stress (reviewed in Adams *et al* 2003). For genes, which have been successfully transferred to the nucleus, and are functional, there is an intermediate (functional) stage where copies of the gene are present in both mitochondria and nucleus (reviewed in Brennicke *et al* 1993, Bonen 2006). Functional transfer occurs when there is activation and expression of the transferred gene in the nucleus (reviewed in Adams *et al* 2003 and Liu *et al* 2009).

Currently, there are only three examples, in plants that show intact potentially functional genes present in both compartments: *cox2* in some legumes (Nugent & Palmer 1991), *rpl5* in wheat (Sandoval *et al* 2004) and *sdh4* in populus (Choi *et al* 2006). *Neurospora*, a filamentous fungi has two distinctive *atp9* copies, but the mitochondrial copy is only expressed in germinating conidia (Bittner-Eddy *et al* 1994). Either the mitochondrial copy or the nuclear copy will be lost once gene transfer occurs, due to deletion events (Bonen 2006). Non-functional gene transfer occurs when the transferred gene in the nucleus cannot be properly expressed or lacks appropriate mitochondrial targeting signals and the result is referred to as a pseudogene (Liu *et al* 2009). Random bits and pieces of mtDNA are known to be integrated into the nuclear genome. Only one functional gene transfer to the nucleus has been reported in animals, the Atp9 in sponge *Amphimedon queenslandica* (Epenbeck *et al* 2007) and *Neurospora crassa* (Brittney-Eddy *et al* 1994)**.** A mitochondrial gene gets transferred to the nucleus as a cDNA intermediate, because there is no C-to-U editing and group II intron splicing machinery in the nucleus. If not, the mitochondrial gene will not code for a functional protein in the nucleus. In order to become active, the transferred gene must have a promoter and other regulatory elements for proper expression. In addition to this, the gene must also acquire a specialized amino (N-terminal) sequence for targeting the protein back to mitochondria (Adams *et al* 2003).

Angiosperm mitochondrial genomes also have DNA that is of nuclear origin, which is made up of transposable elements. Sequences derived from nuclear genes in mitochondrial genomes are pseudogenes. For instance, recently in the mitochondrial genomes of *Brassicaceae* and *Arabidopsis*, an orf164 was found. This ORF is derived, from a nuclear gene called ARF17 that encodes an auxin responsive protein (Qui *et al* 2014). Gene transfer

Figure 1.1- Model for mitochondrial gene transfer to the nucleus.

Evolutionary model for mitochondrial gene transfer to the nucleus. Mitochondrial gene shown in red usually migrates to the nucleus in its cDNA form (which occurs during stress or lysis). Upon transfer the nuclear copy will gain regulatory sequences for proper expression as well as targeting signals to aid the protein back to mitochondria where it is needed. There is a period of time where there will be functional copies in both compartments, which is referred to as the transition stage. The transition stage can come to an end when deleterious mutations occur in either mitochondrial or the nuclear copy.



Adapted from Bonen 2006

also occurs from chloroplast to the nucleus and one particular example is Acetyl-coA carboxylase subunit gene (accD) (Rousseau-Gueutin *et al* 2013).

1.5 Gene Activation in the nucleus

Targeting signals enable the transferred gene in the nucleus to get its protein back to the mitochondria where it is required (Liu *et al* 2009). These signals can be located at the N-terminus (which is most common) or even within the protein itself; however in this case the signal will not be removed after import into the mitochondria, as would an N-terminal signal. One such example is of *rps19* gene in sugar beet whose targeting signal is located within the protein (Matsunaga *et al* 2013). Targeting signals will be recognized by the receptors on the mitochondrial surface by TOM 20, a single translocase of the outer membrane complex and TIM 17 (translocase of inner membrane) in plants (Duncan *et al* 2012). The Tim 17:23 complex is responsible for the import of most mitochondrial proteins in plants. Two Tim 21 proteins have been identified, which are known to be imported, into the mitochondrial membrane (Murcha *et al* 2014). In addition to the Tim 17:23 complex interaction, these proteins are found to be associating with the respiratory chain of complex I (Murcha *et al* 2014).

Comparative analysis was done for 42 transferred genes (includes same genes in different plants) in various angiosperms to determine the nature of targeting sequence acquisition in a study done by Liu *et al* 2009. Half of the genes that have been looked at in this study were found to have gained their targeting sequence from another nuclear gene for a mitochondrial protein. It can thus be said that acquisition of targeting sequence from another existing gene in the nucleus is very common, but there are other mechanisms such as alternative splicing as well, which allow the gene to acquire targeting sequence as well as 5' regulatory elements (Figueroa 2000). Hsp70 mitochondrial chaperone protein targeting sequence has been used for five different genes that have been transferred to the nucleus: *sdh3* in *Ipomoea*, *Populus,* and *Arabidopsis*; *rps10* in *Fushsia*, and *rps19* in maize. Hsp22 targeting sequence has been used by *sdh3* in *Gossypium* and *Triphysaria* and *rps10* in *Daucus*. Genes encoding products that are involved in translation have also been identified as donors for targeting sequences for transferred genes (Liu *et al* 2009).

Direct fusion of N-terminal targeting sequence upstream of a gene has been observed for cases such as: *Sdh3* in *Malus* has been inserted into mitochondrial EF-TU and *rps10* in

Daucus has been fused with *hsp22*. Interestingly, *rps14* in the grasses was inserted into an intron of a succinate dehydrogenase gene (Figuerora 2000). Once the protein is successfully targeted to the mitochondrion, the N-terminal targeting sequence is recognized and cleaved off. There are other cases where the transferred genes do not have any 5' extension and the targeting information is located within the core region of the protein. In this case the transferred mitochondrial gene must already have targeting information within it, so it is unnecessary to undergo further activation in the nucleus. Some examples of genes with intrinsic targeting information are *rps10*, *rps14* and *rps1* (Ueda *et al* 2001). Another study showed that 25% of the transferred ribosomal proteins did not have any targeting presequence (Bonen and Calixte 2006).

Intron location can impact on the way a gene acquires its presequence or regulatory elements. There are five different possible locations for introns: 5'UTR, non-core region including the presequence, between the non-core and core regions, in the core regions, and 3'UTR. Out of 62 transferred genes that have been surveyed in a study done by Liu *et al* 2009, 48% were found to have the introns within the non-core regions making it the most common location. Second most common location for the introns was found to be within the UTRs. Only three genes have been reported to have an intron in the core region: 5'*rpl2* in beta, sdh4 in *Ocimum* and *Euphorbia* (Liu *et al* 2009). Thirty percent of transferred genes have introns at the junction of the targeting sequence, indicating that it must have been acquired through an exon shuffling type event. Introns within the 5'UTR introns enable gene activation in the nucleus via gaining 5'cis regulatory elements such as promoters. Introns within the 3'UTR introns are equally important as they help the gene in acquiring 3' cis regulatory elements and transcript stability (Bonen and Calixte 2006).

Transferred genes in the nucleus among various angiosperms were found to have an increase in the rate of substitution with respect to their mitochondrial copy, especially at silent sites in the nuclear genome. This is illustrated in a study done by Wolfe *et al* where the synonymous substitution rate of nuclear genes in angiosperms is twice as high when compared to chloroplast genes; it is even more drastic between mitochondrial and nuclear genes as nuclear mutation rates are 5 times higher. For each gene synonymous substitution rate (Ks) of all the nuclear copies was higher than the Ks of all mitochondrial copies. Thus both Ks and Ka (non-synonymous substitution rate) increase after a gene is transferred to the nucleus (Wolfe *et al* 1987). Consequently, the rate of nucleotide substitution is lower in mitochondria than it is in nucleus or plastids in seed plants, but the opposite was found to be true in red alga (Smith *et al* 2014).

Codon usage pattern can differ among genes as well as different compartments within the cell with respect to the base composition. For instance, a study done on *Triticum avestivum* by Zhang *et al* 2007 to investigate the codon usage patterns among the mitochondrion, chloroplast and nucleus showed that the GC content in nuclear genes was higher than that of mitochondrion or chloroplast genes. Bias for A and T nucleotides at silent positions were seen for wheat *rpl2* and *rps19* mitochondrial genes whereas their functional nuclear copies were found to have a preference for G and C at the silent positions (Fallahi *et al* 2005, Subramanian and Bonen 2006). Codon bias in the nucleus is a result of strong mutational bias, but in mitochondria and chloroplasts it is influenced by the tRNA pool. There are also few genes in which the GC content of the nuclear copy is lower than that of the mitochondrial copy. This is true for *sdh3*in *Medicago* and *Vaccinum* (Zhang *et al* 2007).

Duplication events can occur after transfer to the nucleus, for instance the *rp16* gene in the rice nucleus was duplicated and has two distinct copies (Kubo *et al* 2008). Similarly rice *rpl10* was duplicated after transfer to the nucleus (Kubo *et al* 2000). Interestingly, both copies have homologous 5’UTR sequences indicating that the gene was duplicated after acquisition of its UTR from other nuclear genes (Kubo *et al* 2000).

1.6 Possible fates of paralogous copies after duplication

Paralogous copies can be beneficial to the plant as a new function can be gained which can introduce variation. Other possible fates for the duplicated gene are: decay through the accumulation of deleterious mutations (pseudogene), and retention of the original function (Wendel 2000). Gain of new function after duplication can be predicted by comparing the amino acid and nucleotide sequences of both copies. One of the copies can undergo very fast rate of amino acid substitution resulting in a new function and is referred to as positive selection. On the other hand both copies can be evolving at a very slow rate. A pair of duplicated genes R and B on chromosomes 10 and 2 which encode helix-loop-helix transcriptional activators have been evolving at the same rate as they both have the same function, which is to differentially regulate purple pigmentation, in maize tissues (Gaut and Doebley 1997).

Gene duplication can also result in preservation of function for both copies. As both copies share the same function, maintaining both copies might be essential for proper protein function. Genes can be involved in pathways with other genes which results in selection against any mutation that would cause its silencing (Wendel 2000). DNA level interactions can change the sequence of one or both duplicated copies.

1.7 Purpose of study

Rps19 and *rpl2* genes are known to be transferred to the nucleus a lot in angiosperms (39 and 41 times respectively) (Adams et al 2001) which makes them interesting candidates to survey in grasses. S19 and L2 proteins are important for ribosome assembly and mitochondrial protein synthesis machinery (Subramanian *et al* 2001). Although my thesis primarily focuses on *rps19* gene, *rpl2* will also be studied because it is only three nucleotides upstream of *rps19* and they are both co-transcribed together in rice with downstream *nad4L* (Subramanian et al 2001). Prior research in our lab (Fallahi et al 2005) showed that the *rps19* gene transfer event occurred in the common ancestor of grasses. This is based on the evidence that the nuclear copies in wheat, barley and maize all acquired the same hsp70 presequence and shared derived amino acid changes that are not present in the native mitochondrial copy. We know that maize lost its mitochondrial copy while rice lost its nuclear copy. Rice (Kubo et al 1996), *Bambusa* (EU365401) and *Ferrocalamus* (iron bamboo) (JN120789.1) also have functional copies of both *rps19* and *rpl2* in the mitochondria. This suggests that there was a transition period where functional copies are kept in both compartments for at least 20 million years. We also know that oats mtDNA sequencing revealed the presence of an early stop codon at the 5'end of *rps19* as well as a 3' truncation event, barley lost its mitochondrial copy and rye and wheat have 5' truncated pseudogenes in the mitochondria (Fallahi et al 2005). This raises some questions regarding what the status of these genes is in the mitochondrial genome of brome. Also we are interested in knowing if the transition period extends much further than the 20 million years time period.

Similar to *rps19*, *rpl2* gene also underwent many gene transfer events. Maize *rpl2* gene is found to be in the nucleus while the mitochondrial copy is known to be lost (Adams *et al* 2001). In wheat, 3’end of *rpl2* pseudogene is found to be in the mitochondria and its functional copy is found as one gene in the nucleus (Fallahi *et al* 2005). The functional copy of *rpl2* is present in the mitochondrial genome of rice but no copy was found in the completely sequenced rice nuclear genome (Kubo *et al* 2006). Even though there is limited information on exact dates of divergence for many members of the grass (Poaceae) family (Kellogg 2001), the phylogenetic relationships are known. Subfamily of Poaceae, known as Pooideae (wheat or rice) and Panicoideae (maize) are known to have diverged 50-60 million years ago (Kellogg 2001). The oat-wheat clade is approximately 25 million years ago (Gaut 2002). Barley-wheat clade is 15 million years ago and brome-wheat clade is approximately 20million years (Pont et al 2011). The divergence time between rye and wheat is around 5 million years ago (Gaut 2002).

1.8 Objectives

To determine the “status” of *rps19 and rpl2* genes in grasses primarily in brome and other grasses such as oats, wheat, barley and maize. I am interested in looking at the status of *rps19* in lineages sharing an even more recent common ancestor with wheat (which has a functional gene present in the nucleus) whose lineages have diverged less than 45 million years ago. Sarah Rampersad, a fourth year honours student did some preliminary work on brome mitochondrial *rpl2* and *rps19* genes. Her clone data showed that the *rpl2* gene is a pseudogene and *rps19* is functional. I further validated her work with my experimental data. My objectives were to identify if there are other functional or pseudogene copies of *rpl2* and *rps19* in the mitochondrial genome of brome. I am also interested in determining the status of these genes in the nuclear genome of brome and other close relatives. The nuclear genes will be studied with respect to gene structure, to gain insight into how the gene was activated in the nucleus and gene expression. My specific objectives are as follows:

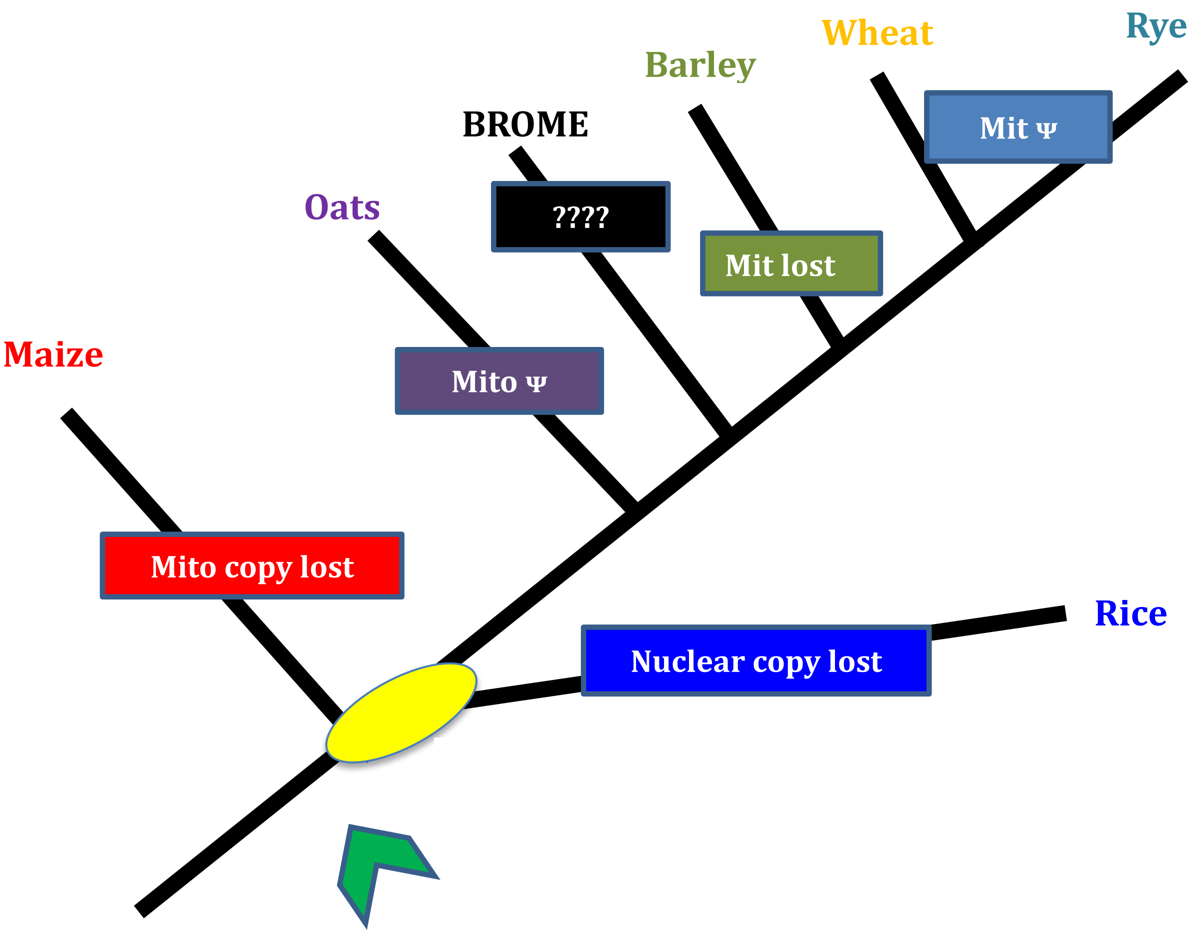
- Determine if there are other functional or pseudogene copies or *rpl2* and *rps19* genes in the mitochondrial genome of brome via Southern Analysis, and inverse PCR.
- Determine the sizes of transcripts coming from *rpl2-rps19-nad4l* locus in brome mitochondrial genome via Northern Analysis.

Figure 1.2:- Schematic representation of the *rps19* gene status among grasses and the purpose of studying brome. (b) Evolutionary status of mitochondrial *rps19* and *rpl2* genes in the mitochondria of different plants

(a) The status of the *rps19* gene in the mitochondria of different grasses is shown in boxes on the phylogenetic tree. Green arrow represents that the *rps19* gene transfer event occurred in the common ancestor of the grasses. Yellow oval represents the transition stage which must have existed for at least ~25 million years as maize lost its mitochondrial copy and rice lost its nuclear copy (Fallahi et al 2005). Black box with white question marks on the brome lineage indicates that the status of the *rps19* and *rpl2* genes are unknown and are the main focus of study in this thesis. This figure has been adapted from Fallahi et al 2005.

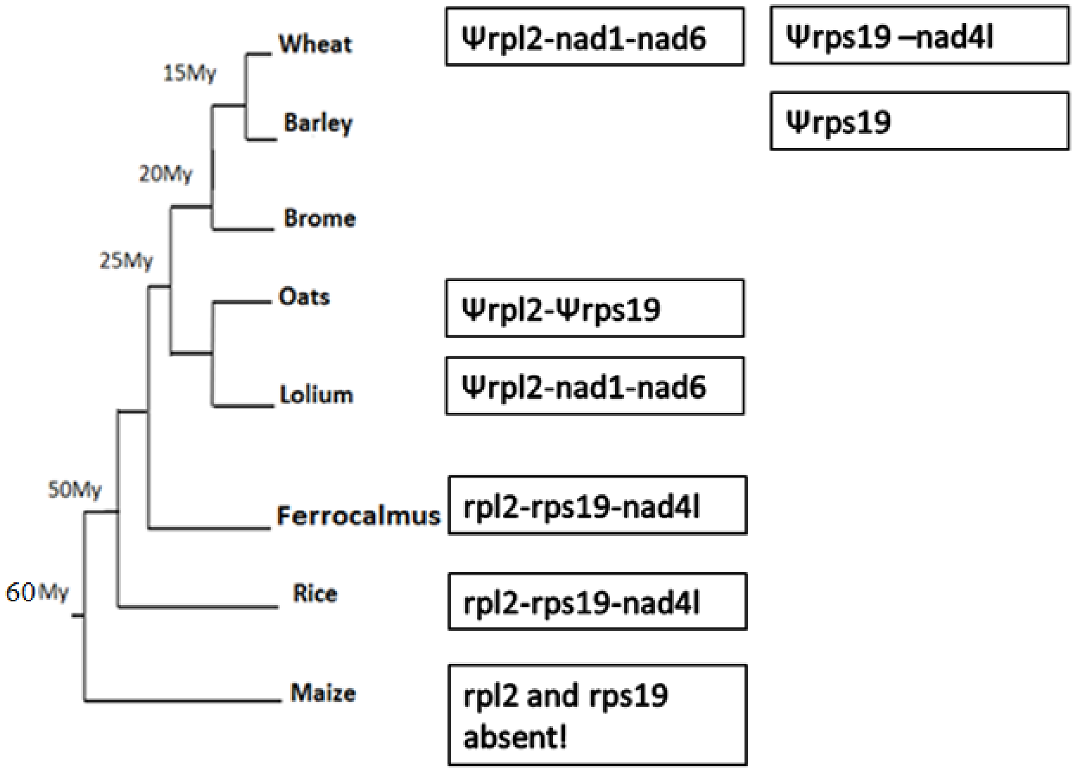
(b) Phylogenic tree showing the gene order and status of the *rps19* and *rpl2* genes in the mitochondria of different grasses (Subramanian et al 2001, Fallahi et al 2005). The divergence times of the grasses is shown to the left of the tree (Gaut 2002; Catalan et al. 2004; Kellogg and Bennetzen 2004; Chalupska et al. 2008). Ferrocalamus is a type of bamboo species. The linkage of the genes are shown in a box and the Ψ represents pseudogene.

(a)



Common transfer of *rps19* to the nucleus

(b)


Wheat
Barley
Brome
Oats
Lolium
Ferrocalmus
Rice
Maize
15My
20My
25My
50My
60My
Ψrpl2-nad1-nad6
Ψrps19 –nad4l
Ψrps19
Ψrpl2-Ψrps19
Ψrpl2-nad1-nad6
rpl2-rps19-nad4l
rpl2-rps19-nad4l
rpl2 and rps19 absent!

- Determine the status of *rpl2* and *rps19* genes in the nuclear genome of brome and determine its functionality using RT-PCR
- Sequence the oats nuclear *rps19* gene if present, and determine if it’s expressed.
- Comparative analysis of nuclear *rps19* gene in brome, barley, wheat, oats, and maize will be done to determine the type of gene activation and adaptation events that occurred.

Materials and Methods

2.1 Plant Material

Brome grass (*Bromus inermis*) seeds were purchased from Ritchie Feed and Seed limited, Ottawa. Wheat (*Triticum aestivum* var Frederick), Barley (*Hordeum vulgare* var OAC Kipper) and oats (*Avena sativa* var AC Goslin) were kindly provided by Dr. R. Pandeya and Dr. TM Choo (Agriculture and Agri-food Canada).

2.2 Mitochondrial RNA and DNA isolation

Thirty grams of brome seeds were surface sterilized in a 1:6 dilution of Javex for five minutes and rinsed a couple of times with distilled water. The seeds were treated with 100mls of 10mM HCl and stirred occasionally for 10 minutes. The seeds were subsequently rinsed 4-5 times in distilled water. For embryo mitochondrial RNA or DNA isolation, the seeds were then plated on petri dishes with autoclaved filter paper. The seeds were watered every 48 hours until day 10. For seedling mtRNA or DNA, the sterilized brome seeds were planted in a large tray with autoclaved vermiculite. Both embryos and seedlings were grown in dark to prevent chloroplast biogenesis. Mitochondrial RNA and DNA were isolated from brome embryos and seedlings using procedures that have been previously described (Subramanian *et al* 2001). Seedlings were separated from the seeds and roots in the vermiculite before use. Mitochondrial RNA was taken through two rounds of phenol extractions (Phenol saturated in TE buffer) and precipitated using 0.1volume of 5M NaCl and 2 volumes of 95% ethanol

2.3 Total DNA and RNA isolation

Total DNA was isolated from brome grass seedlings using the protocol described by Ausubel et al (1990) with modifications as described in (Hazle and Bonen 2007) except for the final reprecipitation using the CTAB method. Total RNA was isolated using Trizol® (Invitrogen, Carlsbad, CA) according to the manufacturer's specifications.

2.4 DNA ligation and Inverse PCR

Mitochondrial DNA was digested using restriction enzyme for 3 hrs at 37°C. It was then purified by phenol extractions and was ligated using T4 DNA ligase overnight at 14°C. Inverse PCR was done using standard PCR conditions.

2.5 Mitochondrial RNA Analysis

Before use, the quality and concentration of the mitochondrial RNA was assessed by its fluorescence profile when run on an RNA gel. The brome mitochondrial RNA sample (500ml) was loaded on to a 1.2% agarose / formaldehyde gel. Gel electrophoresis was carried out at 60 V for approximately 4 hours. Northern Transfer of the RNA gel onto a nylon membrane was done using the standard procedure.

2.6 Northern and Southern hybridization

As described in Li-Pook-Than and Bonen (2006), γ-^{32}P-5' end labelled 20nt oligomers were used for Northern and Southern Analysis. Prehybridization of the RNA blots was done overnight at 42 degrees in 15mls of wash solution (20x SSC, deionized formamide, 10% SDS and yeast tRNA). Overnight hybridization was done for the radioactively labelled blots. The next day blots were washed twice with 20ml wash buffer (20x SSC + 0.1% SDS) for 2x 15 minutes. Blots were phosphoimaged using a molecular imager (X-Biorad).

2.7 RT-PCR and cDNA synthesis

RT-PCR was used to analyze expression of the genes in the mitochondria and the nucleus. If the gene does not have introns the RNA prep was 2x or 3x DNAse treated according to the RQ1 DNAse I protocol from Promega. cDNA synthesis was done using M-MLV reverse transcriptase (Invitrogen) for 2 hrs at 37°C. RT-PCR products were gel purified using Ultra Clean 15 (MoBio Laboratories).

2.8 Cloning and Sequencing

PCR products (Inverse PCR and RT-PCR) were ligated using pGEM T-easy vector system (ProMega) and then cloned using competent TB1 *E.coli.* Plasmid DNA was isolated using the QIA spin mini prep kit (Quiagen) and PCR was done to confirm that the insert has been cloned successfully. Ottawa Hospital Research institute (OHRI) stem core lab performed sequencing of our samples. Direct sequencing of the RT-PCR products was done to look at the population in order to access the degree of editing or nuclear gene copy number and SNPs between copies.

Table 2.1 Genome sequences used for rps19 and rpl2 mitochondrial and nuclear sequence comparison

Organism	**Genome**	**Accession #**	**Publication**
Oryza sativa (Rice)	Mitochondria	BA000029	Notsu et al. 2002
Zea mays (Maize)	Mitochondria	AY506529	Clifton et al. 2004
Bambusa Oldhamii (bamboo)	Mitochondria	EU365401	unpublished
Ferrocalamus (bamboo)	Mitochondria	JN120789.1	Ma et al. 2009
Triticum aestivum (Wheat)	Mitochondria	AP008982	Ogihara et al. 2005
Lolium	Mitochondria	JX99999.1	
Phoenix	Mitochondria		
Hordeum vulgare (Barley)	Nuclear	AC249420 (8486-9101)	NCBI High throughput
Zea mays (Maize*)* chr5	Nuclear	AC215864 (152433-153063)	NCBI High throughput
Zea mays (Maize*)* chr1	Nuclear	AC194145 (60705-61217)	NCBI High throughput
Triticum aestivum (Wheat) chr5	Nuclear	Traes_5BS_693478364	Ensembl.plants
Triticum aestivum (Wheat) chr3b	Nuclear	Traes_3B_32FF5125C	Ensembl.plants
Hordeum vulgare Chr4	Nuclear	4 215818848 to 215819095 (+)	Ensembl.plants

Table 2.2 Oligomers used in the study

Plant	**Name**	**Sequence 5’-3’**	**Gene**
Brome (mito)	Lb459	ACGAGTGAGAAGGATAAGGG	rpl2 5’utr
	lb97	TTGAGTACGGGGAAGTCCGC	rpl2 exon1
	lb62	GTTACAGACTCACTTATCCG	rpl2intron
	lb61	ATAGATTCCCGATGCCGAGC	rpl2exon2
	lb458	TGCCCGAAATCCTGCTTTGG	rpl2exon2
	lb201	CTGCTTCGTTCCTTGACTTC	rpl2intron
	lb163	TGGAAGGGAAGTTTTGTTGATGC	rps19
	lb164	AGCAAACTCTCCAAATTTATGAC	rps19
	lb196	TCGAGTATGAAGAAAAGACC	rps19 3’utr
	lb834	GGGTGGGGAAACGCAGAATT	rpl2exon2
	lb826	GAGGCATGAAGGTAGGAAGG	rpl2exon1
	lb69	GTCTATTAAGGAGAATTCCC	nad4l
	lb600	ATTAGTTCTAACAGTGGCAG	nad4l
	lb598	AGGAGGATTCCCCAAATACC	nad4l
	lb162	TGTATTTGGTAACTCTCCTG	nad4l
	lb63	TTCCACAATGCCAATAGACG	rpl2exon1
	lb72	CTAGTACGATGGATCGAAGG	rpl2exon1
	lb174	AATTCCTCCCTGCGGACTTC	rpl2exon1
Barley/rye (mito)	lb853	AGACCTTCGGTGTCACCTTG	pseudorpl2basedonwheat
	Lb859	CCGAATACAGGTCTAGTAGG	pseudorpl2basedonwheat
	lb56	CTCCATGAGGATGATCCACTGG	pseudorpl2wheat3’
	lb822	AGTAGTAGGGGTGGTGAAAC	pseudorpl2wheat3’
	lb485	CACGAATAGGAATGAAACGG	nad6(wheat)
	lb177	CCATGTTTCCGAAACGGATC	nad1(wheat)
Wheat, Brome (nuclear)	lb836	ATTGCTTCTCTCATCGCCTC	nuclearrps19
	lb837	TTCCGCGTAAAAGCAAACTC	nuclearrps19
	lb862	TGCGGCTGAGAGCAGCCATG	nuclearrps195’end(green)
	lb904	CAATATCGTAACCTCACCTT	nucelarrps193’end(green)
	lb923	TGACGCGCATTCCCTTTTCT	nuclearrps193’end(green)
	lb79	AAGGATCCAGCATACAAGGC	nuclearrpl2
	lb89	GTTGGAGACGATGCCGATGG	nuclearrpl2
	lb849	ATGCAGAAGTTCAAGACCTT	nuclearrpl25'
	lb73	GTGTGCCAGGTCATCGGTGC	ATP-beta synthase
	lb109	CATCCATACCCAAAATAGCAAT	ATP-beta synthase
	lb844	ACTGGTTTGACTGTTGCTGA	ATP-beta synthase
Oats (nuclear)	lb892	GACGTACGCATCACCACCAA	oatsrps195’utr
	lb851	ATTGCCATGGTTCAGGCTAC	oatsrps19
	lb852	TTGTTTGCCCTTCGCTATAG	oatsrp19
	lb860	GGAAGGGAGCTTTCGTTGAT	oatsrps19
	lb867	ATGGGGCCAAGAACCATTCT	oatsrps19
oligodt (3’race)	lb83	AAGCTTTTTTTTTTTG	
	lb84	AAGCTTTTTTTTTTTC	
	lb85	AAGCTTTTTTTTTTTA	

2.9 Bioinformatics

Plant mitochondrial genomes used in sequence comparisons with brome mitochondrial sequences are listed in Table 2.1. Nuclear draft *rps19* sequences used for analysis of nuclear *rps19* and *rpl2* copies are also listed in Table 2.1. Analysis of sequences was done using programs such as MUSCLE, NCBI BLAST, Target P and DNA to protein sequence converter.

Chapter 3 Results: Retention of functional genes for S19 ribosomal protein in both mitochondrion and nucleus for over 60 million years

Sruthi Atluri · Sarah N. Rampersad · Linda Bonen

Comments

This chapter has been written in the format of a manuscript and was submitted for publication.

Research contributions from other students

Sarah Rampersad, a fourth year honours student contributed preliminary data that identified the status of *rpl2-rps19* genes in the mitochondrial genome of brome. Her work has thus provided a solid foundation for this project.

Keywords: Gene transfer, ribosomal protein, mitochondria, brome, grasses

3.1 Abstract

Ribosomal protein genes occasionally undergo successful migration from the mitochondrion to the nucleus in flowering plants and we previously presented evidence that the S19 ribosomal protein gene (*rps19*) had been transferred to the nucleus in the common ancestor of Poaceae grasses. In many lineages, the mitochondrial copy was subsequently lost or pseudogenized, although in rice it was retained and the nuclear copy lost. We have now determined that functional *rps19* genes are present in both the mitochondrion and nucleus in brome grass (*Bromus inermis*). The mitochondrion-located *rps19* gene, which is immediately downstream of an *rpl2* pseudogene, is transcribed and edited. The nuclear-located *rps19* gene is also actively- expressed and it possesses the same intron-containing *hsp70*-type presequence as its counterparts in other grasses, as well as shared-derived amino acids in the S19 core. We conclude that this brome *rps19* gene derives from the same transfer event that occurred in the common ancestor of grasses at least 60 million years ago. In the oat lineage, a subsequent exon shuffling-type event has resulted in novel amino-terminal sequences replacing part of the *hsp70* presequence, and in the barley lineage, there has been an additional DNA-mediated transfer of the mitochondrial *rps19* gene and its flanking sequences, followed by relatively recent loss of the mitochondrion-located copy. The prolonged persistence of functional copies in both compartments, as evidenced by present-day brome, raises interesting questions about their respective roles.

Keywords: Gene transfer, ribosomal protein, mitochondria, brome, grasses

3.1 Introduction

Massive gene transfer from the mitochondrion to the nucleus is known to have occurred early in eukaryotic evolution following endosymbiosis (reviewed in Gray et al. 1999), and migration is still ongoing in plants, albeit sporadically (reviewed in Adams and Palmer 2003). This contrasts with many other eukaryotes where alterations to the mitochondrial genetic code have effectively stopped successful gene movement. In flowering plants, the mitochondrial ribosomal protein genes exhibit a greater propensity for migration to the nucleus than respiratory chain genes, and at most three large subunit ribosomal protein genes and eleven small subunit ones remain in the organelle, although only a subset is typically seen in any given plant species. This is fewer than for the non-vascular plant *Marchantia polymorpha* which still encodes sixteen in the mitochondrion (Takemura et al. 1992) and is only a small proportion of the 54 or so bacterial-type proteins in present-day plant mitoribosomes, the rest being nuclear-encoded (cf. Bonen and Calixte 2006). Initial insights into the extent and variability of gene loss from the mitochondrion among flowering plants came from Southern hybridization surveys (reviewed in Adams and Palmer 2003) and this has been extended by mitochondrial genome sequence data now available (cf. NCBI Organelle Genome Resources website). Information about nuclear-translocated gene copies has for the most part come from individual gene analysis (reviewed in Adams and Palmer 2003; Liu et al. 2009), from surveys of plant nuclear genomes for particular gene sets, such as the ribosomal protein ones (cf. Bonen and Calixte 2006), and from large-scale analyses of genomes representing various eukaryotic lineages (cf. Maier et al. 2013; Kannan et al. 2014). Analysis can be complicated by the presence of non-functional mitochondrial DNA (Numt) sequences in nuclear genomes, although it is worth noting that in plants, successful transfer also requires an RNA intermediate since virtually all mitochondrial coding sequences undergo C-to-U RNA editing to generate the correct amino acid sequence (reviewed in Takenaka et al. 2013) and the nuclear expression system does not contain such editing machinery nor for the splicing of mitochondrial group II introns.

Successful transfer to the nucleus not only requires integration of the mitochondrial gene into a “hospitable” genomic site, but also acquisition of regulatory elements for proper expression and targeting signals so that the protein can be imported back into the mitochondrion. Amino-terminal targeting signals are sometimes acquired from duplicated copies of genes that specify other mitochondrion-destined proteins, sometimes from unknown sources, and in some cases, there is no additional extension. About 25% of the nuclear-located mitochondrial ribosomal protein genes in Arabidopsis and rice fall in this latter category (Bonen and Calixte 2006). In one interesting case in grasses, a mitochondrial *rps14* gene has been inserted within the intron of a gene encoding another mitochondrial protein (*sdh4*) and is expressed through alternative splicing (Figueroa et al. 1999; Kubo et al. 1999). Additional protein-coding domains are sometimes acquired by the translocated gene, such as the RNA-binding domain fused to *rps19* in Arabidopsis (Sanchez et al. 1996). Implicit in evolutionary scenarios of functional mitochondrion-to-nucleus gene transfer is that there will be a period in which active copies are present in both compartments (Brennicke et al. 1993; Adams and Palmer 2003; Bonen 2006), after which the redundant copy degenerates into a pseudogene and is eventually lost. Little is known about what factors influence the length of such periods of co-existence or the eventual outcome (ie. which copy “wins out”). In some cases, the transition period may be extremely short, based on variation seen among families within *Silene vulgaris* for the mitochondrial *rpl5* and *rps13* gene status (Sloan et al. 2012). To our knowledge, there are only four documented cases of active gene copies simultaneously being present in both the mitochondrion and the nucleus: *atp9* in *Neurospora crassa* (van den Boogaart et al. 1982), *cox2* in certain legumes (Nugent and Palmer 1991), *rpl5* in wheat (Sandoval et al 2004) and *sdh4* in *Populus* (Choi et al 2006).

Although unsuccessful mitochondrion-to-nucleus gene transfers may be difficult to detect because the final outcome is as though the event had never happened, it was possible in the case of the mitochondrial *rps19* gene in grasses to infer that the mitochondrial copy “won out” in the rice lineage (Fallahi et al. 2005). In present-day rice, this mitochondrial gene is located immediately downstream and co-transcribed with an intron-containing *rpl2* gene in an ancestral bacterial-type linkage (Kubo et al. 1996), and there is no functional *rps19* gene copy in the nucleus. In maize, the mitochondrial *rps19* gene has been lost, and in wheat mitochondria it is a 5’-truncated pseudogene (Fallahi et al. 2005). There is strong support for the *rps19* gene transfer having occurred in the ancestor of these grasses, because the nuclear-located *rps19* genes in lineages that diverged before and after the rice lineage, namely maize/sugarcane and wheat/barley respectively, share the same acquired amino-terminal targeting sequence derived from the *hsp70* pre-sequence as well as a number of shared amino acids that are absent from the rice mitochondrial *rps19* gene (Fallahi et al. 2005). Such features are most simply explained by a single transfer of the *rps19* gene to the nucleus in the common ancestor of these grasses, which diverged about 60 million years ago (Kellogg and Bennetzen 2004). For these reasons, we were interested in examining the status of *rpl2* and *rps19* in another grass, brome (*Bromus inermis*) which shared a common ancestor with wheat about 20 million years ago (Gaut 2002; Kellogg and Bennetzen 2004), with the aim of gaining insight into the potential length of the period of co-existence and factors that might influence eventual success of transfer.

3.3 Materials and Methods

Mitochondrial DNA and RNA were isolated from 6-9 day etiolated seedlings of brome grass (*Bromus inermis*), barley (*Hordeum vulgare* cv. OAC Kippen), and oats (*Avena sativa* cv. 0A974-1) using previously described procedures (Subramanian et al. 2001). Brome seeds were purchased from Ritchie Feed and Seed Inc (Ottawa Canada) and other seeds were kindly provided by Dr. R. Pandeya (Agriculture and Agri-Food Canada). Total DNA was isolated from brome seedlings using standard protocols with the modifications described by Hazle and Bonen (2007), except that the final CTAB step was omitted. Total RNA was isolated using Trizol® (Invitrogen) according to the manufacturer's specifications.

The brome mitochondrial and nuclear DNA regions of interest were obtained primarily by PCR using primers based on conserved regions from other grasses. Oligomer sequences are given in Supplementary Table S1. The brome mitochondrial DNA region preceding *ψrpl2-rps19* was obtained by inverse-PCR using DNA restricted with *Hin*dIII and circularized with DNA ligase (Invitrogen) prior to PCR amplification. For RNA editing analysis of *rps19*, brome mitochondrial RNA from 6-day etiolated seedlings was treated twice with DNase I (Amersham) prior to cDNA synthesis with MMLV reverse transcriptase (Introgen) at 37°C for 2hrs and subsequent PCR amplification. A similar RT-PCR protocol, using total RNA without DNAse I treatment, was conducted with primers directed at mRNAs for nuclear-located *rps19*, *atpβ* and *rpl2* (Supplementary Table S1). PCR and RT-PCR amplification products from brome, as well as oats, were gel-purified using UltraClean 15 (MoBio Laboratories Inc.) and after corroboration by nested PCR, they were either sequenced directly or cloned into pGemT-Easy plasmid vectors (Promega) prior to sequencing. Sequencing was performed by StemCore Laboratories at the Ottawa Health Research Institute, Ottawa, Canada. Sequences have been deposited in the NCBI databank with accession numbers xxx.

For comparative sequence analysis, *rps19* homologues were retrieved from the NCBI databases (nr, EST, and high throughput genomic) as well as from the wheat genome website (plants.ensembl.org) using BLAST. Sequence alignments were carried out using MUSCLE. Accession numbers are given in Supplementary Table S2. It also includes two additional copies of *rps19* present in the wheat nuclear genome (on chromosome 1) but which were omitted from our analysis because of their close similarity to the copies on chromosome 3 and 5. It should also be noted that the *rpl2* intron/exon junction was mis-annotated in the bamboo *Ferrocalamus rimosivaginus* databank entry (JN120789) and for *Bambusa oldhamii* (EU365401) no *rpl2* annotation was given, although inspection shows a complete *rpl2* gene is present in the database entry. Prediction of protein targeting to the mitochondrion was assessed using TargetP (Emanuelsson et al. 2007) and Predotar (Small et al. 2004).

3.4 Results and Discussion

3.4.1 An expressed rps19 gene is located downstream of rpl2 pseudogene in the brome mitochondrial genome

We identified an intact *rps19* gene in the brome mitochondrial genome using PCR with primers based on the rice counterpart (Kubo et al 1996). It is located three nucleotides downstream of an *rpl2* pseudogene and has a downstream linkage with *nad4L,* as in rice (fig. 1A). The brome *rpl2* homologous sequences are missing 298 bp within exon 1 relative to rice, as well as the extreme 3’ end of exon 1 and most of the intron (fig. 1A). Based on our Southern analysis, there are no additional full-length copies of *rps19* or *rpl2* located elsewhere in the brome mitochondrial genome, although there are short pseudogene segments (data not shown).

The *ψrpl2-rps19* region of the brome mitochondrial genome is actively transcribed as determined by RT-PCR analysis (fig. 1B), and direct sequencing of the products revealed that the *rps19* mRNA undergoes C-to-U editing at five positions, two of which are non-synonymous and result in conserved amino acids (fig. 1C). Synonymous sites are known to be often only partially-edited in the RNA population (reviewed in Takenaka et al. 2013) and this can be seen for codon 42. The extent of *rps19* editing was also less complete in longer transcripts (fig. 1C inset) reflecting their immature state. The same five edits, plus an additional silent one, had been observed in rice *rps19* (Kubo et al. 1996) and overall, the brome and rice mitochondrial *rps19* genes have only one non-synonymous difference and a 9-bp indel preceding the stop codon (fig. 2). Based on our northern analysis, transcripts from the brome *ψrpl2-rps19* region are present at only low steady-state levels in etiolated seedlings (data not shown), although it is worth noting that certain other mitochondrial ribosomal protein genes also exhibit very low levels during this developmental stage in wheat (Li-Pook-Than et al. 2006).

Within the brome *ψrpl2* coding region, there is an edit at the same position as the single site in rice, but not one within the *rpl2-rps19* spacer (Kubo et al. 1996). Our RT-PCR experiments (fig. 1B, lane 3) also showed no evidence for splicing of the truncated *rpl2* intron. This is not surprising since group II introns require intricate RNA folding for splicing competence (Bonen 2008). The region upstream of *ψrpl2* in brome is virtually identical to that of bamboo (Ma et al. 2012), consistent with an ancestral-type promoter driving transcription of both *ψrpl2* and *rps19*. There are other examples of pseudogenes being retained when located very close to functional genes (cf. *rps14*, Ong and Palmer 2006) and such sequences may serve a role in mRNA stability or translation. We had previously presented evidence for *rpl2* gene transfer having occurred in the common ancestor of wheat and barley (Subramanian and Bonen 2006), and our data for brome, which confirms the presence of an *rpl2* gene in its nucleus (Supplemental fig.S1) pushes the date of transfer back to more than 20 million years ago.

3.4.2 Brome also possesses a functional nuclear gene for the mitochondrial S19 ribosomal protein

To determine whether the situation in brome resembles that of rice, where the mitochondrial *rps19* copy has persisted during evolution and the nuclear copy lost, we examined brome nuclear DNA using PCR primers designed from wheat genomic (Brenchley et al. 2012) and wheat/barley EST data. Unexpectedly we found that a functional *rps19* gene is present in the nucleus (fig. 2A) and it resembles counterparts in other grasses; it has a conserved *hsp70*-type presequence containing two introns (fig. 2A) which are located at the same positions as in the *hsp70* gene (cf. fig.3B). At the protein level, nine amino acids within the S19 core region are shared by the other nuclear-located copies but not the mitochondrial ones (fig. 2B, asterisks), so all evidence points to a single gene transfer in the common ancestor of grasses, rather than a recent independent event in the brome lineage. The brome nuclear-located *rps19* gene shares 80% amino acid identity with the mitochondrion-located copy and some of the differences reflect early adaptation events upon transfer to the nucleus (fig.2B, asterisks). The brome nuclear-located *rps19* gene is both transcribed and correctly spliced as determined by RT-PCR sequence analysis (fig. 2A and supplemental fig. S2). The lower amount of *rps19* RT-PCR product compared to *atpβ* when the same amount of RNA template is used (fig. 2A, lane 2 vs. lane 3) likely reflects lower levels of *rps19* mRNA, although it should be noted that the necessity of using an *hsp70*-type oligomer for PCR may have contributed to mis-priming. Direct sequencing of the PCR and RT-PCR products within the core *rps19* revealed several polymorphic positions (supplemental fig.S3), not surprisingly since brome is polyploid (*Bromus inermis* AAAABBBB; cf. Armstrong 1979).

The brome nuclear-located *rps19* gene is more similar to the wheat counterpart located on chromosome 3 than to a paralogous copy on wheat chromosome 5. There are several signature amino acid residues that distinguish the chromosome 3 and chromosome 5 copies in wheat (fig. 2B, open rectangles), and the former more closely resembles the mitochondrial form, as do those from all the other grasses in fig. 2B, except for barley, which is more similar to the wheat chromosome 5 copy. Since the barley genome contains only one such *rps19* gene (based on the nuclear draft genome and EST data), it appears that the duplication event pre-dated the wheat-barley lineage split, with a subsequent loss of the chromosome3-type *rps19* gene in the barley lineage.

Figure 3.1. Organization of brome mitochondrial rpl2-rps19 genomic region and its expression.

[A] Schematic comparing brome and rice *rps19* (yellow) and *rpl2* exons (green). Triangles represent regions missing from brome *rpl2* pseudogene. Black arrows show positions of oligomers (#1-4) used in RT and RT-PCR, and grey arrows (#5-6) for oligomers used in inverse PCR to obtain upstream sequence.

[B] RT-PCR products with brome mitochondrial RNA from etiolated seedlings and oligomer 4 for cDNA synthesis. Lane 1: primers 4+2, lane 2: primers 4+2, no RT, lane 3: primers 4+1, lane 4: primers 4+3. M denotes size markers.

[C] Schematic of brome mitochondrial *rps19* C-to-U editing sites (circles) and chromatograms of direct sequencing of RT-PCR products from lane 4 in panel B. Arrows show editing sites. Inset shows editing status of codon 55 for RT-PCR products from lane 1 (upper chromatogram) and lane 3 (lower chromatogram).

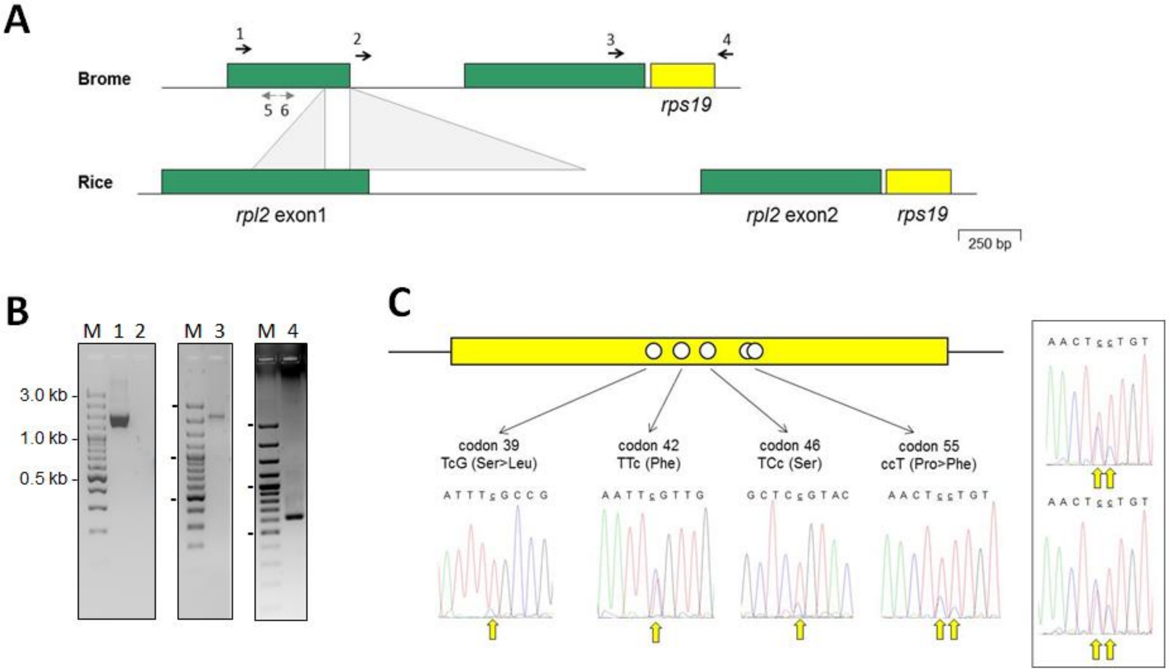
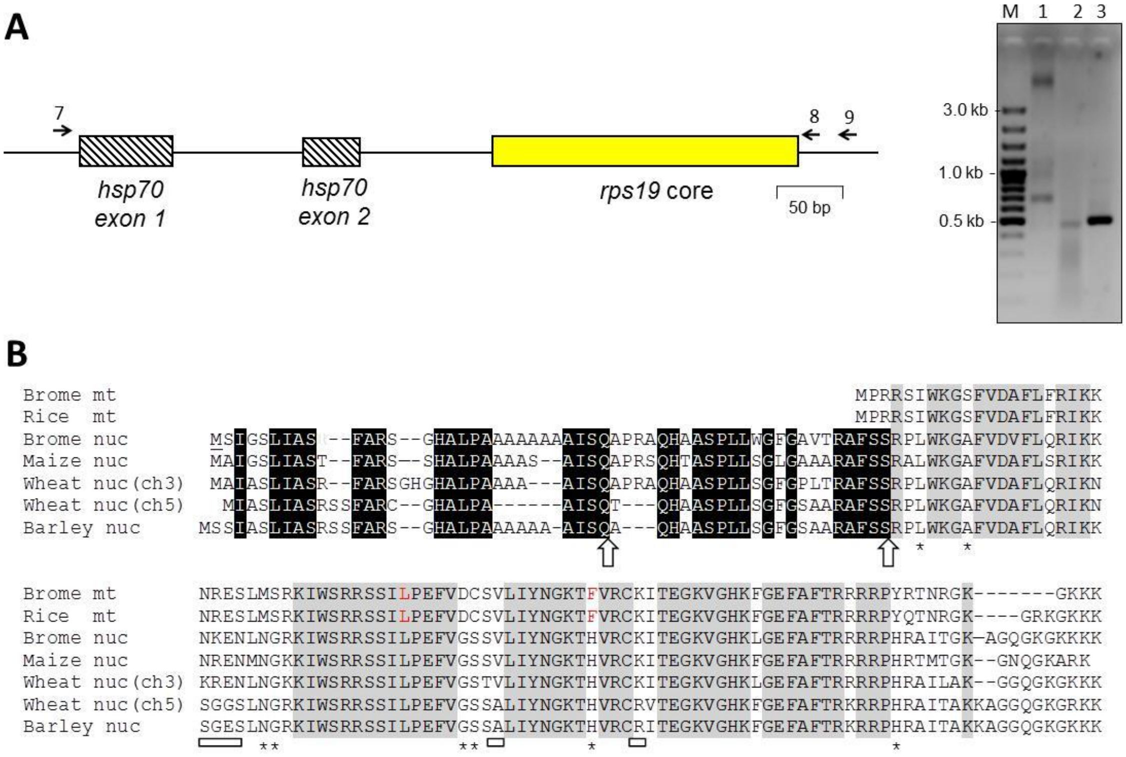

A
1
2
3
4
Brome
5 6
rps19
Rice
rpl2 exon1
rpl2 exon2
rps19
250 bp
B
M 1 2
M 3
M 4
3.0 kb
1.0 kb
0.5 kb
C
codon 39
TcG (Ser>Leu)
codon 42
TTc (Phe)
codon 46
TCc (Ser)
codon 55
ccT (Pro>Phe)
A T T T c G C C G
A A T T c G T T G
G C T C c G T A C
A A C T c c T G T
A A C T c c T G T
A A C T c c T G T

Figure 3.2. . Organization and expression of brome nuclear gene for mitochondrial S19 protein.

[A] Schematic of brome nuclear *rps19* gene with core (yellow) and *hsp70*-type pre-sequence exons (hatched). In gel at right, lane 1: PCR product, primers 7+9, lane 2: RT-PCR product, primers 7+8, lane 3: RT-PCR product with ATP-β primers. Primer sequences are given in Supplemental Table S1. M denotes size markers. Etiolated seedling RNA was used as template. [B] Amino acid alignment of mitochondrial S19 ribosomal protein homologues for the mitochondrion-located genes of brome and rice, as well as the nucleus-located genes in brome, barley, wheat and maize. Accession numbers are given in Supplemental Table S2. Identities within the S19 core and HSP70-type presequence are shaded in grey and black, respectively. Asterisks denote positions shared among the nuclear gene copies, but not the mitochondrial ones. Positions of introns are shown by arrows. The wheat paralogous copies on chromosomes 3 and 5 are shown, and there are also copies similar to these on wheat chromosome 1 (supplemental Table S2). The three open rectangles indicate motifs that distinguish the wheat chromosome 3 vs. chromosome 5 sequences. The maize chromosome 1 copy is shown (see Table S2 for chromosome 5 copies). Changes generated by mitochondrial editing in brome and rice are shown in red letters.



```
Brome mt                                                                MPRRSIWKGSFVDAFLFRIKK
Rice  mt                                                                MPRRSIWKGSFVDAFLFRIKK
Brome nuc          MSIGSLIAS --FARS--GHALPAAAAAAAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGAFVDVFLQRIKK
Maize nuc          MAIGSLIAST--FARS--SHALPAAAAS--AISQAPRSQHTASPLLSGLGAAARAFSSRALWKGAFVDAFLSRIKK
Wheat nuc(ch3)     MAIASLIASR--FARSGHGHALPAAAA---AISQAPRAQHAASPLLSGFGPLTRAFSSRPLWKGAFVDAFLQRIKN
Wheat nuc(ch5)      MIASLIASRSSFARC--GHALPA------AISQT---QHAASPLLSGFGSAARAFSSRPLWKGAFVDAFLQRIKN
Barley nuc        MSSIASLIASRSSFARS--GHALPAAAAAA-AISQA---QHAASPLLSGFGSAARAFSSRPLWKGAFVDAFLQRIKK
                                                    ⇧                       ⇧ *   *

Brome mt          NRESLMSRKIWSRRSSILPEFVDCSVLIYNGKTFVRCKITEGKVGHKFGEFAFTRRRRPYRTNRGK-------GKKK
Rice  mt          NRESLMSRKIWSRRSSILPEFVDCSVLIYNGKTFVRCKITEGKVGHKFGEFAFTRRRRPYQTNRGK----GRKGKKK
Brome nuc         NKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKLGEFAFTRRRRPHRAITGK—AGQGKGKKKK
Maize nuc         NRENMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKFGEFAFTRRRRPHRTMTGK--GNQGKARK
Wheat nuc(ch3)    KRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRCKITEGKVGHKLGEFAFTRRRRPHRAILAK--GGQGKGKKK
Wheat nuc(ch5)    SGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRCRVTEGKVGHKFGEFAFTRKRRPHRAITAKKAGGQGKGRKK
Barley nuc        SGESLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRCRITEGKVGHKFGEFAFTRKRRPHRAITAKKAGGQGKGKKK
                  ▭▭▭▭ **                ** ▭        *  ▭                    *
```

3.4.3 Rearrangement within the hsp70 presequence region of rsp19 in the oat lineage

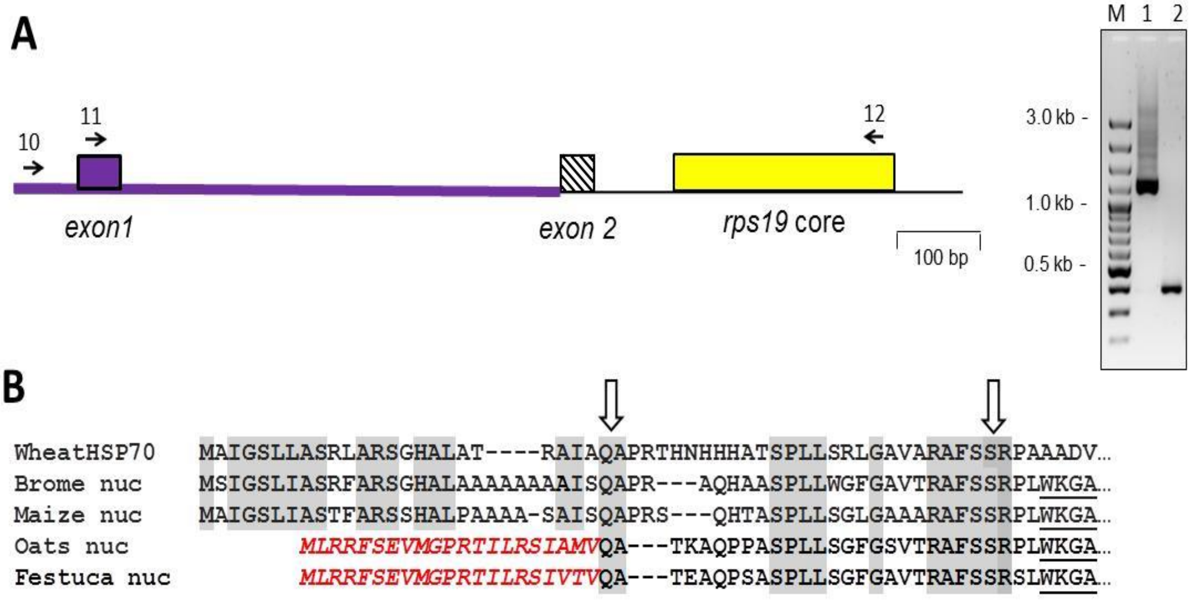
Because the mitochondrion-located *rps19* in oats is known to be a pseudogene (Fallahi et al. 2005), we anticipated that a functional copy would be present in the nucleus. This indeed is the case (fig. 3) and it shares the expected features with nuclear copies in other grasses, except within the extreme amino-terminus where a subsequent rearrangement has replaced the first exon and intron of the *hsp70* presequence with sequences unrelated to any databank entries. Close relatives of oats also have this amino-terminal sequence, as deduced from EST data for *Lolium*, *Festuca* and *Dactylis* (fig. 3B, red and data not shown), so we conclude that this exon shuffling event occurred in their ancestor, that is about 15 million years ago (Catalan et al. 2004; Kellogg and Bennetzen 2004). Both of the potential initiation codons (fig. 3B) are predicted to be able to target the S19 protein to the mitochondrion based on algorithms such as TargetP (Emanuelsson et al. 2007) and Predotar (Small et al. 2004). Our RT-PCR sequencing data confirm that the oat nuclear *rps19* copy is expressed and properly spliced (fig. 3A and supplemental fig.S4). Figure 3B also illustrates the similarity between the *hsp70*-type presequences that were acquired by *rps19* and that of the *hsp70* gene, and it can be seen that they are less well-conserved than the S19 core reigon and have a particularly rapidly-evolving stretch of alanine codons.

3.4.4 Discussion

Our comparative analysis strongly suggests that the nuclear-located *rps19* gene in present-day brome is derived from a gene transfer event which occurred in the common ancestor of Poaceae grasses, prior to divergence of the wheat/maize lineages and that the mitochondrion-located *rps19* gene in brome represents the “native” endosymbiotic-origin form. Both are actively-transcribed and undergo the expected RNA processing events of splicing and editing, respectively. This period of co-existence is much longer than for the three other cases in plants which are in an “intermediate stage” of gene transfer. The *Populus sdh4* mitochondrial and nuclear genes share such high similarity (95% amino acid identity) that this almost certainly indicates a very recent transfer event (Choi et al. 2006). For the *cox2* gene in legumes, transfer was estimated to have occurred during the evolution of the legume subfamily Papilionoideae (Adams et al. 1999) which would be about 25-30 million years ago (Stefanovic et al. 2009). In the case of *rpl5* in wheat (Sandoval et al. 2004), our

Figure 3.3. Organization and expression of oat nuclear gene for mitochondrial S19 protein.

[A] Schematic of oat nuclear gene with *rps19* core (yellow), *hsp70*-type pre-sequence exon 2 (hatched) and novel exon1/intron 1 (purple). In gel at right, lane 1: PCR product with primers 11+12, lane 2: RT-PCR product with primers 10+12. M denotes size markers. [B] Amino acid alignment of presequence regions of nuclear-located *rps19* for oat, *Festuca*, brome and maize and *hsp70* gene in wheat. Underlining shows where the core S19 begins. Identical amino acid positions are shaded in grey. Intron positions are shown by arrows and novel pre-sequences in oat and *Festuca* are in italics.



```
WheatHSP70    MAIGSLLASRLARSGHALAT----RAIAQAPRTHNHHHATSPLLSRLGAVARAFSSRPAAADV…
Brome nuc     MSIGSLIASRFARSGHALAAAAAAAAISQAPR---AQHAASPLLWGFGAVTRAFSSRPLWKGA…
Maize nuc     MAIGSLIASTFARSSHALPAAAA-SAISQAPRS---QHTASPLLSGLGAAARAFSSRPLWKGA…
Oats nuc              MLRRFSEVMGPRTILRSIAMVQA---TKAQPPASPLLSGFGSVTRAFSSRPLWKGA…
Festuca nuc           MLRRFSEVMGPRTILRSIVTVQA---TEAQPSASPLLSGFGAVTRAFSSRSLWKGA…
```

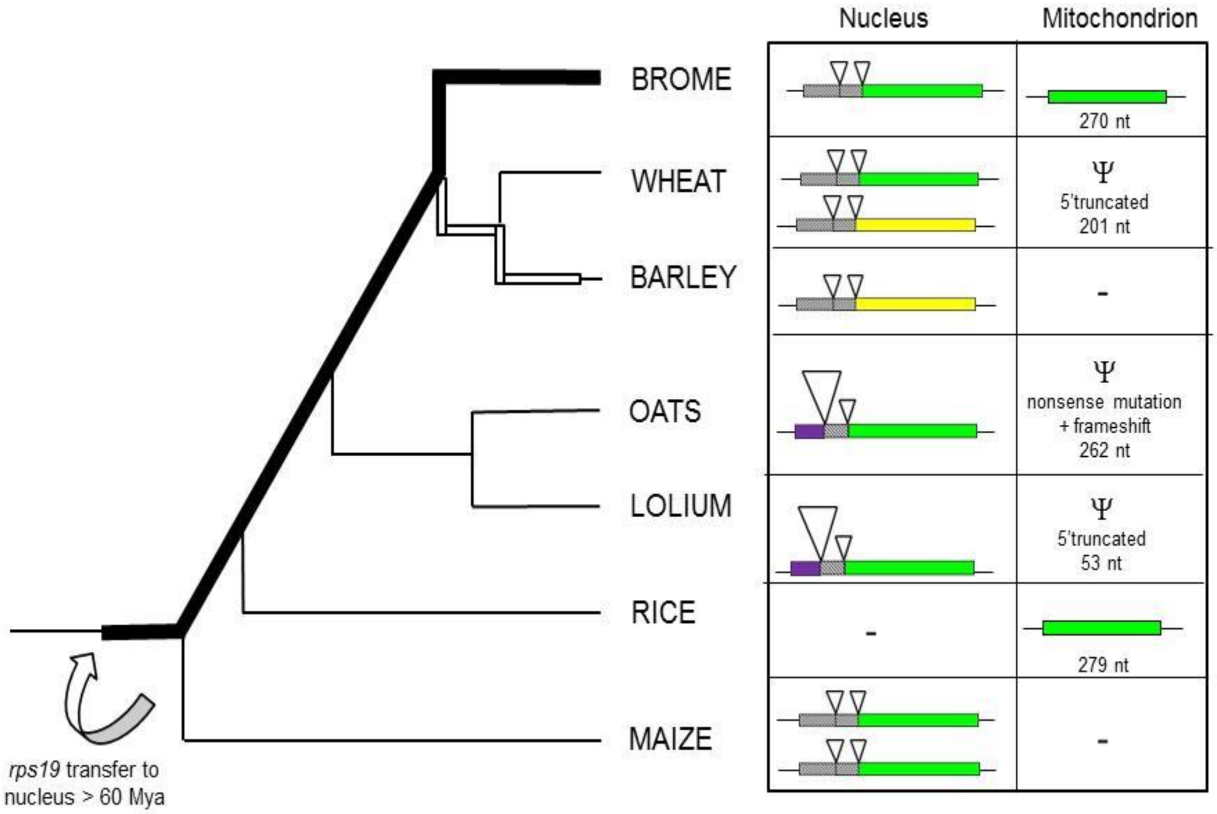
examination of EST data for relatives of wheat supports the view that transfer occurred in the common ancestor of wheat/barley/oats (Supplemental Table S2 and fig.S4), lineages which have a divergence time of approximately 25-30 million years (Kellogg and Bennetzen 2004; Chalupska et al. 2008).

The only other documented case of functional gene copies being present in both compartments, is *atp9* in filamentous fungi, including *Neurospora crassa* (van den Boogaart et al. 1982) and *Aspergillus nidulans* (Brown et al. 1984), which diverged from a common ancestor over 200 million years ago (Taylor and Burbee, 2006). These genes appear to have evolved specialized roles and are expressed either during vegetative growth or in germinating spores (Bittner-Eddy et al. 1994; Dequard-Chablat et al. 2014). Interestingly, in the *Podospora* lineage, duplicated copies of the nuclear-located *atp9* gene perform those two roles and the mitochondrial *atp9* gene copy has been lost (Dequard-Chablat et al. 2011). This illustrates that, even after very long periods of co-existence, events such as gene duplication can lead to redundancy of other paralogues (in this case the mitochondrial one) and its subsequent loss.

Figure 4 presents a scenario for the evolutionary history of the mitochondrial *rps19* gene in grasses. After migration of a copy to the nucleus more than 60 million years ago, in its new environment the gene acquired an *hsp70*-type targeting sequence through exon shuffling as well as amino acid-altering differences from the mitochondrial copy that are shared among the grasses (fig.2B, asterisks), so presumably were acquired early during adaptation to the new environment and then subject to evolutionary constraint. In certain lineages, the mitochondrial *rps19* copy became a pseudogene (eg. wheat, oats, *Lolium*) and in others it was lost (eg. maize, barley). In contrast, in the rice lineage, the mitochondrial gene was retained and the nuclear copy lost. In the oat clade, a lineage-specific rearrangement event conferred a new distal amino-terminal targeting sequence in their ancestor that is about15 million years ago (Kellogg and Bennetzen 2004). The duplication event leading to multiple copies in present-day wheat likely occurred after divergence of the wheat and oat lineages because EST databases did not reveal any evidence for multiple *rps19* copies in the nucleus for members of the oat clade (oat/*Festuca*/*Lolium*/ *Dactylis*). The maize nuclear genome has multiple *rps19* gene copies, however their close similarity suggests either a recent independent duplication event or gene conversion which confounds tracing

Figure 3.4. Scenario showing evolutionary history of the mitochondrial rps19 gene in grasses.

Bold black line on tree indicates period of co-existence of functional *rps19* gene copies in both the mitochondrion and nucleus. Curved arrow indicates mitochondrion-to-nucleus gene transfer in ancestor of grasses. Table beside the phylogenetic tree shows the status of *rps19* genes in both mitochondria and nucleus for these grasses, with the functional genes depicted as in fig.3.2 and 3.3, except that introns are shown as triangles. The *rps19* gene copies most similar to the mitochondrial gene are shown in green, and the wheat chr5 copy (as well as the barley one) are shown in yellow. The presence of mitochondrial DNA containing *rps19* in the barley nucleus, plus the absence of homologous sequences in the mitochondrion, points to a recent DNA-mediated transfer (Numt) in the barley lineage and preceding loss of the mitochondrion-located copy. This relatively long period of co-existence in both compartments in the barley lineage is shown by white bars on tree. Data for *rps19* status in the mitochondrion is from Fallahi et al. 2005; databank information (see Suppl. Table S1) and this study. Numbers indicate the lengths of *rps19* homologous sequences. 'Ψ' indicates pseudogene while – indicates absence. Divergence times for the grasses shown on the tree are maize/wheat ~60 Mya, rice/wheat ~50 Mya, wheat/oat ~25-30 Mya, brome/wheat ~20 Mya, barley/wheat ~15 Mya, and oat/Lolium ~15 Mya (Gaut 2002; Catalan et al. 2004; Kellogg and Bennetzen 2004; Chalupska et al. 2008).


Nucleus
Mitochondrion
BROME
270 nt
WHEAT
Ψ
5'truncated
201 nt
BARLEY
-
OATS
Ψ
nonsense mutation
+ frameshift
262 nt
LOLIUM
Ψ
5'truncated
53 nt
RICE
-
279 nt
MAIZE
-
rps19 transfer to
nucleus > 60 Mya

their evolutionary history (data not shown). Interestingly, a search of the barley draft nuclear genome revealed a block of mitochondrial DNA containing *rps19* and its flanking sequences (>850 bp) and we confirmed its presence by PCR sequencing (supplemental fig.S5). Because it is an unedited form and lacks nuclear-type expression sequences we conclude that this reflects a non-functional DNA-mediated transfer (Numt) to the nucleus. Although the *rps19* reading frame has not been pseudogenized, there are five non-synonymous substitutions. The present-day barley mitochondrial genome lacks an *rps19* gene (based on our Southern hybridization experiments using *rps19* PCR product as well as oligomers as probes), so its loss must have been quite recent. Therefore, it appears that the barley lineage has been in a transition stage (fig.4, white bars on tree) for close to the same length of time as seen for brome (fig.4, thick black line on tree).

It will be of interest to learn whether the brome mitochondrial ribosomes contain both forms of the S19 protein or whether there is variation under certain environmental or developmental conditions. In a broader context, the presence of multiple gene copies, regardless of whether they are located in the same or different genetic compartments, and regardless of whether they are generated by intragenomic DNA duplication or intracellular horizontal transfer, provides the opportunity for specialization or acquisition of new cellular functions. In this regard, it is perhaps worth noting that recent chloroplast-to-mitochondrion gene transfer events have resulted in ten intact ribosomal protein genes, including *rps19*, being present in the *Vitis vinifera* mitochondrial genome (Goremykin et al. 2008). This raises the possibility of their recruitment for a role in mitochondrial translation, analogous to certain chloroplast-origin tRNAs which have replaced endogenous ones (reviewed in Huot et al. 2014), or for the creation of chimeric genes through homologous recombination. Moreover, since ribosomal proteins are well-known for performing additional extra-ribosomal functions (Wool 1996), and considering the complexities of RNA processing events in plant mitochondria (reviewed in Hammani and Giegé 2014), it might not be surprising if certain ribosomal proteins undertook additional “moonlighting” roles.

Acknowledgements

We thank members of the Bonen laboratory for helpful discussions and Dr R. Pandeya (Agriculture and Agri-food Canada) for kindly providing seeds. Financial support from the Natural Sciences and Engineering Research Council of Canada is also gratefully acknowledged.

Electronic supplementary material

Supplementary Table S1. Primer sequences used in PCR and RT-PCR.

Supplementary Table S2. Accession numbers of databank sequences used in comparative analysis.

Chapter 3 addendum:

This chapter includes the supplemental information for chapter 3, which has been written as a manuscript for publication.

3.5.1 Brome mitochondrial genome has pseudo *rpl2* upstream of functional *rps19*

Evolutionary status of *rpl2* and *rps19* genes has been surveyed in brome and other closely related grasses. Brome *rpl2* is missing 311bp of exon 1, creating a frameshift, and 956nt of 5' group II intron relative to rice and bamboo *rpl2*, making it a pseudogene (figure 3.5a and 3.5b) (Appendix A: i). The 5' end that's missing in brome *rpl2* group II intron is very important for splicing. This is because the folding of domain I of group II intron allows the 5'exon-intron junction to undergo first-step catalysis (Maracia and Pyle 2012). Comparative analysis of mt*rpl2* in rice against *ferrocalamus, Bambusa, bromus,* and *Phoenix* shows that both rice and brome are missing 144nt of exon1. Rice *rpl2* intron is missing 150nt, which is present in brome, *bambusa*, *ferrocalamus* and *phoenix* (Appendix A: i). The indel in brome exon 1 *rpl2* is not a multiple of three, thus it causes a frame shift mutation. Comparative analysis between brome and bamboo exon 1 *rpl2*, indicates that there is a direct repeat in *bambusa*, which is shown in red and underlined in (Appendix A, ii). This could be one possible reason as to why the exon 1 *rpl2* gene in brome has a deletion. Direct repeats can be responsible for genetic instability and were shown to cause such deletions and recombination in plant mitochondria (Sloan *et al* 2010) and yeasts (Lobachev *et al* 1998). Unlike exon 1, exon 2 is conserved with rice, bamboo and other grasses. Sequences upstream of the start codon of *rpl2* are completely conserved between brome and bamboo, but the break point in homology between brome and rice occurs 85nt upstream (Appendix B). At the amino acid level, the brome *rpl2* gene has 11 nonsynonymous substitutions compared to rice *rpl2*. Out of these eleven changes, 6 changes are unique to brome and three of them are drastic amino acid changes (shown in red boxes, Appendix A: iii). The indel in brome *rpl2* causes a frameshift mutation shown in grey highlight.

The *rpl2* pseudogene is edited at one position in brome similar to rice (Kubo et al 2006). The edit that is present within the 3nt spacer region, between *rpl2* and *rps19,* in rice (Kubo et al 2006), however, is not present in brome. In (Appendix A: IV b & c) the tracings

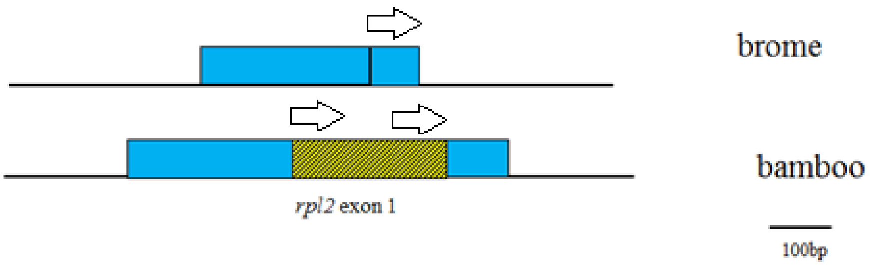
Figure 3.5:- Brome mitochondrial genome has pseudorpl2 upstream of functional rps19.

[A] Schematic shows the missing region in brome *rpl2* exon 1 relative to bamboo mitochondrial *rpl2* gene. Missing region in brome is shown as a yellow box in bamboo. Direct repeats are shown as white arrows

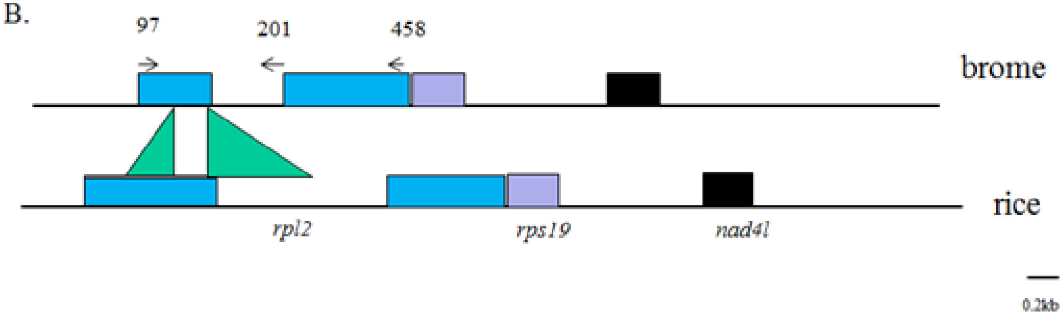
[B] Schematic shows the missing region in brome *rpl2* exon 1 and intron relative to rice mitochondrial *rpl2* gene.

[C] Agarose gel showing the difference in size of PCR products due to indels in brome (B) relative to rice (R) *rpl2* gene. The oligomers used in the gel are shown on the schematic in 3.5b. The sizes of PCR products are shown to the left in kilo bases.

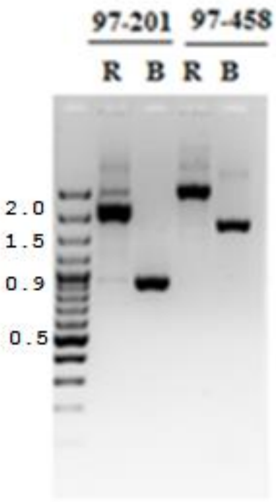
A.


brome
bamboo
rpl2 exon 1
100bp

B.


97
201
458
brome
rice
rpl2
rps19
nad4l
0.2kb

C


97-201
97-458
R
B
R
B
2.0
1.5
0.9
0.5

for the *rpl2* and *rps19* edits are shown for all of the transcripts in brome to highlight that precursors are under edited as mentioned in chapter 3.

3.5.2 Southern analysis of brome mitochondrial *rpl2* and *rps19* genes

Southern analysis was performed to address if brome has pseudo *rpl2* and *rps19* genes in addition to the copy that I have sequenced. Using exon 1 *rpl2* oligomer lb 174, only a single intense 1.2kb HindIII fragment was observed. This suggests that there is not another functional *rpl2* in the brome mitochondrial genome as lb 174 is conserved among *bambusa*, rice and *ferrocalamus*. However it cannot be ruled out that there could be a functional *rpl2* which has an indel in the region where the oligomer lb 174 is located. Interestingly, reprobing the same blot (BAS2) used in the analysis described above with exon 2 *rpl2* and *rps19* oligomer lb 458 and lb188 respectively showed two hybridization signals with HindIII (fig 3.6a panel 2 and 3). Southern analysis suggests that the second *rpl2* and *rps19* copies are not beside each other as the bands are of different sizes. The same blot re-probed with another oligomer lb 164 in the 3' end of *rps19* gave only one intense hybridizing signal with HindIII and BamHI (fig 3.6a panel 4) which was reproducible with another blot (data not shown). The 4.1kb faint fragment seen with Hind III in panel 4 is not of equal intensity and must therefore represent sublimons or repeated sequences within the brome mitochondrial genome. Because two signals of equal proportions are not seen with lb164 it can be said that the second *rps19* copy is missing the 3'end of the gene where lb188 oligomer is present. As wheat, rye, barley and lolium all have pseudogene *rpl2* with *nad1* and *nad6* linkage, I suspected that brome *rpl2* would also have it. Using inverse PCR, I was able to successfully obtain the pseudo *rpl2-nad1-nad6* product in brome, confirming my prediction (Appendix C). Thus brome was found to have another pseudogene copy of *rpl2* with *nad1* and *nad6* linkage as well as a pseudogene *rps19* elsewhere in the genome with a truncated 3'end.

3.5.3 Southern analysis of mitochondrial *rpl2* in barley, and oats

Barley mitochondrial genome was restricted with ECORI to find the pseudo *rpl2* gene. I predicted that, as in wheat which is a close relative, barley should also have the pseudo *rpl2* gene beside *nad1* and *nad6*. Oligomer lb 458 was found to hybridize with barley and is specific to *rpl2* exon 2, which is present in the wheat pseudogene (Subramanian et al 2001). Barley exhibited a 5.4kb signal with Hind III and 7.6 kb signal with BamHI. Using EcoRI enzyme, 0.7kb fragment was seen hybridizing to barley. As predicted, the product from the inverse PCR experiment in barley mapped to pseudo*rpl2* with *nad1* and *nad6* linkage (Appendix C). Oats, on the other hand, showed complete absence of *rpl2* exon 1 and exon2 probes. From prior work in our lab, we know that oats underwent a rearrangement with respect to wheat and its pseudogene piece is dissimilar to wheat's as a result. It can also be said that the absence of signals is not due to poor labelling of the probes as they worked on other blots with brome in figure 3.6a.

3.5.4 Northern analysis of *rpl2-rps19-nad4L* locus in brome mitochondrial genome

RNA blot analysis was conducted for the *rpl2-rps19-nad4l* locus in brome to know the sizes of the transcripts coming from this region. In (figure 3.7, A), we see that *rpl2* exon 1, exon 2 and *rps19* specific oligomers hybridized to a 3.4kb precursor. This precursor is expected to include *nad4l* downstream to *rps19* since the corresponding DNA size from this locus is 3.2kb. A 2kb messenger RNA which corresponds to *rpl2-rps19* cotranscript is also seen with all three oligomer probes. However the exon 2 *rpl2* oligomer (lb458, panel iii) gave a different profile than exon 1 *rpl2* oligomer (lb826, panel i). This could be due to the fact that the exon 2 oligomer lb458 is present in the other pseudo *rpl2* gene copy cotranscribed with *nad1* and *nad6*. It is also worth noting that the levels of the transcripts seen with three different oligomers from *rpl2-rps19* locus in brome are very low. To show that this is the case, the same blot used for *rpl2* exon 1 has been reprobed with cytochrome b specific oligomer (lb 638, panel ii). An intense 2.2kb hybridizing signal was seen with lb 638 indicating that the quality of the RNA used for my northern analysis was good. Panel IV, in figure 3.7 A, shows hybridization profile with *rps19* gene specific oligomer. Low level transcripts (3.4 and 2.0kb) were seen that would be the size of precursors coming from this locus which includes *rps19* itself.

Two different oligomers were used from brome mitochondrial *rps19* gene for Northern analysis (figure 3.7 B), but both attempts have been unsuccessful. Figure 3.7B (panel i) shows a case where the *rps19* specific oligomer looks like it is hybridizing to another gene's messenger as the 1.0kb species was not seen with lb918 *rps19* oligomer (figure 3.7 A panel i). Figure 3.7.B panel ii, shows another case where the *rps19* specific oligomer lb913 was found to be cross hybridizing to ribosomal RNA.

3.5.5. Evidence for functional *rpl2* gene in brome nuclear genome

As I have determined that the mitochondrial *rpl2* gene in brome is a pseudogene, it is expected that the functional gene is in the nucleus. Using wheat and barley nuclear *rpl2* EST sequences, I was able to design primers in conserved regions so they can work on other grasses. Although lb849 worked on wheat and rye (800bp) it did not work on barley and brome. Lb79 and lb89, however, were designed in highly conserved regions of *rpl2* gene so they worked on brome, giving a product of 500bp. This indicates that the *rpl2* gene has been transferred to the nucleus in brome as expected (figure 3.8A).

The nuclear *rpl2* gene does not have introns present within it (based on high-throughput and EST data available for barley and draft genome for wheat), therefore, an RT-PCR experiment was performed using 3xDNAse treated brome total RNA. Brome nuclear *rpl2* gene was shown to be expressed, as an abundant 500bp product was obtained with my +RT sample and no product was seen in –RT lane (figure 3.8b). Direct sequencing tracings of brome *rpl2* gene at the DNA level revealed the presence of multiple copies. Sequencing the brome *rpl2* product obtained from an RT-PCR experiment using 3xDNAse treated RNA revealed that the cDNA sequence was different from the DNA sequences. This could mean that the RNA copy that was sequenced is potentially the most expressed out of all the other DNA copies or they are silent. Tracings in figure 3.9b show the copy 1 and copy 2 of brome nuclear *rpl2* genes sequence at the DNA and RNA level.

Similar to *rps19*, the wheat nuclear *rpl2* gene also has paralogous copies (ensemble.plants.ca) (Appendix D). The paralogous copies of *rpl2* in wheat are on chromosomes 1 and 7 1Ds, and 7Al. Similarly barley was also found to have paralogous copies on chromosomes 3 and 1(Appendix D). The percent identity between my brome nuclear *rpl2* DNA and my RNA copies is 88%. As there are a significant number of positions where the DNA and RNA copies differ from each other, they seem to be paralogous copies rather than homeologous ones. Another interesting thing is that although there are no introns within the nuclear *rpl2* gene itself (figure 3.8c), its 5'UTR was found to have an intron and an exon in wheat and barley. The 5'UTR exon sequences present in wheat and barley are not conserved. This 5' UTR exon and intron for nuclear *rpl2* gene might explain how the gene became functional when it moved to the nucleus. As wheat and barley are close relatives of brome, it could be speculated that brome also might have a UTR exon and intron in the UTR but might not be conserved.

3.5.6 Codon usage pattern of the nuclear located *rpl2* gene in brome versus other mitochondrial genes

Nuclear and mito *rpl2* genes are compared with other mitochondrial and nuclear genes that are available for brome in NCBI (Appendix L). Nuclear and mitochondrial *rpl2* gene in brome have bias for G and C at the third postion (figure 3.10). This is in contrast to other mito genes in general as they usually have higher A and T codon bias (Subramanian thesis, 1999). Thus after transfer to the nucleus, brome nuclear *rpl2* gene has adapted to the GC environment. Wheat is also typically found to have bias for G and C at the third position of the codon in the nucleus (Subramanian thesis, 1999).

3.5.7 RPL2 gene transfer to the nucleus: -Evidence for two independent transfer events

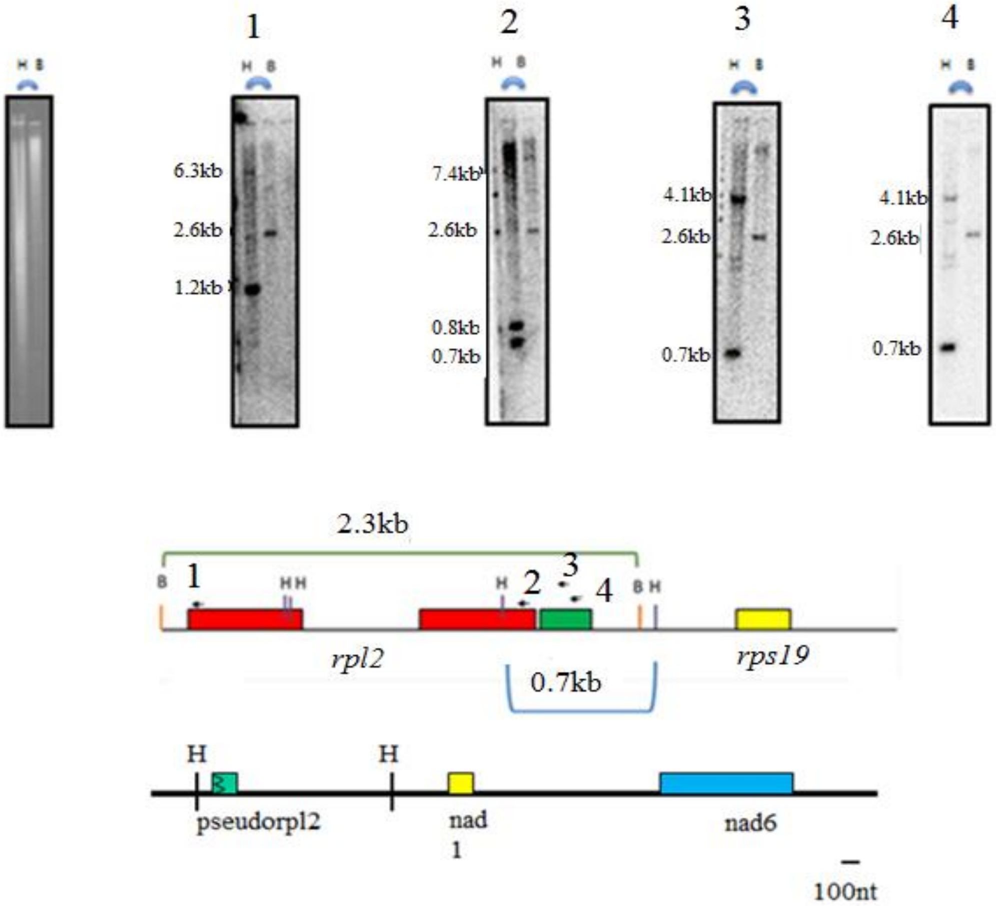
Prior research in our lab by Selvi Subramanian on *rpl2* in wheat and other grasses illustrated that the transfer of *rpl2* gene was not likely a single common transfer event in the ancestor but two independent transfer events (one in the maize lineage and one in the ancestor of wheat and barley) (Subramanian et al 2001). My experimental data (rye and brome) and EST data from festuca, supports the fact that the transfer of *rpl2* must have been in the common ancestor of wheat, rye, barley, brome and extends to oats clade which includes lolium and festuca. The amino acid alignment in figure 3.11a and b highlights that wheat, barley, rye, brome and oats share amino acids that are not present in the "ancestral" rice mitochondrial *rpl2* gene. This suggests that those amino acid changes are shared derived and have been acquired in the nucleus after transfer has occurred.

Figure 3.6:- Southern analysis suggests additional pseudo*rpl2* and *rps19* copies in brome mitochondrial genome *(in addition to the functional rps19).*

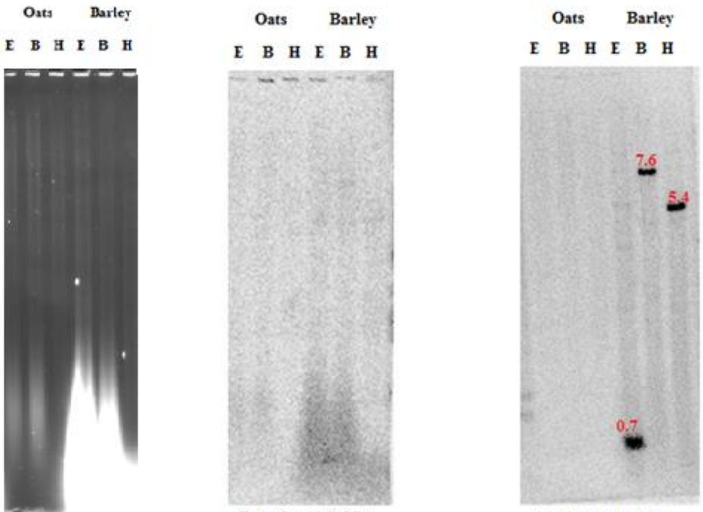
[A] Brome mitochondrial DNA blot BAS2 (prepared by Bronwyn) digested with Hind III and BamHI restriction enzymes has been hybridized with *rpl2* exon 1 specific oligomer lb 174, oligomer 1 (panel 1), *rpl2* exon 2 specific oligomer lb 458, oligomer 2 (panel 2), *rps19* specific oligomers lb 188 , oligomer 3 (panel 3) and lb 164, oligomer 4 (panel 4).

[B] i) Mitochondrial DNA blot (EBS1) with oats and barley digested with restriction enzymes EcoRI, BamHI, and Hind III, hybridized with exon 1 *rpl2* specific oligomer #1, lb 174 ii) same blot that has been reprobed with exon2 *rpl2* oligomer lb 458

i)



ii)



E xon 1 rpl2 lb 174

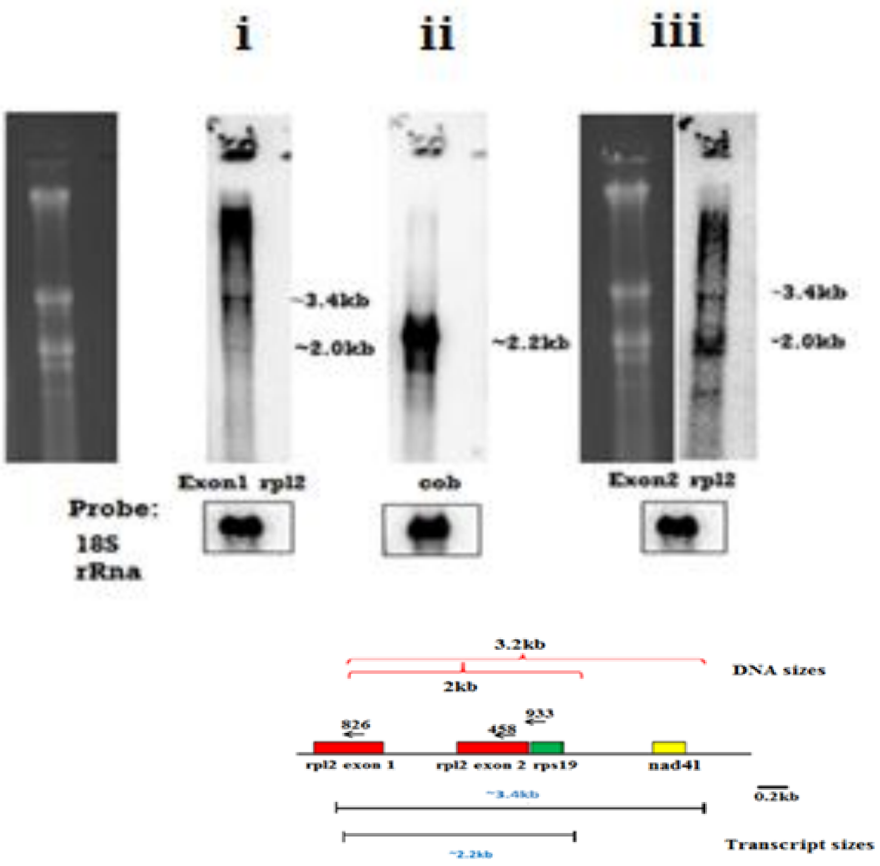
Exon 2 rpl2 lb 458

Figure 3.7:- RNA blot analysis for rpl2-rps19-nad4l locus in brome mitochondria

A. (i) Northern blot (SAN19C) with brome seedling mitochondrial RNA hybridized first with the ^{32}P-end labelled brome *rpl2*-specific oligomer probe lb 826. (ii) the same blot was then hybridized with a cytochrome b (cob) specific oligomer probe lb 638. (iii) Another sister blot (SAN19B) was hybridized with *rpl2* exon specific oligomer lb 458. (iv) BAN5B blot (prepared by Bronwyn) hybridized with brome *rps19* specific oligomer probe lb 933. Fluorescence profiles as well as the 18S rRNA loading controls are indicated below. Size markers in kilobases are shown on the right. Schematic of brome *rpl2-rps19-nad4l* locus to show the sizes of transcripts seen in Northerns and how they compare with the DNA sizes.

B. Northern blot (SAN19A sister blot of san19b and san19c used in figure 3.7 A (panel i) hybridized with brome mitochondrial *rps19* gene specific oligomer lb 164. (ii) Northern blot (SAN18) hybridized with brome mitochondrial *rps19* gene specific oligomer lb 918. Fluorescence profiles have been shown for each blot and the size markers are shown to the right in kilobases.

A.


i
ii
iii
iv
~3.4kb
~2.0kb
~2.2kb
~3.4kb
~2.0kb
~3.4kb
~2.0kb
Exon1 rpl2
cob
Exon2 rpl2
rps19
Probe:
18S rRna
3.2kb
DNA sizes
2kb
826
933
458
rpl2 exon 1
rpl2 exon 2
rps19
nad41
~3.4kb
0.2kb
Transcript sizes
~2.2kb

B.

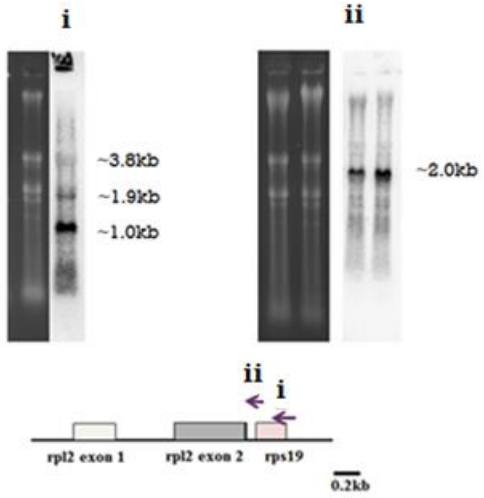
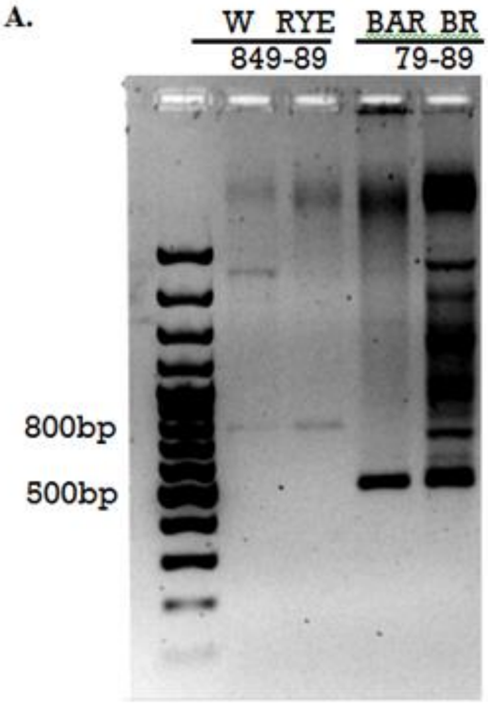
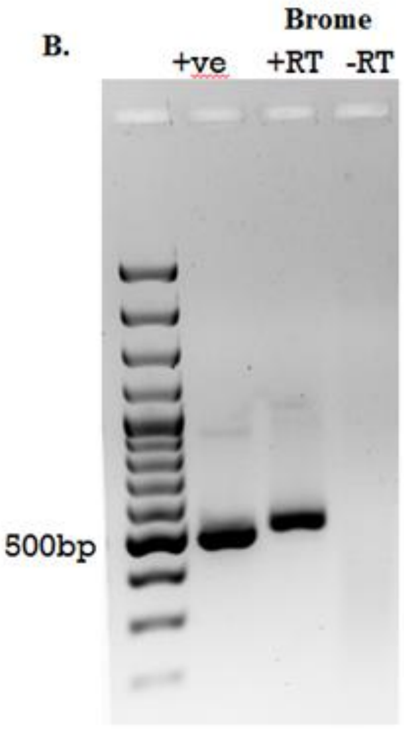

i
ii
~3.8kb
~1.9kb
~1.0kb
~2.0kb
ii
i
rpl2 exon 1
rpl2 exon 2
rps19
0.2kb

Figure 3.8 Organization and expression of brome nuclear rpl2 gene.

A) Agarose gel showing presence of *rpl2* gene in the nucleus of brome, wheat, barley and rye.

B) RT-PCR gel showing brome nuclear *rpl2* is expressed. +ve control is ATP-synthase beta and negative control is without reverse transcriptase enzyme. This gel shows the re-amplified products in all three lanes as the original products were too faint.


A.
W
RYE
BAR
BR
849-89
79-89
800bp
500bp


B.
Brome
+ve
+RT
-RT
500bp

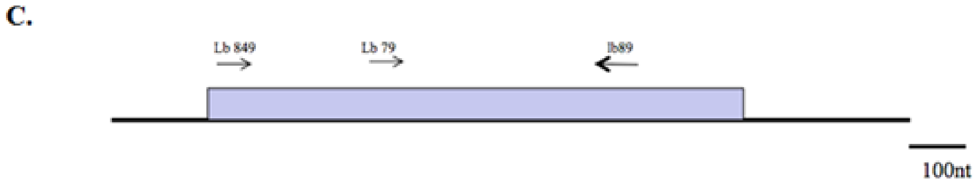

C.
Lb 849
Lb 79
lb89
100nt

Figure 3.9:- Sequence alignments and direct sequencing tracings for the two distinct copies of nuclear rpl2 in brome.

(A) Nucleotide sequence alignment of 3xDNAse treated RT-PCR product (with oligomers lb79 and lb89) and DNA copies. Red box highlights the sequence where the RNA copy has an indel relative to the DNA copies.

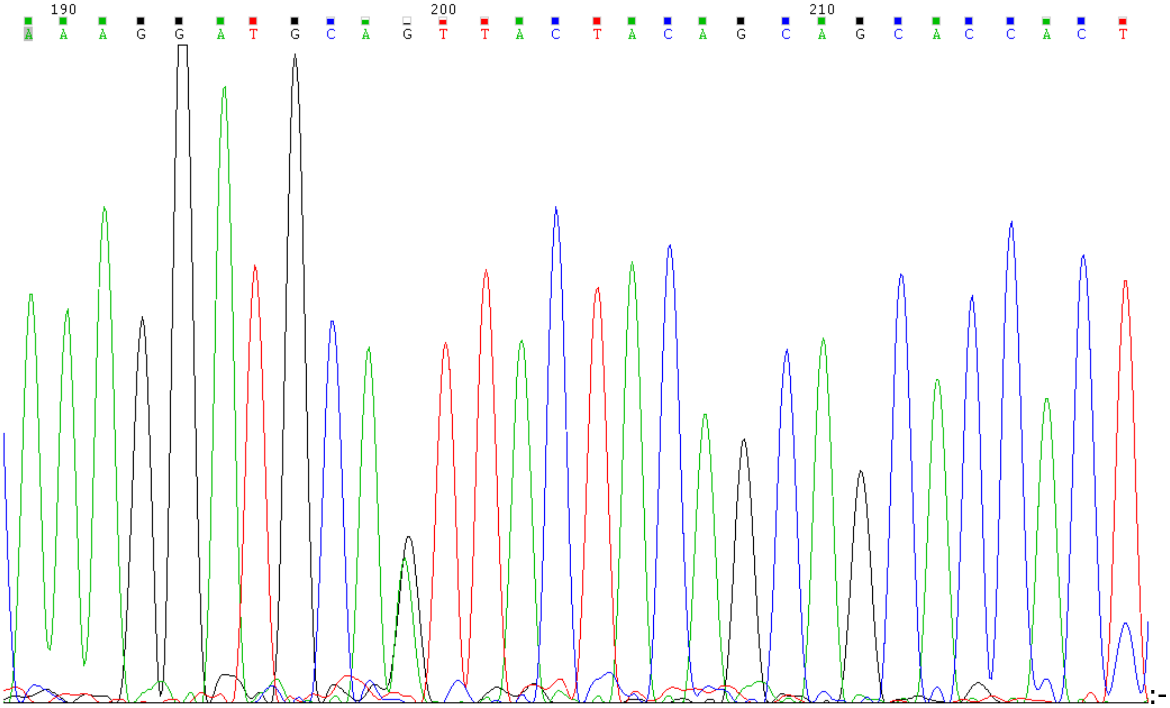
(B) Direct sequencing tracings of PCR and RT-PCR products for brome nuclear *rpl2*. Purple box highlights the three nucleotide indel missing in the DNA copy

A.

```
Brrpl2RNA                         ACAAGGCCGAGCATGCACCTGTCAATTACATAATAGCCAGCCATCAAATGGAACCGGGCA
Brrpl2DNA1                        ---------------------------------------GTCATCAAATGGAACCGGGCA
Brrpl2DNA2                        --------------------------------------AGTCATCAAATGGAACCGGGCA
                                                                         * *******************

Brrpl2RNA                         GCATGGTGGTGAACAGCGACTCCTCCAAACCTTCCACCACCGGCTCCTTGATGCGACCTG
Brrpl2DNA1                        GCATGGTCGTGAACAGCGATTGCTCCAAACCTTCCACAACCGGCTCCCTGCTGCGACCTG
Brrpl2DNA2                        GCATGGTCGTGAACAGCGATTGCTCCAAACCTTCCACAACCGGCTCCCTGCTGCGACCTG
                                  ******* *********** * *************** ********* ** *********

Brrpl2RNA                         CCCACAATGCTGATTCTTTCCTTCGGTTCCAAGAGCTTTTTCGCAAGGCCAGCCAAGTCG
Brrpl2DNA1                        CCCAGAATGCCGATTCTTTCCTTCGCTTCCAAGAGCTTTTTCGCAAAGCCAGTCAAGGTG
Brrpl2DNA2                        CCCAGAATGCCGATTCTTTCCTTCGCTTCCAAGAGCTTTTTCGCAAAGCCAGTCAAGGTG
                                  **** ***** ************** ******************** ***** ****  *

Brrpl2RNA                         AAGAGGGCACTGATGATCAAGCAAAGGATGCAGCGGTCACTACAGCAGCACCGCTCATGC
Brrpl2DNA1                        AAGAGGACACTGATGATCAAGCAAAGGATGCA---GTTACTACAGCAGCACCACTCATGC
Brrpl2DNA2                        AAGAGGACACTGATGATCAAGCAAAGGATGCA---GTTACTACAGCAGCACCACTCATGC
                                  ****** *************************   ** ************** *******

Brrpl2RNA                         CTGCTGATTTGCTGGATCTGAATTCCAAGGTGGGGAACTGCATGCCATTGTCTGACATCC
Brrpl2DNA1                        CTGCCGATCTACTGGATCTGAATTCCAAAGTGGGAAACTGCATGCCACTAGCTGACATCC
Brrpl2DNA2                        CTGCCGATCTACTGGATCTGAATTCCAAAGTGGGAAACTGCATGCCACTAGCTGACATCC
                                  **** *** * ***************** ***** ************ *  *********

Brrpl2RNA                         GTATGGGAACATGGGTGCACAGCATCGAGCTGCGTCACGGCCAAGGGGCGAAGCTCGTCC
Brrpl2DNA1                        GTATGGGAACATGGGTGCACAGCATTGAGTTGCGTCATGGTCAAGGGGCAAAGCTCGTTC
Brrpl2DNA2                        GTATGGGAACATGGGTGCACAGCATTGAGTTGCGTCATGGTCAAGGGGCAAAGCTCGTTC
                                  ************************* *** ******* ** ******** ******** *

Brrpl2RNA                         GGGCCGCTGGAGCGTACGCCAAGGTGATCAAGGAGTCAGCCACGCAGTGCCTTGTGCGGC
Brrpl2DNA1                        GAGCTGCTGGAACTTACGCTAAAGTGGTCAAGGAGTCAGCCTCGCAGTGTCTTGTACGGC
Brrpl2DNA2                        GATCTGCTGGAACTTACGCTAAAGTGGTCAAGGAGTCAGCCTCGCAGTGTCTTGTACGGC
                                  *  * ****** * ***** ** *** ************** ******* ***** ****

Brrpl2RNA                         TGCCGTCGGGCGTCGAGAAGCTG-------------------------------------
Brrpl2DNA1                        TGCCATCGGGTGTTGAGAAACTGATAGACTCCCGATGCAGAGCCACCATCGGCATCGTCT
Brrpl2DNA2                        TACCATCGGGTGTTGAAAAA----------------------------------------
                                  * ** ***** ** ** **
```

B.

COPY 1



DNA

COPY 2:- RNA

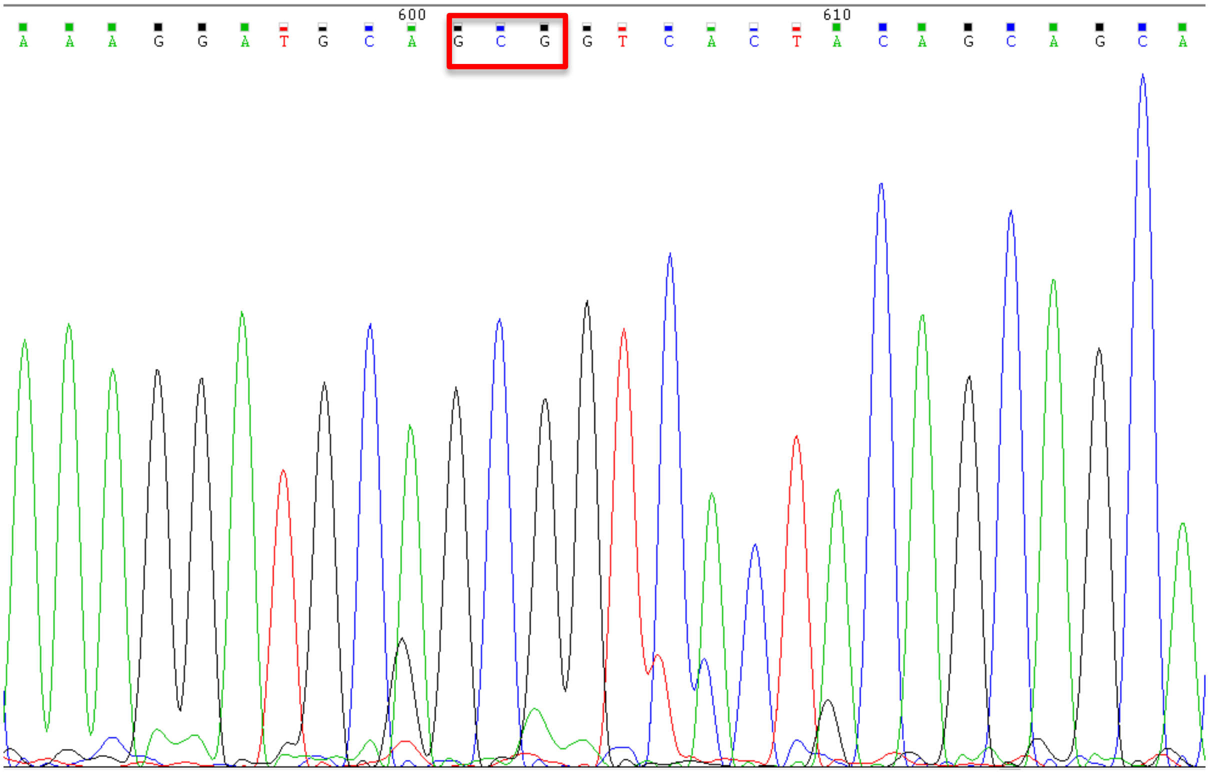


Figure 3.10. Bar graph showing codon bias between mitochondrial pseudo rpl2 and functional rpl2 gene in the nuclear genome of brome

Comparative analysis of codon bias within mitochondrial and nuclear *rpl2* genes in brome is shown in the following bar graph. Data set available for other mitochondrial genes and nuclear genes was used to see if the *rpl2* gene is behaving in the same way. Blue bar represents the mitochondrial *rpl2* pseudogene, red bar represents nuclear *rpl2* gene in brome, green bar represents the dataset of mitochondrial genes used for brome, and purple bar represents the dataset of nuclear genes used for brome. Dataset of other mitochondrial and nuclear genes used for analysis is shown in Appendix L.

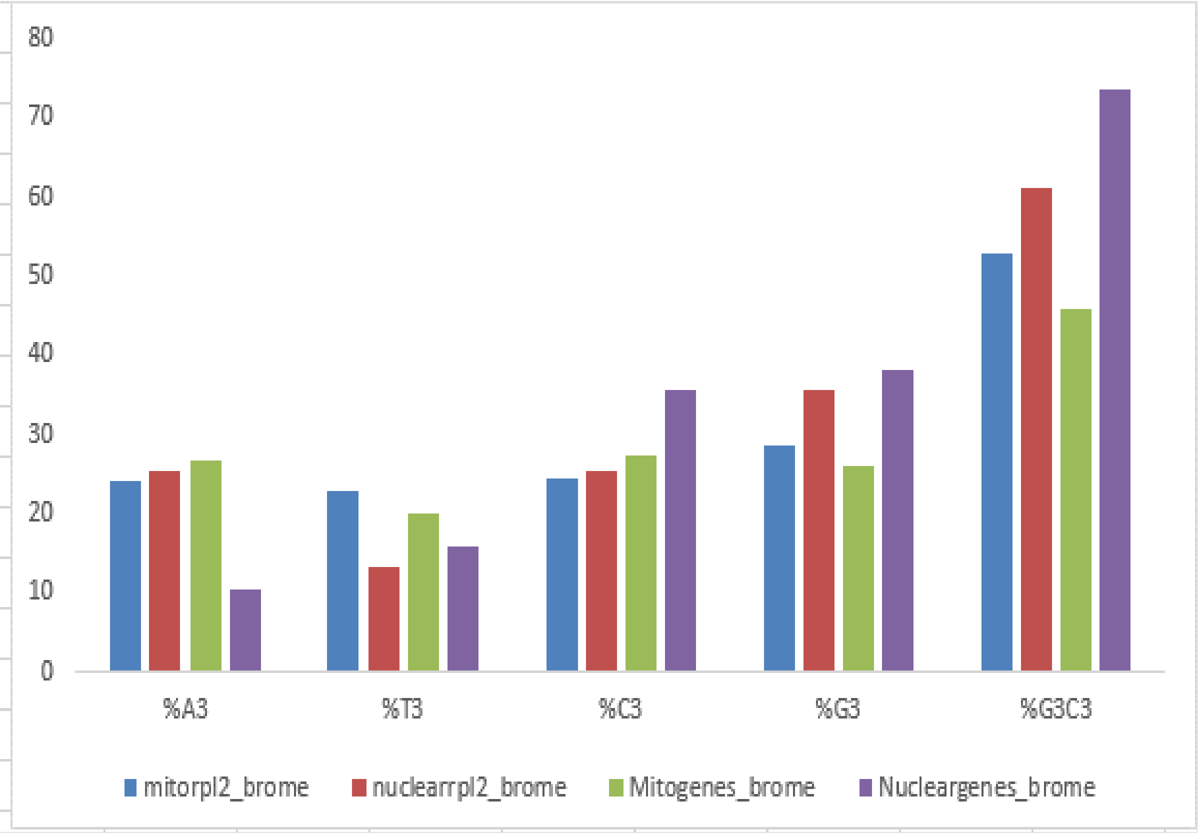
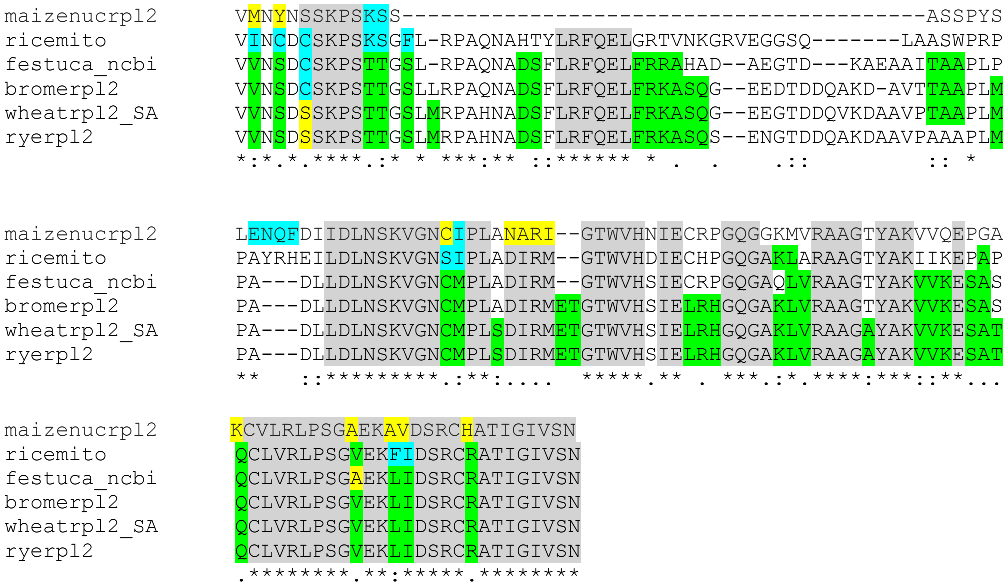

80
70
60
50
40
30
20
10
0
%A3
%T3
%C3
%G3
%G3C3
mitorpl2_brome
nuclearrpl2_brome
Mitogenes_brome
Nucleargenes_brome

Figure 3.11 Amino acid alignment of nuclear rpl2 gene in brome and other grasses to show when the transfer event might have occurred.

A. Amino acid alignment of brome nuclear *rpl2* gene with all the other nuclear *rpl2* genes in (rye, wheat, maize and festuca) along with the mitochondrial copy in rice. Grey highlight shows the conserved regions of *rpl2* among nuclear located copies in all of the plants and the mitochondrial located copy in rice. Blue highlights the amino acid changes within the mitochondrial copy that are not present in the nuclear copies. Green highlight shows the shared derived amino acid changes in the nuclear located copies of different grasses such as wheat, rye, brome, and festuca. Yellow highlights show the amino acid changes that are unique to maize clade which are not present in the nuclear located copies of other grasses or the mitochondrial rice *rpl2* gene.

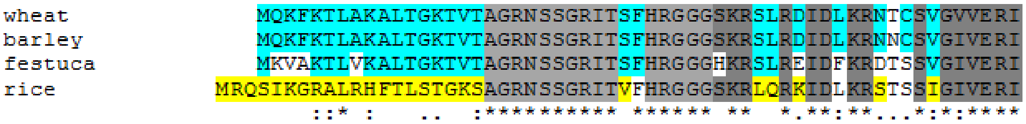
B. Amino acid alignment of the n-terminal sequence of nuclear *rpl2* in wheat, barley, festuca and the mitochondrial copy in rice. Grey highlight shows the conserved amino acid positions while blue highlights the shared derived amino acid positions in wheat, barley and festuca that are different from the mitochondrial rice *rpl2 (*shown in yellow)

(a)



```
maizenucrpl2        VMNYNSSKPSKSS-----------------------------------------ASSPYS
ricemito            VINCDCSKPSKSGFL-RPAQNAHTYLRFQELGRTVNKGRVEGGSQ-------LAASWPRP
festuca_ncbi        VVNSDCSKPSTTGSL-RPAQNADSFLRFQELFRRAHAD--AEGTD---KAEAAITAAPLP
bromerpl2           VVNSDCSKPSTTGSLLRPAQNADSFLRFQELFRKASQG--EEDTDDQAKD-AVTTAAPLM
wheatrpl2_SA        VVNSDSSKPSTTGSLMRPAHNADSFLRFQELFRKASQG--EEGTDDQVKDAAVPTAAPLM
ryerpl2             VVNSDSSKPSTTGSLMRPAHNADSFLRFQELFRKASQS--ENGTDDQAKDAAVPAAAPLM
                    *:*.*.****.:* * ***:** ::****** * .  .    .::         :: *

maizenucrpl2        LENQFDIIDLNSKVGNCIPLANARI--GTWVHNIECRPGQGGKMVRAAGTYAKVVQEPGA
ricemito            PAYRHEILDLNSKVGNSIPLADIRM--GTWVHDIECHPGQGAKLARAAGTYAKIIKEPAP
festuca_ncbi        PA---DLLDLNSKVGNCMPLADIRM--GTWVHSIECRPGQGAQLVRAAGTYAKVVKESAS
bromerpl2           PA---DLLDLNSKVGNCMPLADIRMETGTWVHSIELRHGQGAKLVRAAGTYAKVVKESAS
wheatrpl2_SA        PA---DLLDLNSKVGNCMPLSDIRMETGTWVHSIELRHGQGAKLVRAAGAYAKVVKESAT
ryerpl2             PA---DLLDLNSKVGNCMPLSDIRMETGTWVHSIELRHGQGAKLVRAAGAYAKVVKESAT
                    **   ::*********.:**:....  *****.** . ***.:*.****:***::**...

maizenucrpl2        KCVLRLPSGAEKAVDSRCHATIGIVSN
ricemito            QCLVRLPSGVEKFIDSRCRATIGIVSN
festuca_ncbi        QCLVRLPSGAEKLIDSRCRATIGIVSN
bromerpl2           QCLVRLPSGVEKLIDSRCRATIGIVSN
wheatrpl2_SA        QCLVRLPSGVEKLIDSRCRATIGIVSN
ryerpl2             QCLVRLPSGVEKLIDSRCRATIGIVSN
                    .********.**:*****.********
```

(b)



```
wheat           MQKFKTLAKALTGKTVTAGRNSSGRITSFHRGGGSKRSLRDIDLKRNTCSVGVVERI
barley          MQKFKTLAKALTGKTVTAGRNSSGRITSFHRGGGSKRSLRDIDLKRNNCSVGIVERI
festuca         MKVAKTLVKALTGKTVTAGRNSSGRITSFHRGGGHKRSLREIDFKRDTSSVGIVERI
rice         MRQSIKGRALRHFTLSTGKSAGRNSSGRITVFHRGGGSKRLQRKIDLKRSTSSIGIVERI
                    ::* :   ..  :********** ****** **  *.**:**...*:*:****
```

Figure 3.12 Amino acid sequence alignment of nuclear rpl5 gene in maize, barley, wheat and oats

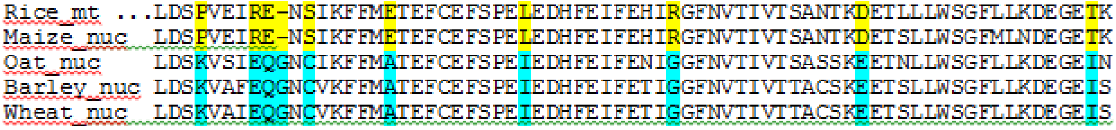
A. Amino acid sequence alignment of nuclear *rpl5* gene in maize, barley, wheat (Sandoval et al 2005) and oats highlighting the pre-sequence. Yellow highlights the different presequence in maize compared to the wheat and oats lineages (Accession numbers are provided in the supplementary table S1).

B. Amino acid alignment showing a small snapshot of core nuclear *rpl5* sequences in maize, barley, wheat (Sandoval et al 2005) and oats. Yellow highlights show amino acids shared between the maize nuclear *rpl5* gene and the mitochondrial *rpl5* gene in rice. Blue highlights show amino acids that are shared derived among oats, barley and wheat.

A.

```
maize_nuc                    MAARNILRGAACGHRFLVPVPSGALAGSGSRSIGAVAQGLRHYAADEA
barley_nuc         MSIRSLIAASRSGHALASATISQAASRAHQHAG-VPHLLSRLAPLARAFSSSPAADDAAS
wheat_nuc          MSIRSLIAASRSGHALAAATISQAASRAHQHAAGVPPLLPRLGPLARAFSSSPETADVDS
oats_nuc           MAIGSLISAARSSRLLASSAISQAS--RLQQHTTISPLLSRLGPVARAFSSAPGAADVDS

maize_nuc          VGNLELHAPKAKRRSYIPTKNGTAMMLPLHIHYEDVLRQDLLLKQNHANIMQVPGLFEVK
barley_nuc         GVPVMEEQKRRVGKKAKGSKSGKAMMFPLHFHYEDVLREDLLLKLNHTNTMEVPGLFEIR
wheat_nuc          GVPVMEVQKRRVGKKAKGAKSGKAMMFPLHFHYEDVLREDLLLKLNHTNTMEVPGLFEIR
oats_nuc           GVSVVEGQTRRVGNKAKGAKSGKAMMFPLHFHYEDVLRQDLLLKMNHTNIMEVPGLFEIR

maize_nuc          LAPKAGSDLKIPIGKMAMEVLSGQRFKEAKIDPFAKARKSSRTNPFIGADKDGSTVFAQP
barley_nuc         LVPKSTTGAKIQFGKLAMEILCGQRCIQAELPAHLKG---KATNAYLGSQKDAVSL--RQ
wheat_nuc          LVPKSTTGAKIQFGKLAMEILCGQRCIQAELPAHLKG---KATNAYLGSQKDAVSL--RQ
oats_nuc           LVPKSTSDAKIQFGKLAMEILSGQKSIQAQLPPHLKAGRSSGSNTFLASQKDATSL--RQ

maize_nuc          TVLRGHAMYNFLVRMLTVMSMLNSRADIRE-NTVKFFMETEFCEFSPELEDHFEIFEHIR
barley_nuc         SIIRGHGMYNFLVRVLTVMSMLDSKVAFEQGNCVKFFMATEFCEFSPEIEDHFEIFETIG
wheat_nuc          SIIRGHGMYNFLVRVLTVMSMLDSKVAIEQGNCVKFFMATEFCEFSPEIEDHFEIFETIG
oats_nuc           SIIRGNGMYNFLVRLLTVMSMLDSKVSIEQGNCIKFFMATEFCEFSPEIEDHFEIFENIG

maize_nuc          GFNVTIVTSANTKDETSLLWSGFMLNDEGETK
barley_nuc         GFNVTIVTTACSKEETSLLWSGFLLKDEGEIS
wheat_nuc          GFNVTIVTTACSKEETSLLWSGFLLKDEGEIS
oats_nuc           GFNVTIVTSASSKEETNLLWSGFLLKDEGEIN
```

B.



```
Rice_mt ...LDSPVEIRE-NSIKFFMETEFCEFSPELEDHFEIFEHIRGFNVTIVTSANTKDETLLLWSGFLLKDEGETK
Maize_nuc  LDSPVEIRE-NSIKFFMETEFCEFSPELEDHFEIFEHIRGFNVTIVTSANTKDETSLLWSGFMLNDEGETK
Oat_nuc    LDSKVSIEQGNCIKFFMATEFCEFSPEIEDHFEIFENIGGFNVTIVTSASSKEETNLLWSGFLLKDEGEIN
Barley_nuc LDSKVAFEQGNCVKFFMATEFCEFSPEIEDHFEIFETIGGFNVTIVTTACSKEETSLLWSGFLLKDEGEIS
Wheat_nuc  LDSKVAIEQGNCVKFFMATEFCEFSPEIEDHFEIFETIGGFNVTIVTTACSKEETSLLWSGFLLKDEGEIS
```

3.5.8 Retention of *rpl5* mitochondrial and nuclear copies can be dated back to ~25 million years

Evidence from sequence alignments (figure 3.12) suggests that the transfer of *rpl5* did not occur in the common ancestor of grasses. This is because barely, oats, and wheat have shared derived amino acid changes (blue highlights) that were acquired after transfer which are not present in maize (yellow highlights) (figure 3.12B. This indicates that the transfer event in maize must have been a lineage specific independent event. Also because maize nuclear *rpl5* gene resembles the mitochondrial copy still present in rice compared to the other grasses suggests that the transfer event in maize must have been a very recent event. Another key evidence that suggests two independent transfer events (one in maize and one in wheat lineage) is that they both acquired different mitochondrial targeting presequences (fig 3.12A). The transition stage for *rpl5* therefore does not date back to maize, as it is a separate event. Thus the transfer of the *rpl5* gene which is in transition in wheat, must have been at least ~25 million years ago.

CHAPTER 4 Comparative Analysis of nuclear *rps19* gene in brome and other grasses

This chapter focusses on the nuclear *rps19* gene in brome and other grasses to gain insight into the evolutionary events that occurred after its transfer. I will be addressing topics such as the type of presequence acquired after transfer among different grasses, copy number (to address if there has been a duplication event) and gene expression.

4.1 Wheat has three paralogous copies of *rps19* gene on chromosome 3, 5 and 1

The wheat nuclear *rps19* gene whose protein is targeted to the mitochondria has EST sequences available through NCBI. Using EST sequences from wheat, barley and maize, I designed oligomers (lb 836 and lb 837) in conserved regions (Fallahi et al 2005). From previous results in chapter 3, we know that the *rps19* gene has mitochondrial targeting presequence from heat shock protein 70, so I ensured one of my oligomers is in the core *rps19*. The Wheat Nuclear Draft Genomic sequence is available through ENSEMBL PLANTS but is incomplete (http://plants.ensembl.org/index.html). The wheat nuclear genome has three distinct copies of *rps19* gene (chr5B, chr3 and chr1As). Chromosome 3 copy is the only one that has all three homeologous copies: 3a, 3b and 3d annotated, but some of the copies such as 3a, 3d and 1as are incomplete (Appendix E). Through my experimental work, I was able to fill in the missing sequence information for chromosome 3 homeologous copies in wheat as well as show expression for both paralogous copies which are located on chromosome 3 and 5B.

The distinct copies of wheat nuclear *rps19* were obtained experimentally using oligomers lb836 and 837 via PCR at the DNA level (figure 4.1). I cloned and sequenced the three bands of the triplet separately and they agreed with the database information. The 590bp product was representative of the chromosome 5 copy and the 610bp product was representative of the chromosome 3 copy. From the database we also know there is another paralogous copy on chromosome 1 (1AS_AC7CAC6F3 which is more homologous to the chromosome 3 copy and 1AS_0706B8808 which is more homologous to the chromosome 5 copy). The size of the *rps19* gene on chromosome 1AS_AC7CAC6F3 is 641bp which is close to the 700bp band I saw on the DNA gel and the cloned sequence data for it agrees (figure 4.1b). Although the paralogous copies were easily distinguishable, it was much harder to designate the chromosome 3 homeologous copies as 3A, 3B and 3D (among my clones and ESTs). This is because there seems to be a lot of variability and they keep interchanging at different nucleotide positions of the sequence. One possible reason for this is that there could be SNPs in addition to being homeologous or exhibit “high heterozygosity” (i.e. there is variability between each seed) (Ranwez et al 2013). It is also not excluded that errors during PCR amplification such as template switching might be the reason for such occurrence (Wu et al 2006). To avoid this I tried using two times less template, but it failed to generate any product. There are ESTs available for the entire chromosome 5, 3 and 1 copies of *rps19* gene in the nucleus so we know that they are all expressed. Like my experimental results, the EST sequences also do not agree 100% with the draft genome data. This reassures that the discrepancy is most likely not due to technical errors.

In figure 4.2, nucleotide sequence alignment of chromosome 5 and chromosome 3 clones, from my experimental work is shown. In this alignment, it is shown that the difference between the two distinct copies is more evident in the intron regions, as mentioned before. The majority of nucleotide substitutions between the distinct copies are synonymous changes (figure 4.3 and 4.4). An RT-PCR experiment was carried out using wheat total RNA with lb 837, (which is specific to the chromosome 5 copy and 17/20 to chromosome 3 copy) as cDNA synthesis oligomer. The nucleotide alignment of my experimental RT-PCR product sequence, for wheat nuclear *rps19* gene, is 100% to that of the band amplified from chromosome 5 (Appendix F). This is expected because the cDNA synthesis oligomer was biased (20/20 nt) more towards this copy. Direct sequencing of the RT-PCR product also shows that the chromosome 3 copy is being expressed and is poorly amplified compared to the chr5 copy. Although the oligomers are not specific to both copies, my direct sequencing tracings confirm expression of both paralogous copies in wheat (figure 4.3b)

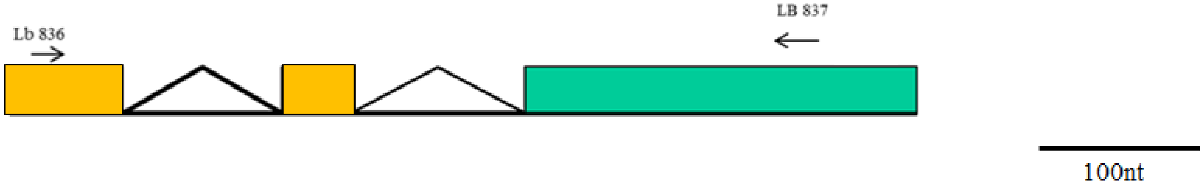
Although there were 21 differences between the wheat paralogous copies at the nucleotide level in the core region, there are only seven changes at the amino acid level. This is expected because the rate of synonymous substitution is higher than the rate of non-synonymous substitution for functional genes (Wendel 2000). Another point of interest is a variable region between the two paralogous copies (figure 4.4), which is serving as a diagnostic section to distinguish the two paralogous copies in other grasses. This variable region has four amino acid changes between the two paralogous copies: SGGS for chromosome 5 (yellow), seen in wheat and barley, whereas KREN is present for

Figure 4.1 Presence and expression of distinct rps19 copies in wheat nuclear genome

A. Schematic of nuclear *rps19* gene in wheat. Yellow squares represent the n-terminal targeting presequence (hsp70); triangles represent the introns and the green rectangle represents the core *rps19* gene.

B. Agarose gel showing PCR and RTPCR products of wheat nuclear *rps19* gene using oligomers 836 and 837 on total DNA and total RNA from etiolated seedlings respectively.

A.



B.

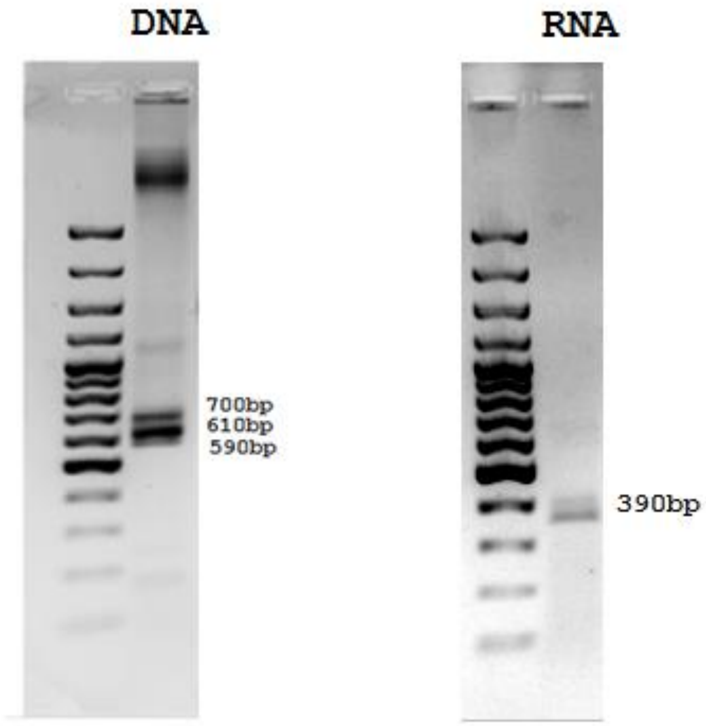


Figure 4.2 Nucleotide sequence alignment of chr5 and chr3 rps19 clones in wheat.

The flanking dinucleotides located at the intron junctions are highlighted in red (GT-AG). The differences in the intron regions between the paralogous copies from chr5 (shown in yellow) and chr3 copies (shown in green) are highlighted.

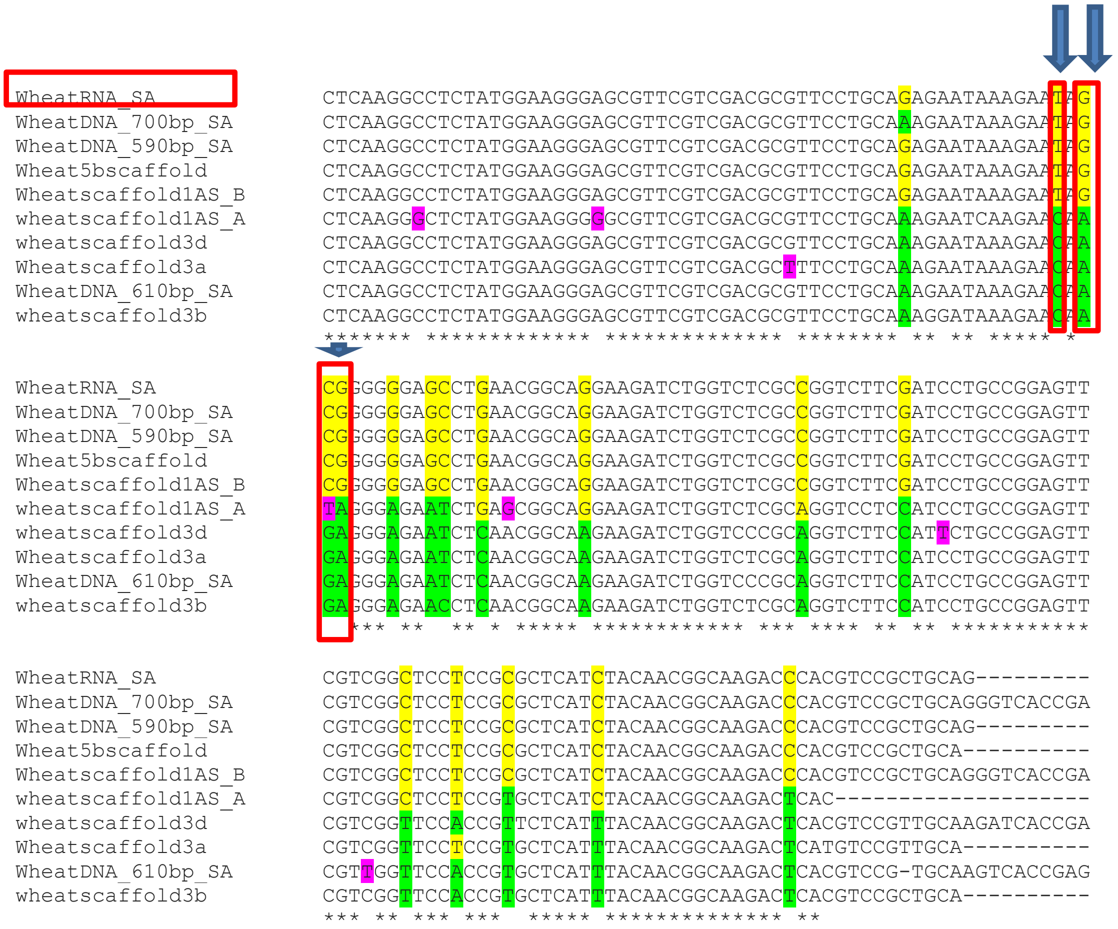
```
Wheatchr5_(590bp)DNA_SA             ------------TTGCTTCTCTCATCGCCTCCAGGTCCAGCTTCGCCAGGTGCGGCCACG
wheatchr3_copy3 (700bp)_SA          TATTGTTTTTTTTTTTATTTTCCAGGTTCTCCAGGTCCGGTCACGGGCACGCCCTCCCCC
Wheatchr3_copy1(610bpDNA_SA         --------------------------------AGGGTATGTCGTACG-----CCCTCCCG
Wheatchr3_copy2(610bp)clone_SA      ------------------------------------------------------------


Wheatchr5_(590bp)DNA_SA             CTCTCCCCGCCGCCATCTCTCAGGTGCAA-------ATCCAAATCTT-GATCCTCGCTCA
wheatchr3_copy3 (700bp)_SA          CCTCCCCCCCGGCCCTGTGTCAAGTTCCCCGTCCCCCCCCCCCCCTTCACCCTTTGTCCA
Wheatchr3_copy1(610bpDNA_SA         CCGCCGCCGCGGCCATCTCTCAGGTGCCGCCTCCCCCTCCAACTCTT-GCCCCTTGCTCA
Wheatchr3_copy2(610bp)clone_SA      ------------------------------------------------------------


Wheatchr5_(590bp)DNA_SA             AATGCAAATCTTCCTTTCCTTACCGCCGGTTGTTCTTGTGCCGATGTTGATCTTGTTTGT
wheatchr3_copy3 (700bp)_SA          AATGCAAAATTT---------------AGTTGTTTCTTGGCCGAGGGTGATTTGATATGT
Wheatchr3_copy1(610bpDNA_SA         AATGCAAATCTT---------------GGTTGTTCCTGAGCCGATGTTGATCTGATTTGT
Wheatchr3_copy2(610bp)clone_SA      ------------------------------------------GAGGTTGATTAGATTTGT
                                                                              ** * ****    * ***

Wheatchr5_(590bp)DNA_SA             TCAAATGCAAATCTTTCTACCGCGCGACC-------------------CCACCAGACCCA
wheatchr3_copy3 (700bp)_SA          T-------GGTTTGTTTCCCCGCCCGCCCGGTGCGGGGGTGGTCCGGGGGCCCGGGGCCA
Wheatchr3_copy1(610bpDNA_SA         T-------GGTTTGTTCCACCGCGCGCGCGAT------CTGGTCAGGCGCCCAGGGCCCA
Wheatchr3_copy2(610bp)clone_SA      T-------GGCTAGTTCCCCAGCGAGCGGGAT------CTGGTCAGGCGCCCAGGGCCCA
                                    *          *  **   * **  *                         *  *  ***

Wheatchr5_(590bp)DNA_SA             GCACGCCGCGTCTCCGCTGCTCTCGGGGTTCGGATCAGCGGCTC-GCGCTTTCAGGTAAC
wheatchr3_copy3 (700bp)_SA          ACCCCGCGCATTTTCGGTGCTTTTCGGCTTTGGGCCCATCCCCTGGCCCTTTCGGGTACC
Wheatchr3_copy1(610bpDNA_SA         GCACGCCGCATCTCCGCTGCTCTCCGCCTTCGGACCAGTCACTC-GCGCCTTCAGGTAAC
Wheatchr3_copy2(610bp)clone_SA      GCACGCCGCATTTCCGCTGCTCTCTGCCTTCGGACCAGTCACTT-GCGCCTTCAGGTAAC
                                     * *  *** * * ** **** *  *  ** **  *     *   ** * *** **** *

Wheatchr5_(590bp)DNA_SA             CACCATTACA---------------------GTCATCACAGGCAGTCTCGCTATGCTGCT
wheatchr3_copy33(700bp)_SA          CCCCTCCCCAAGTTCA-CTCCCTACGTACGGG----CCCAGATCA-CAAGATGATCCGCT
Wheatchr3_copy1(610bpDNA_SA         CACCATCACAAGCTCAGCTCACTACGTATGTGCAAACACAGATAA-CTAGCTGATCCGCT
Wheatchr3_copy2(610bp)clone_SA      CCCCATCACAAGCTCAGCTCACTACGTATGTGCAAACACAGATAA-CTAGCTGATCCGCT
                                    * **    **                     *    * ***     *  * *   * ***


Wheatchr5_(590bp)DNA_SA             CAAC------ATTGACTCCAATGTATG-----CCC--CCAAGCTGC---TGTCTGCGTGC
wheatchr3_copy3 (700bp)_SA          CCACATAACAATTGAATTGAATGGATGCCATACCCTGCCGTCCTGTATGTGTGTGTGTGC
Wheatchr3_copy1(610bpDNA_SA         CAACATTGAAATTGAACCGAATGTATG-----CCCTGCCCTGCTGTGTGTGTGTGTGTGC
Wheatchr3_copy2(610bp)clone_SA      CAACATTGAAATTGAACCGAATGTATG-----CCCTGCCCTCCTGTGTGTGTGTGTGTGC
                                    * **      *****    **** ***     ***  **   ***    *** ** ****

Wheatchr5_(590bp)DNA_SA             AGCTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAAT
wheatchr3_copy3 (700bp)_SA          AGCTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAAC
Wheatchr3_copy1(610bpDNA_SA         AGCTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAAAAC
Wheatchr3_copy2(610bp)clone_SA      AGCTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAAC
                                    *********************************************** ******** *


Wheatchr5_(590bp)DNA_SA             AGCGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAG
wheatchr3_copy3 (700bp)_SA          AAGAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATCCTGCCGGAG
Wheatchr3_copy1(610bpDNA_SA         AAGAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATCCTGCCGGAG
Wheatchr3_copy2(610bp)clone_SA      AAGAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATCCTGCCGGAG
                                    *   *** **  ** ******* ************ *** ******* ************

Wheatchr5_(590bp)DNA_SA             TTCGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG-------
wheatchr3_copy3 (700bp)_SA          TTCGTCGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCGCTGCAAGATCACC
Wheatchr3_copy1(610bpDNA_SA         TTCGTTGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACC
Wheatchr3_copy2(610bp)clone_SA      TTCGTTGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCG-TGCAAG-TCACC
                                    ***** ** *** *** ****** ************** ******** ****
```

Figure 4.3 Evidence for expression of both paralogous rps19 gene copies in wheat

(a) Nucleotide alignment of my experimental sequences for wheat DNA and RNA copies as well as wheat DNA scaffolds sequences from the ensemble website. The red box highlights the RNA sequence of nuclear *rps19*. Pink highlights show the positions where nuclear *rps19* has a mismatch compared to others. The two distinct copies of *rps19* are highlighted in yellow and green. Blue arrows show the positions in the sequence alignment which indicate the presence of two distinct *rps19* copies.

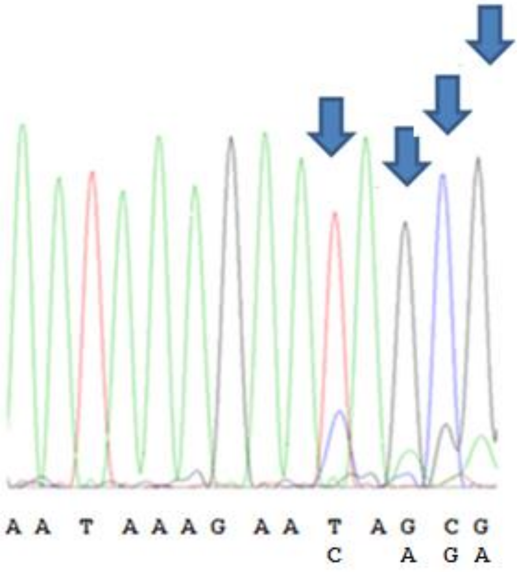
(b) Direct sequencing of wheat RT-PCR product which contains major and minor sequences representing the expression of both paralogous copies. Blue arrows show the positions in the tracing where it is clear that two distinct copies are expressed.

(a)



```
WheatRNA_SA             CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
WheatDNA_700bp_SA       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAATAG
WheatDNA_590bp_SA       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
Wheat5bscaffold         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
Wheatscaffold1AS_B      CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
wheatscaffold1AS_A      CTCAAGGGCTCTATGGAAGGGGCGTTCGTCGACGCGTTCCTGCAAAGAATCAAGAACAA
wheatscaffold3d         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAACAA
Wheatscaffold3a         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAACAA
WheatDNA_610bp_SA       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAACAA
wheatscaffold3b         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGGATAAAGAACAA
                        ******* ************* ************** ******** ** ** ***** *

WheatRNA_SA             CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
WheatDNA_700bp_SA       CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
WheatDNA_590bp_SA       CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
Wheat5bscaffold         CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
Wheatscaffold1AS_B      CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
wheatscaffold1AS_A      TAGGGAGAATCTGAGCGGCAGGAAGATCTGGTCTCGCAGGTCCTCCATCCTGCCGGAGTT
wheatscaffold3d         GAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATTCTGCCGGAGTT
Wheatscaffold3a         GAGGGAGAATCTCAACGGCAAGAAGATCTGGTCTCGCAGGTCTTCCATCCTGCCGGAGTT
WheatDNA_610bp_SA       GAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATCCTGCCGGAGTT
wheatscaffold3b         GAGGGAGAACCTCAACGGCAAGAAGATCTGGTCTCGCAGGTCTTCCATCCTGCCGGAGTT
                          *** **  ** * ***** ************ *** **** ** ** ***********

WheatRNA_SA             CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG---------
WheatDNA_700bp_SA       CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAGGGTCACCGA
WheatDNA_590bp_SA       CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG---------
Wheat5bscaffold         CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCA----------
Wheatscaffold1AS_B      CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAGGGTCACCGA
wheatscaffold1AS_A      CGTCGGCTCCTCCGTGCTCATCTACAACGGCAAGACTCAC--------------------
wheatscaffold3d         CGTCGGTTCCACCGTTCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGA
Wheatscaffold3a         CGTCGGTTCCTCCGTGCTCATTTACAACGGCAAGACTCATGTCCGTTGCA----------
WheatDNA_610bp_SA       CGTTGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCG-TGCAAGTCACCGAG
wheatscaffold3b         CGTCGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCGCTGCA----------
                        *** ** *** ***  ***** ************** **
```

(b)



chromosome 3 homeologous copies (green). The wheat 1AS_A is more homologous to the chromosome 3 copy but it varies from the others as it has NREN rather than KREN, indicating that it might be evolving much faster (blue highlights in figure 4.4). Wheat 1as_b, on the other hand, is more similar to the chromosome 5 copy. It is evident that there has been a duplication event for *rps19* gene in wheat lineage after its transfer to the nucleus. This raises interesting questions as to whether other closely related grasses also have these paralogous copies present, and if they are both functional as in wheat. To address this, bioinformatic analysis was conducted for barley nuclear *rps19* gene using EST library and the high throughput genomic database information.

4.2 Barley nuclear genome consists of chr5 type copy in wheat as well as nuc-mt *rps19* gene

EST alignments of barley *rps19* gene with different cultivars revealed the presence of only one copy. Alignment of barley sequence with other wheat copies that I experimentally sequenced, as well as the scaffolds show that it is more like the chromosome 5 copy than the chromosome 3 one (figure 4.5). This is an interesting observation because it gives rise to two scenarios: the duplication event must have occurred in the ancestor of all grasses and barley lost the chromosome 3 copy or the duplication event only occurred in the wheat lineage. To address this, more grasses need to be surveyed. Thus, oligomers that are specific to the two distinct copies seen in wheat were designed to test on brome.

The barley draft nuclear genome in the Ensembl website was found to have a copy of *rps19* which excluded the hsp70 presequence. This is referred to as nuc-mt as it is a mitochondrial *rps19* gene with *rpl2* upstream of it located in the nuclear genome of barley. This nuc-mt *rps19* gene in barley is missing the edits (S>L and F>H), suggesting that it must have been the result of a recent, independent, DNA mediated transfer of *rps19* to the nucleus (Figure 4.6). This was confirmed experimentally by using total barley DNA with mitochondrial *rps19* oligomers designed on rice. Sequencing data from the product showed an unedited *rps19* gene product confirming this finding. This suggests that there must have been a transition stage of ~60 million years for *rps19* in barley and its nuclear copy must have won out recently. This also cannot be from contaminating mitochondrial DNA in nuclear genome as prior southern analysis data by Magid Fallahi (Fallahi thesis 2000) using a long PCR product as a probe shows absence of *rps19* gene in the mitochondrial genome of barley (Figure 4.6b). Similarly I was able to support Magid's data that barley mitochondrial genome is missing *rps19* gene via a single oligomer probe (lb 188) in Southern analysis. (figure 4.6c). Presence of nuc-mt sequences in the nuclear genome is fairly common in plants and has been reported in various angiosperms (Pont et al 2012, Takenaka et al 2013).

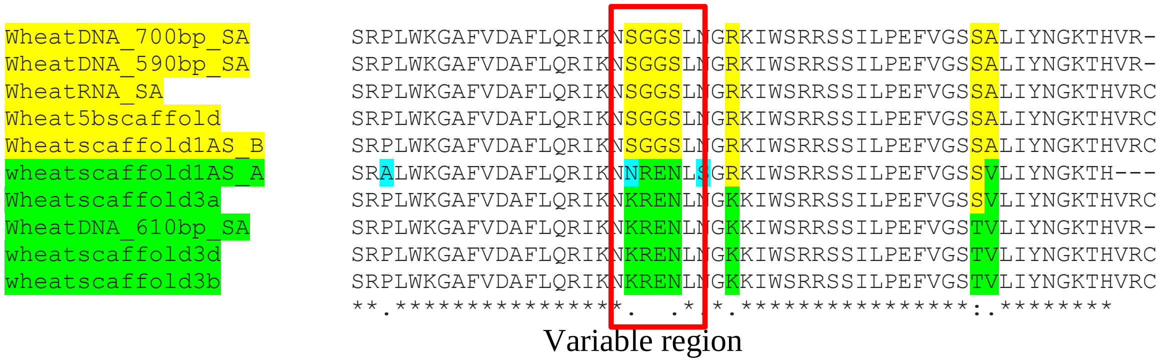
4.3 Brome nuclear genome only has a single wheat-chromosome 3-type *rps19* gene copy

As discussed in chapter 3, brome has functional copies of *rps19* gene in both mitochondria and the nucleus. The brome nuclear rps19 copy was more similar to the wheat chr3 copy. Oligomers based on wheat chr5 copy were used to see if brome also has one in its genome but the experiment was unsuccessful (data not shown). This does not completely rule out the possibility of brome having a chromosome 5 type paralogous copy as its sequence might be a bit different, since brome and wheat are ~20 million years apart while barley and wheat are only ~15 million years apart. Also the oligomer pair 836-837 which gives both chr5 and chr3copies in wheat, gave only one clean sequence in both brome and barley, while the brome sequence was more similar to chr3 type copy, barely was more similar to the chr5 copy (figure 4.7). In figure 4.7, tracings of the variable region are shown to highlight the two distinct copies in wheat, which are absent in brome and barley.

The brome nuclear *rps19* gene is homologous to the chromosome 3 type copy present in wheat (figure 4.8b). This copy was found to be expressed at the RNA level as shown in figure 4.8a (refer to gel in chapter 3, figure 3.2). Direct sequencing tracings of the DNA and RNA products in the core *rps19* region showed three positions with two peaks. The proportion of the peaks varied between positions. Since both forms are seen in the RNA tracings, we know that both forms are being expressed (figure 4.9a). Both forms have been seen in different clones of DNA and RNA products. Two out of the three changes are non-synonymous (figure 4.9b) which is unexpected as usually the rate of synonymous changes is higher as seen among the homeologous copies in wheat. This suggests that the other form of *rps19* we see might be paralogous rather than homoeologous. More than seven different clones were screened at the DNA and RNA level which showed some unique amino acid changes in one or two clones randomly which result in non-synonymous changes. As these are only seen in one clone or two,

Figure 4.4 Amino acid alignment of wheat paralogous copies (from my experimental work at the DNA and RNA level) along with the draft genome sequences

The two distinct copies of *rps19* are highlighted in yellow and green. Red box highlights the variable region among the two distinct copies which is acting as the signature region.



```
WheatDNA_700bp_SA        SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVR-
WheatDNA_590bp_SA        SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVR-
WheatRNA_SA              SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
Wheat5bscaffold          SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
Wheatscaffold1AS_B       SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
wheatscaffold1AS_A       SRALWKGAFVDAFLQRIKNNRENLSGRKIWSRRSSILPEFVGSSVLIYNGKTH---
Wheatscaffold3a          SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRC
WheatDNA_610bp_SA        SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVR-
wheatscaffold3d          SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRC
wheatscaffold3b          SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRC
                         **.****************.  .*.*.****************:.********
```

Figure4.5 Nucleotide and amino acid sequence alignments of nuclear rps19 gene in barley with wheat paralogs.

(A) Nucleotide sequence alignment of core *rps19* gene region with barley, wheat experimental data (PCR and RT-PCR) as well as the scaffolds from the Ensembl website. Yellow highlights represent the chromosome 5 copy and green represents chromosome 3 homeologous copies (3A, 3B and 3D). Stars show the conserved nucleotide positions among the two distinct copies. Red highlight shows lineage specific changes in barely nuclear *rps19* gene. Blue highlights show the distinct amino acid positions that are unique to wheat scaffold 3a and 3d

(B) Amino acid sequence alignment of nuclear *rps19* gene core sequence between barley and wheat. Yellow highlights represent the chromosome 5 copy and green represents chromosome 3 homeologous copies. Stars show the conserved nucleotide positions among the two distinct copies.

A.

```
Barley htgs             CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCCTTCCTGCAGAGAATAAAGAAGAG
WheatDNA_700bp_SA       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAATAG
WheatDNA_590bp_SA       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
WheatRNA_SA             CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
Wheat5bscaffold         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
wheatscaffold3d         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAACAA
Wheatscaffold3a         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAACAA
WheatDNA_610bp_SA       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGAATAAAGAACAA
wheatscaffold3b         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGGATAAAGAACAA
                        ******* ************* ************** ******** ** ** ***** *

Barley htgs             CGGGGAGAGCCTGAACGGCAGGAAGATCTGGTCTCGCAGGTCTTCGATCCTGCCGGAGTT
WheatDNA_700bp_SA       CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
WheatDNA_590bp_SA       CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
WheatRNA_SA             CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
Wheat5bscaffold         CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
wheatscaffold3d         GAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATTCTGCCGGAGTT
Wheatscaffold3a         GAGGGAGAATCTCAACGGCAAGAAGATCTGGTCTCGCAGGTCTTCCATCCTGCCGGAGTT
WheatDNA_610bp_SA       GAGGGAGAATCTCAACGGCAAGAAGATCTGGTCCCGCAGGTCTTCCATCCTGCCGGAGTT
wheatscaffold3b         GAGGGAGAACCTCAACGGCAAGAAGATCTGGTCTCGCAGGTCTTCCATCCTGCCGGAGTT
                          *** **  ** * ***** ************ *** **** ** ** ***********

Barley_htgs             CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGTTGCAGGATCACCGA
WheatDNA_700bp_SA       CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAGGGTCACCGA
WheatDNA_590bp_SA       CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG---------
WheatRNA_SA             CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG---------
Wheat5bscaffold         CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCA----------
wheatscaffold3d         CGTCGGTTCCACCGTTCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGA
Wheatscaffold3a         CGTCGGTTCCTCCGTGCTCATTTACAACGGCAAGACTCATGTCCGTTGCA----------
WheatDNA_610bp_SA       CGTTGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCG-TGCAAGTCACCGAG
wheatscaffold3b         CGTCGGTTCCACCGTGCTCATTTACAACGGCAAGACTCACGTCCGCTGCA----------
                        *** ** *** ***  ***** ************** **
```

B.

```
Barley_htgs             SRPLWKGAFVDAFLQRIKKSGESLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
WheatDNA_700bp_SA       SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVR-
WheatDNA_590bp_SA       SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVR-
WheatRNA_SA             SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
Wheat5bscaffold         SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
Wheatscaffold3a         SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRC
WheatDNA_610bp_SA       SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVR-
wheatscaffold3d         SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRC
wheatscaffold3b         SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRC
                        **.***************:.  .*.*.****************:.********
```

Figure4.6 Barley nuclear genome has nu-mt rps19 gene suggesting a recent independent DNA mediated transfer.

A. Agarose gel showing the PCR product generated with mitochondrial *rps19* gene oligomers lb 834 and lb 196 (designed from rice). Sequence alignment shows comparison between the mitochondrial *rps19* gene in rice relative to the two nuclear-located *rps19* genes in barley. Black highlighting with white letters shows the *hsp70* presequence of one of the nuclear located *rps19* gene and is missing in the nuc-mt suggesting that it is not from the same common transfer event. Blue boxes point towards the unedited sites in barley nuc-mt (highlighted in red) showing that the transfer occurred as a DNA intermediate.

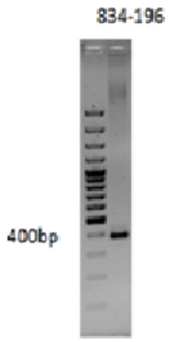
B. Barley mitochondrial genome is missing *rps19* gene. This has been shown by Southern analysis using ^{32}P-labelled PCR product shown by parenthesis in the schematic (figure 4.6D) previously by Majid Fallahi (Fallahi thesis 2000). The fluorescence profile of blot MS23 is shown to the very left. The center blot shows that barley is missing *rps19* in the mitochondria. Hybridization in the wheat lane shows that the probe is good. Also the same blot reprobed with *nad4L* shows hybridization for barley again showing evidence for the absences of *rps19* gene in barley mitochondrial genome.

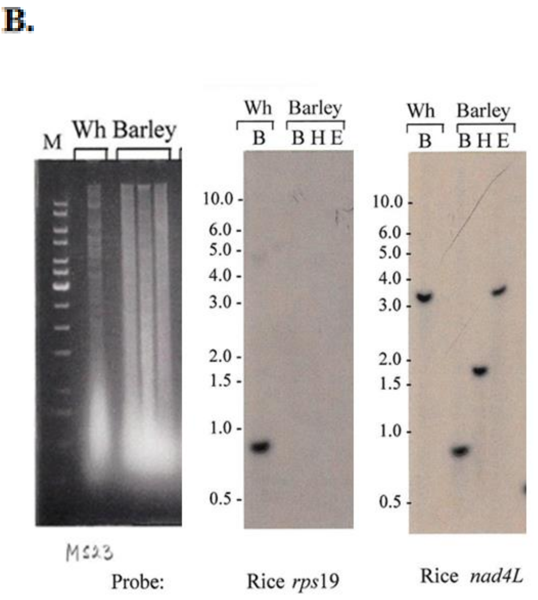
C. Barley mitochondrial genome using single *rps19* probe lb 188 also showed similar results as the long PCR product probe used by Magid. Blot. Fluorescence profile for brome and barley mitochondrial genome digested with HindIII and BamHI is shown in the far left. The middle panel shows that the absence of *rps19* in barley but presence of it in brome which shows that there is no problem with the probe lb 188. The same blot has been reprobed by *rps13* by Bronwyn as a positive control which shows the blot BAS2 has good barley DNA.

D. Schematic of *rps19* gene showing the long PCR product used by Fallahi with a parenthesis bar and the single oligomer probe used in C .

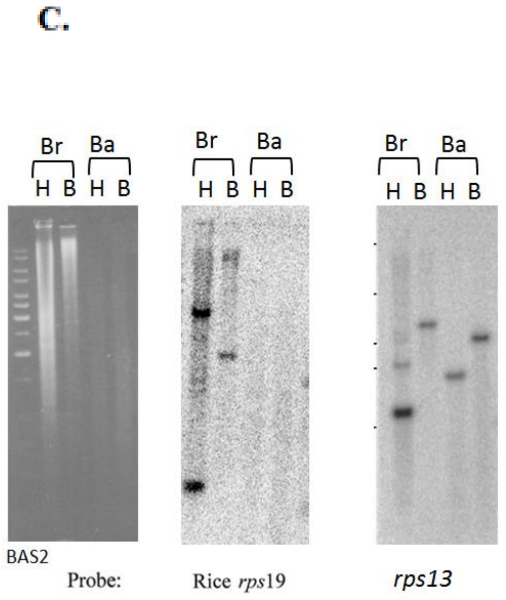
A.

```
barleynuc                 MSSIASLIASRSSFARSGHALPAAAAAAAISQAQHAASPLLSGFGSAARAFSSRPLWKGA
barleynuc-mt                                                                MPRRSIWKGS
bromemit                                                                    MPRRSIWKGS
ricemit                                                                     MPRRSIWKGS
                                                                            :.  *.:***:

barleynuc                 FVDAFLQRIKKSGESLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRCRITEGKVGHKFG
barleynuc-mt              FVNPFQFRIKKNRESLMSRKIWSRRSSISPEFVDCSVLIYNGKTPVRCKITKGKVGHKFG
bromemit                  FVDAFLFRIKKNRESLMSRKIWSRRSSILPEFVDCSVLIYNGKTFVRCKITEGKVGHKFG
ricemit                   FVDAFLFRIKKNRESLMSRKIWSRRSSILPEFVDCSVLIYNGKTFVRCKITEGKVGHKFG
                          **:.*  ****. *** .********** ****..*.******* ***.**:********

barleynuc                 EFAFTRKRRPHRAITAKKAGGQGKGKKK
barleynuc-mt              EFSFTRRRRPYRT----NRG-------K
bromemit                  EFAFTRRRRPYRT----NRG---KGKKK
ricemit                   EFAFTRRRRPYQT----NRGKGRKGKKK
                          **:***.***:.:    : *       *
```





Fallahi Thesis 2000



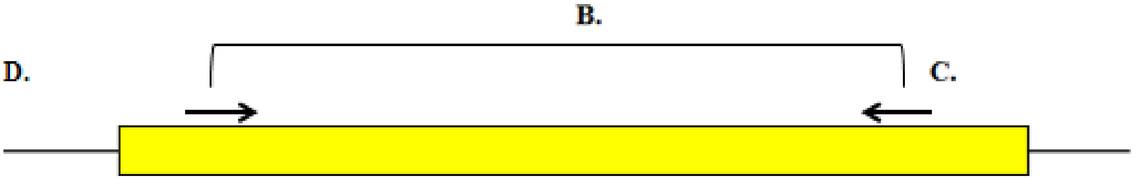
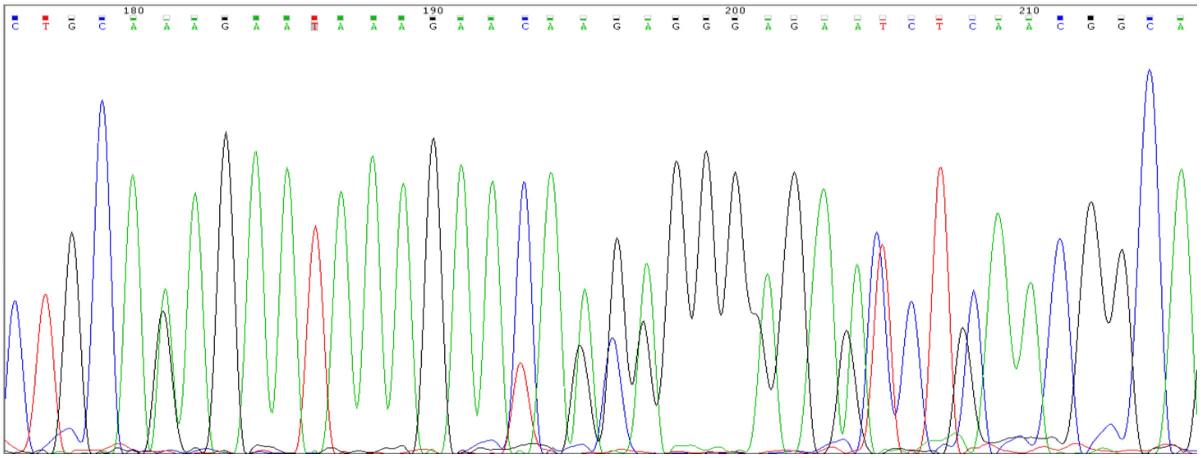


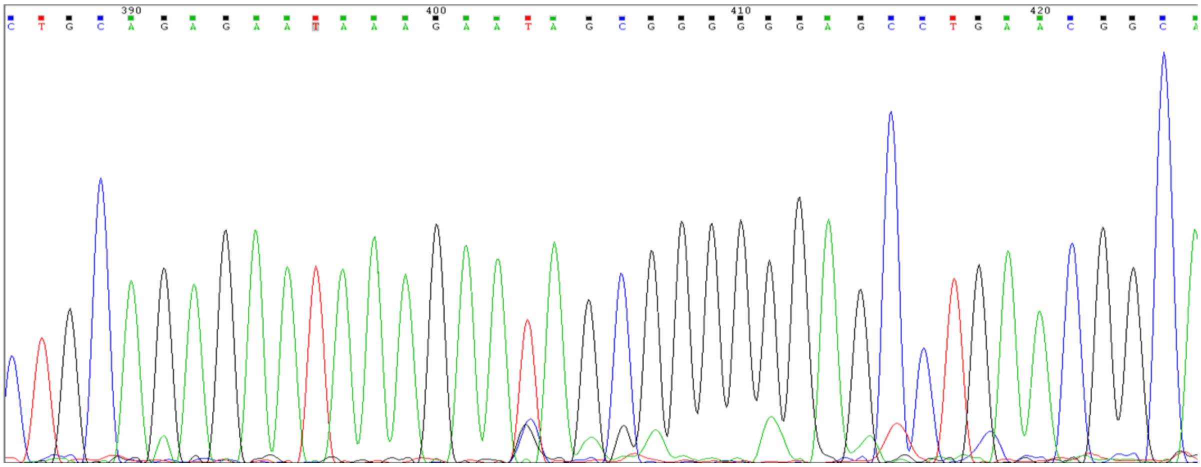
Figure 4.7 Brome is missing chromosome 5 type rps19 copy seen in wheat

Direct sequencing tracings of the diagnostic variable region is shown in wheat, barley and brome. Wheat shows presence of two copies (chr5 and chr3) while barley and brome show one (chr5) and (chr3) respectively.

Wheat 836-837



Brome 836-837



Barley836-837

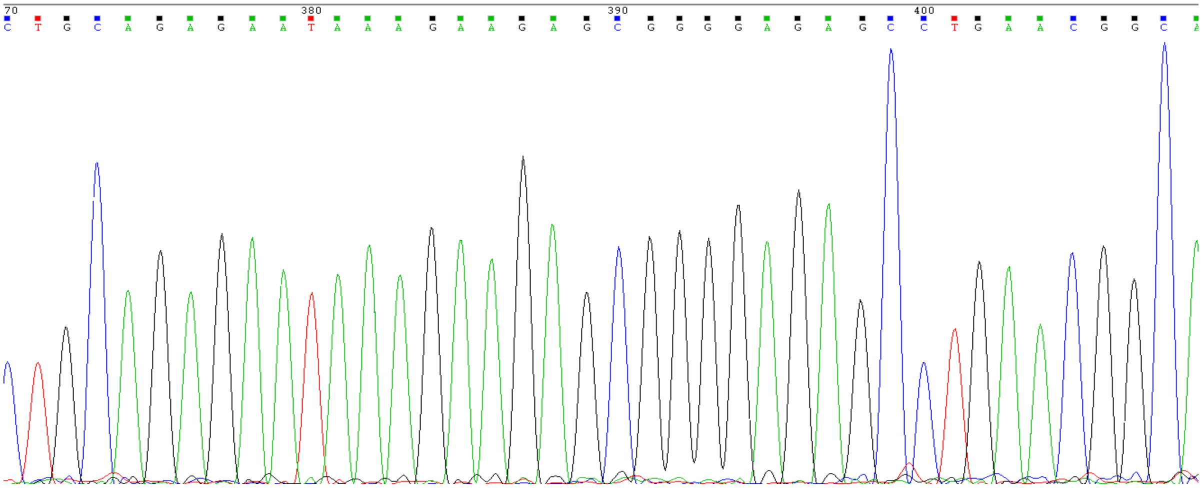


Figure 4.8: Brome nuclear genome has a single chromosome-3-type copy found in wheat

(a) Nucleotide alignment of brome nuclear *rps19* at the DNA and cDNA level.

(b) Amino acid alignment of brome nuclear *rps19* at the DNA and RNA level with other grasses such as barley, and wheat (chr5 and chr3). The diagnostic variable region is highlighted: wheat chr5 and wheat chr5-type-copy in barley as yellow and wheat chr3 copy in wheat and chr3-type-copy in brome as green

(a)

```
br862-904_dna2                  ATGTCGATTGGCTCTCTCATCGCCTCCAGGTTCGCCAGGTCCGGCCACGCCCTCGCCGCC
br862-923cdna_c3dec12th         ATGTCGATTGGCTCTCTCATCGCCTCCAGGTTCGCCAGGTCCGGCCACGCCCTCCCCGCC
                                ****************************************************** *****

br862-904_dna2                  GCCGCCGCCGCCGCCGCCATCTCTCAGGTCCCCCCCTCCCCTCCAACTCTTGCTCCTTG
br862-923cdna_c3dec12th         GCCGCC------------ATCTATCAG---------------------------------
                                ******            **** ****

br862-904_dna2                  CTCAAATGCAAATCTTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCGC
br862-923cdna_c3dec12th         ------------------------------------------------------------


br862-904_dna2                  GGTCTGGTCTGGTCAGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTT
br862-923cdna_c3dec12th         ----------------GCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTT
                                                ********************************************

br862-904_dna2                  CGGAGCAGTGACTCGTGCCTTCAGGTAACTACCATTACAGTCCACCTCTCGCATTCCTTG
br862-923cdna_c3dec12th         CGGAGCAGTGACTCGTGCCTTCAG------------------------------------
                                ************************

br862-904_dna2                  CTCAACATTGAATCGAATGTATGCCCCAAGCTGCTGCTGTGTGCAGCTCAAGGCCTCTAT
br862-923cdna_c3dec12th         ----------------------------------------------CTCAAGGCCTCTAT
                                                                              **************

br862-904_dna2                  GGAAGGGAGCGTTCGTCGACGTTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
br862-923cdna_c3dec12th         GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
                                ********************* **************************************

br862-904_dna2                  ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
br862-923cdna_c3dec12th         ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
                                ************************************************************

br862-904_dna2                  TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
br862-923cdna_c3dec12th         TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
                                ************************************************************

br862-904_dna2                  ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
br862-923cdna_c3dec12th         ATAAGTTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
                                ***** ******************************************************

br862-904_dna2                  AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAA
br862-923cdna_c3dec12th         AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAA
                                ********************************
```

(b)

```
wheatgreen                    -MAIASLIASR--FARSGHGHALPAAAAAISQAPRAQHAASPLLSGFGPLTRAFSSRPLW
bromeDNA                      -MSIGSLIASR--FARSGHALP----AAAIYQAPRAQHAASPLLWGFGAVTRAFSSRPLW
bromeRNA                      -MSIGSLIASR--FARSGHALAAAAAAAAISQAPRAQHAASPLLWGFGAVTRAFSSRPLW
wheatyellow                   --MIASLIASRSSFARCGHALPA-----AISQ---TQHAASPLLSGFGSAARAFSSRPLW
barley                        MSSIASLIASRSSFARSGHALPAAAAAAAISQ---AQHAASPLLSGFGSAARAFSSRPLW
                                 *..*****  ***.**. .      ** *   :******** ***. :*********

wheatgreen                    KGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRCKITEGKVGH
bromeDNA                      KGAFVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGH
bromeRNA                      KGAFVDVFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGH
wheatyellow                   KGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRCRVTEGKVGH
barley                        KGAFVDAFLQRIKKSGESLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRCRITEGKVGH
                              .*****.******:.  .***.****************:.******:****.::******

wheatgreen                    KLGEFAFTRRRRPHRAILAK--GGQGKGKKKK
bromeDNA                      KFGEFAFTRRRRPHRAITGK--AGQGKGKKKK
bromeRNA                      KLGEFAFTRRRRPHRAITGK--AGQGKGKKKK
wheatyellow                   KFGEFAFTRKRRPHRAITAKKAGGQGKGRKK-
barley                        KFGEFAFTRKRRPHRAITAKKAGGQGKGKKK-
                             *:*******.* ****  .*  ..****.**
```

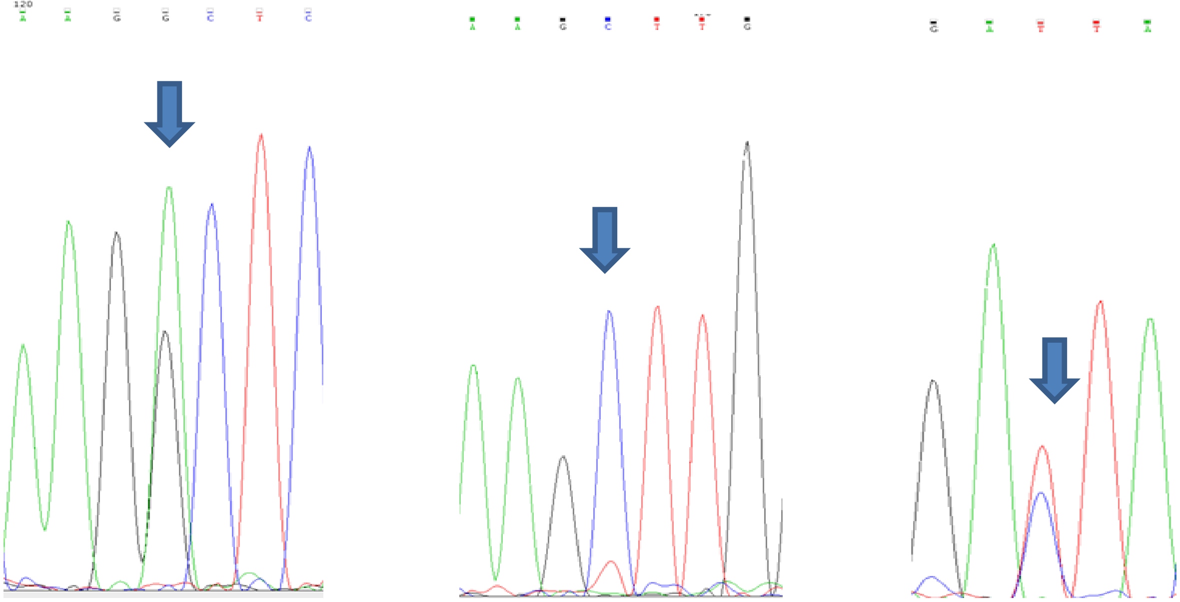
Figure 4.9 Direct sequencing tracings of brome rps19 gene at the DNA and RNA level to show presence and expression of two different forms.

(a) Blue arrows show the corresponding position in the tracing at the DNA and RNA level which shows two peaks. These three positions are from the core *rps19* region and the sequencing primer used was lb 951. The two forms of *rps19* at these three positions are present in both DNA and RNA which indicates that they are both functional

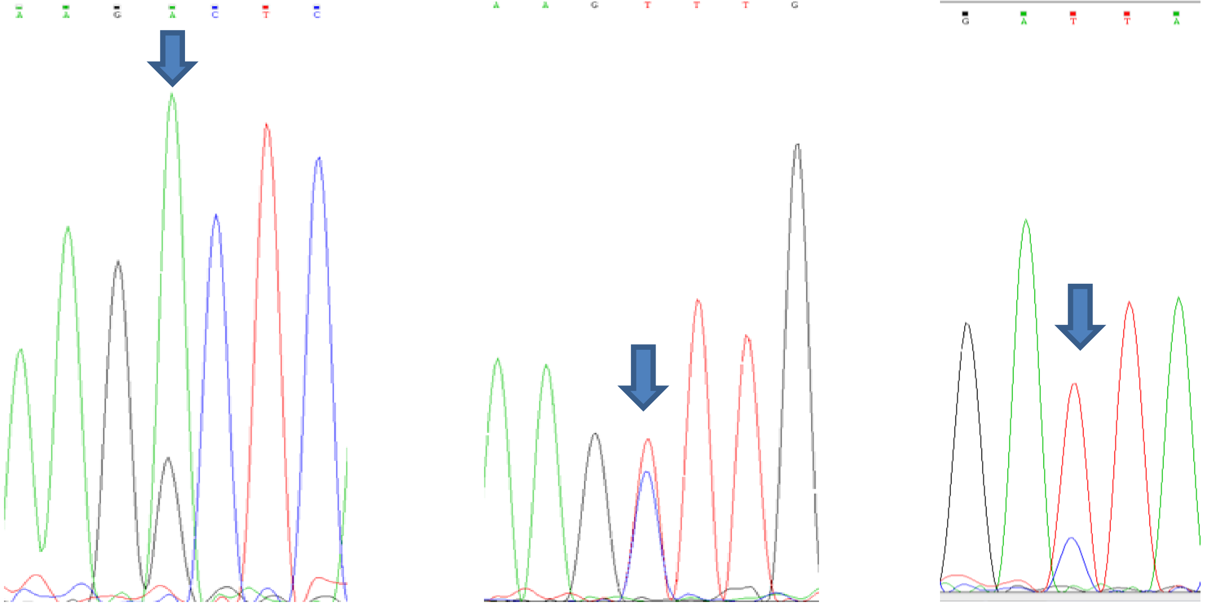
(b) Amino acid alignment of brome DNA and RNA clones show the two different forms of *rps19* (highlighted in green and blue. Red boxes show the amino acid position where the two forms show non-synonymous change.

(a)

DNA (Br951-952)



RNA (Br862-952)



(b)

```
BR862-904_dna_clone4dec         MSIGPLIASRFARSGHALPA----TAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
862-904_c2dnaoct                MSIGSLIASRFARSGHALAAAAAAAAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
brome862-904Clone6jan           MSIGSLIASRFARSGHALPAAAAATAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
brome862-904clone1dnaoct        MSIGPLIASRFARSGHALPAAAAATAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
brome862-904Clone5jan           MSIGPLIASRFARSGHALPAAAAATAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
br862-923cdna_c1                MSIGSLIASRFARSGHALPAAA--TAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
BR862-904_dna_clone3dec         MSIGSLIASRFARSGHALPAA---TAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
brome862-904Clone7jan           MSIGSLIASRFARSGHALPAAA--TAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWRGA
br862-923cdna_c3dec12th         MSIGSLIASRFARSGHALPAA----AIYQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
                                ****.*************.*     ** *****************************.**

BR862-904_dna_clone4dec         FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKIAEGKVGHKF
862-904_c2dnaoct                FVDVFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKL
brome862-904Clone6jan           FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKIAEGKVGHKL
brome862-904clone1dnaoct        FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKL
brome862-904Clone5jan           FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKL
br862-923cdna_c1                FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKAHVRCKITEGKVGHKF
BR862-904_dna_clone3dec         FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKAHVRCKITEGKVGHKL
brome862-904Clone7jan           FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKL
br862-923cdna_c3dec12th         FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKF
                                ***.***************************************:******:*******:

BR862-904_dna_clone4dec         EFAFTRRRRPHRAITGKAGQGKGKKKK
862-904_c2dnaoct                EFAFTRRRRPHRAITGKAGQGKGKKKK
brome862-904Clone6jan           EFAFTRRRGPHRATTGKAAQGKGKKKK
brome862-904clone1dnaoct        EFAFTRRRRPHRATTGKAGQGKGKKK-
brome862-904Clone5jan           EFAFTRRRRPHRATTGKAAQGKGKKKK
br862-923cdna_c1                EFAFTRRRRPHRATTGKAGQGKGKKKK
BR862-904_dna_clone3dec         EFAFTRRRRPHRAITGKAGQGKGKKKK
brome862-904Clone7jan           EFAFTRRRRPHRAITGKAGQGKGKKKK
br862-923cdna_c3dec12th         EFAFTRRRRPHRAITGKAGQGKGKKKK
                                ******** **** ****.*******
```

they might be due to mutations during cloning or UV damage (Appendix H a and b).

4.4 Oats has a single copy of *rps19* gene and a novel presequence

Oats has one EST sequence available in the NCBI database while lolium has two and festuca has three. Using these EST sequences, primers were designed to obtain the nuclear *rps19* gene in oats. EST sequences and my experimental data showed that oats, lolium, and festuca still retained half of the hsp70 presequence which is present in other grasses such as wheat, brome and maize but acquired a novel piece in front (Appendix J a, b and c). The first intron, which is present in this novel presequence of oats, is much bigger than the intron 1 present in hsp70 presequence. Even though the transfer of nuclear *rps19* gene occurred in the common ancestor of all the grasses, the oats clade seems to have undergone lineage specific rearrangements to its presequence over time.

Direct sequencing and clones of my oats PCR and RT-PCR products showed presence of homeologous copies rather than paralogous copies (figure 4.8a). This might not be surprising as in wheat and brome we saw that oligomer bias can play a huge role in which copies get amplified. At the nucleotide and amino acid level, oats nuclear *rps19* gene copy was more similar to the chromosome 3 copy present in wheat and brome (figure 4.8b). The diagnostic region in oats was not really helpful in distinguishing as it had some changes that agree with one or the other and some that are lineage specific (figure 4.9a). To test the possibility that there might be paralogous copies in oats as in wheat, comparative analysis was done with other close relatives of this clade. EST sequence alignments for oats, lolium, festuca and dactylis support the idea that there is only a single copy in oats (figure 4.9b). The amino acid alignment (figure 4.9b) shows that in the database there is evidence for only a single copy of *rps19* gene being present in oats, festuca, lolium and dactylis. However, it could be the case that other copies are not expressed and therefore not present in the ESTs.

4.5 3’UTR exon for nuclear *rps19* is conserved in wheat and barley but not in oats

From the wheat draft genome website and barley high through put genomic sequence, the 3’UTR sequences for the *rps19* gene are known. Wheat and barley nuclear *rps19* genes have a 3’UTR exon that is conserved. As showing brome nuclear *rps19* gene expression was my primary research objective, it is interesting to see if brome also has this UTR exon and if it is conserved. If brome and oats (distant relative of wheat clade) also have this UTR exon, it would suggest that it might have a significant role in RNA processing. However, my 3'RACE attempts remained unsuccessful for brome, but worked for oats *rps19*. A 750 bp RT-PCR product was obtained for oats using lb 860 located within core *rps19* region and oligo(dT) primer. The same experiment was unsuccessful using brome total RNA which could be due to very big 3'UTR. Sequence data from oats oligo (dT) showed that the 3'UTR is different in oats relative to wheat and barley (Appendix k). This could be due to the fact that wheat and barley are more closely related to each other than oats. However the 3'UTR coding sequence of *rps19* gene in oats was similar to its close relatives such as festuca and oats whose ESTs are available in NCBI.

4.6 Maize also has three copies of *rps19* gene

Using barley high throughput genomic sequence as query, I was able to find maize *rps19* gene sequences in the high throughput genomic database. (Appendix L). Two paralogous *rps19* copies were found on chromosome 5 and chromosome 1 in maize (figure 4.10a). These two distinct copies are similar to each other as there are only 19 nucleotide substitutions between them. This indicates that either the duplication event in maize was fairly recent or it occurred in the common ancestor and the paralogs underwent gene conversion with one another. BLAST revealed another *rps19* gene hit for maize on chromosome 1 that is shown in green in the above figure and did not contain introns at the DNA level (figure 4.10a). I was able to find expressed sequences for this in the EST library. This indicates that maize contains a retroprocessed *rps19* gene in the nucleus. Also maize is very distant from wheat so it is not evident if these paralogous copies in maize are the same as those present in wheat.

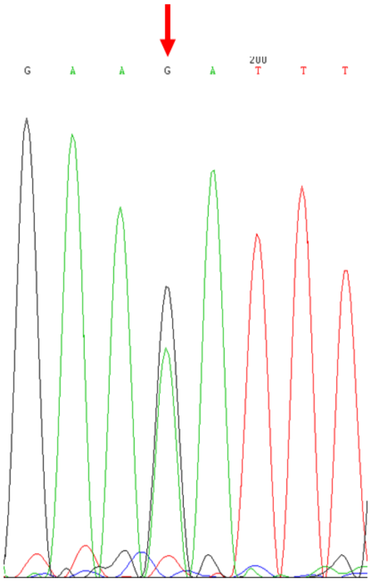
The amino acid alignment comparison of maize *rps19* gene shows some amino acid changes that are not present in wheat copies. This is because maize diverged 60 million years ago; therefore, these are lineage specific changes. At the amino acid level, maize nuclear *rps19* looks more like the wheat chromosome 3 copy, than the chromosome 5 copy (fig 4.10b). Similar phenomenon has also been seen for oats nuclear *rps19* gene copy. It is also true that the chromosome 3 copy is more like the ancestral mitochondrial *rps19* gene. Based on these observations it can be speculated that the chromosome 3 copy is the ancestral *rps19* gene that was transferred to the nucleus in grasses and chromosome 5 copy is the duplicated gene.

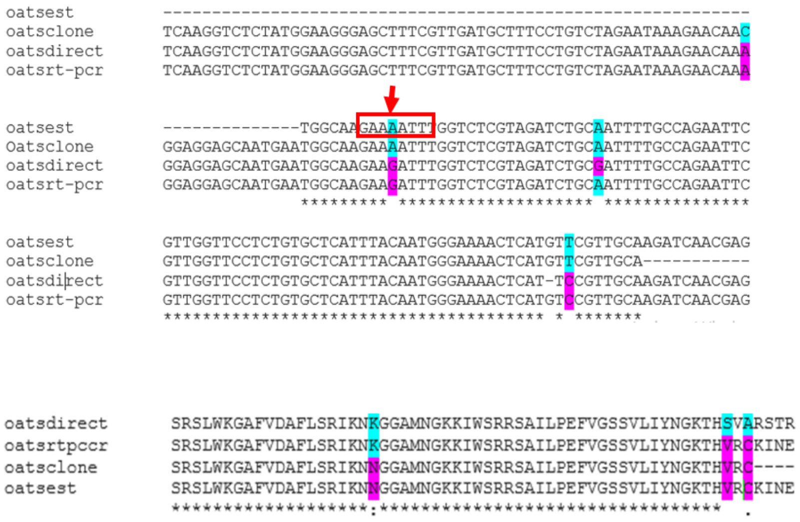
4.7 Codon usage analysis of mitochondrial *rps19* versus nuclear *rps19* gene in brome

As brome has functional *rps19* genes in both the mitochondrion and nucleus, it was interesting to compare the codon usage pattern between the two cell environments. Using the CAIcal database < http://genomes.urv.cat/CAIcal/> codon usage was calculated for nuclear genes and mitochondrial genes in brome (Appendix M), mitochondrial *rps19* gene in brome, nuclear *rps19* gene in brome with and without presequence hsp70, nuclear *rps19* gene in wheat with and without presequence hsp70 and rice mitochondrial *rps19* gene. Mitochondrial genes in general for brome were found to have codon bias for AT compared to nuclear genes which have higher GC%. This has been observed for other plants, such as wheat by Selvi Subramanian.

Figure 4.10 Evidence for nuclear rps19 gene expression in oats

Blue and purple highlights show nucleotide and amino acid positions where we see differences. The red arrow on the sequence tracing where we see two peaks is highlighted on the sequence alignment using a red box.





```
oatsest      ------------------------------------------------------------
oatsclone    TCAAGGTCTCTATGGAAGGGAGCTTTCGTTGATGCTTTCCTGTCTAGAATAAAGAACAAC
oatsdirect   TCAAGGTCTCTATGGAAGGGAGCTTTCGTTGATGCTTTCCTGTCTAGAATAAAGAACAAA
oatsrt-pcr   TCAAGGTCTCTATGGAAGGGAGCTTTCGTTGATGCTTTCCTGTCTAGAATAAAGAACAAA

oatsest      --------------TGGCAAGAAAATTTGGTCTCGTAGATCTGCAATTTTGCCAGAATTC
Oatsclone    GGAGGAGCAATGAATGGCAAGAAAATTTGGTCTCGTAGATCTGCAATTTTGCCAGAATTC
oatsdirect   GGAGGAGCAATGAATGGCAAGAAGATTTGGTCTCGTAGATCTGCGATTTTGCCAGAATTC
oatsrt-pcr   GGAGGAGCAATGAATGGCAAGAAGATTTGGTCTCGTAGATCTGCAATTTTGCCAGAATTC
                           ********* ******************** ***************

oatsest      GTTGGTTCCTCTGTGCTCATTTACAATGGGAAAACTCATGTTCGTTGCAAGATCAACGAG
oatsclone    GTTGGTTCCTCTGTGCTCATTTACAATGGGAAAACTCATGTTCGTTGCA-----------
oatsdirect   GTTGGTTCCTCTGTGCTCATTTACAATGGGAAAACTCAT-TCCGTTGCAAGATCAACGAG
oatsrt-pcr   GTTGGTTCCTCTGTGCTCATTTACAATGGGAAAACTCATGTCCGTTGCAAGATCAACGAG
             *************************************** * *******


oatsdirect   SRSLWKGAFVDAFLSRIKNKGGAMNGKKIWSRRSAILPEFVGSSVLIYNGKTHSVARSTR
oatsrtpccr   SRSLWKGAFVDAFLSRIKNKGGAMNGKKIWSRRSAILPEFVGSSVLIYNGKTHVRCKINE
oatsclone    SRSLWKGAFVDAFLSRIKNNGGAMNGKKIWSRRSAILPEFVGSSVLIYNGKTHVRC----
oatsest      SRSLWKGAFVDAFLSRIKNNGGAMNGKKIWSRRSAILPEFVGSSVLIYNGKTHVRCKINE
             *******************:*********************************  .
```

Figure 4.11 Evidence to show that oats has a single copy of rps19

A. Amino acid alignment of *rps19* gene region which includes paralogous copies seen in wheat, maize, brome and barley relative to oats (yellow (chr5-type-copy in wheat) and green (chr3-type-copy in wheat)). Blue highlighting shows the amino acid changes within the variable region that are unique to oats.
B. Amino acid alignment of EST sequences, for oats, festuca, lolium and dactylis nuclear *rps19* gene available in NCBI. The accession numbers for the ESTs used are in brackets.

A.

```
Oats_sa              SRSLWKGAFVDAFLSRIKNNGGAMNGKKIWSRRSAILPEFVGSSVLIYNGKTHVRC
maize1htgs           SRALWKGAFVDAFLARIKKNRENMNGKKIWSRRSSILPEFVGSTVLIYNGKSHVRC
maize2_htgs          SRALWKGAFVDAFLSRIKKNRENMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRC
wheat3b              SRPLWKGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRC
brome                SRPLWKGAFVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRC
wheatchromosome5     SRPLWKGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
barley_htgs          SRPLWKGAFVDAFLQRIKKSGESLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRC
                     **.*********** ***:.   :**.*******:********:.******:****
```

B.

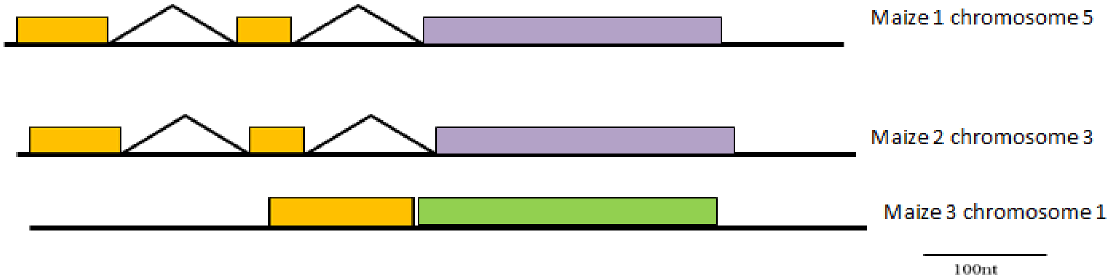
```
oatS19(GO589432)         SNSKMLRRFSEVMGPRTILRSIAMVQATKAQPPASPLLSGFGSVTRSFSSRSLWKGAFVD
festucaS19(GO890465)     ISSEMLRRFSEVMGPRTILRSIVTVQATEAQPSASPLLSGFGAVTRAFSSRSLWKGAFVD
festucaS19(GO801447)     ISSEMLRRFSEVMGPRTILRSIVTVQATEAQPSASPLLSGFGAVTRAFSSRSLWKGAFVD
loliumS19 (GR522935)     --------------PRTILRSAIAVQATKAQHPTSPLIPGFGAVTRAFSSRSLWKGAFVD
dactylis(HO158277)       -------SWFSAMGPRTILRSVVM--GTNAQHFTSPLLSGFGAVTRSFSSRSLWKGAFVD
dactylis(HO160164)       -------SWFSAMGPRTILRSVVM--GTNAQHFTSPLLSGFGAVTRSFSSRSLWKGAFVD
dactylis(HO171082)|      --------------------------GTNAQHFTSPLLSGFGAVTRSFSSRSLWKGAFVD
dactylis(HO169962)       -------------------------GGTNAQHFTSPLLSGFGAVTRSFSSRSLWKGAFVD
                                                   .*:**  :***:.***:***:*************

oatS19(GO589432)         AFLSRIKNNGGAMNGKKIWSRRSAILPEFVGSSVLIYNGKTHVRCKINEGKVGHKFGEFA
festucaS19(GO890465)     AFLSRIKNNGGAMNGKKIWSRRSSILPEFVGSSALIYNGKTHVRCKISEGKVGHKFGEFA
festucaS19(GO801447)     AFLSRIKNNGGAMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKINEGKVGHKFGEFA
loliumS19(GR522935)      AFLSRIKNNGGAMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKINEGKVGHKFGEFA
dactylis(HO158277)       AFLSRIKQNGGAMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKIHEGKVGHKFGEFA
dactylis(HO160164)       AFLSRIKQNGGAMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKIHEGKVGHKFGEFA
dactylis(HO171082)|      AFLSRIKQNGGAMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKINEGKVGHKFGEFA
dactylis(HO169962)       AFLSRIKQNGGAMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKINEGKVGHKFGEFA
                         *******:***************:*********.************* ************

oatS19(GO589432)         FTRRRRPHRTTIAKGKQVKGKK
festucaS19(GO890465)     FTRRRRPHRTTIAKGKQVKGKK
festucaS19(GO801447)     FTRRRRPHRTTIAKGKQVKGKK
loliumS19(GR522935)      FTRRRRPHRTTIAKGKQVKGKK
dactylis(HO158277)       FTRRRRPHRTTIAKGKQVKGKK
dactylis(HO160164)       FTRRRRPHRTTIAKGKQVKGKK
dactylis(HO171082)|      FTRRRRPHRTTIAKGKQVKGKK
dactylis(HO169962)       FTRRRRPHRTTIAKGKQVKGKK
                         **********************
```

Figure 4.12 Maize has three distinct rps19 gene copies.

A.Schematics of three distinct *rps19* gene copies in maize Core *rps19* for maize chromosome 1 and 5 is shown in purple to emphasize that they are very similar to each other while the core *rps19* of retroprocessed chromosome 1 copy is shown in green

B. Amino acid alignment of three distinct copies in maize
Amino acids that are distinct to wheat chromosome 5 type *rps19* copy are highlighted in wheat. Maize *rps19* copies are homologous to the chromosome 3 type copy in wheat and they are highlighted in green.

A.

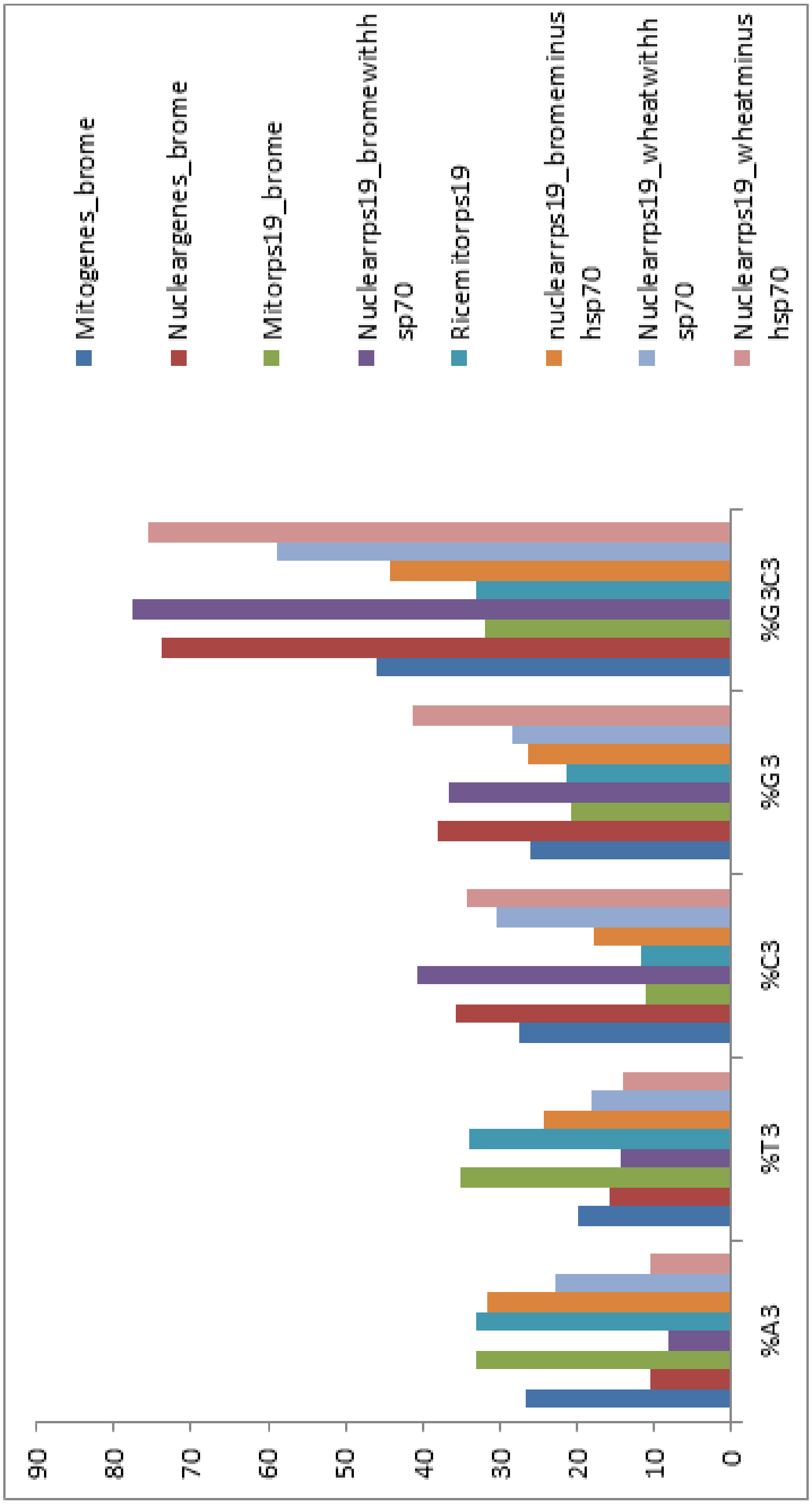


B.

```
Maizechr1      -MAIGSLIAST--FARS--SHALPAAAA----SAISQIPAHRISSTLRPWSSG-SSRALW
maizechr5      -MPIGSLFVSR--LARS--SHALPAAAA----AAISQAPAHRISYILRPWSSG-SSRALW
wheatchr5      MSSIASLIASRSSFARC--GHALPAAIS------QTQHAASPLLSGFGSAARAFSSRPLW
wheatchr3      -MAIASLIASR--FARSGHGHALPAAAAAISQAPRAQHAASPLLSGFGPLTRAFSSRPLW
                 .*.**:.*   :**.  .****** :       :* .*  :   : . : . ***.**

Maizechr1      KGAFVDAFLSRIKKNRENMNGKKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGH
maizechr5      KGAFVDAFLARIKKNRENMNGKKIWSRRSSILPEFVGSTVLIYNGKSHVRCKITEGKVGH
wheatchr5      KGAFVDAFLQRIKNSGGSLNGRKIWSRRSSILPEFVGSSALIYNGKTHVRCRVTEGKVGH
wheatchr3      KGAFVDAFLQRIKNKRENLNGKKIWSRRSSILPEFVGSTVLIYNGKTHVRCKITEGKVGH
               ********* ***:.  .:**.****************:.******:****.:*******

Maizechr1      KFGEFAFTRRRRPHRTMTGK--GNQGKARKK
maizechr5      KFGEFAFTRRRRPHRTITGK--GNQGKGRKK
wheatchr5      KFGEFAFTRKRRPHRAITAKKAGGQGKGRK-
wheatchr3      KLGEFAFTRRRRPHRAILAK--GGQGKGKKK
               *:*******.*****:: .*  *.***..*
```

Figure 4.13- Bar graph showing the codon usage pattern for mitochondrial rps19 and nuclear rps19 genes

The names of the genes that have been used for codon usage analysis of nuclear *rps19* and *rpl2* genes in brome are shown in Appendix J.



90
80
70
60
50
40
30
20
10
0
%A3
%T3
%C3
%G3
%G3C3
Mitogenes_brome
Nucleargenes_brome
Mitorps19_brome
Nuclearrps19_bromewithhsp70
Ricemitorps19
nuclearrps19_bromeminushsp70
Nuclearrps19_wheatwithhsp70
Nuclearrps19_wheatminushsp70

(Subramanian thesis 1999). Brome mitochondrial *rps19* gene and rice mitochondrial *rps19* genes share a codon bias for AT.

Brome nuclear *rps19* gene with hsp70 has a higher codon bias for GC compared to core *rps19* alone. Wheat, whose mitochondrial copy is a pseudogene, must require expression of the nuclear copy at all the times. This is evident from the fact that wheat's core *rps19* has higher GC% than brome. Although wheat and brome nuclear genes in general had high GC content, oats exhibits more bias for T in the third position of a codon (data not shown).

Chapter 5:- General Discussion

5.1 Plant mitochondrial DNA rearrangements impact on *rpl2* and *rps19* genes

Plant mitochondrial genomes undergo frequent rearrangements through homologous recombination, which give rise to insertions and deletions in genomic sequences. Direct repeats can be responsible for genetic instability and were shown to cause deletions and facilitate recombination in plant mitochondria (Sloan et al 2010) and yeasts (Lobachev et al 1998). Similarly, in the case of *rpl2* gene in brome, the indel region in exon 1 is where the direct repeat is present in *Bambusa* (figure 3.2). This points towards the idea that the brome mitochondrial *rpl2* inactivation, occurred due to recombination between direct repeats. My results showed that there is another pseudo *rpl2* gene (short exon 2) with a *nad1* and *nad6* linkage. Thus, over evolutionary time, Ψ*rpl2* gene underwent duplication and both copies were inactivated and have been relocated due to rearrangements. Recombination between short dispersed inverted repeats gives rise to large inversions quite often which cause “scrambling” of gene order within closely related species (Palmer et al 2000). The cluster of genes in the order that is conserved in bacteria is *rps10-rpl2-rps19-rps3-rpl16* (Nissen 2000). This order is not conserved in brome, as there is a Ψ*rpl2* gene beside *rps19* that is upstream of *nad4l* and Ψ*rpl2* gene beside *nad1* and *nad6*. These types of rearrangements have not just been seen in brome, but also in wheat, for instance, whose Ψ*rpl2* gene is beside nad1 and *nad6*. Similarly, Ψ*rps19* is beside *nad4l* elsewhere in the genome (Subramanian et al 2001)**.** In addition to the functional *rps19* gene in brome, my Southern analysis showed that there is another pseudogene copy of *rps19* gene elsewhere in the genome (figure 3.4a). This pseudogene piece in brome is not like the one present in wheat as it is missing the 3’end of the gene rather than the 5’end. This, again, is an example of rearrangements occurring in mitochondrial genomes which result in the relocation of genes. Novel chimeric genes in plant mitochondria have been shown to be a result of frequent duplication and inversion events. Some notable examples of these chimeric genes are known to give rise to cytoplasmic male sterility (Palmer et al 2000).

5.2 Gene loss from mitochondria and transfer to the nucleus

Many interesting questions arise when we think about mitochondrial gene transfer to the nucleus. One particular question is: what causes an increase in the rate of gene transfer in some lineages? This has been addressed in a study by Adams et al 20002, where 40 mitochondrial protein genes were surveyed using southern blot hybridization in 280 angiosperm species. During evolutionary time some lineages were found to remain immobile in terms of gene transfer as they kept all 40 protein coding genes while some other lineages only lost six. One example of such lineage is the family magnoliidae which include all of the earliest angiosperms. This lineage shows very little loss as 27 of the 33 magnoliids did not lose any genes and the other six only lost one or two genes (Adams et al 2002). Some other lineages however are experiencing recent rapid rates of gene loss. For example, plants such as Lachnocaulon, Allium, and Erodium have lost all or most of their ribosomal protein genes (Adams et al 2002).

This thesis has contributed information with respect to gene loss for two protein coding genes *rpl2* and *rps19* among different grasses. This difference in the rate of gene loss at different periods of time is what Palmer refers to as a "Punctuated pattern of evolution of mitochondrial gene content". Palmer suggests that the high rate of gene transfer can be due to an increase in reverse transcription of mitochondrial ribosomal protein genes. The second reason is that polyploidy and whole genome duplication effectively lead to an increase in the amplification of nuclear genes that encode mitochondrial proteins, thus increasing the availability of mitochondrial targeting peptides and regulatory elements. Finally, an increase in gene transfer to the nucleus can occur if the mitochondria are under stress and are more inclined to release DNA (Adams et al 2002).

The presence of the mitochondrial *rps19* gene in the nuclear genome of barley provides additional evidence for the occurrence of DNA mediated transfer of genes between organelles in addition to the known mechanism involving RNA mediated transfer. Although RNA mediated transfer usually results in functional gene transfer DNA mediated transfer can result in an unsuccessful one because it will be unedited and the group II introns within the mitochondrial genes cannot undergo splicing via the nuclear splicing machinery. Similarly the nuc-mt rps19 gene in the nucleus of barley is unedited and lacks eukaryotic-type expression signals therefore it looks like it is non-functional.

5.3 Coexpression of mitochondrial and nuclear *rps19* copies in brome

Brome *rps19* was found to be expressed in both the mitochondrion and the nucleus, which is quite rare (Nugent & Palmer 1991, Sandoval *et al* 2004, Choi *et al* 2006 and Bittner-Eddy *et al* 1994). My results show that the *rps19* gene has been functional in mitochondria and the nucleus for about 60 million years (Kellog 1998), which is the longest period among plants that such a transition stage has been reported to date. Gene transfer of the mitochondrial *rps19* gene in brome might have occurred as an RNA intermediate, but my data suggests that there is a possibility for DNA-mediated transfer, as well. This is because the mitochondrial *rps19* only has two edits that are non-synonymous and one of the edits is not retained after transfer to the nucleus.

My experimental results show that the levels of transcripts for both mitochondrial and nuclear copies are low in brome seedling RNA. Northern analysis of mitochondrial *rps19* gene failed to give an abundant signal for *rps19* gene with seedling RNA. During the seedling stage of development in wheat, ribosomal protein genes like *rps2, rps3 and rps7* show a decrease in relative mRNA abundance especially during seedling stages compared to respiratory chain genes such as *coxI* and *atp6,* which show relatively steady state levels of mRNA over the course of plant development (Li-Pook-Than et al 2004). These observations are in keeping with the fact that during embryonic germination, transcription and translation of ribosomal proteins is very important as it is a period during which there is a rapid mitochondrial biogenesis and high oxygen consumption (Li-Pook-Than et al, 2004). For this reason, I tried to isolate brome embryo RNA in order to see if the mitochondrial *rps19* transcripts would be more abundant in my northern hybridizations. However there were some challenges associated with this because brome embryos are relatively smaller in size compared to wheat embryos. Due to this reason less mitochondrial RNA was isolated for an equal mass of seeds since the proportion of seed: embryo is higher for brome than for dissected wheat embryos. Also due to the small size of brome seeds, it was not very feasible to remove the seed coat from each individual seed before isolation. As the seed coats remained intact during isolation, the isolated brome mitochondrial embryo RNA had higher bacterial contamination than the wheat embryo RNA.

Once transferred to the nucleus, the *rps19* gene needs to be adapted to the new environment. Nuclear *rps19* sequences allowed me to look at codon usage patterns in brome and compare them with other grasses codon usage patterns. Prior work by Selvi Subramanian in our lab showed that, in the nucleus, wheat has a bias for G and C nucleotides at the third position of the codon compared to the mitochondrion, which has bias for A and T nucleotides (Subramanian thesis 1999). The brome *rps19* gene in the nucleus without the heat shock protein presequence retains its codon bias for A or T at the third position just like the mitochondrial copy. This is interesting because although the transferred *rps19* gene has been in the nucleus for about 60 million years it has not adapted to the nuclear environment like other nuclear genes in brome. The nuclear *rps19* gene with the presequence on the other hand has higher GC bias as expected because the hsp70 presequence has been in the nucleus for a longer time. The codon usage pattern for oats nuclear *rps19* gene is again different than that of wheat and brome as it has bias for T at the third position

5.4 When might the duplication of *rps19* gene have occurred?

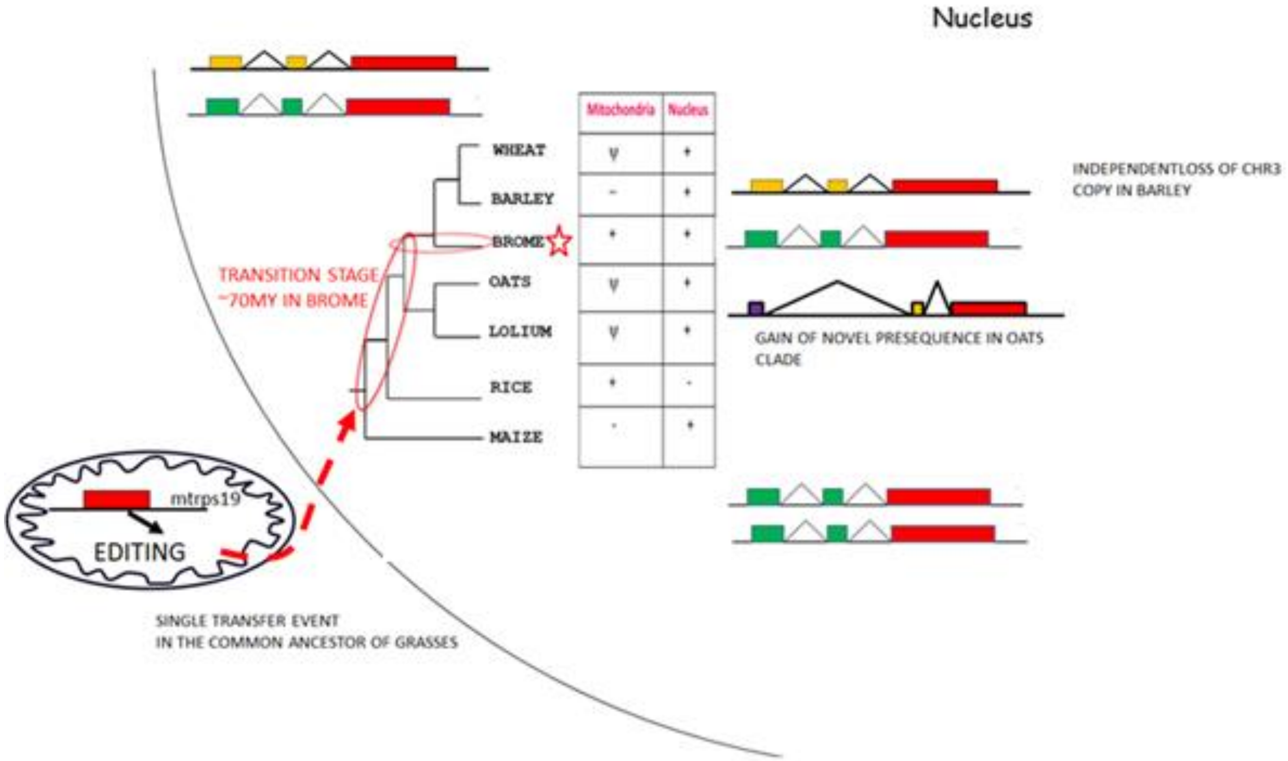
From my experimental results and NCBI high throughput genomic searches we know that wheat, and maize have three distinct copies of the *rps19* gene. Barley only has one of the paralogous copies (closer to the chromosome 5 copy seen in wheat) so it must have lost the other copy (chromosome 3 copy seen in wheat and brome) over evolutionary time. Although only a single copy of *rps19* was identified in brome and oats (lolium and festuca), one cannot exclude the possibility that they might also have paralogous copies which are dissimilar and hard to detect with the oligomers being used. Paralogous copies in maize more closely resemble the chromosome-3-type copy seen in wheat and brome, which closely reflects the mitochondrial type *rps19* gene. This suggests that the chromosome3-type-copy is likely the transferred copy and the chromosome 5 copy may have arisen through duplication. The paralogs in maize also have more nucleotide and amino acid changes relative to the *rps19* gene in wheat, brome and barley, which could be due to the fact that maize is a more distant relative. Nevertheless, the two distinct copies in maize are similar to each other (compared to the retroprocessed copy), even in the intron regions. Thus, the duplication event in maize might have been a recent, independent event or it may have occurred in the common ancestor and later underwent gene conversion. The retroprocessed *rps19* gene in maize must be yet from another independent duplication event in maize. Consequently, my experimental data and bioinformatic analysis suggest that the *rps19* gene duplication most likely occurred in the common ancestor of barley and wheat. Additionally, barley must have lost its chromosome-3-type copy which is seen in wheat and brome. More information from the oats and brome lineages is needed to say whether or not the duplication event might have occurred in the common ancestor of maize and rice or that of wheat and brome (figure 5.1). Also, the fact that the maize nuclear *rps19* gene copies have the same presequence hsp70 as brome suggests that it is a common transfer event (Fallahi et al 2005) and we can say the transition period is about 60 million years (Kellog 1998).

5.5. Future directions

Future studies can include identifying the composition of mitochondrial ribosomes for the *rps19* protein. As there are two functional copies in brome, it would be interesting to find out which copy is being used the most. This can be done using proteomics and western blot analysis by designing antibodies that are specific to nuclear and mitochondrial *rps19* proteins in brome. It would also be interesting to see if the transfer event of *rps19* gene to the nucleus could have extended past grasses such as maize and extends into eudicots. The nuclear *rps19* gene structure gave us insight into the nature and origin of mitochondrial targeting presequence. The novel presequence found in oats clade, which includes *Festuca* and *Lolium* is most likely due to rearrangements via exon shuffling. Western blot analysis can be used to confirm that this novel presequence is mitochondrial targeting. It would be also very interesting to investigate how most genes in the nucleus gain their targeting signals from these heat shock proteins. Understanding how transferred genes are integrated and expressed in the nucleus can help us in creating better transgenic plants using methods such as VLPs (virus like particles)for transient expression, *Agrobacterium tumefaciens* for nuclear integration, biolistics for random integration into the nucleus and organelles. Transgenic plants can produce large quantities of recombinant proteins and inexpensive biomass. In order to introduce foreign DNA and optimize its expression, it is useful to understand the gene structure of the host (Serres-Giardi et al 2012). The translation efficiency of the inserted gene is higher when one understands the sequences upstream of translation start site. For instance in *Arabidopsis thaliana*, a highly conserved 21nt upstream region was shown to cause variation in translation up to 20times (Kim et al 2014). Gene engineering tools such as zinc-finger-nucleases (ZFNs), transcription-activator-like effector nucleases (TALENs), clustered regulatory interspaced short palindromic repeats (CRISPRs) can be used to take advantage of this 5'UTR sequence in order to improve transformation in plants. Introns also have ability to have an impact on nuclear gene expression as they can change the transcription initiation rate by interacting with polyadenylation. For example, a 5'UTR located intron was shown to increase gene expression in *Arabidopsis thaliana* (Akua and Shaul 2013). Similarly, the 3'UTR introns in *rps19* nuclear gene seen in wheat and barley could also have an important role. Another important aspect of using a transgene to produce proteins is to ensure that the codon usage matches that of the host nuclear genome (Ullrich et al 2015). This is pertinent as my results for both *rpl2* and *rps19* show that the nuclear genes tend to have higher GC bias similar to 200 other seed plants that have been surveyed (Chevance et al 2014).Forthcoming studies can unravel if stress has a key role in initiating gene transfer. There are many examples of gene transfer from mitochondria and chloroplasts to the nucleus but there are not many instances of nuclear genes being integrated into other organelles. Thus, it is interesting, to know the reason to why such transfer events do not occur very often. One explanation is that the chloroplasts genome is small compared to the mitochondrial genome, so if the transferred gene gets integrated into the organelle's chromosome, it is more likely to be inactivated. Scientists use this to their benefit as one study shows how targeted inactivation of the plastid *nhB* gene in tobacco causes an increase in photosynthesis when there are low levels of stomatal closure (Horvath et al 2000). Mitochondrial genomes however are bigger than chloroplast genomes so the gene has more chance to be integrated into the non-coding regions (Gualberto et al 2014). It would be interesting to see if there are more possible ways to transform chloroplasts and mitochondria which could be beneficial for agriculture industry. Cytonuclear interactions such as retrograde signalling have been found to be a key player in restoring cytoplasmic male sterility (CMS) in plants. (Chase 2007).

Figure 5.1 Summary diagram of possible scenario by which nuclear rps19 gene becomes active in different grasses

Red ovals represents the transition period in brome where a copy of *rps19* gene has been maintained in both mitochondria and the nucleus. Star beside brome indicates that its *rps19* gene is in transition stage. Yellow (chr5) and green (chr3) colouring for the presequence shows the two distinct paralogous copies found in wheat. Brome only has the green copy while barley only has the yellow copy. The novel presequence in oats is shown by a purple exon 1 and a novel intron 1 sequence. The pre-sequence of the paralogous copies in maize are coloured in green as they most closely resemble the green copy in wheat and brome and are very similar to each other.


Nucleus
Mitochondria
Nucleus
WHEAT
ψ
+
BARLEY
-
+
BROME
+
+
OATS
ψ
+
LOLIUM
ψ
+
RICE
+
-
MAIZE
-
+
INDEPENDENTLOSS OF CHR3 COPY IN BARLEY
TRANSITION STAGE ~70MY IN BROME
GAIN OF NOVEL PRESEQUENCE IN OATS CLADE
mtrps19
EDITING
SINGLE TRANSFER EVENT IN THE COMMON ANCESTOR OF GRASSES

References

Adams KL, Chuan Ong H, Palmer JD. 2001. Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Mol Biol Evol.*18:2289-2297.

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phyl Evol.* 29:380-395.

Adams Kl, Percifield R, Wendel JF. 2004. Organ-Specific Silencing of Duplicated Genes in a Newly Synthesized Cotton Allotetraploid. *Genet.*168:2217-2226

Adams KL, Qui, Y-L, Stoutemyer M, Palmer JD. 2002. 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.*15: 9905-9912.

Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD. 1999. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. *Proc. Natl. Acad. Sci.*96:13863-13868.

Aguileta G, Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, Gabaldon T. 2014. High variability of mitochondrial gene order among fungi. *Genome Biol Evol.* 6:451-465.

Akua T, Shaul O. 2013. The Arabidopsis thaliana MHX gene includes an intronic element that boosts translation when localized in a 50 UTR intron. *J Exp Bot*. 64:4255-4270.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering Plant Arabisopsis thaliana. *Nature*. 408:796-815.

Armstrong KC. 1979. A and B genome homoeologies in tetraploid and octaploid cytotypes of *Bromus inermis*. *Can J Genet Cytol.* 21:65–71.

Ausubel FM, Brent R, Kinston RE, Moore DD, Seidman JG, Smith JA, Struhl K. 1990. Current protocols in molecular biology. Current protocols, Boston.

Bendich AJ, McCarthy BJ. 1970. DNA composition among barley, oats, rye and wheat. *Genet.* 65:545-565.

Binder S, Brennicke A. 2003. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. Philos. Trans. R. *Soc. Lond. Biol. Sci*. 358:181-188.

Bittner-Eddy P, Monroy AF, Brambl R. 1994. Expression of mitochondrial genes in the germinating conidia of *Neurospora crassa*. *J Mol Biol.* 235:881–897.

Bonen L. 2006. Mitochondrial genes leave home. *New Phytol.* 172:397-381.

Bonen L. 2008. Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion*. 8:26-34.

Bonen L, Calixte S. 2006. Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. *Mol Biol Evol.* 23:701-712.

Bosler DM, Kerhornou A, Walts B, Kersey P. 2015. Triticeae Resources in Ensembl Plants. *Plant and Cell Phys*. 3:1-11.

Brenchley R1, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature.* 491:705-710.

Brennicke A, Grohman L, Hiesel R, Knoop V, Shuster W. 1993. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEB Lett.* 325:140-145.

Brown TA, Ray JA, Waring RB, Scazzocchio C, Davies RW. 1984. A mitochondrial reading frame which may code for a second form of ATPase subunit 9 in *Aspergillus nidulans*. *Curr Genet*. 8:489-92.

Calixte S, Bonen L. 2008. Developmentally specific transcripts from the *ccmFN-rps1* locus in wheat mitochondria. *Mol Gen Genom.* 280:419-26.

Carnahan H L, Hill HD. 1960. The nature of polyploidy in smooth brome grass (Bromus inermis). *Heredity* 51:43-44.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world plant mitochondrial nuclear interactions. *Trends Genet*. 2:81-90.

Chevance F F, Le Guyon S, Hughes KT. 2014. The effects of codon context on in vivo translation speed. *PLoS Genet*. 10:1004392.

Choi B, Acero M, Bonen L. 2012. Mapping of wheat mitochondrial mRNA termini and comparison with breakpoints in DNA homology among plants. *Plant Mol Biol*. 80:539-552.

Choi C, Liu Z, Adams KL. 2006. Evolutionary transfers of mitochondrial genes to the nucleus in the *Populus* lineage and coexpression of nuclear and mitochondrial *Sdh4* genes. *New Phytol*. 172:429–439.

Covello PS, Gray MW. 1989. RNA editing in plant mitochondria. *Nature* 341:662–66.

Déquard-Chablat M, Sellem CH, Golik P, Bidard F, Martos A, Bietenhader M, di Rago JP, Sainsard-Chanet A, Hermann-Le Denmat S, Contamine V. 2011. Two nuclear life cycle-regulated genes encode interchangeable subunits c of mitochondrial ATP synthase in *Podospora anserina*. *Mol Biol Evol.* 28:2063-2075.

Dombrowski S. 1997. 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *EMBO J*. 16: 5069-5076.

Duncan O, Monica WM, James W. 2013. Unique components of the plant mitochondrial protein import apparatus. *Biochmica et Biophysica Acta*. 1833: 304-313.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols* 2:953-971.

Erpenbeck D, Voigt A, Adamski OM, Adamska MJ, Hooper NA, Wo¨rheide GB, Degnan MS 2007. Mitochondrial Diversity of Early-Branching Metazoa Is Revealed by the complete mt Genome of a Haplosclerid Demosponge. *Mol. Biol. Evol*. 24:19–22.

Eva MH, Peter SO, Joet T, Rumeau D, Cournac L, Horvath G.V. 1999. Targeted inactivation of the plastid ndhB gene in Tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure 1. Biology research center. Hungary. 148.

Fallahi M, Crosthwait J, Calixte S, Bonen L. 2005 Fate of mitochondrially located S19 protein genes after transfer of a functional copy to the nucleus in grasses. *Mol Gen Genom.* 273:76-83.

Fallahi M, 2000. Status of mitochondrial S19 ribosomal protein gene in various grasses. Thesis for Masters of Science. University of Ottawa.

Fica SM, Tuttle N, Novak T, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. 2013. RNA catalysis nuclear pre-mRNA splicing. *Nature*. 503: 229-234.

Figueroa P, Gómez I, Holuigue L, Araya A, Jordana X. 1999. Transfer of rps14 from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J.* 18:601-609.

Forner J. 2007. Mapping of mitochondrial mRNA termini in Arabidopsis thaliana: t-elements contribute to 5' and 3' end formation. *Nucleic Acid Res*. 35: 3676-3692.

Gaut BS. 2002. Evolutionary dynamics of grass genomes. *New Phytol.* 154: 15-28.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp, Japonica). *Science*. 296:92-100.

Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. *Mol Biol Evol.* 26:99-110.

Gray MW. 1992. The endosymbiont hypothesis revisited. Int Rev Cytol. 141:233-357.

Gray MW. 2014. The Pre-Endosymbiont hypothesis: A new perspective on the origin and evolution of mitochondria.*Cold Spring Harb Perspect Biol* .6:16097.

Gray MW, BurgerG, Lang BF. 1999. Mitochondrial evolution. *Science.* 283:76-83.

Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. *Annu. Rev. Genet*. 38:477-524.

Grewe F, Viehoever P, Weisshaar B, Knoop V. 2009. A trans-splicing group I intron and tRNA hyper editing in the mitochondrial genome of the lycophyte Isoetes engelmanni. *Nucl. Acids Res*. 5093– 104.

Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-lofti F, Dietrich A. 2014. The plant mitochondrial genome: Dynamics and maintenance. *Biochime*. 100:107-120.

Hammani K, Giegé P. 2014. RNA metabolism in plant mitochondria. *Trends Plant Sci.* 19:380-389.

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

Horvath EM, Peter SO, Joet T, Rumeau D, Cournac L, Horvath GV, Kavanagh TA, Schafer C,Peltier G, Medgyesy P. 2015. Targeted inactivation of the plastid ndhB gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. *Plant Phys*. 123:1337-1349.

Huang S, Sirikhachornkit A, Faris, JD, Su X, Gill BS, Haselkorn R, Gornicki P. 2002. Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant Miol Biol*. 48: 805-820.

Huot JL, Enkler L, Megel C, Karim L, Laporte D, Becker HD, Duchêne AM, Sissler M, Maréchal-Drouard L. 2014. Idiosyncrasies in decoding mitochondrial genomes. *Biochimie*. 100:95-106.

International Wheat Genome Sequencing Consortium (IWGSC). 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*. 345: 1251788.

Islam MS, Studer B, Byrne SL, Farrell JD, Panitz F, Bendixen C, Moller IM, Asp T. 2013. The genome of perennial ryegrass mitochondria. *BMC Genomics*. 14:202.

Kannan S, Rogozin IB, Koonin EV. 2014. MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. *BMC Evol Biol*. 14:237.

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

Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C. 2014. Ensembl Genomes 2013: scaling up access to genome-wide data.*Nucl. Acids Res*. 42:D546-D552.

Kim Y, Lee G, Jeon E, Sohn EJ, Lee Y, Kang H, Lee DW, Kim DH, Hwang I. 2014. The immediate upstream region of the 5'UTR from the AUG start codon has a pronounced effect on the translational efficiency in Arabidopsis thaliana. *Nucleic Acids Res*. 42: 485-498.

Klein M, Eckert-Ossenkopp U, Schmiedeberg P, Brandt M, Unsled A, Brennicke W, SchusterW. 1994. Physical mapping of the mitochondrial genome of Arabidopsis thaliana by cosmid and YAC clones. *Plant J.* 6:455-447.

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

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 N, Jordan X, Ozawa K, Zanlungo S, Harada K, Sasaki T, Kodowaki K. 2000. Transfer of the mitochondrial *rps10* gene in the nucleus in rice: acquisition of the 5'untranslated region followed by gene duplication. *Mol Gen Genet*. 4:733-9

Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K. 2003. RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Res*. 31:2417–23

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.

Liere K, Weihe A, Borner T. 2011. The transcription machineries of plant mitochondria and chloroplasts: Composition, function and regulation. *J. Plant Physiol.* 168: 1345-1360.

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

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*. 26:275-891.

Li W, Zhang XC, Zhao J, Shi Y, Zhu XP. 2015. Complete mitochondrial genome of Cura trifasciata (Chinese three-striped box turtle), and a comparative analysis with other box turtles. *Gene.* 555: 169-177.

Lobachev KS, Shor BM, Tran HT, Taylor W, Keen J, Resnicl MA, Gordenin DA. 1998. Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cervisiae *Genet.* 148:1507-1524.

Ma PF, Guo ZH, Li DZ. 2012. Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. *PLoS One.* 7:e30297.

Maffey L, Degand H, Boutruy M, 1997. Partial purification of mitochondrial ribosomes from broad bean and identification of proteins encoded by the mitochondrial genome. *Mol Gen Genet*. 254:365-371

Maracia M, Pyle AM. 2012. Visualizing group II intron catalysis through stages of splicing. *Cell*.153: 497-507.

Martin W, Roettger M, Kloesges T, Thiergart T, Wohele C, Gould S, Dagan T. 2006. Mordern endosymbiotic theory: getting lateral gene transfer into the equation. *J Endocyt and Cell Res*. 23:1-5.

Matsunaga M, Takahashi Y, Kurino RY, Mikami T, Kubo T. 2013. Evolutionary aspects of a unique internal mitochondrial targeting signal in nuclear-migrates *rps19* of sugar beet (Beta vulgaris L). *Gene.* 517:19-26.

Maier UG1, Zauner S, Woehle C, Bolte K, Hempel F, Allen JF, Martin WF. 2013. Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. *Genome Biol Evol.* 5:2318-2329.

Mayer K.F.X, Waugh R, Brown J.S.W, Schulman A, Langridge P, Platzer M. 2012. International Barley Genome Sequencing Consortium. *Nature*. 2012. 491: 711-71.

Mehrotra S, Goyal V. 2014. Repetitive sequences in plant nuclear DNA: Types, Distribution, Evolution and Function.*Genomics, Proteomics and Bioinformatics,*12: 164-171.

Murcha M.W, Kubiszewski-Jakubiak S, Wang, Y, Whelan, J. 2014. Evidence for Interactions between the mitochondrial apparatus and respiratory chain complexes via Tim21-like proteins in Arabisopsis. *Front. Plant Sci*. 5:1-10.

Nugent JM, Palmer JD. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell.* 66:473-81.

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

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qui Y-L, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proc. Natl. Acad. Sci.* 97: 6960-6966.

Pinel C, Douce R, Mache R. 1996. A study of mitochondrial ribosomes from the higher plant Solanum tuberosum. *Mol Biol Rep*. 11:93-97.

Pont C, Murant F, Confolent C, Balzergue S, Salse J. 2011. RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum L.)*2011. *Genome Biol*. 12:R119.

Puigbo P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct*. 3:38.

Qui Y, Filipenko SJ, Darracq A, Adams KL. 2014. Expression of a transferred nuclear gene in a mitochondrial genome. *Curr Plant Biol.* 1:68-72.

Reddy ASN. 2007. Alternative splicing of Pre-Messenger RNAs in Plants in the genomic Era. Annu. Rev. *Plant Biol*. 58: 267-94.

Rosseau-Guetin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN. 2013. Potential functional replacement of the plastidic Acetyl-CoA Carboxylase subunit (accD) Gene by recent transfers to the nucleus in some angiosperm lineages. *Plant phys*. 161:1918-1929.

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

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.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S. 2009. The B73 maize genome:complexity, diversity and dynamics. *Science*. 326: 1112-1115.

Serres-Giardi L, Belkhir K, David J, Glemmin S. 2012. Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell*. 24:1379-1.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012.

Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondrial with exceptionally high mutation rates. PLOS biology. 1:e1001241

Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics.* 4:1581-1590.

Sloan DB, Alverson AJ, Storchova H, Palmer JD, Taylor DR. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm silene latifolia. *BMC Evol Biol*. 10:274.

Smith DR, Arrigo KR, Alderkamp A, Allen AE. 2014. Massive difference in synonymous substitution rates among mitochondrial, plastid and nuclear genes of Phaeocystis algae *Mol. Phyl Evol*. 71: 36-40.

Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ 2009. Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. *Syst Bot*. 34:115–128.

Subramanian S, 1999. Plant mitochondrial L2 ribosomal protein genes and their expression Thesis for Doctor of philosophy. University of Ottawa.

Subramanian S, Fallahi M, Bonen L. 2001. Truncated and dispersed *rpl2* and *rps19* pseudogenes are co-transcribed with neighbouring downstream genes in wheat mitochondria. *Curr Genet* 39: 264-272

Subramanian S, Bonen L. 2006. Rapid evolution in sequence and length of the nuclear-located gene for mitochondrial L2 ribosomal protein in cereals. *Genome.* 49:275-81

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genome*. 272: 603-615.

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

Takenaka M, Zehrmann A, Verbitsky D, Hartel B, Brennicke A. 2013. RNA editing in plants and its evolution. *Annu Rev Genet*. 47:335-52.

Taylor JW, Berbee ML. 2006. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia.* 98:838-49.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Genetics*. 5: 123-135.

Ullrich KK. Hiss M, Rensing SA. 2015. Means to optimize protein expression in transgenic plants. *Curr Opin Biotechnol*. 32: 61-67.

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

Vincent R, Yan H, Gautier S, Morgane A, Sylvain S, Sylvain G, Muriel T, Jacques D. 2013. Disentangling homeologous contigs in allo-tetraploid assembly: application to durum wheat. *BMC Bioinformatics*. 14:S15

Wendel JF. 2000. Genome Evolution in Polyploids. Plant Mol Biol. 42: 225-249.

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

Wool IG. 1996. Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci.* 21:164-165.

Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. *Curr Opin MicrobioL*. 22: 38-48.

Zhang WJ, Zhou J, Li Z, Wang L, Gu Xu, Zhong Y. 2007. Comparative analysis of codon usagepatterns among mitochondrion, chloroplast and nuclear genes in Triticum aestivum L. J *Journ Integ. Plant. Biol.* 49: 246-254.

APPENDIX

Appendix A. Mitochondrial *rpl2* and *rps19* genes in brome and other grasses

i) Nucleotide sequence comparison of *rpl2* gene in grasses such as brome, rice, ferrocalmus130, bambusa and phoenix

```
phoenix          ATGAGACAAAGCCTAAAGGGGAGAGCACTTAGACATTTCACTTTGAATACGGGGAAGTCC
brome            ATGAGACAAAGCCTAAAGGGGAGAGCACTTAGACATTTCACTTTGAGTACGGGGAAGTCC
rice             ATGAGACAAAGCATAAAGGGGAGAGCGCTTAGACATTTCACTTTGAGTACGGGGAAGTCC
ferrocalmus      ATGAGACAAAGCATAAAGGGGAGAGCACTTAGACATTTCACTTTGAGTACGGGGAAGTCC
bambusa          ATGAGACAAAGCATAAAGGGGAGAGCACTTAGACATTTCACTTTGAGTACGGGGAAGTCC
                 ************ ************* ******************* *************

phoenix          GCAGGGAGGAATTCCTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
brome            GCAGGGAGGAATTCTTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
rice             GCAGGGAGGAATTCTTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
ferrocalmus      GCAGGGAGGAATTCTTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
bambusa          GCAGGGAGGAATTCTTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
                 ************** *********************************************

phoenix          TTGCAGCGAAAAATTGACCTGAAACGAAGCACTTCGTCTATTGGCATTGTAGAAAGGATC
brome            TTGCTGCGAAAAATTGACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATC
rice             TTGCAGCGAAAAATTGACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATC
ferrocalmus      TTGCAGCGAAAAATTGACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATC
bambusa          TTGCAGCGAAAAATTGACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATC
                 **** *************** ***************************** *********

phoenix          GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAGGGGGTGCTG
brome            GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTG
rice             GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTG
ferrocalmus      GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTG
bambusa          GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTG
                 ************************************************** *********

phoenix          CTACGCCGCCAGAGGAAATGCAACACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTC
brome            CCCGGCCGCCAGAGGAAATTCAAGACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTC
rice             CCCGGCCGCCAGAGGAAATTCAAGACGATAGAAGAGTTCGCTCTGCCGCGTAAGATCCTC
ferrocalmus      CCCGGCCGCCAGAGGAAATTCAAGACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTC
bambusa          CCCGGCCGCCAGAGGAAATTCAAGACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTC
                 *   *************** *** ******************* ****************
```

xon 1 rpl2 sequence missing in brome and rice

Exon 1 rpl2 sequence missing in brome

```
phoenix          GAACCTACCACGGCCACCATCTTTTGCCTTTTTTCGTTCTCTTCCCTGCCCGGGAAAGTG
brome            GAATCCACCACGGCCACTATCTTTTGCCTTTTTTCGTTCTCTTCCCTGCCCGGGAAAGTG
rice             GAATCCACCACGGCCACTATCTTTTGCCTTTTTTCGTTCTCTTCCCTG------------
ferrocalmus      GAATCCACCACGGCCACTATCTTTTGCCTTTTTTCGTTCTCTTCCCTGCCCGGGAAAGTG
bambusa          GAATCCACCACGGCCACTATCTTTTGCCTTTTTTCGTTCTCTTCCCTGCCCGGGAAAGTG
                 *** * *********** ******************************
phoenix          GATCAAAGAAAGGTAGCTTGCTCCCAAGGCAAGTGCTTGCTTACGCTTTATGTAGTGGTC
brome            GATCAAAGA---------------------------------------------------
rice             ------------------------------------------------------------
ferrocalmus      GATCAAAGAAAGGTAGCTTGCTCCCAAGGCAAGTGCTTGCTTACGCTTTATGTAGTGGTC
bambusa          GATCAAAGAAAGGTAGCTTGCTCCCAAGGCAAGTGCTTGCTTACGCTTTATGTAGTGGTC


phoenix          GGCCTTCCTACCAGAATGCCTCCTTGGTTGAAGAGCCAAGCCTGCGCAGGAAGCAAACAA
brome            ------------------------------------------------------------
rice             ------------------------------------------------------------
ferrocalmus      GGCCTTCCTACCTTCATGCCTCCTTGGTCGAAGAGCCAAGCCTGCGCAGGAAGCAAACAA
bambusa          GGCCTTCCTACCTTCATGCCTCCTTGGTCGAAGAGCCAAGCCTGCGCAGGAAGCAAACAA


phoenix          ACTTGCGCGAAGGACGTTTTCTTCTCTGCCCTGTCCTCTCCCTTGGCCAAGGGAGAGACT
brome            ------------------------------------------------------------
rice             ---------------------------------TCCTCTCCCTTGGCCCAGGGAGAGACT
ferrocalmus      ACTTGCGCGAAGGACGTTTTCTTCTCTGCCCTGTCCTCTCCCTTGGCCCAGGGAGAGACT
bambusa          ACTTGCGCGAAGGACGTTTTCTTCTCTGCCCTGTCCTCTCCCTTGGCCCAGGGAGAGACT


phoenix          GCATCCCTTTCCTTCGGTAGCTCTTTGGGTTTCCCAAGGATAGCGGTAGCTGGGGCAAAG
brome            ------------------------------------------------------------
rice             GCATCCCTTTCCTTCGGCAGCTCTTTAGGTTTCCCAAGGATAGCGGTAGCTGGGGCAAAG
ferrocalmus      GCATCCCTTTCCTTCGGTAGCTCTTTAGGTTTCCCAAGGATAGCGGTAGCTGGGGCAAAG
bambusa          GCATCCCTTTCCTTCGGTAGCTCTTTAGGTTTCCCAAGGATAGCGGTAGCTGGGGCAAAG


phoenix          CCCGCTTTCTTCGCTCCGCGAATGAGAGAGAAACTCATCGGAAAGAAGACGTTTTCTCTT
brome            ------------------------------------------------------------
rice             CCCGCTTTCTTCGCGGAGCGAATGAGAGAGAAGAAAATCGGAAAAAAGACGTTTTCTCTT
ferrocalmus      CCCGCTTTCTTCGCTCCGCGAATGAGAGAGAAGAAAATCGGAAAAAAGACGTTTTCTCTT
bambusa          CCCGCTTTCTTCGCTCCGCGAATGAGAGAGAAGAAAATCGGAAAAAAGACGTTTTCTCTT


phoenix          TGCGAGATCCGAAAGTGGAGAACGCATTGCGTTCTCTGGGCACATAGGATCAAACGTAAA
brome            ------------------------------------------------------------
rice             TGCGAGATCCGAAAGTGGAGAACGCATTGCGTTCTCTGGGCACATAGGATCAAACGTAAA
ferrocalmus      TGCGAGATCCGAAAGTGGAGAACGCATTGCGTTCTCTGGGCACATAGGATCAAACGTAAA
bambusa          TGCGAGATCCGAAAGTGGAGAACGCATTGCGTTCTCTGGGCACATAGGATCAAACGTAAA


phoenix          GCAGCGCTTTCTTGGCAGAGTTTGAGGCAGCAAGAAACTTTAGGGCTTGTTGGAGCTGCT
brome            ------------------------------------------------------------
rice             GCAGCGCTTTCTTGGCAGAGTTTGAGGCAGCAAAAAACTTTAGAACTTGTTGGAGCTGCT
ferrocalmus      GCAGCGCTTTCTTGGCAGAGTTTGAGGCAGCAAAAAACTTTAGAGCTTGTTGGAGCTGCT
bambusa          GCAGCGCTTTCTTGGCAGAGTTTGAGGCAGCAAAAAACTTTAGAGCTTGTTGGAGCTGCT


phoenix          GAGCATAACGAATCGAAGCCGAAGGCGGATCAAGGTAGCTTGCTCCCAAGGCAAGTGCTT
brome            -------------------------------AAGGTAGCTTGCTCCCAAGGCAAGTGCTT
rice             GAGCATAACGAATCGAAGCTGAAGGCGGATCAAGGTAGCTTGCTCCCAAGGCAAGTGCTT
ferrocalmus      GAGCATAACGAATCGAAGCTGAAGGCGGATCAAGGTAGCTTGCTCCCAAGGCAAGTGCTT
bambusa          GAGCATAACGAATCGAAGCTGAAGGCGGATCAAGGTAGCTTGCTCCCAAGGCAAGTGCTT
                                                ****************************
```

```
phoenix          GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCAGAATGCCTCTAGAAGCTTCTACAAA
brome            GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCTTCATGCCTCTAGAAGCTTTGTAGAA
rice             GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCTTCATGCCTCTAGAAGCTTCTACAAA
ferrocalmus      GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCTTCATGCCTCTAGAAGCTTCTACAAA
bambusa          GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCTTCATGCCTCTAGAAGCTTCTACAAA
                 **********************************   ****************    **

phoenix          GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCTGCCAA---GCCTATAGGC
brome            GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCTGCCAA-------------
rice             GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCCGCCAAGCCGCCTATAGGC
ferrocalmus      GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCTGCCAAGCCGCCTATAGGC
bambusa          GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCTGCCAAGCCGCCTATAGGC
                 ***************************************** *****

phoenix          GAAGGGCCGAAGGATGGAGCGTGCAAAGTCGATCGTGCACCTGTCGTGTGACCCGTTGGT
brome            ------------------------------------------------------------
rice             GAAGGGCCGAAGGATGGAGCGTACAAAGTCGATCGTGCACCTGTCGTGTGA-CCGTTGGT
ferrocalmus      GAAGGGACGAAGGATGGAGCGTACAAAGTCGATCGTGCACCTGTCGTGTGACCCGTTGGT
bambusa          GAAGGGACGAAGGATGGAGCGTACAAAGTCGATCGTGCACCTGTCGTGTGACCCGTTGGT

phoenix          CCTAAGCAATGTCTTTCGCGAAGCGACCCACCTAGAAAGAGCTCTCCTTTTAGGGAGGCA
brome            ------------------------------------------------------------
rice             CCTAAGCAATGTCTTTCGCGAAGCGACCCACCTAGAAACAGCTCTCCTTTTA-GGAGGCA
ferrocalmus      CCTAAGCAATGTCTTTCGCGAAGCGACCCACCTAGAAACTGCTCTCCTTTTA-GGAGGCA
bambusa          CCTAAGCAATGTCTTTCGCGAAGCGACCCACCTAGAAACAGCTCTCCTTTTA-GGAGGCA


phoenix          AAAAAATGGAACTTCGATCTGATGCGGGTATCCAATCCCGCCGCTGAGATGTCCAGCGGA
brome            ------------------------------------------------------------
rice             AAAAAATGGAACTTCGATCTGATGCGGGTATCCAATCCCGCCGCTGGGATGTCCAGCGGA
ferrocalmus      AAAAAATGGAACTTCGATCTGATGCGGGTATCCAATCCCGCCGCTGGGATGTCCAGCGGA
bambusa          AAAAAATGGAACTTCGATCTGATGCGGGTATCCAATCCCGCCGCTGGGATGTCCAGCGGA


phoenix          TTCCGGGGCCTTGACGAATGGCCGGCCACCATTGTTGTTAGAAGAGCAAAAGGCCCGGGG
brome            ------------------------------------------------------------
rice             TTCCGGGGCCTTGACAAATGGCCGTTCACCATTGTTGTTAGAAGAGCAAAAGGCCCGGGG
ferrocalmus      TTCCGGGGCCTTGACAAATGGCCGGTCACCATTGTTGTTAGAAGAGCAAAAGGCCCGGGG
bambusa          TTCCGGGGCCTTGACAAATGGCCGGTCACCATTGTTGTTAGAAGAGCAAAAGGCCCGGGG


phoenix          CATAGCAGGATGAACCAATGTGAATGAGTGTAAGCTTCGTTGCCCGAACACGATTGGTGC
brome            ------------------------------------------------------------
rice             CATAGCAGGATGAACCAATGTGAATGAGTGTAAGCTTCGTTG-CCGAACACGATTGGTGC
ferrocalmus      CATAGCAGGATGAACCAATGTGAATGAGTGTAAGCTTCGTTGCCCGAACACGATTGGTGC
bambusa          CATAGCAGGATGAACCAATGTGAATGAGTGTAAGCTTCGTTGCCCGAACACGATTGGTGC


phoenix          TGACCACACTAGGTGCTACCGTGGTAGCAAGAGAGGCCAGGCAGTGACAATTGAGAGGTT
brome            ------------------------------------------------------------
rice             TGACCACACTAGGTGCTACCGTGGTAGCAAGAGAGGCCCGGCGGTGACAATTGAGAGGTT
ferrocalmus      TGACCACACTAGGTGCTACCGTGGTAGCAAGAGAGGCCAGGCGGTGACAATTGAGAGGTT
bambusa          TGACCACACTAGGTGCTACCGTGGTAGCAAGAGAGGCCAGGCGGTGACAATTGAGAGGTT


phoenix          GTCACTGAGCATTCCGTCTCACACGGGAAGAGAGGTCAAATGGCAAGGCAAAAGGCCATA
brome            ------------------------------------------------------------
rice             GTCACTGAGCATTCCGTCTCACACGGGAAGAGAGGTCAAATGGCAAAGCAAAAGGCCATA
ferrocalmus      GTCACTGAGCATTCCGTCTCACACGGGAAGAGAGGTCAAATGGCAAAGCAAAAGGCCATA
bambusa          GTCACTGAGCATTCCGTCTCACACGGGAAGAGAGGTCAAATGGCAAAGCAAAAGGCCATA
```

5’ rpl2 intron missing in brome

```
phoenix          CGCCCGTGTGGCTCCTCGCGGAGTATAGCTCACATCCAAACATCTGATTGGGGAACGGGG
brome            ------------------------------------------------------------
rice             CGCCCGTGTGGCTCCTCGCGGAGTATAGCTCACATCCAAACATCTGATTGGGGAACGGGG
ferrocalmus      CGCCCGTGTGGCTCCTCGCGGAGTATAGCTCACATCCAAACATCTGATTGGGGAACGGGG
bambusa          CGCCCGTGTGGCTCCTCGCGGAGTATAGCTCACATCCAAACATCTGATTGGGGAACGGGG

phoenix          CAACGCCCATGAAGCTCCGGCGGAAAGGGAAGGCCTGCCAGGCCGTATGCCCATGGGTGC
brome            ------------------------------------------------------------
rice             CAACGCCCATGAAGCTCCGGC--AAAGGGAAGGCCTGCCAGGCCGTATGCTCATGGGTGC
ferrocalmus      CAACGCCCATGAAGCTCCGGCGGAAAGGGAAGGCCTGCCAGGCCGTATGCCCATGGGTGC
bambusa          CAACGCCCATGAAGCTCCGGCGGAAAGGGAAGGCCTGCCAGGCCGTATGCCCATGGGTGC

phoenix          AGGATTCTTCGAAAAAGCGCGGGCTGACTCGGAGACCTGGGACCTTGGCTTAGCAACGAA
brome            ------------------------------------------------------------
rice             AGGATTCTTCGAAAAAGCGCGGGCTGACTCGGAGACCTGAGACCTTGGCTTAGCTACGAA
ferrocalmus      AGGATTCTTCGAAAAAGCGCGGGCTGACTCGGAGACCTGAGACCTTGGCTTAGCTACGAA
bambusa          AGGATTCTTCGAAAAAGCGCGGGCTGACTCGGAGACCTGAGACCTTGGCTTAGCTACGAA

phoenix          TGAAGGG----AAGCTCGAAGAGCTTTCTCCGCCAGCGGCTTATGTAGTGGTCGGCCAAC
brome            ------------------------------------------------------------
rice             TGAAGGGGAGAAAGCTCTTCGAGCTTTCTCCGCCAGCGGCTTATGTAGTGGTCGGCCAAT
ferrocalmus      TGAAGGGGAGAAAGCTCTTCGAGCTTTCTCCGCCAGCGGCTTATGTAGTGGTCGGCCAAT
bambusa          TGAAGGGGAGAAAGCTCTTCGAGCTTTCTCCGCCAGCGGCTTATGTAGTGGTCGGCCAAT

phoenix          TAAAGCTCGCTAAGCTTCGCTTCCCTCCCTTATGAAACTTCATGAAGTAGTCGGCATTCT
brome            ------------------------------------------------------------
rice             TAAAGCTCGCTAAGCTTCGCTTCCC-CCCTTATGAAAA-----AAAGTAGTCGGCATTCT
ferrocalmus      TAAAGCTCGCTAAGCTTCGCTTCCC-CCCTTATGAAAA-----AAAGTAGTCGGCATTCT
bambusa          TAAAGCTCGCTAAGCTTCGCTTCCC-CCCTTATGAAAA-----AAAGTAGTCGGCATTCT

phoenix          ATAAGCGACTTGTCGAGTTTACAAAGCTTTGCTTCTTGTAGTCGGCCCTCCATGCCTCCC
brome            ------------------------------------------------------------
rice             ATAAGCGACTTGTCGAGTTTACAAAGCTTTGCTTCTTGTAGTTGGCCCTCCATGCCTCCC
ferrocalmus      ATAAGCGACTTGTCGAGTTTACAAAGCTTTGCTTCTTGTAGTTGGCCCTCCATGCCTCCC
bambusa          ATAAGCGACTTGTCGAGTTTACAAAGCTTTGCTTCTTGTAGTTGGCCCTCCATGCCTCCC

phoenix          TTCATTTGCTTGCCTCCTTTCAGCCCAGAATGCCTACTGACGTTAAGCTAACCGCCAGAA
brome            ------------------------------------------------------------
rice             TTCATTTGCTTGCCTCCTTCCAGCCCAGAATGCCTACTGACGTTAAGCTAACCGCCAGAA
ferrocalmus      TTCATTTGCTTGCCTCCTTCCAGCCCAGAATGCCTACTGACGTTAAGCTAACCGCCAGAA
bambusa          TTCATTTGCTTGCCTCCTTCCAGCCCAGAATGCCTACTGACGTTAAGCTAACCGCCAGAA

phoenix          GCGACACCCGAAGGGTGAGCTTCTGGCGCAGGAGGCCAAGCATTCTGCCAGACGTCCCGG
brome            ------------------------------------------------------------
rice             GCGACACCCGAAGGGTGAGCTTCTGGCGCAGGAGGCCAAGCATTCTGCCAGACGCCCC--
ferrocalmus      GCGACACCCGAAGGGTGAGCTTCTGGCGCAGGAGGCCAAGCATTCTGCCAGACGTCCC--
bambusa          GCGACACCCGAAGGGTGAGCTTCTGGCGCAGGAGGCCAAGCATTCTGCCAGACGTCCC--

phoenix          CCTGGGAGCCCCGTTCGCCTTTACTATAGTCTTTAGTTTTTCTTCATTTTATTA-AGTAG
brome            ------------------------------------------------------------
rice             ---GGGAGCCCCGTTTGCCTTTACTAGAGTCTTGAGTTTTTCTTTCTTTTTTCATTATAG
ferrocalmus      ---GGGAGCCCCGTTTGCCTTTACTAGAGTCTTGAGTTTTTCTTTCTTTTGTCATTATAG
bambusa          ---GGGAGCCCCGTTTGCCTTTACTAGAGTCTTGAGTTTTTCTTTCTTTTTTCATTATAG
```

```
phoenix          CTAAGTTTCAAAAGCAAAGGTGAACAGCCCCCTGTCCTTTCTTTATA-GTAGTCAAGGGC
brome            ------------------------------------------------------------
rice             CTAAGTTTCAAAAGCAAAGGTGAACAGCCCCCTGTCCTTTAGATTTGCATAGTCAAGGGC
ferrocalmus      CTAAGTTTCAAAAGCAAAGGTGAACAGCCCCCTGTCCTTTAGATTTGCATAGTCAAGGGC
bambusa          CTAAGTTTCAAAAGCAAAGGTGAACAGCCCCCTGTCCTTTAGATTTGCATAGTCAAGGGC


phoenix          TTATAAGAAAAGTCAGGAAGGAGGCATGGAGGTGGGAAGGCCGAC-----CACTACACGG
brome            ------------------------------------------------------------
rice             TTATAAGAAAAGTAAGGAAGGAGGCATGGAGATGGGAAGGCCGACCACTACACTACACGG
ferrocalmus      TTATAAGAAAAGTAAGGAAGGAGGCATGGAGATGGGAAGGCCGACCACTACACTACACGG
bambusa          TTATAAGAAAAGTAAGGAAGGAGGCATGGAGATGGGAAGGCCGACCACTACACTACACGG


phoenix          ATAGATTTCTATACCAAGTCATGAGCGATAGCGAAGCCAAGCCGCCGCCTATAGGCGAAG
brome            -------------------------------------------GCCGCCTATAGGCGAAG
rice             AGAAATTTCTATACCAAGTCATGAGCGATAGCGAA--------GCCGCCTATAGGCGAAG
ferrocalmus      AGAGATTTCTATACCAAGTCATGAGCGATAGCGAAGCCAAGCCGCCGCCTATAGGCGAAG
bambusa          AGAGATTTCTATACCAAGTCATGAGCGATAGCGAAGCCAAGCCGCCGCCTATAGGCGAAG
                                                            *****************

phoenix          GGACGAAAGCCCGGCAAAGGTTAGACCTGATGGGAAAAAGTACGAAAAAAAAATAGTACG
brome            GGCCGAAAGCCC-------ATTAGACCTGATGGGAAAAAGTATGCAAAA-----AGTACG
rice             GGACGAAAGCCC-------GTTAGACCTGATGGGAAAAAGTATGAAAAA-----AGTACG
ferrocalmus      GGACGAAAGCCC-------GTTAGACCTGATGGGAAAAAGTATGAAAAA-----AGTACG
bambusa          GGACGAAAGCCC-------GTTAGACCTGATGGGAAAAAGTATGAAAAA-----AGTACG
                 ** *********        ********************** * ****     ******
```

Missing region in rice only

```
phoenix          TTAAGGTTACGAAGTCACTTAGCCGAGCCGAGGCTAAGGTTATGGAATTACAGAGTAGGT
brome            TTAAGGTTACGAAGTCACTTAGCCTCGGCTCGGCTAAAGTTATGAAATTACAGAGTAGGT
rice             TAAAGGTTAC--------------------------------------------------
ferrocalmus      TTAAGGTTACGAAGTCACTTAGCCTCGGCTCGGCTAAGGTTATGAAATGACAGAGTAGGT
bambusa          TTAAGGTTACGAAGTCACTTAGCCTCGGCTCGGCTAAGGTTATGAAATGACAGAGTAGGT
                 * ********

phoenix          TTTAGGTAAAGTGAAACTAAGGAACTGGCTTAGCTGTTCTACAAAGGAGAAAAGCTTTTG
brome            TTTTTTT-AAGTAATACTAAAGAACTGGAAAAACAGTTCGGAAAAGGAGAAAAGCTAAAG
rice             ------------------------------------------------------------
ferrocalmus      TTTTTGT-AAGTAATACTAAGGAACTGGAAAAACAGTTCGGAAAAGGAGAAAAGCTAAGG
bambusa          TTTTTGT-AAGTAATACTAAGGAACTGGAAAAAAAGTTCGGAAAAGGAGAAAAGCTAAGG
```

```
phoenix          TTAAAGAGTCACTTATCCGTCTACAAAGGGAAAGGCGTCGGTACGGAGTCACGTCAGCTG
brome            TTACAGACTCACTTATCCGTCTACAAAGGGAAAGGCGTCGGTACGGAGTCACGTCAGCTG
rice             ----AGACTCACTTATCCGTCTACAAAGGGAAAGGCGTCGGT--------ACGTCAGCTG
ferrocalmus      TTACAGACTCACTTATCCGTCTACAAAGGGAAAGGCGTCGGTACGGAGTCACGTCAGCTG
bambusa          TTACAGACTCACTTATCCGTCTACAAAGGGAAAGGCGTCGGTACGGAGTCACGTCAGCTG
                     *** **********************************        **********

phoenix          TGGATATAGACTAGGCGTCGGAACGGAGTCTTAAAGTATTCACCGAGACTACACTAAGGA
brome            TGGATATAGACTAGGCTAAGGAACGGAGTCTTAAAGTATTTACCGAGACTACACTAAGGA
rice             TGGATATAGACTAGGCTAAGGAACGGAGTCTTAAAGTAGTTACCGAGACTACACTAAGGA
ferrocalmus      TGGATATAGACTAGGCTAAGGAACGGAGTCTTAAAGTATTTACCGAGACTACACTAAGGA
bambusa          TGGATATAGACTAGGCTAAGGAACGGAGTCTTAAAGTATTTACCGAGACTACACTAAGGA
                 ****************   ******************* * *******************

phoenix          AGAAGTCAGCGAAAGTGAGCTCGTAACTAAGAAAGCCGGTATCAGAAACGAAGCCCTCTA
brome            AGAAGTCAGCGAAAGTGAGCTCGTAACTAAGAAAGCCGGTATCAGAAACGAAGCCCTT--
rice             AGAAGTCAGCGAAAGTGAGCTCGTAACTAAGAAAGCCGGTATCAGAAACGAAGCCCTT--
ferrocalmus      AGAAGTCAGCGAAAGTGAGCTCGTAACTAAGAAAGCCGGTATCAGAAACGAAGCCCTT--
bambusa          AGAAGTCAGCGAAAGTGAGCTCGTAACTAAGAAAGCCGGTATCAGAAACGAAGCCCTT--
                 **********************************************************
```

```
phoenix          TAATAAAAAAAAGCAAAGAAGTCAAGGAACGAAGCTGCTTCTCTAATAACCCCGTTGAAT
brome            CCAAAATCAAAAGCAAAGAAGTCAAGGAACGAAGCTGCTTCTCTAATAGCCCCGTTGAAT
rice             CCAAAATAAAAAGCAAAGAAGTCAAGGAACGAAGCAGCTTCTCTAATAACCCCGTTGAAT
ferrocalmus      CCAAAATCAAAAGCAAAGAAGTCAAGGAACGAAGCTGCTTCTCTAATAGCCCCGTTGAAT
bambusa          CCAAAATCAAAAGCAAAGAAGTCAAGGAACGAAGCTGCTTCTCTAATAGCCCCGTTGAAT
                   * **  *************************** ************ ***********

phoenix          AGGAGGGCGAAGGCTTTTTCAAAAAGTTTGATAG-AGGTGGGCTTCGACCTTCTTAGGAA
brome            AGGAGGGCGAAGGCTTTTTCGAACTTTTTTATAG-AGGGGGGCTTCGACCTTCTTAGGAA
rice             AGGAGGGCGAAGGCTTTTTCGAACTTATTTATAG--AGGGGGCTTCGACCTTCTTAGGAA
ferrocalmus      AGGAGGGCGAAGG---TTTCGAACTTTTTTATAG-AGGGGGGCTTCGACCTTCTTAGGAA
bambusa          AGGAGGGCGAAGG---TTTCGAACTTTTTTATAGAAGGGGGGCTTCGACCTTCTTAGGAA
                 *************   **** **    ** ****   * *********************

phoenix          GAGCCGTACGAGGCAGCTCACGTACGGTTCGGGAGCCGAGCCACCGCACAGGGGCTTAGG
brome            GAGCCGTACGAGGCTGCTCACGTACGGTTCGGGAGCCGAGCCACTGCGCAGGGGCTTAGG
rice             GAGCCGTACGAGGCTGCTCACGTACGGTTCGGGAGCCGAGCCACTGCACAGGGGCTTAGG
ferrocalmus      GAGCCGTACGAGGCTGCTCACGTACGGTTCGGGAGCCGAGCCACTGCACAGGGGCTTAGG
bambusa          GAGCCGTACGAGGCTGCTCACGTACGGTTCGGGAGCCGAGCCACTGCACAGGGGCTTAGG
                 ************** ***************************** ** ************

phoenix          TCAACACTTATATATTAGCCAGTCATCAATTGGAAGCGGGTAAGATGGTGATGAATTGCG
brome            TCAACACTTATATATTAGCCAGTCATCAATTGGAAGCGGGTAATATGGTGAGAAATTGCG
rice             TCAACACTTATATATTAGCCAGTCATCAATTGGAAGCGGGTAATATGGTGATAAATTGCG
ferrocalmus      TCAACACTTATATATTAGCCAGTCATCAATTGGAAGCGGGTAATATGGTGATAAATTGCG
bambusa          TCAACACTTATATATTAGCCAGTCATCAATTGGAAGCGGGTAATATGGTGATAAATTGCG
                 ******************************************* *******  *******

phoenix          ATTGGTCCAAACCTTCTAAAAGCGGCTTCTTGCGACCTGCCCAGAATGCCCATACATACC
brome            ATTGCTCCAAACCTTCTAAAAGCGGCTTCTTGCGACCTGCCCAGAATGCTCATACATACC
rice             ATTGCTCCAAACCTTCTAAAAGCGGCTTCTTGCGACCTGCCCAGAATGCTCATACATACC
ferrocalmus      ATTGCTCCAAACCTTCTAAAAGCGGCTTCTTGCGACCTGCCCAGAATGCTCATACATACC
bambusa          ATTGCTCCAAACCTTCTAAAAGCGGCTTCTTGCGACCTGCCCAGAATGCTCATACATACC
                 **** ******************************************** **********

phoenix          TTCGGTTTCAAGACCTTGTTCGCACA---GCGAATAAAGGTCGGGTTGAAGGGGGCAGTC
brome            TTCGGTTCCAAGAGCTTGTTCGCACAGCGGCGAATAAAGATCGGGTTGAAGGGGGCAGTC
rice             TTCGGTTCCAAGAGCTTGGACGCACA---GTGAATAAAGGTCGGGTTGAAGGGGGCAGTC
ferrocalmus      TTCGGTTCCAAGAGCTTGTTCGCACA---GCGAATAAAGGTCGGGTTGAAGGGGGCAGTC
bambusa          TTCGGTTCCAAGAGCTTGTTCGCACA---GCGAATAAAGGTCGGGTTGAAGGGGGCAGTC
                 ******* ***** ****  ******   * ******** ********************

phoenix          AGCTGGCTGCTTCTTGGCCACGCCCCCCTGCTTATAGATACGAGATACTTGATCTAAATT
brome            AGCTGGCTGCTTCTTGGCCACGCCCCCCTGCTTATAGACACGAGATATTGGATCTCAATT
rice             AGCTGGCAGCTTCTTGGCCACGCCCCCCTGCTTATAGACACGAGATATTGGATCTAAATT
ferrocalmus      AGCTGGCTGCTTCTTGGCCACGCCCCCCTGCTTATAGATACGAGATATTGGATCTAAATT
bambusa          AGCTGGCTGCTTCTTGGCCACGCCCCCCTGCTTATAGATACGAGATATTGGATCTAAATT
                 ******* ****************************** ******** * ***** ****

phoenix          CTAAAGTAGGAAATTGCATACCATTAGCTGATATACGTATGGGAACATGGGTACATGATA
brome            CCAAAGTAGGAAATAGCATACCATTAGCTGATATACGTATGGGAACATGGGTACATGATA
rice             CAAAAGTAGGAAATAGCATACCATTAGCTGATATACGTATGGGAACATGGGTACATGATA
ferrocalmus      CAAAAGTAGGAAATAGCATACCATTAGCTGATATACGTATGGGAACATGGGTACATGATA
bambusa          CAAAAGTAGGAAATAGCATACCATTAGCTGATATACGTATGGGAACATGGGTACATGATA
                 * ************ *********************************************

phoenix          TTGAATGTCATCCAGGTCAAGGCGCAAAGCTGGCTCGAGCCGCAGGAACTTATGCTAAAA
brome            TTGAATGTCATCCAGGTCAAGGCGCAAAGCTGGCTCGGGCCGCAGGAACTTATGCTAAAA
rice             TTGAATGTCATCCAGGTCAAGGCGCAAAGCTGGCTCGGGCCGCAGGAACTTATGCTAAAA
ferrocalmus      TTGAATGTCATCCAGGTCAAGGCGCAAAGCTGGCTCGGGCCGCAGGAACTTATGCTAAAA
bambusa          TTGAATGTCATCCAGGTCAAGGCGCAAAGCTGGCTCGGGCCGCAGGAACTTATGCTAAAA
                 ************************************* **********************
```

```
phoenix          TAATTAAGGAGCCAGCCCCACAATGTCTTGTGCGGCTACCTTCGGGTGTTGAAAAACTCA
brome            TAATTAAGGAGCCAGCCCCACAATGTCTTGTGCGGCTACCATCGGGTGTTGAAAAACTCA
rice             TAATTAAGGAGCCAGCCCCACAATGTCTTGTGCGGCTACCATCGGGTGTTGAAAAATTCA
ferrocalmus      TAATTAAGGAGCCAGCCCCACAATGTCTTGTGCGGCTACCATCGGGTGTTGAAAAACTCA
bambusa          TAATTAAGGAGCCAGCCCCACAATGTCTTGTGCGGCTACCATCGGGTGTTGAAAAACTCA
                 **************************************** *************** ***

phoenix          TAGATTCCCGATGCCGAGCTACTATTGGTATAGTTTCCAATCCCAACCATGGTGCACGTA
brome            TAGATTCCCGATGCCAAGCTACTATTGGTATAGTTTCCAATCCCAACCATGGTGCACGTA
rice             TAGATTCCCGATGCCGAGCTACTATTGGTATAGTTTCCAATCCCAACCATGGTGCACGTA
ferrocalmus      TAGATTCCCGATGCCGAGCTACTATTGGTATAGTTTCCAATCCCAACCATGGTGCACGTA
bambusa          TAGATTCCCGATGCCGAGCTACTATTGGTATAGTTTCCAATCCCAACCATGGTGCACGTA
                 *************** ********************************************

phoenix          AGCTTAGAAAAGCAGGACAAAGCCGGTGGTTAGGCAGACGCCCCATTGTTCGTGGTGTTG
brome            AGCTTAGAAAAGCAGGACAAAGCCGGTGGTTAGGCAGACGCCCCATTGTTCGTGGTGTTG
rice             AGCTTAGAAAAGCAGGACAAAGCCGGTGGTCAGGCAGACGCCCCATTGTTCGTGGTGTTG
ferrocalmus      AGCTTAGAAAAGCAGGACAAAGCCGGTGGTTAGGCAGACGCCCCATTGTTCGTGGTGTTG
bambusa          AGCTTAGAAAAGCAGGACAAAGCCGGTGGTTAGGCAGACGCCCCATTGTTCGTGGTGTTG
                 ****************************** *****************************

phoenix          CAATGAATCCAGTGGATCATCCTCATGGAGGAGGTGAGGGGCGCACGAAAGGAGGTAGAC
brome            CAATGAATCCAGTGGATCATCCTCATGGAGGAGGTGAGGGGCGCACGAAAGGAGGTAGAC
rice             CAATGAATCCAGTGGATCATCCTCATGGAGGAGGTGAGGGGCGCACGAAAGGAGGTAGAC
ferrocalmus      CAATGAATCCAGTGGATCATCCTCATGGAGGAGGTGAGGGGCGCACGAAAGGAGGTAGAC
bambusa          CAATGAATCCAGTGGATCATCCTCATGGAGGAGGTGAGGGGCGCACGAAAGGAGGTAGAC
                 ************************************************************

phoenix          CTTCGGTGTCACCTTGGGGGAAGCCCACCAAAGCAGGATTTCGGGCAGTAGTGGGGGTGG
brome            CTTCGGTGTCACCTTGGGGGAAGCCCACCAAAGCAGGATTTCGGGCAGTAGTGGGGGTGG
rice             CTTCGGTGTCACCTTGGGGGAAGCCCACCAAAGCAGGATTTCGGGCAGGAGTGGGGGTGG
ferrocalmus      CTTCGGTGTCACCTTGGGGGAAGCCCACCAAAGCAGGATTTCGGGCAGTAGTGGGGGTGG
bambusa          CTTCGGTGTCACCTTGGGGGAAGCCCACCAAAGCAGGATTTCGGGCAGTAGTGGGGGTGG
                 ************************************************ ***********

phoenix          GGAAACGCAGAATTTAG
brome            GGAAACGCAGAATTTAG
rice             GGAAACGCAGAATTTAG
ferrocalmus      GGAAACGCAGAATTTAG
bambusa          GGAAACGCAGAATTTAG
                 *****************
```

(ii) Brome *rpl2* exon 1 region comparison between brome and bambusa. Inverted repeats present in bambusa and brome are shown in red and underlined.

```
brome           ATGAGACAAAGCCTAAAGGGGAGAGCACTTAGACATTTCACTTTGAGTACGGGGAAGTCC
bamboo          ATGAGACAAAGCATAAAGGGGAGAGCACTTAGACATTTCACTTTGAGTACGGGGAAGTCC
                ************ ***********************************************

brome           GCAGGGAGGAATTCTTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
bamboo          GCAGGGAGGAATTCTTCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGA
                ************************************************************

brome           TTGCTGCGAAAAATTGACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATC
bamboo          TTGCAGCGAAAAATTGACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATC
                **** *******************************************************

brome           GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTG
bamboo          GAATATGACCCTAATCGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTG
                ************************************************************

brome           CCCGGCCGCCAGAGGAAATTCAAGACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTC
bamboo          CCCGGCCGCCAGAGGAAATTCAAGACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTC
                ************************************************************

brome           GAATCCACCACGGCCACTATCTTTTGCCTTTTTTCGTTCTCTTCCCTGCCCGGGAAAGTG
bamboo          GAATCCACCACGGCCACTATCTTTTGCCTTTTTTCGTTCTCTTCCCTGCCCGGGAAAGTG
                ************************************************************

brome           GATCAAAGA---------------------------------------------------
bamboo          GATCAAAGAAAGGTAGCTTGCTCCCAAGGCAAGTGCTTGCTTACGCTTTATGTAGTGGTC
                *********

brome           ------------------------------------------------------------
bamboo          GGCCTTCCTACCTTCATGCCTCCTTGGTCGAAGAGCCAAGCCTGCGCAGGAAGCAAACAA

brome           ------------------------------------------------------------
bamboo          ACTTGCGCGAAGGACGTTTTCTTCTCTGCCCTGTCCTCTCCCTTGGCCCAGGGAGAGACT

brome           ------------------------------------------------------------
bamboo          GCATCCCTTTCCTTCGGTAGCTCTTTAGGTTTCCCAAGGATAGCGGTAGCTGGGGCAAAG

brome           ------------------------------------------------------------
bamboo          CCCGCTTTCTTCGCTCCGCGAATGAGAGAGAAGAAAATCGGAAAAAAGACGTTTTCTCTT

brome           ------------------------------------------------------------
bamboo          TGCGAGATCCGAAAGTGGAGAACGCATTGCGTTCTCTGGGCACATAGGATCAAACGTAAA

brome           ------------------------------------------------------------
bamboo          GCAGCGCTTTCTTGGCAGAGTTTGAGGCAGCAAAAAACTTTAGAGCTTGTTGGAGCTGCT

brome           -------------------------------AAGGTAGCTTGCTCCCAAGGCAAGTGCTT
bamboo          GAGCATAACGAATCGAAGCTGAAGGCGGATCAAGGTAGCTTGCTCCCAAGGCAAGTGCTT
                                               *****************************

brome           GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCTTCATGCCTCTAGAAGCTTTGTAGAA
bamboo          GCTTACGCTTTATGTAGTGGTCGGCCTTCCTACCTTCATGCCTCTAGAAGCTTCTACAAA
                *****************************************************     **

brome           GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCTGCCAAGCCGCCTATAGGC
bamboo          GCTTTGCTTCCGGTAGAAGCTAGTCGCTTCGGTAGCTTGCCTGCCAAGCCGCCTATAGGC
                ************************************************************
```

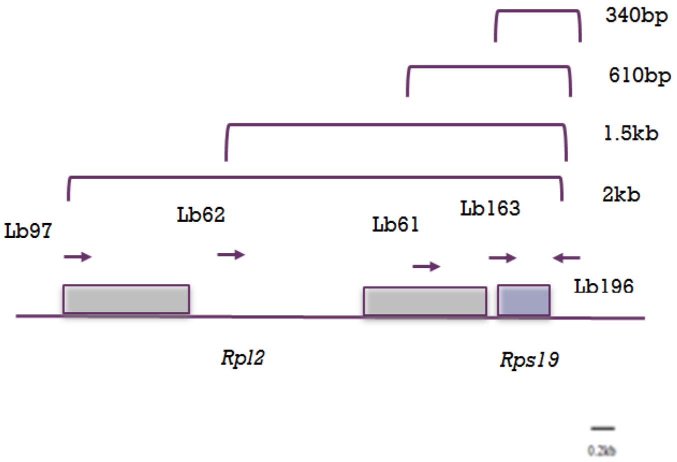
(iii) Amino acid alignment of pseudo*rpl2* gene in brome and functional *rpl2* in other grasses.

```
bromerpl2       MRQSLKGRALRHFTLSTGKSAGRNSSGRITVFHRGGGSKRLLRKIDLKRSTSSIGIVERI
ricerpl2        MRQSIKGRALRHFTLSTGKSAGRNSSGRITVFHRGGGSKRLQRKIDLKRSTSSIGIVERI
bambusarpl2     MRQSIKGRALRHFTLSTGKSAGRNSSGRITVFHRGGGSKRLQRKIDLKRSTSSIGIVERI
                ****:************************************ ******************

bromerpl2       EYDPNRSSRIALVRWIEGVLPGRQRKFKTIEEFAPPRKILESTTATIFCLFSFSSLPGKV
ricerpl2        EYDPNRSSRIALVRWIEGVLPGRQRKFKTIEEFALPRKILESTTATIFCLFSFSSL----
bambusarpl2     EYDPNRSSRIALVRWIEGVLPGRQRKFKTIEEFAPPRKILESTTATIFCLFSFSSLPGKV
                ********************************** *********************

bromerpl2       DQRKVACSQGKCLLTLYVVVGLPTFMPLEAL
ricerpl2                                                           SSPLAQGET
bambusarpl2     DQRKVACSQGKCLLTLYVVVGLPTFMPPWSKSQACAGSKQTCAKDVFFSALSSPLAQGET


bromerpl2
ricerpl2        ASLSFGSSLGFPRIAVAGAKPAFFAERMREKKIGKKTFSLCEIRKWRTHCVLWAHRIKRK
bambusarpl2     ASLSFGSSLGFPRIAVAGAKPAFFAPRMREKKIGKKTFSLCEIRKWRTHCVLWAHRIKRK


bromerpl2
ricerpl2        AALSWQSLRQQKTLELVGAAEHNESKLKADQGSLLPRQVLAYALCSGRPSYLHASRSFYK
bambusarpl2     AALSWQSLRQQKTLELVGAAEHNESKLKADQGSLLPRQVLAYALCSGRPSYLHASRSFYK


bromerpl2                                          TYILASHQLEAGNMVRNCDCSKPSK
ricerpl2        ALLPVEASRFGSLPAKPPIGEGPKDGAYKVDRAPVTYILASHQLEAGNMVINCDCSKPSK
bambusarpl2     ALLPVEASRFGSLPAKPPIGEGTKDGAYKVDRAPVTYILASHQLEAGNMVINCDCSKPSK
                  :*:**                           :*************** *********

bromerpl2       SGFLRPAQNAHTYLRFQELVRTAANKDRVEGGSQLAASWPRPPAYRHEILDLNSKVGNSI
ricerpl2        SGFLRPAQNAHTYLRFQELGRT-VNKGRVEGGSQLAASWPRPPAYRHEILDLNSKVGNSI
bambusarpl2     SGFLRPAQNAHTYLRFQELVRT-ANKGRVEGGSQLAASWPRPPAYRYEILDLNSKVGNSI
                ******************* **  .**.*******************:*************

bromerpl2       PLADIRMGTWVHDIECHPGQGAKLARAAGTYAKIIKEPAPQCLVRLPSGVEKLIDSRCQA
ricerpl2        PLADIRMGTWVHDIECHPGQGAKLARAAGTYAKIIKEPAPQCLVRLPSGVEKFIDSRCRA
bambusarpl2     PLADIRMGTWVHDIECHPGQGAKLARAAGTYAKIIKEPAPQCLVRLPSGVEKLIDSRCRA
                ****************************************************:*****.*


bromerpl2       TIGIVSNPNHGARKLRKAGQSRWLGRRPIVRGVAMNPVDHPHGGGEGRTKGGRPSVSPWG
ricerpl2        TIGIVSNPNHGARKLRKAGQSRWSGRRPIVRGVAMNPVDHPHGGGEGRTKGGRPSVSPWG
bambusarpl2     TIGIVSNPNHGARKLRKAGQSRWLGRRPIVRGVAMNPVDHPHGGGEGRTKGGRPSVSPWG
                *********************** ************************************

bromerpl2       KPTKAGFRAVVGVGKRRI
ricerpl2        KPTKAGFRAGVGVGKRRI
bambusarpl2     KPTKAGFRAVVGVGKRRI
                ********* ********
```

*Blue highlights show the amino acid changes that are shared between brome and bambusa; green highlight shows shared amino acids between rice and bamboo. Red boxes show drastic amino acid changes between brome and other grasses. “MPLEA” in brome shown in grey highlight shows the frame shift region due to the indel.

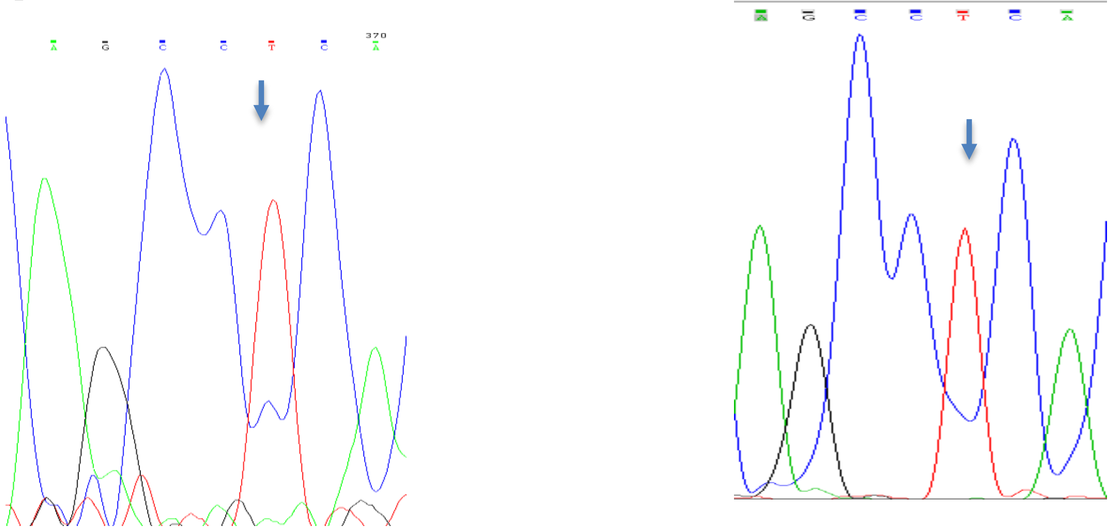
(iv)Brome mitochondrial *rps19* and pseudo*rpl2* are edited. Chromatogram tracings of edit positions within *rpl2* and *rps19* transcripts

a) Schematic of the *rpl2-rps19* locus with oligomers used for the RT-PCR experiment and its corresponding length of transcripts are shown.



b) The tracings from two different transcripts 97-196 and 62-196 show the single edit position in *rpl2.*

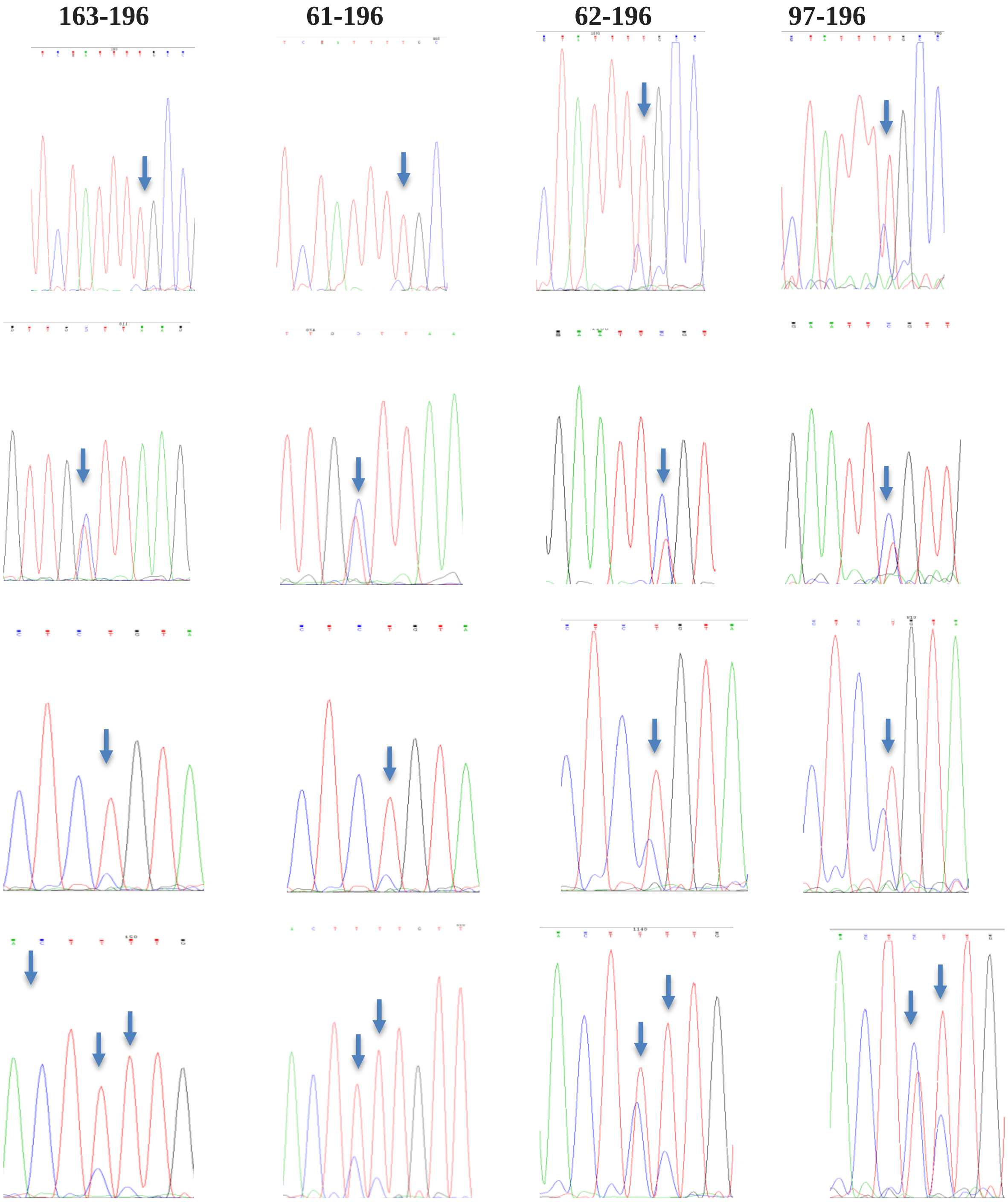
Rpl2 edit



97-196

62-196

C) Tracings of all four edits present in *rps19* gene are shown from each transcript (163-196; 61-196; 62-196; and 97-196). Edit positions in the tracings are shown by a purple arrow



163-196 61-196 62-196 97-196

Appendix B: Break point in homology between brome and rice *rpl2* gene upstream of start codon

```
brome           GAACTCGTCAGTGTAGGATGTGGTGCTCCAAGAGTGACGGAGTGGGAGTCACATTCGCTG
rice            -----------------------CGATCTAGCAATTATGGAATCTATATTCTGTTTACTA
                                        * ** *  * * * *** *     *    **  **

brome           GTCTCGACTAAAGGTTCGACGGGTTTACATCCTGATCTGCTATGCTCGTT----------
rice            AATCACATGAAATTCTCGGGAACTCCACATACGTATTTCATATATGTATTTCATACATAT
                      *  ***   ***     *  **** *  ** *  ***     **

brome           --ATGCGGACGACTTCCTTGTAGCTATCAAGAGCCCCATCGCCTCTACTTTCACGGTTTG
rice            GAATAGAGACAATTTC---------AACGAAAGTTCGA-----------------ATTTG
                  **   *** * ***         * * * **  * *                  ****

brome           CCGGCGAGCATTGGAAATGTTGAGTCCATTTGGTTATGACATTCTGGCCTTGGTTGGTAA
rice            CCAG-----------------GGGTCCAAATGG-AATGAAA-------ATGGGTATATAA
                ** *                 * *****  ***  **** *        * ***   ***

brome           AACATCTCAGGGGATCCGGACCCTACGAGTGAGAAGGATAAGGGGGAGTAGTAGGAGGAG
rice            AAGACCAACGAAA---------CGACGAGTGAGAAGGATAAGGGGGAGTAGTAGGAGGAG
                ** * *   *            * ************************************

brome           CTTTTATTGAAGTAGAGGTTTCATCGAGTGATGGCGAGAGATAGAATGAGACAAAGCCTA
rice            CTTTCATTGAAGTAGAGGTTTCATCGAGTGATGACGAGAGATAGAATGAGACAAAGCATA
                **** **************************** *********************** **

brome           AAGGGGAGAGCACTTAGACATTTCACTTTGAGTACGGGGAAGTCCGCAGGGAGGAATTCT
rice            AAGGGGAGAGCGCTTAGACATTTCACTTTGAGTACGGGGAAGTCCGCAGGGAGGAATTCT
                *********** ************************************************

brome           TCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGATTGCTGCGAAAAATT
rice            TCCGGGCGTATTACTGTTTTTCACCGAGGGGGTGGATCGAAGCGATTGCAGCGAAAAATT
                ************************************************* **********

brome           GACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATCGAATATGACCCTAAT
rice            GACCTTAAACGAAGCACTTCGTCTATTGGCATTGTGGAAAGGATCGAATATGACCCTAAT
                ************************************************************

brome           CGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTGCCCGGCCGCCAGAGG
rice            CGTTCTTCTCGGATCGCTCTAGTACGATGGATCGAAGGGGTGCTGCCCGGCCGCCAGAGG
                ************************************************************

brome           AAATTCAAGACGATAGAAGAGTTCGCTCCGCCGCGTAAGATCCTCGAATCCACCACGGCC
rice            AAATTCAAGACGATAGAAGAGTTCGCTCTGCCGCGTAAGATCCTCGAATCCACCACGGCC
```

Appendix C: Nucleotide sequence alignment of pseudo-*rpl2-nad1-nad6* linkage in brome, barley and rye

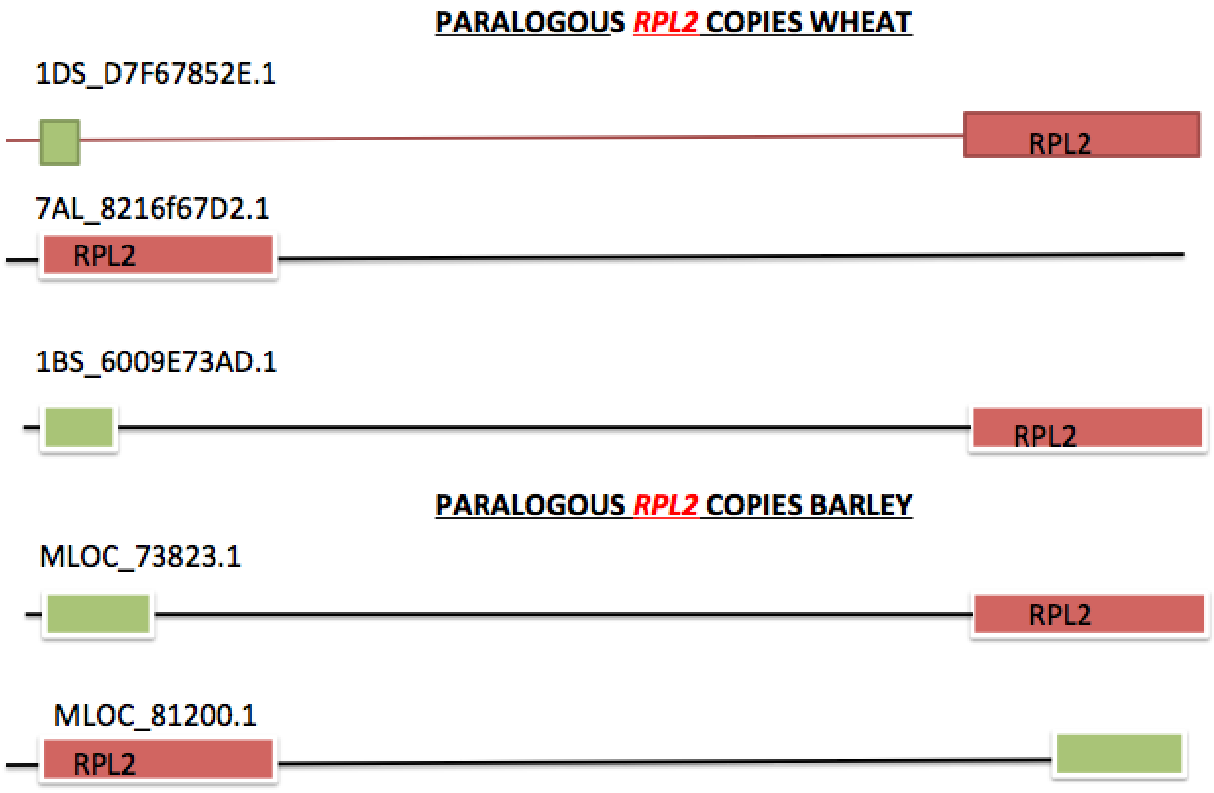
```
 brome853-485clone       GGGGGTGGGGAAAGTTTGGATTTCTGAAATGAGAAACTTTTGCATTCTCAAGATAGAACG
Barley822-485(lb        ------------------------------------TTTTTCTGCTC---AGTTAGAACG
Rye822-485(lb           -------------------------------------TTTTCTGCTC---AGTTAGAACG
                                                             ****    **   ** *******

brome853-485clone       TTTTTTTTCTAAGTTAATTGGCCAGTTCCTTTTAAAAGGAGACTGG---GAGTTCGTCAG
Barley822-485(lb        TAGTTTTTCTAAGTTAATTGGCCAGTTCCTTTTAAAAGGAGACTGGTTTCTTTTTATAAG
Rye822-485(lb           TAGTTTTTCGAAGTGAATTGGCCAGTTCCTTTTAAAAGGAGACTGGTTTATTTTTATAAG
                        *  ****** **** *******************************      **  * **

brome853-485clone       CGACTGGTGGCCGTACTTTCTGTTATTTTGTACGATAATTCTTATTATCTTTCTTTTCTT
Barley822-485(lb        CGAT---TGGCTTCACTTTA---GATTTGTTATCGTAATTTCTCTTTTATTTATGGGTTG
Rye822-485(lb           CGAT---TGGCTTCACTTTA---GATTTGTTATCGTAATTTCTCTTTTATTTATGGGTTG
```

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

brome853-485clone      TTGGTTCCTCTGTGGTTTCCGGAAAAACCAAAAGAGAAATCCAATCGGGAAGGGGCTAAT
Barley822-485(lb       TTATTTATTATTTCGATTTTTGATAAACGAAACG---AATCCAAGCCCGACAAGTAGAAT
Rye822-485(lb          TTATTTATTATTTCGATTTTTGATAAACGAAACG---AATCCAAGCCCGACAAGTAGAAT
                       **  **  * * * * **   ** **** *** *   ******* *  **   *   ***

brome853-485clone      GGCTTTTGGTGGCATCCTGGGTGTCTTTTTCCGTCTGGATGACGA------CGCGTCTCC
Barley822-485(lb       TCCATTCGGTACCTTTGGGGTTTTCTTTTTCCTTCTGGATGACGACGTCACCGCGTCTCC
Rye822-485(lb          TCCATTCGGTACCTTTGGGGTTTTCTTTTTCCTTCTGGATGACGACGTCACCGCGTCTCC
                         * ** ***  * *   ** * ********* ************      *********

brome853-485clone      TTATTCTTGTTGCGATGGTGACAAGGAAGAAAGCGCCGGAGAGATCATCGAAATTCCCTC
Barley822-485(lb       TTATTGTTATTGCGATGGTGACGAGGAACAAAGGGCCGGGGAGATCATCGAAATTACCTC
Rye822-485(lb          TTATTGTTATTGCGATGGTGACGAGGAACAAAGGGCCGGGGAGATCATCGAAATTACCTC
                       ***** ** ************* ***** **** ***** *************** ****

brome853-485clone      TAGCCCAGAGCCCGAGGG---------CACAACCCCACGCATGGTGGTGGGGCACTTCGA
Barley822-485(lb       TAGCCCGGAGCCAGAGGGGGTCGTCGACACAACGCCAACCCTGGTGGTGGGGCACTTCGA
Rye822-485(lb          TAGCCCGGAGCCAGAGGGGGTCGTCGACACAACGCCAACCCTGGTGGTGGGGCACTTCGA
                       ****** ***** *****         ****** ***  * *******************

brome853-485clone      AAGTCTGGGGCTCCTCCTGGAGATTCCCCCTTTACCGGAAACGCCACCCGCTGGTCCTTC
Barley822-485(lb       AGGTCTGGAGACCCTCCTTGAGATTCCCCCTTTACCGGAAACGCCACCCGCAGGTCCTTC
Rye822-485(lb          AGGTCTGGAGACCCTCCTTGAGATTCCCCCTTTACCGGAAACGCCACCCGCAGGTCCTTC
                       * ****** *  ****** ******************************** ********

brome853-485clone      CTTAGCTCCAGAACCAGAAGAAAGGGAGCAAGCCCCCTCTGTTGGGCTAAACCCCGATCC
Barley822-485(lb       CTTAGCT------CCAGAAGAAAGGGAGCAAGCCCCCTCTGTTGGCCTAAACCTCGATCC
Rye822-485(lb          CTTAGCT------CCAGAAGAAAGGGAGCAAGCCCCCTCTGTTGGCCTAAACCTCGATCC
                       *******      ******************************** ******* ******

brome853-485clone      ATCCCAGAGTAGCCGCCGGCTGGACGGCGGCGAACGGGGGGGAAACTCCTCAAAACGATG
Barley822-485(lb       ATCCCAGAGTAGCCGCCGGCTGGACGGCGGCGAACGGGGGGCAAACTCCTCAAAACGAAG
Rye822-485(lb          ATCCCAGAGTAGCCGCCGGCTGGACGGCGGCGAACGGGGGGCAAACTCCTCAAAACGAAG
                       ***************************************** **************** *

brome853-485clone      TTTG--CGTGTC-------GAGATGGAGGAATCCGACAGGCGGTGGAAAGTTGCGAGACA
Barley822-485(lb       TTCTACTTTGTCTTTGGAAGAATTGGAAGAATCGGACTCTAAGTTTAAAGTAGGAGTAGA
Rye822-485(lb          TTCTACTTTGTCTTTGGAAGAATTGGAAGAATCGGACTCTAAGTTTAAAGTAGGAGTAGA
                       **      ****       **  **** ***** ***     **  ***** *    * *

brome853-485clone      GTCTGCGGAGGCAAGCTTCGAAGGAAATAGGCGGGCAGAGCTACTGCAAGCGCTCGAGGC
Barley822-485(lb       GTCTTTAGAGGCTCGCTTAGAAGGAATAAAGAGAAAAAAGCAACTGCAAGCGTTCGGGGC
Rye822-485(lb          GTCTTTAGAGGCTCGCTTAGAAGGAATAAAGAGAAAAAAGCAACTGCAAGCGTTCGGGGC
                       ****   *****  **** *******  * * *   * *** ********** *** ***

brome853-485clone      TTCTCCTTTTGAGACCGCGAAAGCAGAACTGGTTACCGGAATTGGGGCTAATTCCGAGTT
Barley822-485(lb       TTCTCCTTTTGAGACCGCGAAAGCAGAGCTGGTTACGTCCATTGGGGAAAATTCCCAGTT
Rye822-485(lb          TTCTCCTTTTGAGACCGCGAAAGCAGAGCTGGTTACGTCCATTGGGGAAAATTCCCAGTT
                       *************************** ********    *******  ****** ****

brome853-485clone      TAGACTGGCACTAATTAGGTCTCTCTTTATCA---GCCCCTTGACGGGGCAACGGGGGGC
Barley822-485(lb       TCGTGTGGGACTGCTTTTGTCTCTGGATGTCAAATGCCCCTTGACGGGCAAACGGGAGGC
Rye822-485(lb          TCGTGTGGGACTGCTTTTGTCTCTGGATGTCAAATGCCCCTTGACGGGCAAACGGGAGGC
                        *  *  *** ***  **  ******   * ***   *************   ***

brome853-485clone      TCTCGAAAATACGATAGAG-----------------------------------------
Barley822-485(lb       TCTCGAAAAGACGATAGAGGCCATCTTGAAATGCCAAGGGAAGAAGCCCACAGACTCCCT
Rye822-485(lb          TCTCGAAAAGACGATAGAGGCCATCTTGAAATGCCAAGG---GAAGCCCACAGACTCCCT
                       ********* *********
```

Appendix D: Codon usage pattern of the nuclear located *rpl2* gene in brome, wheat, and rye versus mitochondrial *rpl2* gene in rice

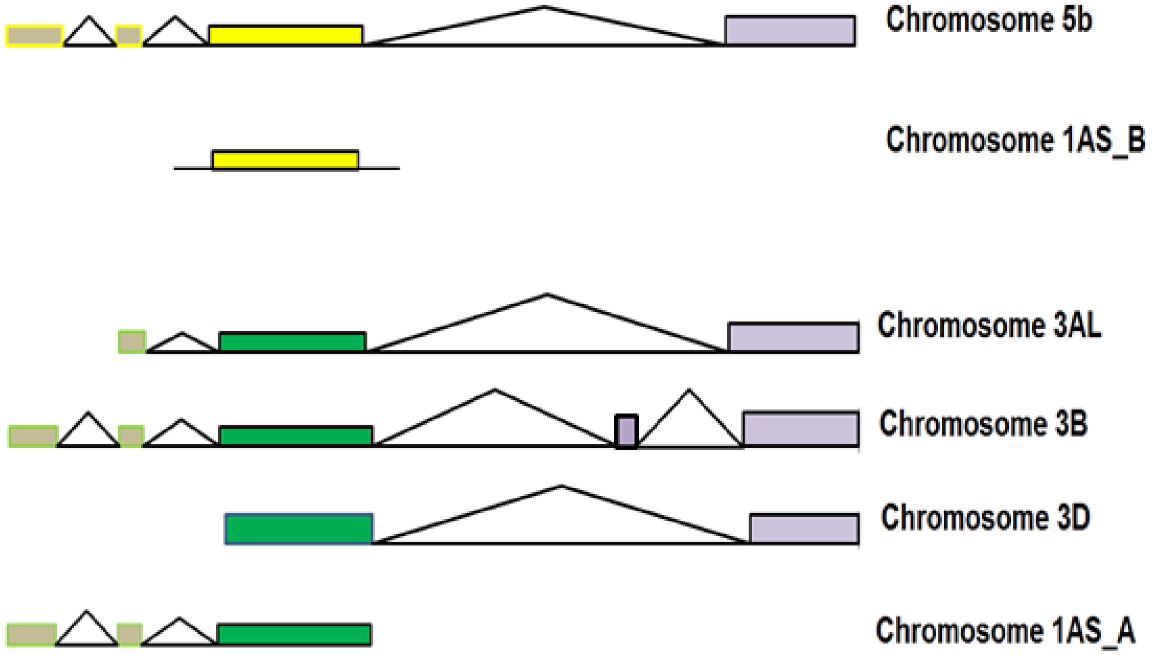
```
(a)
ricemitorpl2          GAAATAGCATACCATTAGCTGATATACGTATGGAACATGGGTACATGATATTGAATGTC
festucarpl2_ncbi      GGAACTGCATGCCGTTAGCTGATATTCGTATGGAACGTGGGTACACAGCATCGAGTGTC
bromerpl2_SA          GAAACTGCATGCCACTAGCTGACATCCGTATGGAACATGGGTGCACAGCATTGAGTTGC
ryerpl2_SA            GAAACTGCATGCCGTTGTCTGACATCCGTATGGAACCTGGGTGCACAGCATTGAACTGC
wheatrpl2_SA          GAAACTGCATGCCGTTGTCTGACATCCGTATGGAACATGGGTGCACAGCATCGAACTGC
barleyrpl2_NCBI       GAAACTGCATGCCATTGTCTGACATCCGCATGGGAACATGGGTGCACAGCATCGAGCTGC
                      * **  **** **  *  **** ** ** ******** ***** **    ** **    *
ricemitorpl2          ATCCAGGTCAAGGCGCAAAGCTGGCTCGGGCCGCAGGAACTTATGCTAAAATAATTAAGG
festucarpl2_ncbi      GGCCTGGCCAAGGGGCACAGCTGGTTCGAGCTGCTGGAACTTACGCTAAGGTGGTCAAGG
bromerpl2_SA          GTCATGGTCAAGGGGCAAAGCTCGTTCGAGCTGCTGGAACTTACGCTAAAGTGGTCAAGG
ryerpl2_SA            GTCATGGCCAAGGCGCAAAGCTCGTCCGGGCCGCTGGAGCCTACGCCAAGGTGGTCAAGG
wheatrpl2_SA          GTCACGGCCAGGGGGCGAAGCTCGTCCGAGCCGCTGGAGCCTACGCCAAGGTGGTCAAGG
barleyrpl2_NCBI       GTCACGGCCAGGGGGCGAAGCTCGTCCGGGCCGCTGGAGCGTACGCCAAGGTGGTGAAGG
                        *  ** ** ** **  **** *  ** ** ** *** * ** ** **  *  * ****

ricemitorpl2          AGCCAGCCCCACAATGTCTTGTGCGGCTACCATCGGGTGTTGAAAAATTCATAGATTCCC
festucarpl2_ncbi      AGTCAGCCTCCCAGTGTCTTGTACGGCTGCCATCAGGTGCTGAGAAACTTATAGATTCCC
bromerpl2_SA          AGTCAGCCTCGCAGTGTCTTGTACGGCTGCCATCGGGTGTTGAGAAACTGATAGACTCCC
ryerpl2_SA            AGTCAGCCACGCAGTGCCTTGTGCGTCTGCCGTCGGGAGTCGAGAAGCTGATCGACTCCC
wheatrpl2_SA          AGTCAGCCACGCAGTGCCTTGTGCGTCTGCCGTCGGGCGTCGAGAAGCTGATAGACTCCC
barleyrpl2_NCBI       AGTCAGCCACGCAGTGCCTTGTGCGGCTGCCGTCAGGCGTCGAGAAGCTGATAGACTCCC
                      ** ***** * ** ** ***** ** ** ** ** ** *  ** **  * ** ** ****
```

Appendix E: Paralogous copies of nuclear *rpl2* gene in wheat and barley from ensemble plants website



- Green box represents UTR exo

Appendix F Schematics of Wheat scaffolds for chromosome 5, 3AL, 3B, 3D, 1As_a and 1As_b from <plants.ensembl> website



Appendix G: Alignment of PCR and RT-PCR sequences of wheat chromosome 5 *rps19* gene copy

```
Wheatchr5_(590bp)DNA_SA      -TTGCTTCTCTCATCGCCTCCAGGTCCAGCTTCGCCAGGTGCGGCCACGCTCTCCCCGCC
WheatRTPCRLB837_RC           ATTGCTTCTCTCATCGCCTCCAGGTCCAGCTTCGCCAGGTGCGGCCACGCTCTCCCCGCC
                              ***********************************************************

Wheatchr5_(590bp)DNA_SA      GCCATCTCTCAGGTGCAAATCCAAATCTTGATCCTCGCTCAAATGCAAATCTTCCTTTCC
WheatRTPCRLB837_RC           GCCATCTCT---------------------------------------------------
                             *********

Wheatchr5_(590bp)DNA_SA      TTACCGCCGGTTGTTCTTGTGCCGATGTTGATCTTGTTTGTTCAAATGCAAATCTTTCTA
WheatRTPCRLB837_RC           ------------------------------------------------------------


Wheatchr5_(590bp)DNA_SA      CCGCGCGACCCCACCAGACCCAGCACGCCGCGTCTCCGCTGCTCTCGGGGTTCGGATCAG
WheatRTPCRLB837_RC           --------------CAGACCCAGCACGCCGCGTCTCCGCTGCTCTCGGGGTTCGGATCAG
                                           **********************************************

Wheatchr5_(590bp)DNA_SA      CGGCTCGCGCTTTCAGGTAACCACCATTACAGTCATCACAGGCAGTCTCGCTATGCTGCT
WheatRTPCRLB837_RC           CGGCTCGCGCTTT-----------------------------------------------
                             *************

Wheatchr5_(590bp)DNA_SA      CAACATTGACTCCAATGTATGCCCCCAAGCTGCTGTCTGCGTGCAGCTCAAGGCCTCTAT
WheatRTPCRLB837_RC           -------------------------------------------CAGCTCAAGGCCTCTAT
                                                                        *****************

Wheatchr5_(590bp)DNA_SA      GGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAGCGGGGGGAGCCTGA
WheatRTPCRLB837_RC           GGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAGCGGGGGGAGCCTGA
                             ************************************************************

Wheatchr5_(590bp)DNA_SA      ACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTTCGTCGGCTCCTCCG
WheatRTPCRLB837_RC           ACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTTCGTCGGCTCCTCCG
                             ************************************************************

Wheatchr5_(590bp)DNA_SA      CGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG
WheatRTPCRLB837_RC           CGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAG
                             *************************************
```

Appendix H: Experimental clone data for presence and expression of homeologous nuclear *rps19* copies in brome

(a) Nucleotide alignment of all the brome nuclear *rps19* homeologous copies at the DNA level

```
BR862-904_dna_clone4dec      ---------------------------ATGTCGATTGGCCCTCTCATCGCCTCCAGGTTC
brome862-904Clone6jan        ACTAGTGATTTGCGGCTGAGAGCAGCCATGTCGATTGGCTCTCTCATCGCCTCCAGGTTC
862-904clone2oct             ------GATTTGCGGCTGAGAGCAGCCATGTCGATTGGCTCTCTCATCGCCTCCAGGTTC
BR862-904_dna_clone3dec      ACTAGTGATTTGCGGCTGAGAGCAGCCATGTCGATTGGCTCTCTCATCGCCTCCAGGTTC
brome862-904Clone7jan        ACTAGTGATTTGCGGCTGAGAGCAGCCATGTCGATTGGCTCTCTCATCGCCTCCAGGTTC
brome862-904Clone5jan        -----CGATTTGCGGCTGAGAGCAGCCATGTCGATTGGCCCTCTCATCGCCTCCAGGTTC
brome862-904clone1oct        ACTAGTGATTTGCGGCTGAGAGCAGCCATGTCGATTGGCCCTCTCATCGCCTCCAGGTTC
                                                        ************ ********************

BR862-904_dna_clone4dec      GCCAGGTCCGGCCACGCCCTCC------------CCGCCACGGCCATCTCTCAGGTC---
brome862-904Clone6jan        GCCAGGTCCGGCCACGCCCTCCCCGCCGCCGCCGCCGCCACGGCCATCTCTCAGGTC---
862-904clone2oct             GCCAGGTCCGGCCACGCCCTCGCCGCCGCCGCCGCCGCCGCCGCCATCTCTCAGGTC---
BR862-904_dna_clone3dec      GCCAGGTCCGGCCACGCCCTCC---------CAGCCGCCACGGCCATCTCTCAGGTC---
brome862-904Clone7jan        GCCAGGTCCGGCCACGCCCTCC------CCGCCGCTGCCACGGCCATCTCTCAGGTC---
brome862-904Clone5jan        GCCAGGTCCGGCCACGCCCTCCCCGCCGCCGCCGCCGCCACGGCCATCTCTCAGGTCCGT
brome862-904clone1oct        GCCAGGTCCGGCCACGCCCTCCCCGCCGCCGCCGCCGCCACGGCCATCTCTCAGGTCCGT
                             *********************             * *** * ***************

BR862-904_dna_clone4dec      ----------------CCGCCACCCCCCTCCAACTCTTGCTCCTTGCTCAAATGCAAATC
brome862-904Clone6jan        ----------------CCGCCACCCCCCTCCAACTCTTGCTCCTTGCTCAAATGCAAATC
862-904clone2oct             ----------------CCCCCCCTCCCCTCCAACTCTTGCTCCTTGCTCAAATGCAAATC
```

```
BR862-904_dna_clone3dec      ------------------------------CCGACTCTTGCTCCTTGCTCAAATGCAAATC
brome862-904Clone7jan        ------------------------------CCGACTCTTGCTCCTTGCTCAAATGCAAATC
brome862-904Clone5jan        CCGTCCGTCCGTCCGTCCCCCCTCCTCCTCCAACTCTTGCTCCTTGCTCAAATGCAAATC
brome862-904clone1oct        CCGTCCGTCCGTCCGTCCCCCCTCCTCCTCCAACTCTTGCTCCTTGCTCAAATGCAAATC
                                                           ** ***************************

BR862-904_dna_clone4dec      TTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCTCGGTCTGGTCTGGTC
brome862-904Clone6jan        TTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCTCGGTCTGGTCTGGTC
862-904clone2oct             TTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCGCGGTCTGGTCTGGTC
BR862-904_dna_clone3dec      TTCCTTCCTGCTGGTTGATCTGAGTTGTTCCTTCGTTCTACCGCGCGGTCTGGTCTGGTC
brome862-904Clone7jan        TTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCGTTCTACCGCGCGGTCTGGTCTGGTC
brome862-904Clone5jan        TTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCGCGGTCTGGTCTGGTC
brome862-904clone1oct        TTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCGCGGTCTGGTCTGGTC
                             ****** *************************** ********* ***************

BR862-904_dna_clone4dec      AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
brome862-904Clone6jan        AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
862-904clone2oct             AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
BR862-904_dna_clone3dec      AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
brome862-904Clone7jan        AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
brome862-904Clone5jan        AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
brome862-904clone1oct        AGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTTCGGAGCAGTGACTC
                             ************************************************************

BR862-904_dna_clone4dec      GTGCCTTCAGGTAACTACCATTACAAGCTCACAGTCCACCTCTTGCATTCCTTGCTCAAC
brome862-904Clone6jan        GTGCCTTCAGGTAACTACCATTACAAGCTCACAGTCCACCTCTTGCATTCCTTGCTCAAC
862-904clone2oct             GTGCCTTCAGGTAACTACCATT--------ACAGTCCACCTCTCGCATTCCTTGCTCAAC
BR862-904_dna_clone3dec      GTGCCTTCAGGTAACCACCATTACAAGCTCACAG--------TCGCATTCCTTGCTCAAC
brome862-904Clone7jan        GTGCCTTCAGGTAACTACCATTACAAGCTCACAG---------CGCATTCCTTGCTCAAC
brome862-904Clone5jan        GTGCCTTCAGGTAACTACCATTACAAGCTCACAG--------TCGCATTCCTTGCTCAAC
brome862-904clone1oct        GTGCCTTCAGGTAACTACCATTACAAGCTCACAGTCCACCTCTCGCATTCCTTGCTCAAC
                             *************** ******        ****          ****************

BR862-904_dna_clone4dec      ATTGAATCGAATGTATGCCCCAAGCTGCTG------CTGTGTGCAGCTCAAGGCCTCTAT
brome862-904Clone6jan        ATTGAATCGAATGTATGCCCCAAGCTGCTG------CTGTGTGCAGCTCAAGGCCTCTAT
862-904clone2oct             ATTGAATCGAATGTATGCCCCAAGCTGCTG------CTGTGTGCAGCTCAAGGCCTCTAT
BR862-904_dna_clone3dec      ATTGGATCGAACGTATGCCCCAAGCTGCTGCTG---CTGTGTGCAGCTCAAGGCCTCTAT
brome862-904Clone7jan        ATTGAATCGAATGTATG-CCCAAGCTGCTGCTGTGTCTGTGTGCAGCTCAAGGCCTCTAT
brome862-904Clone5jan        ATTGGATCGAACGTATGCCCCAAGCTGCTGCTG---CTGTGTGCAGCTCAAGGCCTCTAT
brome862-904clone1oct        ATTGAATCGAATGTATGCCCCAAGCTGCTG------CTGTGTGCAGCTCAAGGCCTCTAT
                             **** ****** ***** ************      ************************

BR862-904_dna_clone4dec      GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
brome862-904Clone6jan        GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
862-904clone2oct             GGAAGGGAGCGTTCGTCGACGTTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
BR862-904_dna_clone3dec      GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
brome862-904Clone7jan        GGAGGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
brome862-904Clone5jan        GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
brome862-904clone1oct        GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
                             *** ***************** **************************************

BR862-904_dna_clone4dec      ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
brome862-904Clone6jan        ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
862-904clone2oct             ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
BR862-904_dna_clone3dec      ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
brome862-904Clone7jan        ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
brome862-904Clone5jan        ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
brome862-904clone1oct        ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
                             ************************************************************

BR862-904_dna_clone4dec      TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCGCCGAAGGGAAGGTCGGCC
brome862-904Clone6jan        TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCGCCGAAGGGAAGGTCGGCC
862-904clone2oct             TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
BR862-904_dna_clone3dec      TGCTCATTTACAACGGCAAGGCTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTTGGCC
brome862-904Clone7jan        TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
brome862-904Clone5jan        TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
brome862-904clone1oct        TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
                             ******************** ******************** ************* ****

BR862-904_dna_clone4dec      ATAAGTTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
brome862-904Clone6jan        ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGGGGCCCCATCGGGCGACTACGGGGA
```

```
862-904clone2oct             ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
BR862-904_dna_clone3dec      ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
brome862-904Clone7jan        ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
brome862-904Clone5jan        ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGACTACGGGGA
brome862-904clone1oct        ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGACTACGGGGA
                             ***** ***************************** *************** ********

BR862-904_dna_clone4dec      AGGCTGGTCAAGGAAAGGGGAAGAAGAAGTAA----------------------------
brome862-904Clone6jan        AGGCCGCTCAAGGAAAGGGGAAGAAGAAGTAAGAAAAGGGAATGTGCGTCAAAGGTGAGG
862-904clone2oct             AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAAGAAAAGGGAATGTGCGTCAAAGGTGAGG
BR862-904_dna_clone3dec      AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAAGAAAAGGGAATGTGCGTCAAAGGTGAGG
brome862-904Clone7jan        AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAAGAAAAGGGAATGTGCGTCAAAGGTGAGG
brome862-904Clone5jan        AGGCCGCTCAAGGAAAGGGGAAGAAGAAGTAAGAAAAGGGAATGTGCGTCAAAGGTGAGG
brome862-904clone1oct        AGGCCGGTCAAGGAAAGGG---GAAGAAGTAAGAAAAGGGAATGCGCGTCAAAGGTGAGG
                             **** * ************   **********
```

Appendix I: Nucleotide and amino acid alignments of brome PCR and RT-PCR sequences for one of the paralogous copy (Brome nuc 1 or chr3 type copy seen in wheat)

(A).

```
br862-904_dna2               ATGTCGATTGGCTCTCTCATCGCCTCCAGGTTCGCCAGGTCCGGCCACGCCCTCGCCGCC
br862-923cdna_c3dec12th      ATGTCGATTGGCTCTCTCATCGCCTCCAGGTTCGCCAGGTCCGGCCACGCCCTCCCCGCC
                             ****************************************************** *****

br862-904_dna2               GCCGCCGCCGCCGCCGCCATCTCTCAGGTCCCCCCCCTCCCCTCCAACTCTTGCTCCTTG
br862-923cdna_c3dec12th      GCCGCC------------ATCTATCAG --------------------------------
                             ******            *********

br862-904_dna2               CTCAAATGCAAATCTTCCTTACTGCTGGTTGATCTGAGTTGTTCCTTCTTTCTACCGCGC
br862-923cdna_c3dec12th      ------------------------------------------------------------


br862-904_dna2               GGTCTGGTCTGGTCAGGCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTT
br862-923cdna_c3dec12th      ----------------GCGCCCAGGGCCCAGCACGCCGCGTCTCCGCTGCTCTGGGGCTT
                                             ********************************************

br862-904_dna2               CGGAGCAGTGACTCGTGCCTTCAGGTAACTACCATTACAGTCCACCTCTCGCATTCCTTG
br862-923cdna_c3dec12th      CGGAGCAGTGACTCGTGCCTT---------------------------------------
                             *********************

br862-904_dna2               CTCAACATTGAATCGAATGTATGCCCCAAGCTGCTGCTGTGTGCAGCTCAAGGCCTCTAT
br862-923cdna_c3dec12th      -------------------------------------------CAGCTCAAGGCCTCTAT
                                                                        *****************

br862-904_dna2               GGAAGGGAGCGTTCGTCGACGTTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
br862-923cdna_c3dec12th      GGAAGGGAGCGTTCGTCGACGCTTTCCTGCAAAGAATAAAGAAGAACAAGGAGAATTTGA
                             ********************* **************************************

br862-904_dna2               ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
br862-923cdna_c3dec12th      ACGGCAGGAAGATCTGGTCCCGCAGGTCTTCGATCCTGCCGGAGTTCGTCGGTTCCTCCG
                             ************************************************************

br862-904_dna2               TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
br862-923cdna_c3dec12th      TGCTCATTTACAACGGCAAGACTCACGTCCGTTGCAAGATCACCGAAGGGAAGGTCGGCC
                             ************************************************************

br862-904_dna2               ATAAGCTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
br862-923cdna_c3dec12th      ATAAGTTTGGGGAGTTTGCTTTCACGCGGAGGCGGAGGCCCCATCGGGCGATTACGGGGA
                             ***** ******************************************************

br862-904_dna2               AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAA
br862-923cdna_c3dec12th      AGGCCGGTCAAGGAAAGGGGAAGAAGAAGTAA
                             ********************************
```

(b)

```
br862-904_DNA2         MSIGSLIASRFARSGHALAAAAAAAAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
BR862-904_DNA1         MSIGPLIASRFARSGHALPAAAAATAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
brome862-923RNAc2      MSIGSLIASRFARSGHALPAA----AIYQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
brome862-923RNAc1      MSIGSLIASRFARSGHALPAA--ATAISQAPRAQHAASPLLWGFGAVTRAFSSRPLWKGA
                       ****.*************.**    ** ********************************

br862-904_DNA2         FVDVFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKLG
BR862-904_DNA1         FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKLG
brome862-923RNAc2      FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKTHVRCKITEGKVGHKFG
brome862-923RNAc1      FVDAFLQRIKKNKENLNGRKIWSRRSSILPEFVGSSVLIYNGKAHVRCKITEGKVGHKFG
                       ***.***************************************:**************:*

br862-904_DNA2         EFAFTRRRRPHRAITGKAGQGKGKKK-
BR862-904_DNA1         EFAFTRRRRPHRATTGKAGQGKGKK--
brome862-923RNAc2      EFAFTRRRRPHRAITGKAGQGKGKKKK
brome862-923RNAc1      EFAFTRRRRPHRATTGKAGQGKGKKKK
                       ************* ***********
```

Appendix J: Analysis of the novel *rps19* gene presequence in oats

a) Nucleotide alignment of the n-terminal targeting sequence in oats and lolium with wheat and barley to show that the sequences are non homologous

```
lolium                                                                 GGCCAA
oats                                       ATGCTGCGGAGGTTTAGCGAGGTTATGGGGCCAA
festuca                                    ATGCTGCGGAGGTTTAGTGAGGTTATGGGGCCAA
wheat_chr3b      ATGGCG---ATTGCTTCTCTCATCGCCTCC------AGGTTCGCCAGGTCC---GGCCAC
barley           ATGTCGTCCATTGCTTCTCTCATCGCCTCCAGGTCCAGCTTCGCCAGGTCC---GGCCAC
                                                                       *****

lolium           GGACCATTCTTCGCTCAGCCATCGCAG-------TTCAGGCTACCAAGGCCCAGCACCCC
oats             GAACCATTCTCCGGTCAATTGCCATGG-------TTCAGGCTACCAAGGCCCAGCCCCCT
festuca          GGACCATTCTCCGGTCAATCGTCACGG-------TTCAGGCTACCGAGGCACAACCCTCA
wheat_chr3b      GGGCACGCCCTCCCCGCCGCCGCGGCCATCTCTCAGGCGCCCAGGGCCCAGCACGCC
barley           GCTCTCCCCGCCGCCGCCGCCGCCGCCATCTCTCA---------GGCCCAGCACGCC
                 *  *    *  *    *      *   *       ***         *** ** * * *

lolium           ACATCTCCTTTGATCCCGGGGTTTGGAGCAGTGACTCGTGCTTTCAGCTCAAGGTCTCTC
oats             GCGTCTCCTCTGCTCTCGGGATTTGGATCAGTGACTCGTGCTTTCAGCTCAAGGTCTCTA
festuca          GCATCTCCTCTGCTCTCAGGATTTGGAGCAGTGACTCGTGCTTTCAGTTCAAGGTCTTTA
wheat_chr3b      GCATCTCCGCTGCTCTCCGGCTTCGGACCACTCACTCGCGCCTTCAGCTCAAGGCCTCTA
barley           GCGTCTCCGCTGCTTTCGGGGTTCGGATCAGCGGCTCGTGCTTTCAGCTCAAGGCCTCTA
                  * *****  ** *  * ** ** *** **    **** ** ***** ****** ** *

lolium           TGGAAGGGAGCTTTCGTTGATGCTTTCCTGTCTAGGATAAAGAACAATGGAGGAGCAATG
oats             TGGAAGGGAGCTTTCGTTGATGCTTTCCTGTCTAGAATAAAGAACAACGGAGGAGCAATG
festuca          TGGAAGGGAGCTTTCGTTGATGCTTTCCTGTCTAGAATAAAGAACAATGGAGGAGCAATG
wheat_chr3b      TGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAAAGGATAAAGAACAAGAGGGAGAACCTC
barley           TGGAAGGGAGCGTTCGTCGACGCCTTCCTGCAGAGAATAAAGAAGAGCGGGGAGAGCCTG
                 *********** ***** ** ** ******   ** ******** *   * *      *
```

b) Amino acid alignment of the nuclear *rps19* presequence in oats clade with that of wheat and barley

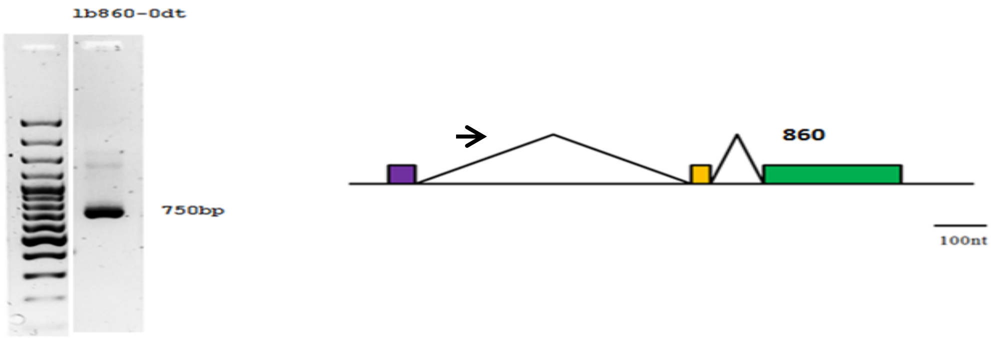
```
oats                        MLRRFSEVMGPRTILRSIAMVQATKAQPPASPLLSGFGSVTRAFSSRSL
lolium                                PRTILRSAIAVQATKAQHPTSPLIPGFGAVTRAFSSRSL
festuca                     MLRRFSEVMGPRTILRSIVTVQATEAQPSASPLLSGFGAVTRAFSSRSL
barley           -MSSIASLIASRSSFARSGHALPAA---AAAAAISQAQHAASPLLSGFGSAARAFSSRPL
wheat_chr3b      --MAIASLIASR--FARSGHGHALPAAAAAISQAPRAQHAASPLLSGFGPLTRAFSSRPL
                           .*  : **                . ** .:***:.***. :******.*

oats             WKGAFVDAFLSRIKNNGGAMNGK
lolium           WKGAFVDAFLSRIKNNGGAMNGK
oats             WKGAFVDAFLSRIKNNGGAMNGK
festuca          WKGAFVDAFLSRIKNNGGAMNGK
barley           WKGAFVDAFLQRIKKSGESLNGR
wheat_chr3b      WKGAFVDAFLQRIKNKRENLNGK
                 **********.***:.   :**.
```

c) Nucleotide alignment of intron 1 and exon 2 junction between oats, wheat and barley nuclear rps19 genes to see where the break point in homology occurs. Yellow highlight represents the novel exon 2

```
oats            TGCTGGTTGTTCTTGTTCAGATGTAGACATAATTTTTTCTATATGTTGTACTGTGAGATC
wheat           -------------------------------------------------ACCGCGCGACC
barley          ------------------------------------------------------GGGATC
                                                                      * ** *

oats            AGCCCAGGCTCCCGAGCCCCAGCCCTGTGTGTCTTCTGTGCTCTCGGGGTTGGGCTCAGT
wheat           CCACCAGA---------CCCAGCACGCCGCGTCTCCGCTGCTCTCGGGGTTCGGATCAGC
barley          TCGTCAGG---------CCCAGCACGCCGCGTCTCCGCTGCTTTCGGGGTTCGGATCAGC
                    ***          ****** *   * **** *  **** ******** ** ****

oats            GACTCGTGCTTTCAGGTAAACGTAGTAAAAACCTCAGTTCATATGCAGATAATTGTCAAC
wheat           GGCTCGCGCTTTCAGGTAACCACCATTACAGTC----ATCACAGGCAG------------
barley          GGCTCGTGCTTTCAGGTAACTACCATCACAGTC----ATCACAGGCAG------------
                * **** ************      * * *  *     *** * ****

oats            GCTAGGTGGCGCATATATCCCAGTCCGCCTCTCGCATTGCTTGTTGAGTATTGAATTGTA
wheat           -----------------------------TCTCGCTATGC-TGCTCAACATTGACTCCAA
barley          -----------------------------TCTCATTATGT-TGCTCAAAATTGAATCCAA
                                             ****    **  ** * *  ***** *   *

oats            TGTA-----------TGCTGT-TGTGTGCAGCTCAAGGTCTCTATGGAAGGGAGCTTTCG
wheat           TGTATGCCCCCAAGCTGCTGTCTGCGTGCAGCTCAAGGCCTCTATGGAAGGGAGCGTTCG
barley          TGTATGCCCCCAAGCTGCTGTCTGCGTGCAGCTCAAGGCCTCTATGGAAGGGAGCGTTCG
                ****           ****** ** ************* **************** ****
```

Appendix K: - 3'UTR exon sequences comparison among grasses for the nuclear *rps19* gene

(a)



b) 3'UTR exon comparison between wheat *chr5* and barley for the nuclear *rps19* gene

```
wheatutr        TAGGAAGCGTGAAAGGTGGGTTTAAGGATGGTGGGATGAGGGCGTGGTTGAAGCACAAGC
barleyutr       T---AAACTTTTCAGGTGGGTTTAAGGACGGTGGGATGAGGGCGTGGTTG----------
                *   ** * *   *************** *********************          

wheatutr        ACTTGAATCACCTTAATTTTCTGGTCTGCATTTTCTTAGAAGAATAATTATAGCAGATCT
barleyutr       --CTGAATCACCTTAATTCTCTGGACTGCATTTTCTTAGAAGAATAATTATAGCAGATCT
                   *************** ***** ***********************************

wheatutr        TACTTTGGACTCTGTTTACTCTGCTAGTGGTTCCAGCATAACTTTCGAGTACATCTTCTA
barleyutr       TACTTTGGACTCTGTTTACTCTGCTAGTGGTTCCAGCATAACTTTCGAGTACATCTTCTA
                ************************************************************

wheatutr        TTCACTGAAGCCATTGCTCTTATATTTTTTTGGCCTTACTGAGGTCTGATGCCAAGTGTC
barleyutr       TTCACTGAAGCCATTGCTCTTATATTTTTTTGGCCTTACTGAGGTCTGATGCCAAGTGTC
                ************************************************************

wheatutr        TTTACACCTGGTTCTGTGAGTGATGATTGTAAGTTTTGCTTCAAAGGTTGTTTTCAGCCA
barleyutr       TTTACACCTGGTTCTGTGAGTGATGATTGTAAGTTCTGCTTCGAAGGTTGTTTTCAGCCA
                *********************************** ****** *****************

wheatutr        AATAGTGATGTGTATCTTGCTACAGTTCCCTTGCAATAGGAGTATCTTGCTGATTGCTGC
barleyutr       AATAGTGATGTGTATCTTGCTACAGTTCCCTTGCAATAGGAGTATCTTGCTGATTGCTGC
                ************************************************************

wheatutr        TTTGCAT---TCAGGATGTTTT-----------TTTTTGCATATAATTAGCCCCTTTCGA
barleyutr       TTTGCATCGGCTAGGATGGTTGATATATATATATTTTTGC--ATAATTAGCCCCTTTCGA
                *******     ****** **            *******  ******************

wheatutr        GTCTTTATTAACAGGCTTAGCCCCCTTAAGAATGTATCACAAAATTTAGTCGCTTGCGCT
barleyutr       GTCTTTATTAACAGGCTTAGCCCCTTTAAGAATGTATTACAGAATTTAGTTGCTTGCGCC
                ************************ ************ *** ******** ******** 

wheatutr        AACGGCGACATGTGTCCTCTGTTGGCAGCTGACAGATTTAGCCTCTGTTTTTGATTTGTT
barleyutr       GACAGTGACATGTGTCCTCTGTGGGCACCTGACAGATTTAACCTCTGTTTTTTATTTGTT
                 ** * **************** **** ************ *********** *******

wheatutr        TGGTTAAC
barleyutr       TGGTTAAC
                ********
```

(c) Nucleotide sequence comparison including oats nuclear 3'UTR exon sequence along with wheat and barley

```
oatsoligodt     ----------------------------------------------GTAGTATAGTAGAC
wheat5b         CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCGTTCCTGCAGAGAATAAAGAATAG
barleyutr       CTCAAGGCCTCTATGGAAGGGAGCGTTCGTCGACGCCTTCCTGCAGAGAATAAAGAAGAG
                                                                * ** ** * *

oatsoligodt     AAGGAGGAGCATGAATGGCAAGAAGATTTGGTCTCGTAGATCTGCAATTTTGCCAGAATT
wheat5b         CGGGGGGAGCCTGAACGGCAGGAAGATCTGGTCTCGCCGGTCTTCGATCCTGCCGGAGTT
barleyutr       CGGGGAGAGCCTGAACGGCAGGAAGATCTGGTCTCGCAGGTCTTCGATCCTGCCGGAGTT
                  **  **** **** **** ****** ********  * *** * **  **** ** **

oatsoligodt     CGTTGGTTCCTCTGTGCTCATTTACAATGGGAAAACTCATGTCCGTTGCAAGATCAACGA
wheat5b         CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGCTGCAGGGTCACCGA
barleyutr       CGTCGGCTCCTCCGCGCTCATCTACAACGGCAAGACCCACGTCCGTTGCAGGATCACCGA
                *** ** ***** * ****** ***** ** ** ** ** ***** **** * *** ***

oatsoligodt     GGGGAAGGTTGGCCACAAGTTTGGAGAGTTTGCTTTTACACGGAGACGGAGACCCCATCG
wheat5b         AGGGAAGGTCGGCCATAAGTTCGGGGAGTTTGCTTTTACGCGGAAACGGAGGCCCCATCG
barleyutr       AGGGAAGGTCGGCCATAAGTTCGGGGAGTTTGCTTTTACGCGGAAACGGAGGCCCCATCG
                 ******** ***** ***** ** ************** **** ****** ********

oatsoligodt     CACAACTATAGCAAAG------GGCAAGCAAGTAAAGGGCAAAAAGTAAAAAAAAAAATC
wheat5b         CGCGATTACGGCGAAGAAGGCCGGTGGTCAAGGAAAGGGGAGGAAGTAGTAGGAAG----
barleyutr       CGCGATTACGGCGAAGAAGGCCGGCGGTCAAGGAAAGGGGAAGAAGTAAA----------
                * * * **  ** ***      **    **** ****** *  *****

oatsoligodt     TTCATGAAAGGTTGGTTCAAGGACGACGGATGATAAGGGACTACAAGAAGAACAACAACT
wheat5b         --CGTGAAAGGTGGGTTTAAGGATGGTGG--GATGAGGGCGTGGTTGAAGCACAAGCACT
barleyutr       --CTTTTCAGGTGGGTTTAAGGACGGTGG--GATGAGGGCGTGGTTG------------C
                  * *   **** **** ***** *  **  *** ****  *    *

oatsoligodt     TGAATCACCTTATTTTCTTATAATCTGGACTGGGTTTGCTTGTAATAATAACAATAAGAA
wheat5b         TGAATCACCTTAATTT-------TCTGGTCTGCATTTTCTTAGAAGAATAATTATAG---
barleyutr       TGAATCACCTTAATTC-------TCTGGACTGCATTTTCTTAGAAGAATAATTATAG---
                ************ **        ***** ***  *** ***  ** *****  ***

oatsoligodt     GAGGTGGCCATGTCAAGATTTGCCTGTGGACGCTGTT---TCGCCAAGTGGTTCC-----
wheat5b         --------------CAGATCTTACTTTGGACTCTGTTTACTCTGCTAGTGGTTCCAGCAT
barleyutr       --------------CAGATCTTACTTTGGACTCTGTTTACTCTGCTAGTGGTTCCAGCAT
                               **** *  ** ***** *****   **  * *********

oatsoligodt     -----------------------------------------------------TCCTTGC
wheat5b         AACTTTCGAGTACATCTTCTATTCACTGAAGCCATTGCTCTTATATTTTTTTGGCCTTAC
barleyutr       AACTTTCGAGTACATCTTCTATTCACTGAAGCCATTGCTCTTATATTTTTTTGGCCTTAC
                                                                      **** *

oatsoligodt     C----TCTGGTGCCAGAT-----------TCATTTTGGAGGTGA------TCCATTTTGC
wheat5b         TGAGGTCTGATGCCAAGTGTCTTTACACCTGGTTCTGTGAGTGATGATTGTAAGTTTTGC
barleyutr       TGAGGTCTGATGCCAAGTGTCTTTACACCTGGTTCTGTGAGTGATGATTGTAAGTTCTGC
                     **** *****  *           *  ** **   ****      *   ** ***

oatsoligodt     T------GTTCTTTTTTTCCAGC---TTATATCTGTTTTGATAC----------------
wheat5b         TTCAAAGGTTGTTTTCAGCCAAATAGTGATGTGTATCTTGCTACAGTTCCCTTGCAATAG
barleyutr       TTCGAAGGTTGTTTTCAGCCAAATAGTGATGTGTATCTTGCTACAGTTCCCTTGCAATAG
                *      *** ****   ***     * ** * * * *** ***

oatsoligodt     ------CTCGC---CTACCACATTG-----------------------------TCTCTG
wheat5b         GAGTATCTTGCTGATTGCTGCTTTGCAT---TCAGGATGTTTT-----------TTTTTG
barleyutr       GAGTATCTTGCTGATTGCTGCTTTGCATCGGCTAGGATGGTTGATATATATATATTTTTG
                      ** **    * *  * ***                             * * **
```

Core *rps19*

3'utr

Appendix L:- Nucleotide alignment of the three *rps19* gene copies present in maize

```
maizechromosome5clone              ATGCCGATTGGTTCTCTTTTCGTGTCGAGGCTGGCCAGGTCTAGCCATGCCCTCCCCGCT
maizechromosome1clone              ATGGCGATTGGTTCTCTGATCGCCTCTACTTTCGCCAGATCTAGCCATGCCCTCCCCGCT
maizechromosome1lacksintrons)      ATGGCGATTGGTTCTCTCATCGCCTCTACTTTCGCCAGATCTAGCCATGCCCTCCCCGCT
                                   *** *************  ***  ** *   * ***** *********************

maizechromosome5clone              GCCGCCGCCGCGGCCATCTCTCAGGTGCCACCT--------CCCCATCCCTCACTTTCGT
maizechromosome1clone              GCCGCCGCTTCGGCCATCTCTCAGGTGCCACCTCCCCCCTCCCCCCTCCCCCTCTTCCGC
maizechromosome1lacksintrons)      GCCGCCGCTTCGGCCATCTC----------------------------------------
                                   ********  **********

maizechromosome5clone              ATTGTTGGTTGTCCATGTGTCGATGTTGATCTGATCAATTTGTATGTTCCACCACGTGGT
maizechromosome1clone              GTCGTTGGTTGTTCGTGTGTCGATGTTGATCTGTCCAATTTGTATGTTCCACCGTGTGGT
maizechromosome1lacksintrons)      ------------------------------------------------------------


maizechromosome5clone              CTGATCAGGCGCCCAGGCCCCAGCACATCGCATCTCCTATATTCTCCGGCCTTGGAGCAG
maizechromosome1clone              CTGATCAGGCTCCCAGATCCCAGCACACCGCATCTCCTCTACTCTCCGGCCTTGGAGCAG
maizechromosome1lacksintrons)      ----TCAGGCTCCCAGATCCCAGCACACCGCATCTCCTCTGCTCTCCGGCCTTGGAGCAG
                                       ****** *****  ********* ********** *  ******************

maizechromosome5clone              CGGCTCGTGATTTCAGGTAAATATAC--CTAGTAGTTTCAGGTAAATATAGCTACGAGCT
maizechromosome1clone              CGGCTCGTGCTTTCAGGTAAATATATAACTACTAGGATACAACTACTAGGAGTACAAGCT
maizechromosome1lacksintrons)      CGGCTCGTGCTTT-----------------------------------------------
                                   ********* ***

maizechromosome5clone              TGTTATGTATGCGGATCATCACCAACACTAACTTGTGCTTCTATGTCAGTCTTCCTCCTG
maizechromosome1clone              TAGTATGTATGCGGATCATCACCAACGCTAGATTGTGCTTCTATGCCAAC----------
maizechromosome1lacksintrons)      ------------------------------------------------------------


maizechromosome5clone              CATTGCTTGTTGATTGTTGAATTGTATGTATCTACTGCTGTGTGCAGCTCAAGGGCTCTA
maizechromosome1clone              CATTGCTTGTTGAGTGTTGAATTGTATGTGTTTGCTGCTGTGTGCAGCTCAAGGGCTCTA
maizechromosome1lacksintrons)      --------------------------------------------CAGCTCAAGGGCTCTA
                                                                               ****************

maizechromosome5clone              TGGAAGGGAGCATTCGTCGATGCCTTCCTGGCTAGAATAAAGAAGAACAGAGAAAACATG
maizechromosome1clone              TGGAAGGGAGCATTCGTTGACGCTTTCCTGTCTAGAATTAAGAAGAACAGAGAAAATATG
maizechromosome1lacksintrons)      TGGAAGGGAGCATTCGTTGACGCTTTCCTGTCTAGAATTAAGAAGAACAGAGAAAATATG
                                   ***************** ** ** ****** ******* ***************** ***

maizechromosome5clone              AATGGGAAGAAGATTTGGTCTCGTAGGTCTTCTATTTTGCCGGAATTTGTTGGTTCCACT
maizechromosome1clone              AATGGCAAGAAGATTTGGTCTCGTAGATCCTCTATTTTGCCGGAATTTGTTGGTTCCTCT
maizechromosome1lacksintrons)      AATGGCAAGAAGATTTGGTCTCGTAGATCCTCTATTTTGCCGGAATTTGTTGGTTCCTCT
                                   ***** ******************** ** *************************** **

maizechromosome5clone              GTGCTCATTTACAATGGGAAAAGTCATGTCCGGTGCAAGATCACTGAAGGGAAGGTTGGC
maizechromosome1clone              GTGCTCATTTACAATGGGAAAACTCATGTCCGGTGCAAGATCACTGAAGGGAAGGTTGGC
maizechromosome1lacksintrons)      GTGCTCATTTACAATGGGAAAACTCATGTCCGGTGCAAGATCACTGAAGGGAAGGTTGGC
                                   ********************** *************************************

maizechromosome5clone              CATAAATTTGGAGAGTTTGCTTTTACACGGAGACGAAGGCCCCATCGAACAATAACAGGA
maizechromosome1clone              CATAAATTTGGAGAGTTTGCTTTTACACGGAGACGAAGGCCCCATCGAACAATGACAGGA
maizechromosome1lacksintrons)      CATAAATTTGGAGAGTTTGCTTTTACACGGAGACGAAGGCCCCATCGAACAATGACAGGA
                                   ***************************************************** ******

maizechromosome5clone              AAGGGAAATCAAGGAAAGGGCAGAAAGTAA
maizechromosome1clone              AAGGGAAATCAAGGAAAGGCAAGAAAGTAA
maizechromosome1lacksintrons)      AAGGGAAATCAAGGAAAGGCAAGAAAGTAA
                                   *******************  *********
```

Appendix M:- Genes used for codon usage analysis of nuclear *rps19* and *rpl2* genes

Mitochondrial genes for brome whose sequence data was available in our lab was used (*rps7, rps3, rps1A, and ccmFN*). The mitochondrial gene data set for brome consisted of 4,066 nucleotides in total. Nuclear genes for brome whose sequence data was available for brome consisted of 8,907 nucleotides (WSI18-like protein, peroxiredoxin, SAR DNA binding protein, embryonic abundant protein, LEA protein, glycoprotein, BiRP2 mRNA for ribosomal protein, BiRP1 mRNA for ribosomal protein, puroindoline b gene, puroindoline a gene, stress-inducible membrane pore protein (BG15, aldose reductase-related protein).

Supplementary tablesS1:- Accession numbers of EST entries for nuclear *rps19*, nuclear *rpl2* and nuclear *rpl5* genes in different grasses as well as the sequenced mitochondrial genomes.

Plant	EST accession number	Gene or genome
Barley	GH228636	nuclear rps19
	BI959815	nuclear rps19
	BM372663	nuclear rps19
Wheat chr5	BQ903089	nuclear rps19
	CJ657426	nuclear rps19
	CJ549719	nuclear rps19
Wheat chr3	BJ290519	nuclear rps19
	AL808871	nuclear rps19
	CK19319	nuclear rps19
Maize chr1	FM181476	nuclear rps19
	EE292591	nuclear rps19
	CF042837	nuclear rps19
Maize chr5	FM195171	nuclear rps19
	EE191481	nuclear rps19
	EE016930	nuclear rps19
Lolium	GR522935	nuclear rps19
Oats	AY669854	mitochondrial rps19
Oats	GO589432	nuclear rps19
Festuca	GO801447	nuclear rps19
	GO890465	nuclear rps19
	GO890465	nuclear rps19
Dactylis	HO158277	nuclear rps19
	HO160164	nuclear rps19
	HO171082	nuclear rps19
Wheat hsp70	BJ277769	wheathsp70
	GO890465	nuclear rps19
Dactylis	HO158277	nuclear rps19
	HO160164	nuclear rps19
Wheat	DQ057345	nuclear rpl2
	CJ805209	nuclear rpl2
	BJ224681	nuclear rpl2
Barley	BI951050	nuclear rpl2
	BU998563	nuclear rpl2
	CA024196	nuclear rpl2
Festuca	GO856499	nuclear rpl2
	GO883981	nuclear rpl2
	GO873545	nuclear rpl2
Wheat	BJ269357	nuclearrpl5
	BQ806344	nuclearrpl5
	BJ269331	nuclearrpl5
Barley	BM368977	nuclearrpl5
	BU973461	nuclearrpl5
	CV053501	nuclearrpl5
Oats	CN821212	nuclearrpl5
Maize	AW066145/ AI833526	nuclearrpl5
	CD435692/	nuclearrpl5

	AI711617	
	BE344589/ CF009652	
***Oryza sativa* (rice)**	BA000029	mitochondrial genome
***Triticum aestivum* (wheat)**	AP008982	mitochondrial genome
***Bambusa oldhamii* (bamboo)**	EU365401	mitochondrial genome
***Ferrocalamus* (iron bamboo)**	JN120789	mitochondrial genome
***Lolium prenne* (rye grass)**	JX99999	mitochondrial genome
***Oryza sativa* (rice)**	BA000029	mitochondrial genome
***Triticum aestivum* (wheat)**	AP008982	mitochondrial genome

S2 List of oligomers used in chapter 3 results

Oligomer number	**Oligomer sequence 5'-3'**	**locus**	**oligomer name**
1	TTGAGTACGGGGAAGTCCGC	brome mito rpl2 exon 1	lb97
2	GTTACAGACTCACTTATCCG	brome mito rpl2 intron	lb62
3	ATAGATTCCCGATGCCGAGC	brome mito rpl2 exon 2	lb61
4	TCGAGTATGAAGAAAAGACC	brome mito rps19 3'UTR	lb196
5	TTCCACAATGCCAATAGACG	brome mito rpl2 exon 1	lb63
6	CTAGTACGATGGATCGAAGG	brome mito rpl2 exon 1	lb72
7	TGCGGCTGAGAGCAGCCATG	brome nuclear rps19 exon 1	lb862
8	TGACGCGCATTCCCTTTTCT	brome nuclear rps19 3'UTR	lb923
9	CAATATCGTAACCTCACCTT	brome nuclear rps19	lb904
10	GACGTACGCATCACCACCAA	oats nuclear rps19 5'UTR	lb892
11	ATGGGGCCAAGAACCATTCT	oats nuclear rps19 exon 1	lb867
12	TTGTTTGCCCTTCGCTATAG	oats nuclear rps19	lb852
13	ACTGGTTTGACTGTTGCTGA	atp synthase beta	lb844
14	CATCCATACCCAAAATAGCAAT	atp synthase beta	lb109
15	AAGGATCCAGCATACAAGGC	nuclear rpl2	lb79
16	GTTGGAGACGATGCCGATGG	nuclear rpl2	lb89
17	GGGTGGGGAAACGCAGAATT	mitochondrial *rpl2* exon 2 (based on rice)	lb834