INVESTIGATIONS INTO THE CAUSE OF POLLEN ABORTION IN MAIZE CMS-C

_____

A Dissertation
presented to
the Faculty of the Graduate School
University of Missouri

_____

In Partial Fulfillment
of the Requirements for the Degree
Doctorate of Philosophy

_____

by
LOUIS J. MEYER

Dr. Kathleen Newton, Dissertation Supervisor

MAY 2009

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

INVESTIGATIONS INTO THE CAUSE OF POLLEN ABORTION IN MAIZE CMS-C

Presented by Louis Meyer

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.

_____

Professor Kathleen Newton

_____

Professor James Birchler

_____

Professor Emmanual Liscum

_____

Professor Jay Thelen

_____

Professor Georgia Davis

ACKNOWLEDGMENTS

I would like to thank all those that have contributed to my education; professor Kathleen Newton for her advice and continued guidance, the members of my committee for there counsel in the direction of my research, and the members of the Newton lab, past and present, for there support.

I would also like to acknowledge my wife and family for all of their patience and encouragement throughout my academic carrier.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1: LITERATURE REVIEW

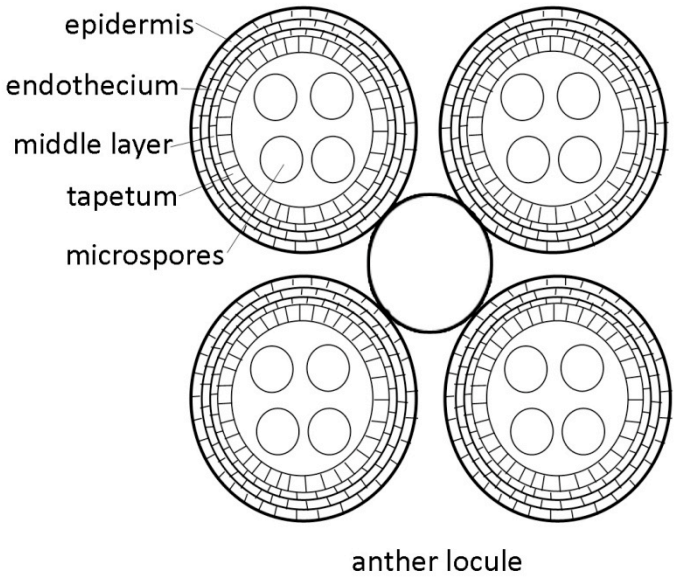
1.1 Pollen Development

Pollen development occurs in the locule of the anther surrounded by four layers: the epidermis, the endothecium, the middle layer, and the tapetum (Fig. 1; Bedinger, 1992). Within the locule, meiocytes (Bedinger, 1992) undergo meiosis to produce four haploid microspores that are encased in a callose wall (tetrad stage). The callose wall then degrades and releases the young microspores into the locule. After release, the microspores mature by first filling with multiple small vacuoles, which then fuse to form one large vacuole. At this point the tapetum starts to degenerate. During microspore development outer pollen wall formation, comprissed mainly of sporopollenin , occurs. (Meuter-Gerhards, 1999; Scott, 1994; Steiglitz, 1977).

Each microspore undergoes mitosis to produce a vegetative and a generative cell. The generative cell then undergoes a second mitosis to produce two sperm cells. Once a microspore cell completes mitosis to produce one vegetative and two sperm cells, it is considered young pollen. Mitosis is followed by accumulation of starch granules within the young pollen. Mature pollen is finally formed when the young pollen dehydrates, reducing its water content by 40 to 58% (Barnabas, 1985). Dehiscence then occurs and the pollen granules are released.

Figure 1: Diagram of a maize anther cross section

Graphical representation of a anther cross section with the major cell layers labeled. The tapetum serves as a nurturive layer for the developing microspores. As the anther develops the tapetum will die and become part of the outer wall of the mature pollen grain.



During pollen development the tapetum plays a critical role (Pacini et al., 1985). At meiosis the tapetum differentiates into binucleate, polar secretory cells that lack a cell wall. Tapetal cells serve a nutritive role in microspore developmentand secrete callase to release young microspores from the callose wall surrounding the meiotic tetrad (Steiglitz, 1977). As the tapetum dies cell remnants are deposited on the maturing pollen surface that serve as precursors to development of the outer pollen wall (Bedinger, 1992).

1.2 Mitochondrial Electron Transport

The mitochondrial electron transport system is comprised of five complexes (I-V) located at the mitochondrial inner membrane (Taiz, 2002). Electrons are transferred from NADH and succinate to complexes I-IV, and electron flow is coupled to proton pumping into the mitochondrial intermembrane space, which creates a pH and electrochemical gradient. ATP is then synthesized from ADP and inorganic phosphate, as protons pass through Complex V.

Complex I (NADH dehydrogenase) is the major complex that contributes electrons to the ubiquinone pool. There are actually two types of NADH dehydrogenases in plants; a rotenone-sensitive type I and a rotenone-insensitive type II (Marres et al., 1991; Moller et al., 1993; Yagi, 1991). The rotenone-sensitive complex I is a 1000 kDa complex in Arabidopsis, making it the largest in the electron transport chain. Complex I NADH dehydrogenases span the inner membrane, while type II are associated with the inner

membrane, either on the intermembrane or matrix side (Moller, 2001; Rasmusson et al., 1998a; Rasmusson et al., 1998b; Rasmusson et al., 2004). Both types of NADH dehydrogenases oxidize NADH and transfer the electrons to the ubiquinone pool. As the electrons are being transferred to the ubiquinone pool from the type I NADH dehydrogenase, protons are pumped into the intermembrane space. However, the type II NADH dehydrogenase does not have proton pumping coupled to the electron transfer. The mitochondrial genome codes for only a small subset of proteins that comprise Complex I, and no proteins of the type II NADH dehydrogenases.

There are four nuclear encoded subunits that make up complex II (succinate dehydrogenase) in Arabidopsis: a flavoprotein, two subunits containing iron-sulfur clusters, and two membrane anchor subunits (Figueroa et al., 2002). Succinate dehydrogenase uses covalently bound flavin, flavinadenine dinucleotide (FAD) to transfer electrons from succinate to the ubiquinone pool but does not couple this to proton pumping into the intermembrane space.

The core of complex III (cytochrome c reductase) is comprised of three subunits, nuclear encoded Rieske iron-sulfer protein, cytochrome c_1 and a mitochondrial encoded cytochrome *b* (reviewed by Crofts, 2004). Cytochrome reductase uses a process called the Q cycle to transfer electrons from the ubiquinone pool to cytochrome c while pumping protons into the intermembrane space (Crofts et al., 2003; Mitchell, 1976). The Q cycle involves two sites, Q_i and Q_o. Two quinols are oxidized in the Q_o site leading to the transfer of two electrons directly to cytochrome c and two electrons to the Q_i

site. When the electrons are transferred to the Q_i site from complexes I and II, protons are pumped to the inner membrane space and ubiquinone is reduced (reviewed by Crofts, 2004).

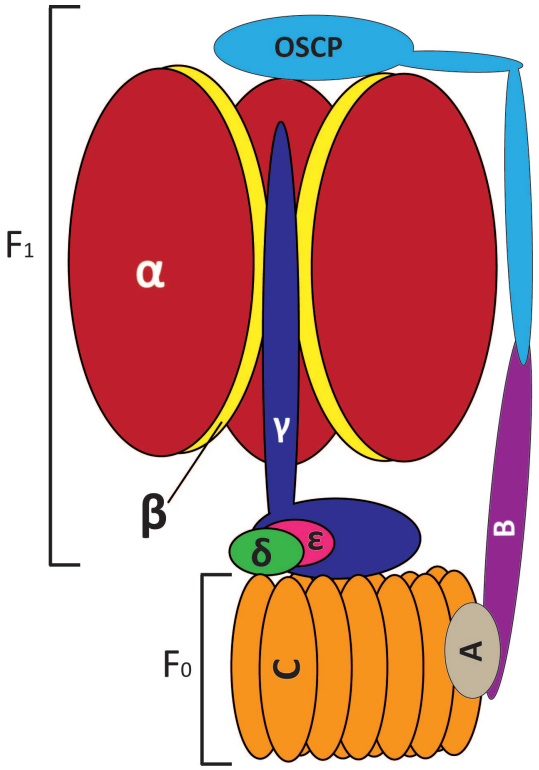
There are three subunits that comprise the core enzyme of complex IV (cytochrome oxidase), all of which are encoded by mitochondrial DNA. There are also six to seven nuclear encoded subunits (Millar et al., 2004). Electrons are transferred from cytochrome c to the active site within subunit I to reduce oxygen; this is coupled to protons entering from the matrix through two channels (D and K). Approximately half of the protons are used within the active site to produce two water molecules, while the remaining protons are pumped into the intermembrane space. The functioning of cytochrome oxidase results in protons being pumped into the intermembrane space and the synthesis of water (reviewed by Namslauer and Brzezinski, 2004; Papa et al., 2004; Pecina et al., 2004).

The proton gradient generated by complex I-IV is then used by complex V (ATP synthase) to synthesize ATP from ADP and P_i. The ATP synthase is comprised of two units; the membrane bound F_0 and the matrix located F_1 unit (Fig. 2). The F_0 unit is comprised of three major proteins; the mitochondrial encoded ATPa (*atp6*), ATPc (*atp9*), and ATPb (*atp4*). The F_1 unit contains five major proteins (α, β, γ, δ', and ε), of which only ATPα (*atp1*) is encoded by the mitochondrial genome (Weber and Senior, 2003). Protons enter the ATP synthase complex between ATPa and ATPc, which triggers a conformation change in ATPc and causes the ring of ATPc proteins to turn (reviewed by

Fillingame et al., 2003). The ring of ATPc proteins act as a motor to rotate the ATPγ and ATPε proteins. This rotation drives ATP synthesis within the ATPα and ATPβ subunits. The net result is ATP being synthesized as protons are transferred back into the matrix.

Figure 2: ATP synthase complex

A model of the ATP synthase proteins present in plant mitochondria. The F_1 subunit lies within the matrix and consists of six core proteins. ATPα (*atp1*) is the only F_1 protein encoded by the mitochondrial genome. All three of the inner membrane bound F_0 subunits are encoded by the maize mitochondrial genome; ATPa is encoded by the *atp6* gene, ATPb by *atp4*, and ATPc by *atp9*.



	F1					
ATP synthase protein	α	β	γ	δ	ε	OSCP
Mitochondrial encoded gene	*atp1*	-	-	-	-	-

	F0		
ATP synthase protein	a	b	c
Mitochondrial encoded gene	*atp6*	*atp4*	*atp9*

1.3 Plant Mitochondrial Genomes

The first plant mitochondrial genome to be sequenced was that of a non-vascular plant, *Marchantia polymorpha*. The 186,608 base pairs (bp) genome contains 66 identified genes, including ribosomal and transfer RNAs (Oda et al., 1992). Subsequently, *Arabidopsis thaliana* mitochondrial genome was sequenced and found to be approximately 367 kb, almost twice as large as that of *Marchantia polymorpha*. Although larger, the Arabadopsis mitochondrial genome contains only 59 identified genes (Unseld et al., 1997). Unseld *et al.* (1997) also reported that the *Arabidopsis* mitochondrial genome contains 84 unidentified open reading frames (ORFs) larger than 100 codons. More recently, rapeseed (*Brassica napus*), a relative of Arabidopsis, was sequenced. Rapeseed contains a nearly identical set of genes to Arabidopsis but its genome is smaller, only 222 kb in length and has little conservation with Arabidopsis outside of the genes (Handa, 2003). The only difference in protein-gene content between the Arabidopsis and rapeseed mitochondrial genomes is that the *rps14* gene is functional in rapeseed while present only as a pseudogene in Arabidopsis. Sugar beet (*Beta vulgaris*), a dicot more evolutionally distant from Arabidopsis, contains a similar set of mitochondrial genes and has a similar genome size at 369 kb. Even the monocot rice (*Oryza sativa*) has a similar gene content to the other sequenced plant mitochondrial genomes despite its larger genome size of approximately 491 kb (Notsu et al., 2002). The largest mitochondrial genome sequenced to date is maize CMS-C which is ~740 kb (*Zea mays;* Allen et al., 2007).

Based upon the sequenced plant mitochondrial genomes, there is a high degree of variability in genome size that occurs independently of the number of identified genes. Part of the variability in mitochondrial genome size is generated by DNA transfer events, in which fragments of chloroplast and possibly nuclear DNA have been transferred into the mitochondrial genome. The DNA fragments of non-mitochondrial origin contain some genes, although only a few transfer RNAs of chloroplast origin have been shown to be expressed in mitochondria (Marechal-Drouard et al., 1993). It is also clear that mitochondrial DNA can transfer to the nucleus (Lough et al., 2008). Chloroplast sequences have moved in to mitochondrial genome of rice and subsequently transferred into the nuclear genome as mitochondrial sequences (Notsu et al., 2002). Over the course of evolution, DNA transfer between genomes has caused the majority of the original bacterial endosymbiont genes to be lost or transferred to the nucleus, leaving only a small number of RNAs and proteins encoded by mitochondrial DNA (Adams et al., 2000; Palmer et al., 2000).

The high rate of recombination in plant mitochondrial genomes is another factor that contributes to the variability in genome size. Mitochondrial DNA (mtDNA) recombination generates large repeats that occur in almost all plant mitochondria and range in size from 1 to 14 kb (Clifton et al., 2004; Fauron et al., 1995; Hanson and Folkerts, 1992). Recombination within mitochondrial genomes also causes rearrangements and deletions, which result in a lack of conservation in the linear order of mitochondrial genes. The rearrangements and deletions also generate the majority

of mitochondrial mutations, since mitochondrial genes have a slower rate of nucleotide substitution than nuclear or chloroplast genes (Palmer and Herbon, 1988).

Recombination in mtDNA can generate chimeric genes which have been identified as the cause of cytoplasmic male sterility (CMS) in some plants. A chimeric gene consists of an open reading frame (ORF) that contains at least one fragment of a mitochondrial gene placed in tandem with pieces of unique sequences. Transcription of the chimeric gene can occur if it is located distal to a functional promoter. The expression of chimeric genes is the reported cause of S and T type CMS in *Zea mays (Dewey et al., 1986; Zabala et al., 1997)*.

The other class of characterized maize mitochondrial mutations, non-chromosomal striping (NCS), is missing parts of known genes. NCS5 and NCS6 have been identified as having a non-functional *cox2* gene, a cytochrome oxidase (complex IV) gene. NCS2 is missing an intact *nad4* gene, which is a component of complex I of the electron transport chain (Karpova and Newton, 1999; Lauer et al., 1990; Marienfeld and Newton, 1994; Newton et al., 1990). Two other mutations, NCS3 and NCS4, have reduced mitochondrial protein synthesis due to two different deletions of the ribosomal protein gene *rps3* (Hunt and Newton, 1991; Newton et al., 1996). Because these are essential genes for normal mitochondrial function, the mutations cannot exist in a homoplasmic state. Rather they survive heteroplasmically, where normal mitochondria are present with the mutants in order to compensate for their defects (Gu et al., 1994).

The *Z. mays* B37 normal cytoplasm (B37N) mitochondrial genome was sequenced and annotated (Clifton et al., 2004). The genome contains a total of 51 genes: 34 protein coding genes, three ribosomal RNAs, and 22 tRNAs, one of which (tRNA-trp) is carried on a linear 2.3 kb plasmid. There are also 132 previously unidentified ORFs of at least 300 nucleotides that begin and end with a start and stop codon identified in the B37N genome. The total number and identity of genes found in B37N is comparable to the gene content of previously sequenced mitochondrial genomes. The 51 genes in the *Z. mays* B37N mitochondrial genome account for only 6.2% of the total genome. Another 4.4% of the B37N genome is composed of DNA transferred from the chloroplast genome, and these sequences contain a few tRNAs. Thus, the majority of the maize B37N mitochondrial genome has no identified function or known origin.

When the total mitochondrial genome sequences of maize B37N, rice and Arabidopsis were compared, using cut-offs of a 20 bp minimum length and at least 90% identity, it was discovered that none of the inter-genic region was conserved (Clifton et al., 2004). *Zea mays* B37N had only 27.9% sequence identity with rice, a grass 50 million years evolutionally divergent from maize (Clifton et al., 2004). When B37N was compared to the eudicot Arabidopsis with over 130 million years of divergence, only 9.5% sequence identity was observed, most of which is accounted for by genes (Devos et al., 1999).

Four other maize mitochondrial genomes have been sequenced: another normal type of mitochondria (NA) and the three cytoplasmic male sterile types (CMS-C, CMS-T, and CMS-S; Allen et al., 2007). These four genomes range in size from 535, 825 bp (CMS-T) to 739,719 bp (CMS-C). The majority of this variation is due to large duplications. However, the genome complexity (size of the genome with only one copy of any repeat) only ranges from 506,760 bp (CMS-C) to 537,180 bp (NA). To further study the variability of maize mitochondria, Allen *et al.* (2007) compared the genome complexities of the five sequenced genomes with each other and determined the percentage of the genome that is absent. This analysis found a range of 0.48% of the CMS-S genome missing in NA to 7.42% of the CMS-T genome absent in B37N. The CMS-C genome in particular ranged from having 0.8% to 4.12% of its genome absent in the other genomes and lacked 4.97% to 6.73% of other mitochondrial genomes. The number of identified ORFs in each genome appears to correlate with the variability in genome size; however, the number of functional genes does not differ from the previously identified 51 in NB. Because the ORFs do not have a high rate of conservation among plant mitochondria they are unlikely to be expressed, this is supported by expression analysis performed in B37N (Meyer, 2004).

1.4 Transcription in Plant Mitochondria

Little is known about gene transcription from plant mitochondrial DNA. Functional promoters have been identified at various distances upstream of genes and initiation can occur at more than one site (Binder et al., 1996; Mulligan et al., 1991; Rapp and Stern, 1992). In maize, six transcription initiation sites have been identified for the *ATPase subunit 9* (*atp9*) gene (Mulligan et al., 1988) and five for *cytochrome oxidase subunit 2* (*cox2*) (Lupold et al., 1999b). Consensus promoter sequences have been deduced for some of the promoters identified for mRNA, tRNA, and rRNA genes (Binder et al., 1996; Rapp and Stern, 1992), consisting of a core CRTA sequence upstream of the transcription start site. However, several promoters do not contain this conserved motif (Fey and Marechal-Drouard, 1999; Newton et al., 1995).

The expression of *cox2* was analyzed in maize using an *in vitro* system (Lupold et al., 1999a; Lupold et al., 1999b). The *cox2* coding sequence is located downstream of a recombination repeat and can be controlled by two different promoters as a result of the recombination events. Further, expression analysis using quantitative PCR and RT-PCR determined one of the promoters had more transcriptional activity when upstream of *cox2* and less activity when upstream of a non-coding region. This observation indicated that genomic context influences promoter strength and suggests a role for genomic recombination in regulation of gene expression in the maize mitochondria.

The stability of mitochondrial transcripts can be affected by a 3' secondary structure that forms a single or multiple stem-loops. This effect on stability has been shown *in vivo* for *apocytocrome B* (*cob*) in both rice and alloplasmic wheat lines (Kaleikau et al., 1992; Saalaoui et al., 1990) and *in vitro* for *atp9* in pea (Dombrowski et al., 1997; Kuhn et al., 2001).

An active mechanism for plant mitochondrial mRNA degradation has been discovered, which is related to known degradation pathways in eubacteria and chloroplasts (Carpousis et al., 1999; Hayes et al., 1999; Kudla et al., 1996; Lisitsky et al., 1996; Regnier and Arraiano, 2000; Schuster et al., 1999). This pathway involves polyadenylation of transcripts which targets them for rapid exoribonucleolytic degradation. Five mitochondrial transcripts have been reported to be polyadenylated in plants. In sunflower a chimeric transcript, *atp1 – orf522*, is tissue-specifically polyadenylated and preferentially degraded during pollen development when the restorer of fertility gene product is present (Gagliardi and Leaver, 1999; Moneger et al., 1994). In maize, *cox2* transcripts can also be polyadenylated but there is only a slight destabilization effect *in vitro* (Lupold et al., 1999c). Polyadenylation of the *atp9* transcript in pea and potato and the *atp1* transcript in *Oenothera* also occurs (Gagliardi et al., 2001; Kuhn et al., 2001). Even though increased degradation due to polyadenylation of those RNA transcripts was observed, the double stem-loop structures of the transcripts of *atp9* in pea and *atp1* in *Oenothera* counterbalanced this destabilization (Kuhn et al., 2001).

Multiple states of *atp9* and *atp1* transcripts were discovered where polyadenylation occurred 3' of the double stem-loop after sequential truncation of the 3' sequence. This suggested that repeated polyadenylation/degradation may destabilize the double stem-loop structure (Kuhn et al., 2001). The stem-loop structure within the potato *atp9* transcripts did not inhibit degradation in contrast to the degradation seen for pea *atp9* and *Oenothera atp1* transcripts, suggesting that an RNase can degrade structured RNA (Gagliardi et al., 2001).

Analyzing mitochondrial RNA levels is further complicated by the rapid rearrangement of the genome. These rearrangements can cause promoters to be switched among genes and can also generate multiple copies of some genes (Allen et al., 2007). Potentially this can cause stochiometric imbalance among the transcripts and proteins. However in pea, transcript levels do not appear to correlate with gene copy number or protein stoichiometry indicating that there might be some gene dosage compensation occurring (Woloszynska et al., 2006).

1.5 Cytoplasmic Male Sterility

The first reported case of cytoplasmic male sterility (CMS) involved two strains of flax that produced sterile progeny when crossed in one direction but not the other (Bateson, 1921). Subsequently, CMS has been identified in several plant species along with nuclear *restorer of fertility* genes (Duvick, 1959). Since its discovery, CMS has been used extensively for hybrid seed production. Cytological work on CMS has identified abortion occurring at every stage of anther development, from meiosis (*Z. mays* CMS-C) to binucleate pollen (*Z. mays* CMS-S; reviewed by Laser, 1972). The cause of CMS is best understood in the petunia and maize CMS-T systems, which serve as models for identifying CMS causing genes in other species.

The petunia CMS-causing gene, termed *pcf*, was generated through mitochondrial rearrangements that produced a fusion of *atp9* and *cox2* coding sequences followed by an unidentified reading frame denoted as *urf-S* (Folkerts and Hanson, 1991; Young and Hanson, 1987). This gene is co-transcribed with the downstream *nad3* and *rps12* genes and produces transcripts with three 5' termini at -121, -266, and -522 (Pruitt and Hanson, 1991). It is not known if these transcripts result from processing or from multiple transcription initiation sites. The *pcf* gene was discovered by forming recombinant mitochondrial genomes through generation of somatic hybrids and looking for segregation of DNA fragments with the CMS phenotype (Hanson, 1984). Later transcription of the *pcf* gene was observed and the transcript altered in the presence of the *Rf* gene (Young and Hanson, 1987).

Translation of the *pcf* transcript produces a 43 kDa precursor protein which is processed into a stable 25 kDa protein representing the *urf-S* portion of the *pcf* transcript that associates with the membrane (Nivison et al., 1994). When the petunia CMS is crossed to a line containing the *restorer of fertility* gene, the abundance of the -121 5' termini *pcf* transcript and 25 kDa protein is reduced (Nivison and Hanson, 1989; Pruitt and Hanson, 1991). Although the *pcf* gene and the resulting PCF protein are correlated with CMS, the mechanism of pollen abortion has yet to be identified.

The petunia *restorer of fertility* gene (*Rf*) was the first RNA-altering restorer gene to be cloned. It was identified as a pentatricopeptide repeat (PPR) protein capable of binding RNA (Bentolila et al., 2002). Subsequently two *Rf* genes in radish and one in rice were also determined to be PPR proteins (Brown et al., 2003; Desloire et al., 2003; Kazama and Toriyama, 2003; Koizuka et al., 2003; Komori et al., 2004; Wang et al., 2006).

1.6 Cytoplasmic Male Sterility in *Zea mays*

Three types of CMS have been identified in maize. These types were identified by the lines/genes that confer fertility (Rogers, 1952). CMS-T is restored by lines containing both *Rf1* and *Rf2*, the major CMS-S restorer is *Rf3*, and CMS-C is restored by *Rf4*. Several other genes have been identified that confer fertility or partial fertility to CSM-S or CMS-C (Gabay-Laughnan et al., 2004).

1.6.1 Cytoplasmic Male Sterility Type T (CMS-T)

CMS-T was first described in 1952 by Rogers and Edwardson (1952). CMS-T is caused by a chimeric gene denoted T-*urf13* (Dewey et al., 1991), which consists of portions of the 26S rRNA sequence and a segment of unknown sequence driven by the *atp6* promoter. T-*urf13* is thought to have been generated by mitochondrial rearrangements (Dewey et al., 1991). This chimeric gene encodes a 13 kDa protein called URF13 that is inserted into the inner mitochondrial membrane during pollen development (Dewey et al., 1986; Wise, 1987). Unlike other types of CMS, CMS-T requires dominant alleles of two *restoration of fertility* genes, *Rf1,* and *Rf2,* for sporophytic recovery of pollen development (reviewed by Laughnan, 1983). Normally, T-*urf13* produces both a 2 kb and 1.8 kb stable RNA transcripts; however, in the *Rf1 rf2* nuclear background, a 1.6 kb transcript is also present. The presence of the *Rf1* gene also causes a reduction in the production of URF13 (Dewey et al., 1986). Even though *Rf1* clearly alters the expression of T-*urf13,* it is not enough to restore fertility; the presence of *Rf2* is also necessary. It is not clear how *Rf2* restores fertility, but it has been identified as an aldehyde dehydrogenase gene and shown to be necessary for normal tassel development even in non-CMS cytoplasms (Cui et al., 1996; Liu et al., 2001).

1.6.2 Cytoplasmic Male Sterility Type S (CMS-S)

Another type of CMS in maize, CMS-S, has a phenotype correlated with a region that contains two ORFs, a unique *ORF355* and *ORF77, ORF77* has segments of sequence similarity to *atp9* (Zabala et al., 1997). The two ORFs are co-transcribed producing two stable transcripts of 2.8 kb and 1.6 kb. When the CMS-S lines are crossed to a line containing *Rf3,* the 1.6 kb transcript is reduced in abundance and an increase in a 1.1 kb transcript is observed (Wen and Chase, 1999; Zabala et al., 1997). *Rf3* has been mapped to the long arm of chromosome 2 and is distinct from the Rf genes for C and T in that its restoration occurs at the stage of gamete development (gametophytic; Kamps et al., 1996). However, a protein associated with the CMS-S phenotype has yet to be identified.

1.6.3 Cytoplasmic Male Sterility Type C (CMS-C)

CMS-C was first identified in the Charrua maize variety by Beckett (1971). The restoration of CMS-C is sporophytic. It can be restored to fertility by one of three nuclear restorer of fertility genes; *Rf4*, *Rf5*, and *Rf6* (Josephson, 1978).

Cytological studies of CMS-C identified two different stages at which pollen abortion occurs (Lee, 1979). In the first type of pollen abortion (Type I), the tapetum at the tetrad stage had a denser cytoplasm, was binucleate, and had several small vacuoles which are similar to normal pollen development. At the young to intermediate microspore stage, the tapetal layer contained larger vacuoles and had an increased

radial thickness. Subsequently, the tapetal layer became less dense, which led to microspore abortion in the intermediate to late microspore stage (mononuclear).

In the type II pollen abortion described by Lee *et al.* (1979) anthers had a more vacuolated tapetum with an increased radial thickness beginning at the early tetrad stage. The tapetum continued to thicken, filled most of the locule and had increasing vacuolation until the late tetrad stage. At this point the tetrads aborted and the locule collapsed. Pollen development aborts in type II before type I has its first physical malformations.

Both of these types of pollen abortion were occasionally seen in different anthers from the same plant but rarely seen in different locules of the same anther. It is possible that the stage of tapetal morphological changes determines the type of pollen abortion occurring (Lee *et. al.*, 1979).

Proteins synthesized "in organello" in mitochondria isolated from etiolated seedling shoot were analyzed by 1D SDS-PAGE (Forde and Leaver, 1980). In CMS-C, a novel 17.5 kDa protein was expressed in CMS-C and it lacked a 15.5 kDa protein present in normal cytoplasm. However, no changes in protein expression were identified between CMS-C and CMS-C restored etiolated shoots. This means that the change in protein expression either could not be the cause of CMS-C pollen abortion or that a change in protein profile due to restoration could be tissue specific.

Previously, Dewey et al. (1991) identified three chimeric genes in the CMS-C mitochondrial genome, *atp6*, *cox2,* and *atp9* (denoted from now on as *atp9-2*). These

genes have different Northern banding patterns than the original genes from the normal cytoplasm (Dewey et al., 1991). However, there were no differences detected in the transcript pattern of the chimeric genes between sterile and restored genotypes (Dewey et al., 1991).

The chimeric *atp6* gene is a fusion of sequences from *atp9*, an unknown region, and *atp6* (Fig 3b). This chimeric gene would potentially be translated into a 429 amino acid chimeric peptide. However, the peptide may be proteolytically cleaved in a similar manner to wheat (Begu, 1988), between the glutamine and serine residue in the *atp6* sequence to produce a core peptide. If this cleavage occurs, then the core peptide produced would be identical to that synthesized in normal maize mitochondria. Further analysis of the *atp6* transcript in CMS-C seedlings showed that it has the same RNA editing as the normal *atp6* gene (Levings and Siedow, 1992).

The chimeric *cox2* gene in CMS-C is comprised of the normal *cox2* gene with the *atp6* 5' flanking region, 143 codons of the *atp6* amino extension, and 43 codons that show homology to the core *atp6* protein (Fig. 3c). The normal *cox2* gene contains the same similarity to the 43 codons of the *atp6* core protein; however, the extended *cox2* gene (not part of the core protein) and the 5' sequence have no homology to the *atp6* gene (Dewey et al., 1986; Dewey et al., 1991). Western blots containing seedling mitochondrial protein from CMS-T, CMS-C, and CMS-C restored were probed with a COX2 antiserum. The results detected no differences in COX2 expression between the three genotypes.

The chimeric *atp9-2* gene in CMS-C was characterized based on Southern analysis as having a normal *atp9* coding region plus 119 bp upstream and 122 bp downstream with similarity to the normal *atp9* gene. The remainder of the studied 5' and 3' sequences of *atp9-2* had no similarity to the normal gene (Dewey *et al*. 1991; Fig. 3a). No further work has been performed to analyze this gene.

Chapter 2: CANDIDATE GENES CAUSING CMS-C

2.1 Introduction

Cytoplasmic male sterility (CMS) has been found in several plant species (reviewed by Hanson and Bentolila, 2004). Most of the CMS-causing genes identified to date are unique to the mitochondrial genomes and contain fragments of known genes in their coding region (chimeric). All CMS genes are expressed because they are either close to a known protein coding gene or have undergone a rearrangement that placed the upstream sequence, containing the functional promoter of a known gene, in front of the CMS causing gene. All of the CMS genes have also been found to contain membrane spanning motifs, which means they most likely insert into the mitochondrial membrane. This has been shown to occur in the tapetum of developing anthers in maize CMS type T (Rhoads, 1995).

There are several CMS causing genes that contain not only the upstream sequence of a known gene but also the first few codons of the associated functional gene. Three of these, *orf224*, *orf222*, and *orf522* from *Brassica pol*, *Brassica nap*us, and sunflower respectively, use the *atp8* promoter and contain several of its codons (Brown, 1999; Sabar et al., 2003). Interestingly *orf224* not only uses the *atp8* promoter but is also upstream of, and cotranscribed with, the *atp6* gene (Brown, 1999). The wheat *orf256*

and Boro rice *orf79* have the first few codons of *cox1*; however, while wheat uses the *cox1* promoter, *orf79* is co-transcribed with the *atp6* gene (Akagi et al., 1994; Song and Hedgcoth, 1994). The other two CMS genes that begin with protein-coding sequence are *orf107* in A3 sorghum and the petunia *Pcf* gene. Both *orf107* and *Pcf* start with the coding region of *atp9* and uses its promoter. The petunia *Pcf* gene also contains a chimeric region in its coding sequence from *cox2* (Boeshore, 1985; Tang et al., 1996).

The remaining CMS-causing genes are co-transcribed with known protein coding genes. Orf263 in *Brassica tour* is co-transcribed with *atp6* and is chimeric for two regions, one from *atp6* and the other from *nad5a* (Landgren et al., 1996). *Urf13* in maize CMS type T is co-transcribed with the *atp4* gene, using a duplicated promoter from *atp6* (Dewey et al., 1986). Two other CMS-causing genes, *orf138* from *Brassica ogura* and *orf125* from Kosena radish, are not chimeric (i.e. do not contain fragments of known genes) and are co-transcribed with the *atp8* gene (Iwabuchi et al., 1999; Makaroff et al., 1989; Makaroff and Palmer, 1988).

S-type cytoplasmic male sterility in maize is a unique case where the transcript is known; however, there are two genes on it: *orf355* and *orf77*. *Orf77* is chimeric with the *atp9* gene while *orf355* is not novel (Zabala et al., 1997). The sterility-associated transcript uses a promoter from within an inverted repeat of a mitochondrial plasmid but it is not known which ORF is causing pollen abortion.

This pool of knowledge on CMS-causing genes can be used to determine criteria for searching the newly sequenced CMS-C genome for candidate CMS genes.

2.2 Methods

2.2.1 Identification of Chimeric Genes

Chimeric genes were identified by a BLAST (http://blast.ncbi.nlm.nih.gov) search of all the genes in the B37N (Gen Bank No. NC007982) genome plus 2 kb upstream and downstream flanking sequence against the CMS-C mitochondrial genome (Gen Bank No. DQ645536). Any genes with their flanking and/or coding sequence that did not exist as a single continuous sequence in CMS-C were noted as chimeric. The genes labeled as chimeric were further analyzed using BLAST searches against the remaining sequenced maize mitochondrial genomes; NA (Gen Bank No. DQ490952), CMS-T (Gen Bank No. DQ490953) and CMS-S (Gen Bank No. DQ645536). The structure of chimeric CMS-C genes was diagrammed based upon the sequence location in NB.

2.2.2 Identification of chimeric ORFs

ORFs larger than 50 codons were identified (Allen et al., 2007) and compared using BLAST against the B37N genome. All fragments of homology were compared to the annotation of the B37N genome and the fragments located within protein coding and ribosomal RNA genes were noted. All ORFs that contained a fragment of sequence from an identified gene were denoted as chimeric. Each of the chimeric ORFs was further

analyzed by BLAST comparisons against all of the sequenced maize mitochondrial genomes (B37N, NA, CMS-S, and CMS-T) to determine if it is unique to CMS-C.

In order to determine the potential for expression, the flanking sequence, up to 4 kb upstream and downstream of the chimeric ORFs, was analyzed using BLAST and the CMS-C published annotated genome. Any ORF that had a known gene within 4 kb of its upstream flanking sequence was noted.

2.2.3 ORFs at Rearrangement Points

Previous work identified the major rearrangement points in CMS-C as compared to the B37N mitochondrial genome (Allen et al., 2007). The locations of these rearrangement points were used and all ORFs within 2 kb of the rearrangement points were denoted for further analysis. The ORFs were further examined by BLAST comparisons against all sequenced maize mitochondrial genomes (B37N, NA, CMS-S, and CMS-T) to determine if they were unique to CMS-C or occurred in other genomes.

In order to determine the potential for expression the flanking sequence up to 4 kb upstream and downstream of the chimeric ORFs were analyzed using BLAST and the CMS-C published annotated genome. Any ORFs that had known genes within range of their upstream flanking sequence were noted.

2.2.4 Transmembrane Domain Identification

Candidate CMS causing ORFs were analyzed for transmembrane domains using TMBETA-NET and SOSUI. TMBETA-NET identifies β strands while SOSUI identifies α-helix transmembrane motifs (Gromiha and Suwa, 2005; Hirokawa et al., 1998). Analysis was performed using the default paramaters.

2.3 Results

2.3.1 Identification of Chimeric Genes

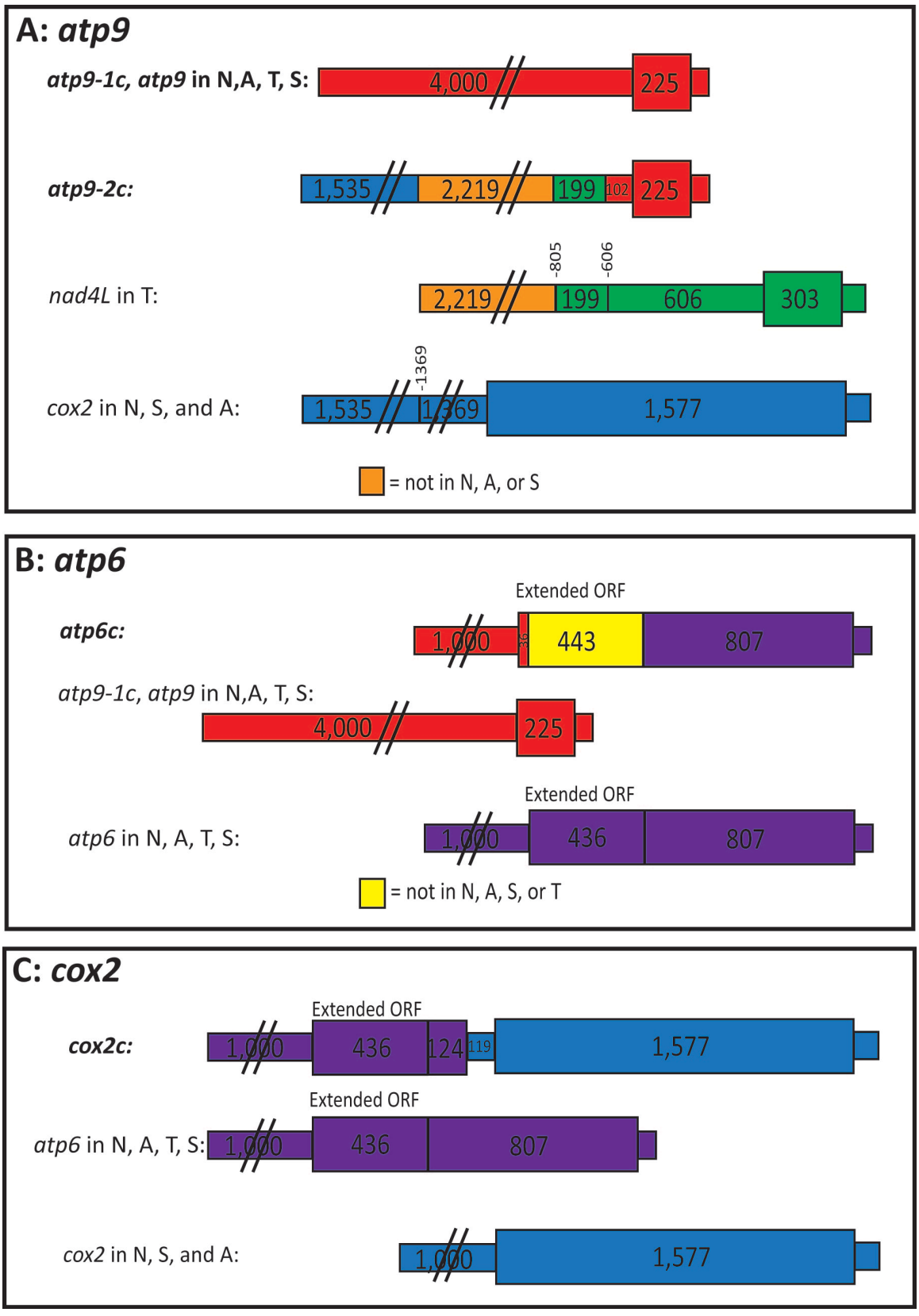
Three chimeric genes in the CMS-C mitochondrial genome were previously identified by Dewey et al. (1991). A chimeric *cox2* gene in CMS-C has the same complete coding region and 3' flanking region of *cox2* from the B37N, NA, and CMS-S genomes (Fig 3c). The CMS-C *cox2* also shares the first 119 bases of the 5' flanking sequence with B37N, NA, and CMS-S, after which *cox2* ends and homology to *atp6* from B37N, NA, CMS-T, and CMS-S begins. This homology extends out to at least 4,000 bases upstream of the *cox2* translation start codon, and contains the extended ORF and first 124 codons of *atp6* (Fig. 3C).

The chimeric *atp6* gene in CMS-C contains the functional region of *atp6*. The functional coding region is preceded by an extended ORF that is unique to CMS-C and contains the first 36 bases of the *atp9* coding region. The 5' flanking region is from *atp9* up to 4kb upstream of the start codon (Fig. 3B).

Figure 3: Chimeric Genes in the CMS-C Genome

Diagram depicting the chimeric genes present in the CMS-C genome. N stands for the mitochondrial genome in B37N, A is found in A188 and also known as NA, T and S represent the CMS-T and CMS-S genomes respectively. Coding regions for the genes are designated by being a larger box. A. The two *atp9* genes present in CMS-C are bolded and at the top of the diagram. The a*tp9-1* copy is normal while *atp9-2* has upstream sequence from *nad4L* and *cox2*. B. The chimeric *atp6* gene in CMS-C contains a unique extended orf with the first 36 bases of the coding region of *atp9* along with its upstream sequence. C. The upstream sequence, extended ORF and first 124 codons of *atp6* are upstream of the *cox2* gene in CMS-C.

Figure 3 Chimeric Genes in the CMS-C Genome



The final chimeric gene in the CMS-C genome is an *atp9* gene with a rearranged 5' flanking sequence (Fig 3A). The chimeric *atp9* gene in CMS-C, denoted as *atp9-2*, is present in addition to a previously unidentified non-rearranged *atp9* gene (*atp9-1*). The a*tp9-2* contains the same coding region and 3' flanking sequence as the *atp9* genes from B37N, NA, CMS-T, and CMS-S until the end of the mapped 3' untranslated region. The first 102 bases of the 5' flanking sequence are also shared with the normal version of *atp9.* Then homology to the 5' flanking sequence of *nad4L* in CMS-T begins and continues for 2,418 bases. Further upstream, sequences corresponding to the 5' flanking sequence from *cox2* are found (Fig 3A).

2.3.2 Identification of Chimeric ORFs

ORF analysis identified fourteen chimeric ORFs in the CMS-C genome: orf410, orf246a, orf191c, orf186a, orf163b, orf149a, orf147b, orf101c, orf90, orf86, orf70, orf62, orf61, and orf59b (Fig 4 and table 1). Orf410 contains 116 base pairs (bp) from *atp6* at bp 473-588 that are out of frame in relation to the *atp6* gene. One α-helix transmembrane domain was identified within orf410. Orf246a is close to the *rps4* gene and contains four short regions of homology to *atp6* at bp 369-394, 538-561, 628-650, and 662-690, all of which are out of frame (Fig 4). There is also a 21 bp stretch of homology to *rps3* exon2 at bp 704-724 in orf246a. The chimeric region of orf246a overlaps with an alpha helix domain that lies at 645-711 bp. In orf191c there is only a small 30 bp out-of-frame region from *rps3* at bp 304-333. Orf191c is upstream of *nad7* and is predicted to have six β-segments and two α-helices. Orf186a is flanking *nad2* and

transmembrane analysis predicted seven β-segments and two α-helices. Orf186a also contains two regions of homology, one at bp 66-145 with homology to *cox2* exon1, the other at bp 304-333 with homology to *rps2B* (Fig 4); both of these regions are out of frame in orf186a. The start of orf163b consists of the first 74 bp of *rps3* exon 1 followed by 209 bp with homology to *rps12* from bp 259-467; this region also contains three predicted α-helices. Orf149a contains only one small 19 bp region of homology at 349-367 to the *rrn26* gene. In orf147b, which is adjacent to *cox1,* there are three regions of homology to three different genes. The first occurs after the only predicted α-helix at bp 301-324 and has homology to *atp4* in reverse orientation. This is immediately followed by 19 bp from *cox1*. The final region of homology is from *atp9* in the reverse orientation and resides at bp 382-420. The chimeric orf101c occurs within the intron of *nad2* and starts with 26 bp from *rps3* exon2 with reverse orientation. Orf90 has one fragment of *rps3* that is 31 bps long. The fragment of *atp6* that is in orf86 is slightly larger at 67 bp long. There is only one 42 bp chimeric region in orf70 from *nad7*. Orf62, flanking *rps4*, is one of two ORF's unique to the C genome and lies within orf246a sharing several small chimeric regions: two fragments from *atp6* 23 and 29 bp long and one from *rps3* 21 bp long. The other chimeric ORF that is unique to the C genome is orf59b; it contains one 29 bp fragment from *atp9* and three predicted β segments. The final chimeric ORF, orf61, has two chimeric regions: one that is 56 bp long from *atp4*, the other is from *atp1* and is 31 bp long. These results are summarized in table 1.

Figure 4: Chimeric ORFs in CMS-C genome

Diagram depicting all of the chimeric ORFs larger than 50 amino acids present in the CMS-C genome; ORFs are to scale and all chimeric regions are color coded based upon the gene from which the chimeric region came.

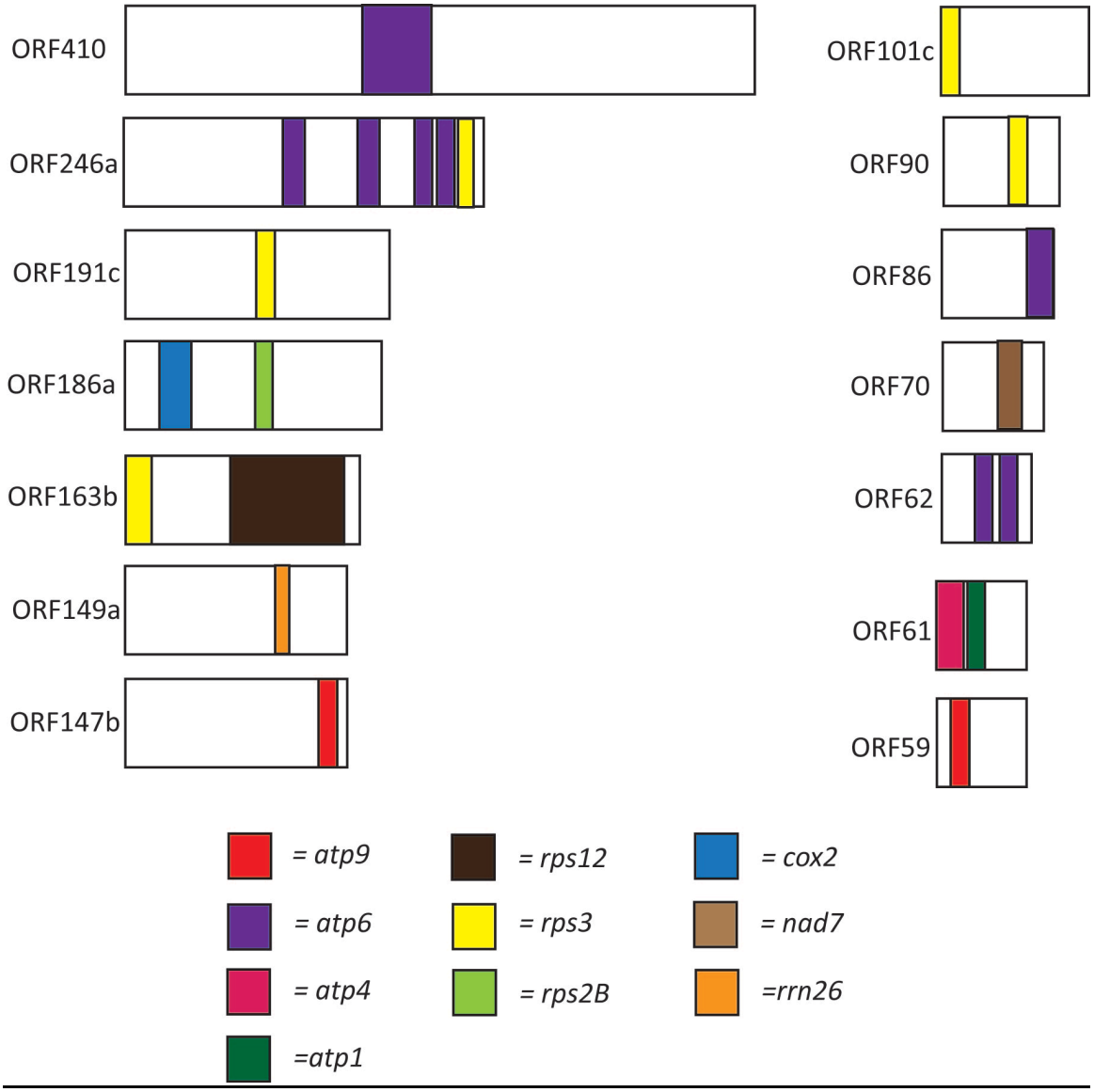


Table 1: Candidate chimeric CMS causing ORFs

All Chimeric ORFs present in the CMS-C genome are listed followed by the size and origin of the chimeric region of each ORF. An ORF was considered unique to C if it did not exist in any form within the other sequenced maize mitochondrial genome. Orf410, orf246a, and orf61 were present in at least one other genome, however they used either a different start or stop codon making the ORF a different size. An ORF was considered to have transmembrane motifs if they were predicted to contain β-segments that could form a β-barrel, or α-helices based upon TMBETA-NET and SOSUI. It was also noted if any of the ORF's were close to the upstream sequence of a known gene, a known gene itself, or internal to a gene/ORF.

Table 1: Candidate chimeric CMS causing ORFs

ORF	Chimeric Region	Unique to C Genome	Transmembrane Motifs	Within 2 kb of Promoter/Gene
410	112bp *atp6* (NB extended ORF)	as 410[1]	Yes	No
246a	26, 24, 23, 29bp *atp6* (C extended ORF); 21bp *rps3*	as 246a[2]	Yes	*rps4* gene
191c	30bp *rps3*	No	Yes	*nad7* gene
186a	80bp *cox2*; 28bp *rps2B*	No	Yes	*nad2* gene
163b	264bp *rps3*; 209bp *rps12*	No	Yes	No
149a	35bp *trnR*	No	No	No
147b	24bp *atp4*; 39bp *atp9*; 19bp *cox1*	No	Yes	*cox1* gene
101c	26bp *rps3*	No	No	Intronic *nad2*
90*	31bp *rps3*	No	No	*nad7* gene Internal (Internal orf191c)
89	28bp plastid *rrn23*	No	No	No
86*	67bp *atp6*	No	No	No (Internal orf410)
70	42bp *nad7*	No	No	No
62*	23, 29 *atp6*; 21 *rps3*	yes	No	*rps4* gene (Internal orf246a)
61	56 *atp4*; 31 *atp1*	Yes	No	No
59b	29bp *atp9*	Yes	Yes	No

Table 1: Candidate chimeric CMS causing ORFs

* ORFs that occur internal to the larger ORF indicated
[1]similar to orf396 in NB genome
[2]similar to orf248 in NB

2.3.3 ORFs at Rearrangement Points

There are fourteen ORFs in the CMS-C genome that lie within 2 kb of rearrangement points when compared to the B37N genome (Table 2): orf307a, orf264a, orf247, orf227a, orf193a, orf186a, orf140d, orf137a, orf126a, orf120a, orf117a, orf112b, orf110a. orf105a. Only one of these ORFs is unique to the C genome: orf264a. There is also one ORF, orf186a, which was also identified as a chimeric ORF. Orf117a is intronic to *nad2-ex1,* and *2* and exists in this state in other mitochondrial genomes. Two ORFs overlap so that all but 22 bp of orf137a is part of orf247. These ORFs are close to the *ccmB* and *trnH* genes. There are two genes close to *nad4L*, orf110a and orf126a, while orf140d is close to both the *nad3* and *nad5* genes. The two remaining ORFs close to known genes are orf120a which is upstream of the *ccmB* gene and close to *atp1* is orf193a. These results are summarized in table 2.

There were two ORFs at rearrangement points that were predicted to contain at least one α-helix. Orf112b contained one α-helix at its N termini and orf105a had a predicted α-helix at the N and C termini. Five other ORFs contained predicted β-sheets. TMBETA-NET predicted 11 β-segments in orf227a, 10 for orf193a, 6 for orf140d and orf20a, and four for orf126a.

Table 2: Candidate CMS causing ORF's at rearrangement points

All ORFs are listed that are within 2 kb of rearrangement points when comparing the CMS-C genome to that of NB. An ORF was denoted as unique to the C genome if it was not present in any of the other sequenced maize mitochondrial genomes. The potential for transmembrane motifs was predicted based upon TMBETA-NET and SOSUI. The ORFs that lie close to known genes or the upstream sequence of known genes was also denoted.

ORF	**Unique to C Genome**	**Transmembrane Motifs**	**Within 2kb of Promoter/Gene**
307a	No	No	No
264a	Yes	No	No
247	No	No	*ccmB*
227a	No	Yes	No
193a	No	Yes	*atp1*gene
186a	No	Yes	*nad2* gene
140d	No	Yes	*nad3* and *nad5* genes
137a	No	No	*ccmB*
126a	No	No	*nad4L*gene
120a	No	Yes	*ccmB* gene
117a	No	No	No
112b	No	Yes	No
110a	No	No	*nad4L* gene
105a	No	Yes	No

2.4 Discussion

When the known CMS-causing genes are compared, several commonalities are observed. All of them are transcribed because they are either close to a known functional gene and are co-transcribed with it, or the CMS gene has a chimeric upstream sequence from a protein-coding gene and its promoter is driving CMS gene expression. Most of these CMS-causing genes are also unique to the genome and are chimeric with sequences from known protein-coding genes within the translated region. A majority of the chimeric CMS-causing genes have rearrangements with ATP synthase complex genes or are co-transcribed with them (Hanson and Bentolila, 2004). Another major commonality among the CMS genes is the occurrence of membrane-spanning motifs. These criteria can be used to help identify a candidate gene causing pollen abortion in CMS-C.

The ORFs that are at least 50 amino acids long in the CMS-C genome were first analyzed to identify fourteen chimeric ORFs. Since not all CMS-causing genes are chimeric with pieces of known genes the list was expanded to include thirteen more ORFs that are larger than 100 amino acids and are within 2 kb of a rearrangement point in CMS-C when compared to the B37N genome. This generated an initial list of 27 candidate CMS-causing genes in CMS-C (Fig. 4, Tables 1 and 2).

Over half of these potential CMS genes, 14 out of 27, have membrane-spanning motifs. However, this does not mean they are expressed and can insert into the

membrane. Eleven of the candidate ORFs lie close to known genes and could be co-transcribed with the adjacent gene. In addition there are three ORFs that are unique to the CMS-C genome; orf264a, orf62, and orf59b.

Although any of these ORFs could be potentially causing male sterility, the strongest candidate was orf186a. Orf186a is chimeric, contains predicted transmembrane domains, is located at a rearrangement point, and is adjacent to the *nad2* gene. However, orf186a is not unique to CMS-C. There are three other ORFs, orf246a, orf191c, and orf147b, that are chimeric, close to known protein-coding genes, and which contain transmembrane domains. However, none of these genes is at a rearrangement point nor is any of them unique to CMS-C; they exist in the same state in the normal B37N genome.

Only one of the ORFs that are unique to CMS-C, orf62, is close to a known gene (*rps4),* with which it could be co-expressed. One other ORF that is unique to CMS-C, orf59b, is not close to any functional gene, nor to any upstream sequence of a known gene, but it does have three potential β-segments. There were no CMS-C unique ORFs that were strong candidates as CMS-causing genes. Therefore, the known protein coding genes were further analyzed to determine if they have any rearrangements around them. This is important since many of the CMS causing genes are in chimeric regions upstream of known protein-coding regions.

One case of CMS is caused by the extended ORF of a chimeric *atp6* gene (Yamamoto et al., 2005). Extended ORFs occur in front of all plant *atp6* genes, although their

compositions vary. It is believed that the extended ORF is translated and then immediately proteolytically cleaved and degraded to produce the mature ATP6 protein. In the Owen cytoplasm of sugar beet, the extended ORF is proteolytically cleaved but not degraded. It then forms concatomers of itself and inserts into the membrane to cause pollen abortion (Yamamoto et al., 2005).

There are three genes in CMS-C that were identified as chimeric, *atp6*, *atp9*, and *cox2*, because they contained rearrangements in their upstream un-translated sequence (Fig. 3). The *atp6* gene does have a chimeric extended ORF that contains the beginning of the *atp9* gene and was thought to be the most likely cause of pollen abortion in CMS-C. However this extended ORF does not contain transmembrane domains and so would not function like other CMS genes (Allen *et al*., 2007). It is still possible that the extended ORF is not being degraded, but instead of inserting into the membrane it is associating with normal ATP9 proteins and impeding normal function. Another possibility is the extended ORF in CMS-C is not being proteolytically cleaved, generating a chimeric *atp6* protein that is not fully functional. In both of these hypotheses, the function of *atp9* or *atp6* cannot be greatly impaired in non-tassel material or a defective plant phenotype would most likely be seen. There is still the possibility of a chimeric ATP6 protein being produced by this transcript or tassel specific effects and restoration.

In addition to the previously identified chimeric *atp9* gene (*atp9-2*) the analysis of the CMS-C genome identified a normal copy (*atp9-1*) and so *atp9-2* does not need to be transcribed or translated for normal mitochondrial function. However, previous

research showed an altered RNA pattern for *atp9* (Dewey et al., 1991) which would not be expected if *atp9-1* is functioning normally.

The best studied chimeric gene in CMS-C is *cox2*. Studies by Dewey et al. (1991) determined that the chimeric transcript is not altered in the presence of the *restorer of fertility* gene (*Rf4*) in seedlings. Furthermore, no chimeric COX2 protein is observed, and COX2 protein levels are not affected (Dewey et al., 1991).

Chapter 3: ANALYSIS OF TRANSCRIPTS FROM CANDIDATE CMS-C CAUSING GENES

3.1 Introduction

The first CMS-causing gene, *urf13*, was identified in maize CMS-T. A unique RNA transcript was found in CMS-T (compared to normal maize mitochondria) that was also altered by the CMS-T *restorer of fertility* (*Rf*) genes (Dewey et al., 1986). This type of comparative analysis has been successfully implemented to identify over eight other CMS-associated genes (Hanson and Bentolila, 2004).

A similar method, hybridizing seedling RNA with clones containing known mitochondrial genes, identified the chimeric genes, *atp6*, *atp9*, and *cox2* in the maize CMS-C genome. Their RNA transcripts were altered relative to the corresponding transcripts from normal mitochondria (Dewey et al., 1991). However, these transcripts were not affected when the *Rf4* allele was introduced. This means that (1) these genes are not involved in pollen abortion; (2) restoration occurs in a tassel specific manner; or (3) restoration of fertility does not occur at the RNA level.

A comparative analysis approach was used to test the candidate CMS-C causing genes. Expression of candidate genes were tested on gel blots containing normal and CMS-C RNA from etiolated seedling shoots and meiotic tassels. CMS-C in the restored

background was also analyzed for alteration of the stable RNA transcripts. Since each CMS-causing gene identified to date is the result of rearrangements and is associated with a novel transcript when compared to related fertile cytotypes (Hanson and Bentolila, 2004), this approach should identify the CMS causing gene in CMS-C.

3.2 Materials and Methods

3.2.1 Mitochondrial-Enriched RNA Isolation

The protocol used to isolate mitochondrial-enriched RNA was a modification from the protocol outlined by (Stern and Newton, 1985), using TRIzol reagent (Invitrogen Cat#15596-026). All solutions used were autoclaved and all isolation steps were conducted at 4° C.

Tissue was harvested from 6-day-old etiolated seedling shoots of *Zea parviglumis*, *Zea perennis*, A188 (NA type mitochondria), B37T (CMS-T type mitochondria), B37S (CMS-S type mitochondria), B37N (NB type mitochondria), B37C (CMS-C type mitochondria), and B37C/Ky21 (CMS-C mitochondria crossed to restorer line), along with pre-emergent tassels from B37N, B37C, B37C/Ky21, and B37N/Ky21.

The pre-emergent tassel material was determined to have entered into meiosis by staging the older anthers on the tassel using aceto-carmine as follows (Freeling and Walbot, 1994). The anthers were removed from the male floret and fixed in a 1.5 mL Eppendorf tube containing 3 parts 95% ethanol to 1 part glacial acetic acid. The anthers

were allowed to soak for a minimum of 10 minutes before being placed on a microscope slide. After one drop of aceto-carmine was added on top of the anthers they were smashed using rusty probes. A cover slip was placed over the smashed anthers followed by gentle heating over an ethanol lamp. The anthers were then examined under a 100x magnification to determine the stage of development.

The tissue was weighed, surface sterilized in 10% sodium hypochlorite, and rinsed in cold Diethyl Pyrocarbonate (DEPC)-treated de-ionized water. The tissue was then placed in a cooled blender jar containing cold homogenization buffer (a ratio of 4 mL buffer per 1 gm tissue). The homogenization buffer contained 0.35 M sorbitol, 50 mM Tris-HCl pH 8, and 5 mM EDTA; 20 mM 2-mercaptoethanol and 0.15% (w/v) bovine serum albumin was added just prior to use. Pulses of two seconds each at low and then high speed was used to homogenize the tissue in a Waring blender. The homogenate was filtered through a sterile funnel lined with four layers of cheese cloth on top of one layer of Miracloth into autoclaved 250 mL centrifuge bottles on ice. Nuclei, plastids, starch, and cellular debris were removed from the homogenate through two differential spins in an RC-5B Sorvall centrifuge at 4°C, with a pre-cooled GSA rotor. First, the filtrate was centrifuged at 2500 rpm (1000 x g) for 5 minutes. The supernatant was carefully decanted into sterile, cooled 250 mL bottles. The supernatant was then centrifuged again at 4000 rpm (2000 x g) for 10 minutes to remove additional plastids. The supernatant was transferred into a sterile, cooled 250 mL centrifuge bottle and centrifuged at 8000 rpm (10,000 x g) for 15 minutes to pellet the mitochondria. The

supernatant was carefully poured off and the pellet resuspended in 2 mL of RNA homogenization buffer.

The resuspended mitochondria were transferred to a sterile cooled 15 mL corex tube and centrifuged for 15 minutes in a cooled Sorvall HB-6 rotor at 8000 rpm (10,000 x g). The supernatant was poured off and the pellet resuspended in TRIzol reagent (Invitrogen, cat no. 15596-026). After incubating the resuspend mitochondria at room temperature for 5 minutes 0.8 mL of chloroform was added and vortexed for 15 minutes. The mitochondria were then incubated at room temperature for 10 minutes followed by centrifugation in a Sorval HB-6 rotor at 8500 rpm (12,000 x g) for 15 minutes.

The aqueous phase was transferred to a new tube and the RNA precipitated by adding 1.5 mL isopropyl alcohol and incubating at room temperature for 10 minutes. The RNA was pelleted by centrifuging at 8500 rpm (12,000 x g) for 10 minutes in a Sorval HB-6 rotor. The supernatant was then poured off and the pellet washed with 3 mL of 75% ethanol. The RNA was pelleted again by centrifuging at 6500 rpm (7,000 x g) for 5 minutes. The supernatant was poured off and the pellet air dried for 10 minutes. The pellets were resuspended in 2 µL formamide (Fisher Scientific cat# BP228-100) per gram starting material and stored at -80°C until use.

3.2.2 RNA Gel and Blotting

The RNA samples were quantified using a NanoDrop 1000. The RNA was then prepared for gel electrophoresis by bringing 10 µg of each RNA sample up to 11 µL in formamide (Maniatis et al., 1989). The samples were then heated at 65°C for 10 minutes. During the 10 minute incubation a master mix consisting of 2 µL 10x RNA running buffer (0.4M MOPS, 100 mM sodium acetate, 10 mM EDTA pH7.0), 3.3 µL 37% formamide, 1.7 µL 0.5 mg/mL ethidium bromide, and 2 µL loading dye per sample was made up. After incubating the samples were placed immediately on ice and 9 µL of the master mix was added for a final concentration of 1x RNA running buffer, 6% formamide, and 0.04mg/mL ethidium bromide. The samples along with the Fermentas ready-to-use High Range RNA ladder (Fermentas cat#SM0423) were then loaded onto a 30 cm long 1.2% agarose gel containing 1x RNA running buffer and 6% formaldehyde. The gel was run at 80 volts for 24 hours in a buffer containing 1x RNA running buffer and 6% formaldehyde.

In order to blot the RNA gel a glass plate was placed across a glass baking dish filled with 20x SSC (Maniatis et al., 1989). Two pieces of Whatman paper slightly wider than the gel were laid over the glass plate into the 20x SSC and the air bubbles removed. The gel was placed back side up on top of the Whatman paper followed by a piece of nylon membrane (Amersham, cat.# RPN303B) the size of the gel. Four pieces of Whatman paper the size of the gel were wetted in 20x SSC and placed on top of the membrane rolling out air bubbles between each piece. This was followed with a stack of paper towels and glass plates. The RNA was allowed to transfer to the membrane overnight.

The apparatus was disassembled and the membrane was crosslinked using a BioRad GS Gene Linker at 125mJ.

3.2.3 Probe Generation

Probe generation was performed on PCR products representing the genes or ORFs prepared previously (Appendix 1) using primer extension with the Stratagene Prime it II Kit (Maniatis et al., 1989). 5 µL of random primer was added to 25 ng of PCR product and brought up to 10 µL using ultrapure water. The sample was then boiled for 1-2 minutes and placed on ice immediately. 5 µL of 5x dCTP buffer and 1 µL of Klenow were added to the sample. The sample was then moved to the radiation bench and 5 µL of ^{32}P dCTP was added. The sample was placed in a 37° C water bath for 1 hour.

After the 37° C incubation, unincorporated nucleotides were removed from the probe reaction with Qiagen's nucleotide removal kit (Qiagen, cat no. 28304). 250 µL of Buffer PN was added to the 25 µL probe reaction and mixed. The sample was then pipetted into a spin column inserted into a 2 mL collection tube and centrifuged for 1 minute at 6000 rpm. The spin column was placed into a new collection tube. The samples were washed by adding 500 µL of Buffer PE and centrifuging for 1 minute at 6000 rpm. The flow-through was discarded and the spin column placed back into the 2 mL collection tube. Another 500 µL of Buffer PE was added and centrifuged for 1 minute at 6000 rpm. The flow-through was discarded and the spin column placed back into the 2 mL collection tube. To remove the residual ethanol, the column was centrifuged for 1 minute at 13,000 rpm. The spin column was placed into a clean 1.5 mL

centrifuge tube and the probe eluted twice with the addition of 100 µL Buffer EB and centrifuging for 1 minute at 13,000 rpm.

3.2.4 Hybridization

Hybridization of the northern blots was performed according to Maniatis *et al.* (1989) by placing the blot in a hybridization tube with a pre-hybridization solution containing 20 mL of formamide, 10 mL of 20 x SSC, 4 g dextran sulfate, 4 mL of 100x Denhardt's (2% (w/v) each of ficoll, polyvinylpyrrolidone, and BSA), and 2 mL of 10% (w/v) SDS (final concentration equaled 50% (v/v) formamide, 5x SSC, 10% (w/v) dextran sulfate, 10x Denhardt's, and 0.5% (w/v) SDS). After the dextran sulfate was in solution, 800 µL of a 10 mg/mL sonicated salmon sperm solution was placed in boiling water for 1 minute and added to the pre-hybridization solution (final concentration of 0.2 mg/mL salmon sperm). The pre-hybridization solution was added to the hybridization bottles and the blots pre-hybridized for 4 hours at 42° C. After the pre-hybridization was completed the probes generated were placed in boiling water for 1 minute and then added to the hybridization tube. The hybridization tubes were placed back into the hybridization oven and hybridized for 20 hours at 42° C.

After hybridization, the blots were washed with a solution containing 0.2 x SSC and 0.1% (w/v) SDS. The hybridization solution was poured into the radiation waste and the blots rinsed by adding 40 mL of wash solution. Another 40 mL of wash solution was added and the bottles placed back into the hybridization oven until the temperature reached 70° C. The wash solution was poured into the radiation waste and 40 mL of

wash solution was again added and the tubes washed at 70° C for 15 minutes. The wash solution was emptied into the radiation waste and two more washes were done with 40 mL of wash solution for 30 minutes at 70° C. The blots were removed from the tubes and a final one hour wash was performed at 70° C with 200 mL of wash solution in a Tupperware container. After the washes were complete the blots were removed, wrapped in Saran wrap, exposed onto a phospho-imager plate for 48 hours and scanned using Fujifilm FLA-2000 fluorescent image analyzer.

The blots were hybridized with a probe for the 18S and 26S ribosomal RNA (rRNA) genes after hybridization with the ORF or gene. The ladder, ORF, and rRNA images were aligned and sizes estimated.

3.3 Results

3.3.1 RNA Gel Blot Analysis of Candidate ORFs

Eighteen different probes were generated to test for expression of twenty candidate CMS-causing ORFs; two ORFs (orf62 and orf137a) overlapped other ORFs. Four of the seven probes representing the chimeric ORFs did not produce a detectable signal (ORFs 410, 191c, 163b, 59b). Another five of the eleven probes to ORFs located near rearrangement points in CMS-C compared with B37N did not produce a detectable signal on the northern blots (ORFs 307a, 264a, 227a, 120a, and 110a). The remaining

probes yielded detectable RNA transcripts on the Northern in all tissues and cytotypes tested.

The bands of the Fermentas ladder did not migrate similarly in all of the blots in relation to the rRNA probes. Furthermore, the sizes of rRNAs estimated from the ladder did not match the known sizes of 26S rRNA (3.5kb) and 18S rRNA 2kb. Thus the estimated sizes for the transcripts reported below may be inaccurate.

Three of the chimeric ORF probes detected a stable mtRNA transcript; orf246a/orf62, orf186a, and orf147b. Four major bands were observed with the orf246a/orf62 probe (Fig. 5B); the smallest is at ~2.5 kb followed by a ~2.9 kb band and then a ~4.5 kb band. The highest band varies among different cytotypes. The sizes can only be roughly estimated because the top band on the RNA ladder is only 6 kb. In the NA (A188) and CMS-C (B37C) cytotypes it is a ~8 kb band. *Z. perennis* produce a smaller transcript at around ~7 kb, while in *Z. parviglumis*, CMS-S (B37S), and B37N the transcript is larger at around ~9.5 kb, and CMS-T (B37T) has the largest transcript at ~10 kb. None of the transcripts in CMS-C are altered in the presence of the *restorer of fertility* (*Rf4*) gene (B37C/Ky21).

Due to a gel anomaly on orf186a (Fig. 6), no data was obtained for *Z. parviglumis*, *Z. perennis*, CMS-T, or CMS-S cytotypes. However, the seedling and tassel material for the B37N and CMS-C cytotypes had four major observable bands in both the B37N and CMS-C cytotypes one at 1.2 kb, the next at 2.8 kb, followed by a 3.2 kb band, and lastly a 3.9 kb band (Fig. 6B). Restoration of fertility did not affect any of these transcripts.

The orf147b probe had strong signals at 2.9 kb and 2.7 kb for all cytotypes except *Z. perennis* in which case the bands were at 3.0 kb and 2.7 kb (Fig. 7). Weak bands were also detected in all cytotypes at 0.3 kb, 0.4 kb, 2 kb, 2.2 kb, and 3.8 kb. As with the other probes for chimeric ORFs, *Rf4* (B37C/Ky21) did not alter expression of the detectable RNA transcripts.

Of the eleven probes used to test for the expression of the ORFs around rearrangement points in the CMS-C genome, six hybridized to RNA transcripts on the northern blots. The orf247c/orf137a probe hybridized to three bands at 4.5 kb, 2.9 kb, and 2.7 kb in all tissues of all tested cytotypes with the presence of the *Rf4* having no effect (Fig. 8A).

Hybridizations with the orf193a probe detected a 5.8 kb, 4.2 kb, and 3.5 kb band in all samples. There was also a weak 1.8 kb band in some of the samples (Fig. 9A). Introduction of *Rf4* to both the CMS-C and normal cytoplasm (B37C/Ky21 and B37N/Ky21) caused a reduction in these transcripts.

With the orf140d probe five bands were detected in all samples at 4.8 kb, 4 kb, 3.2 kb, 2.2 kb, and .6 kb (Fig. 10A). Only one RNA transcript hybridized to the orf126a probe at 2.5 kb in all samples and no change occurred in the presence of *Rf4* (Fig. 11A). The orf112b probe weakly detected two transcripts at 3.9 kb and 2.1 kb in all samples tested (Fig. 12A). The last probe that had detectable hybridization on the northern was orf105a with weak bands in all samples at 4.2 kb and 2.5 kb (Fig. 13A). As with the

chimeric ORFs that showed expression, introduction of *Rf4* had no effect on any of the observed transcripts.

3.3.2 Analysis of Mitochondiral Genes Transcripts

The expression of six different genes was tested, which represented the chimeric genes in CMS-C as well as individual members of different mitochondrial complexes. The non-chimeric genes tested were: *rps13*, *nad7*, and *cob*. The *rps13* probe detected six major transcripts for the gene at 5.9 kb, 3.8 kb, 3.3 kb, 2.8 kb, 2.2 kb, and 1.9 kb in all cytotypes with no differences detected in the presence of the restorer background (Fig. 14A). The *nad7* probe hybridized to one major band in all cytotypes tested at 4.4 kb (Fig. 15A). There were also several weaker signals detected only one of which varies in any of the genomes, this band is at 5.9 kb in all samples except for *Z. perennis*, in which case the band is larger at 6.1 kb. The last gene to be tested, *cob*, hybridized to several bands across all of the cytotypes (Fig. 16A). The only variation is that the A188 cytotype has an extra band at 3.5 kb. Interestingly the introduction of the restorer nuclear background to the CMS-C and B37N genomes yielded an extra band at 2.3 kb with no effect on the other RNA bands.

Transcripts for the three chimeric genes present in the CMS-C genome, *cox2*, *atp6*, and *atp9-2*, were analyzed. The probe used for *cox2* detected several stable RNA transcripts that varied greatly in intensity and size across the different cytotypes (Fig. 17B). All cytotypes produce a 2.1 kb, 2.4 kb, 3.8 kb, and 4.8 kb transcript. A 2.7 kb

transcript is produced only in *Z. perennis*, while A188, B37S, B37N, and B37C produce a 2.8 kb transcript. A188 also produces a unique 3.1 kb transcript and a 3.5 kb transcript shared only with B37N. There were also several larger minor transcripts that varied among the cytotypes.

The *atp6* probe produced results that are comparable to those previously published (Dewey et al., 1986), with all cytotypes except for *Z. perennis* and CMS-C yielding a 4.5 kb, 4.0 kb, 1.8 kb, and 1.6 kb (Fig. 18B). *Z. perennis* had only one major transcript at 1.6 kb, while CMS-C had a strong 2 kb transcript. The probe also detected another weaker band in the CMS-C genome at 4.6 kb. The presence of the restorer gene did not appear to affect the *atp6* transcripts in the CMS-C genome.

Several probes were used to test *atp9* expression. The first probe was to the coding region of the gene (Fig. 19B) and yielded results similar to those published by Dewey et al. (1991). There were three major transcripts in all cytotypes except for CMS-C at 1 kb, 0.8 kb, and 0.6 kb, while CMS-C only produced one major transcript at 5 kb. The probe designed to the upstream sequence of *atp9-1* in the CMS-C genome did not hybridize to any signals in any of the samples. The probe specific to *atp9-2* hybridized to one strong band at 5 kb in only the C cytotypes and also picked up weak bands in the CMS-T cytoplasm (Fig. 20B). A probe downstream of the mapped transcriptional stop site starting at 109 bases after the translation stop site of *atp9* did not detect any RNA transcripts (data not shown).

Figure 5: Transcripts hybridized with orf246 probe

A. Diagram of the chimeric orf246 and orf62 with the 5' ends labeled and the chimeric region colored according to Fig. 2. Line above the diagram indicates the region amplified for probe. The probe lies within the hypothetical coding region of ORF246, and is also downstream of orf62; numbering is based upon the first base of the theoretical translated region of orf246. B. Northern blot hybridized with the orf246/orf62 probe. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C genomes along with both of these genomes with the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. C. Loading Control: The blot from part B hybridized with probes for the 26S, and 18S rRNA.

Figure 5: Transcripts hybridized with orf246 probe

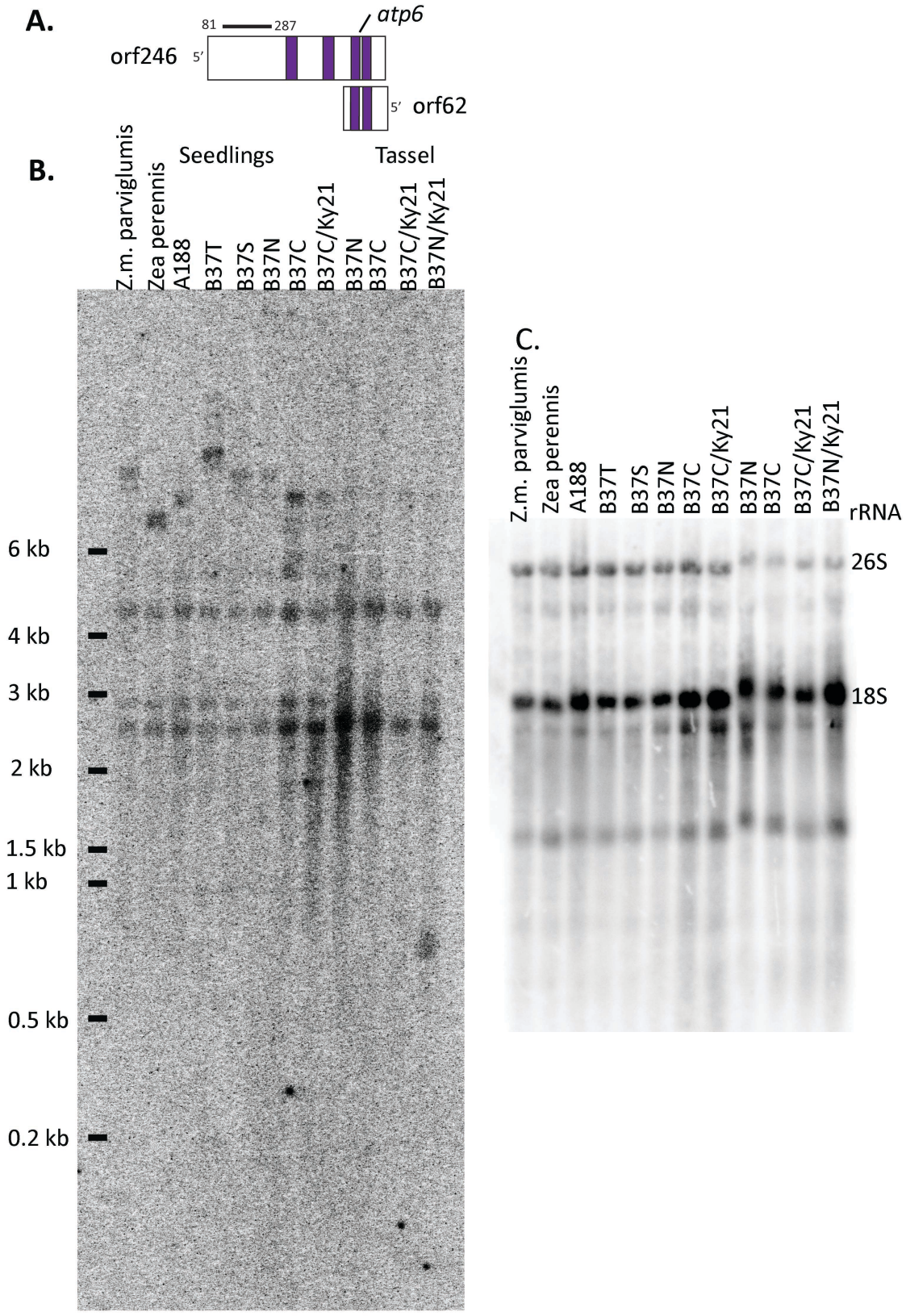


Figure 6: Transcripts hybridized to orf186a probe

A. Diagram of the chimeric orf186a, which contains fragments from *cox2* and *rps2B* (see also fig 2). Line above the diagram indicates the region amplified for probe. Numbering is based upon the first base of the translated region. B. Northern blot hybridized with the orf186a probe. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C genomes along with both of these genomes with the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. C. Loading Control: The blot from part B hybridized with probes for the 26S and 18S rRNA.

Figure 6: Transcripts hybridized to orf186a probe

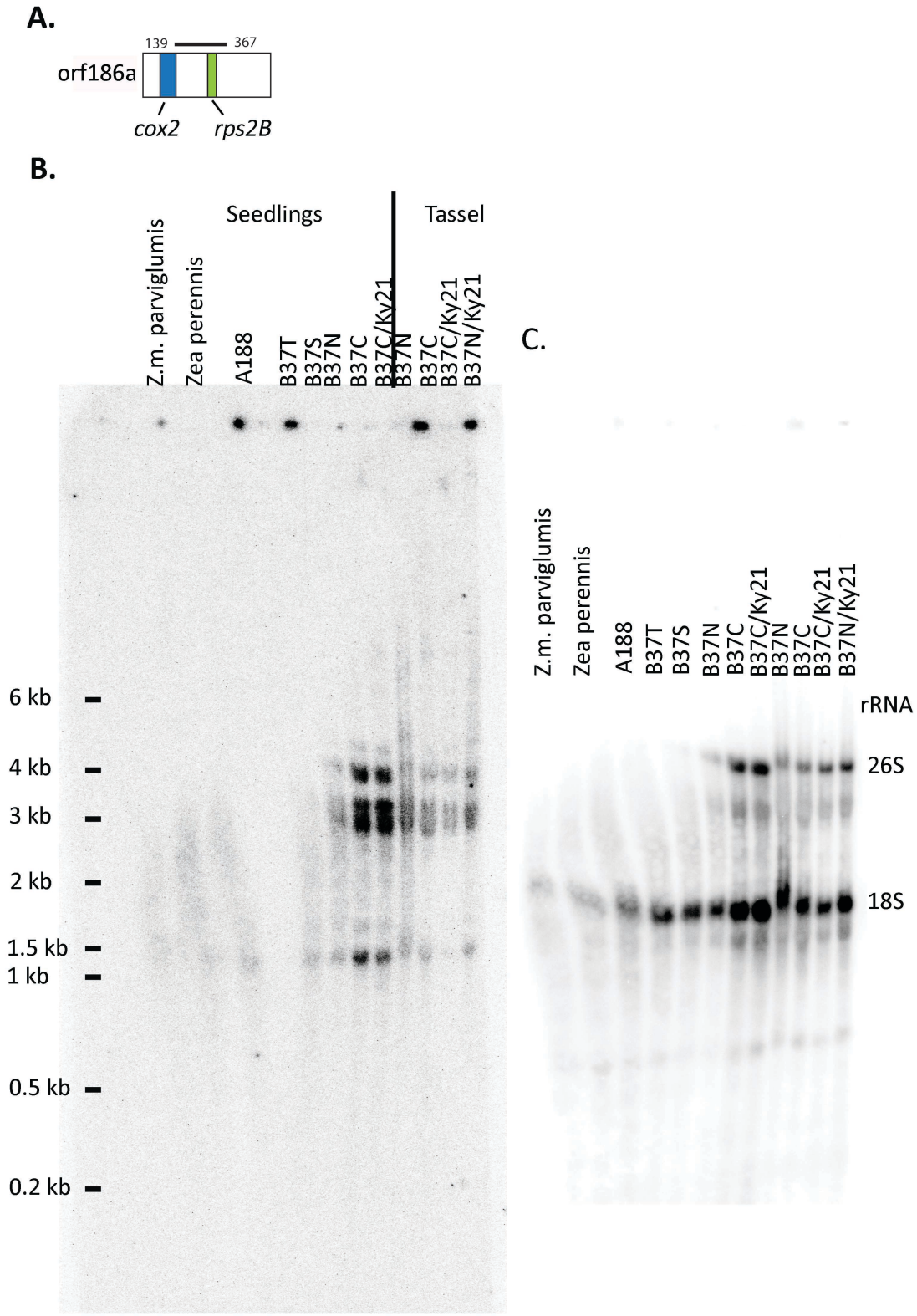


Figure 7: Transcripts hybridized with orf147b probe

A. Diagram of the chimeric orf147b containing fragments from *atp9* (see also Fig. 2). The probe used for the northern is drawn above the hypothetical coding region of the ORF; numbering is based upon the first base of the translated region. B. Northern blot hybridized with the orf147b probe. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C genomes along with both of these genomes with the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. C. Loading Control: The blot from part B hybridized with probes for the 26S and 18S rRNA.

Figure 7: Transcripts hybridized with orf147b probe

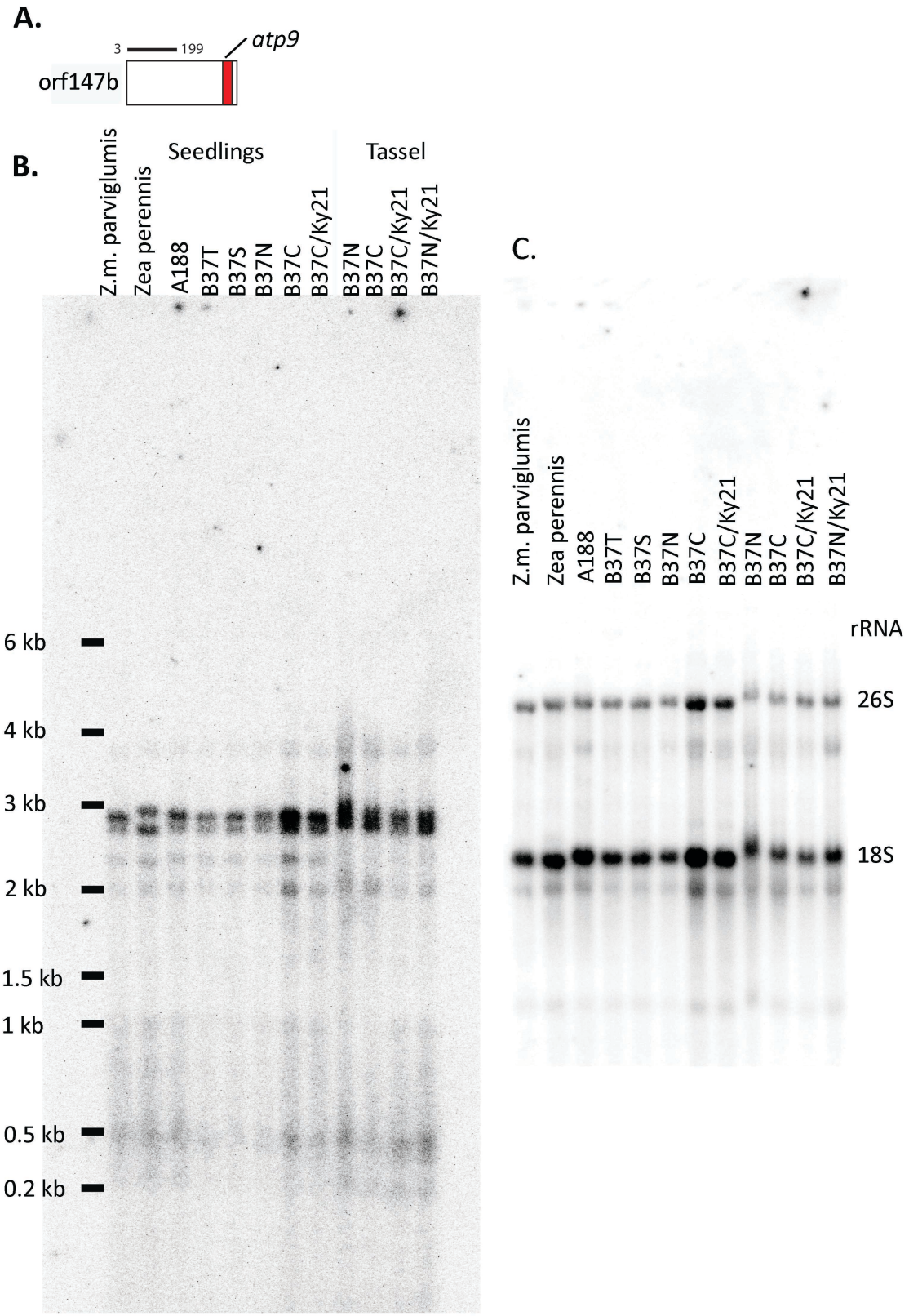


Figure 8: Transcripts hybridized with orf247c/orf137a probe

A. Northern blot hybridized with a probe to the hypothetical coding region of orf247c and orf137a. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes contain tassel mitochondrial RNA from the N and C cytotypes in a non restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 8: Transcripts hybridized with orf247c/orf137a probe

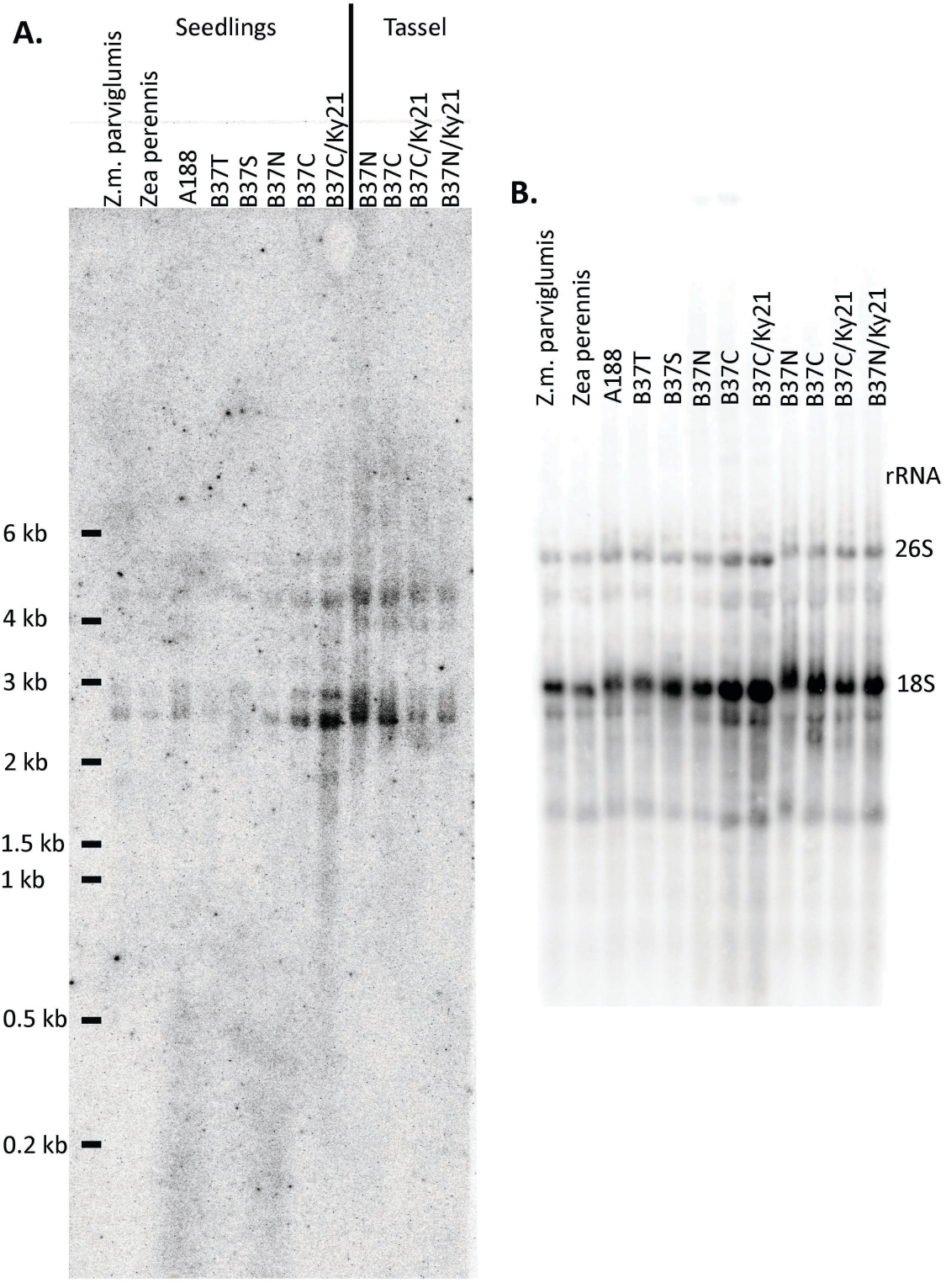


Figure 9: Transcripts hybridized with orf193a probe

A. Northern blot hybridized with a probe to the hypothetical coding region of orf193a. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C genomes along with both of these genomes with the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 9: Transcripts hybridized with orf193a probe

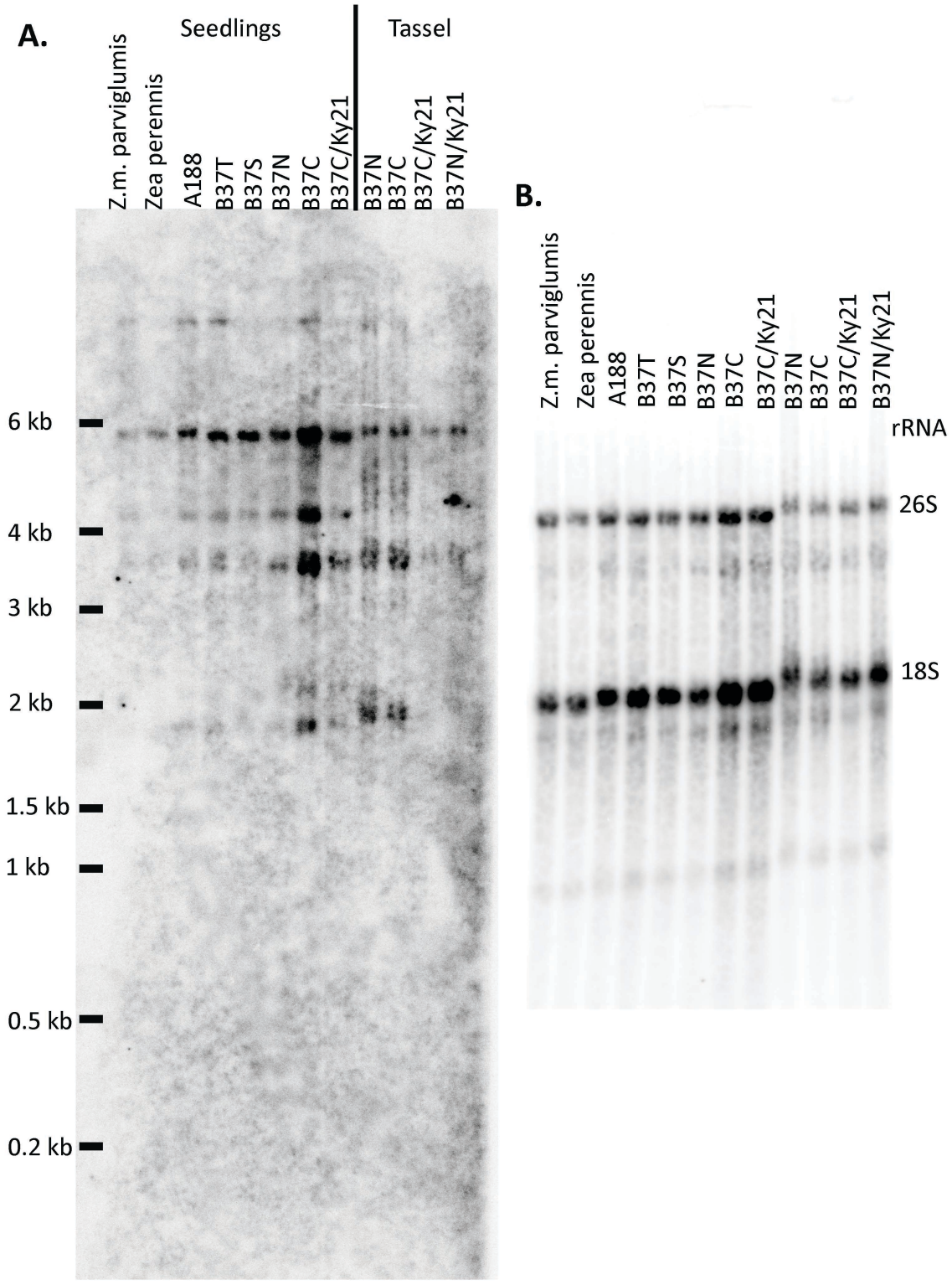


Figure 10: Transcripts hybridized with orf140d probe

A. Northern blot hybridized with a probe to the hypothetical coding region of orf140d. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C genomes along with both of these genomes with the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 10: Transcripts hybridized with orf140d probe

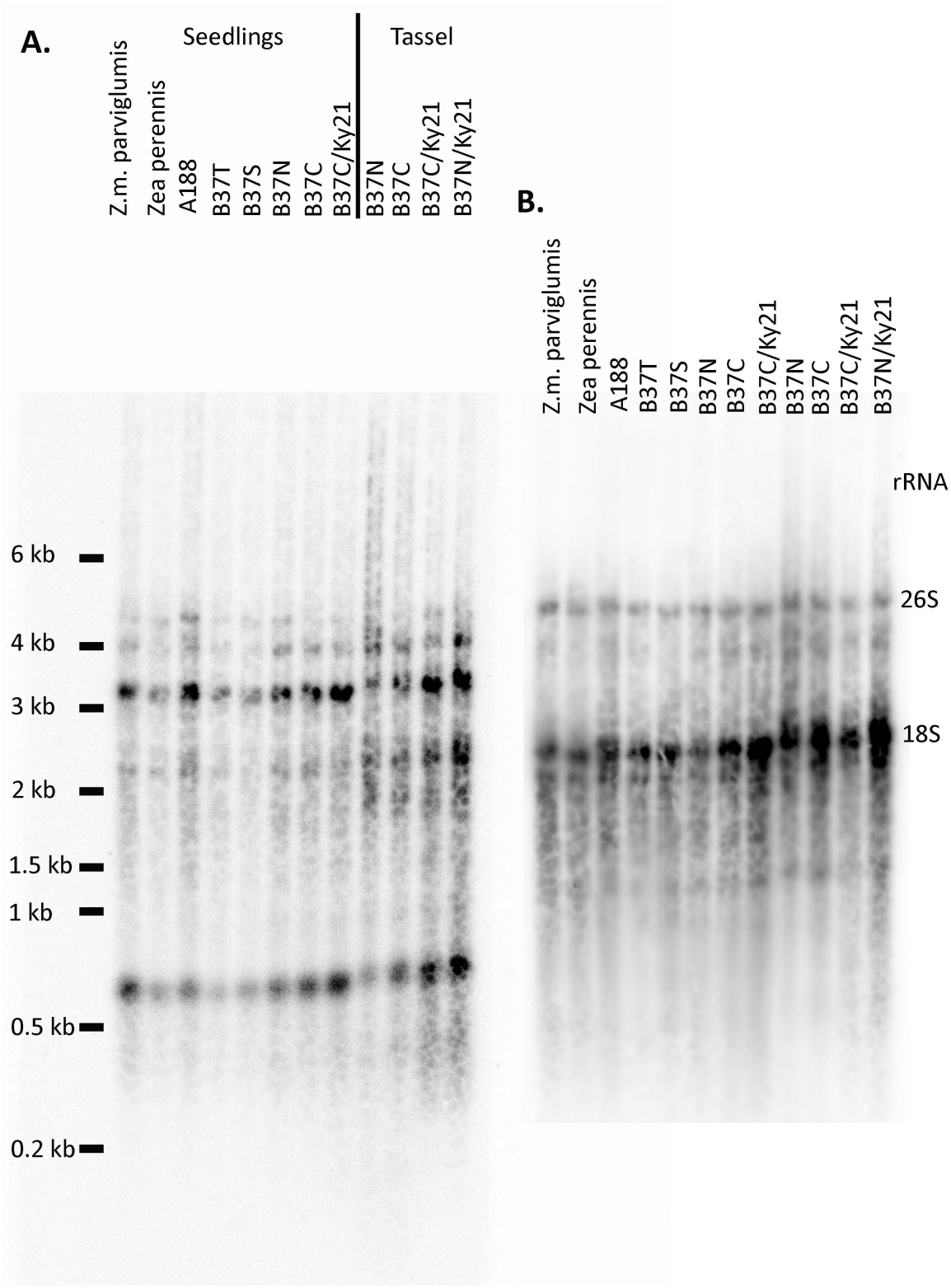


Figure 11: Transcripts hybridized with orf126a probe

A. Northern blot hybridized with a probe to the hypothetical coding region of orf126a. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C genomes along with both of these genomes with the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 11: Transcripts hybridized with orf126a probe

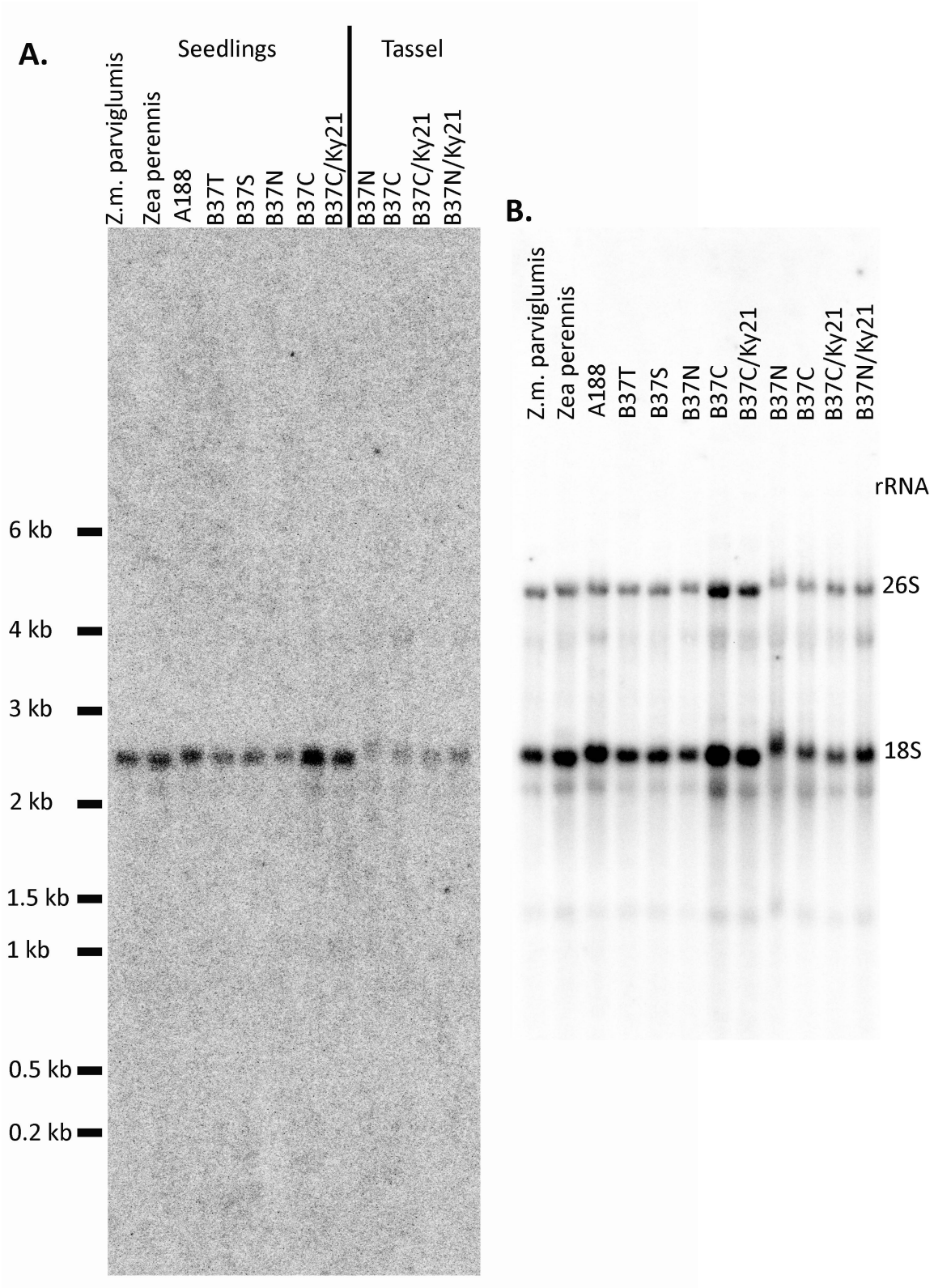


Figure 12: Transcripts hybridized with orf112b probe

A. Northern blot hybridized with a probe to the hypothetical coding region of orf112b. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 12: Transcripts hybridized with orf112b probe

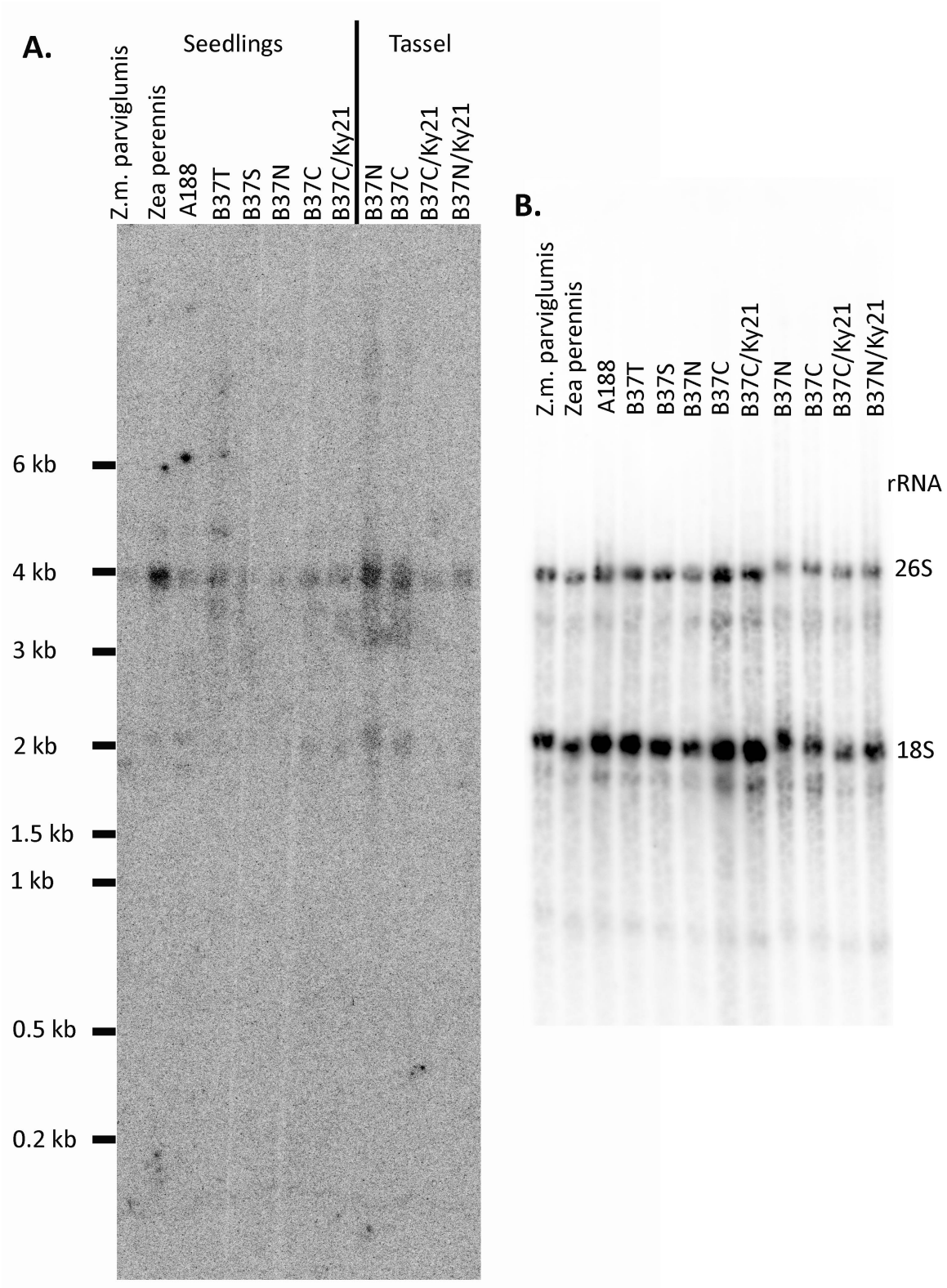


Figure 13: Transcripts hybridized with an orf105a probe

A. Northern blot hybridized with a probe to the hypothetical coding region of orf105a. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 13: Transcripts hybridized with an orf105a probe

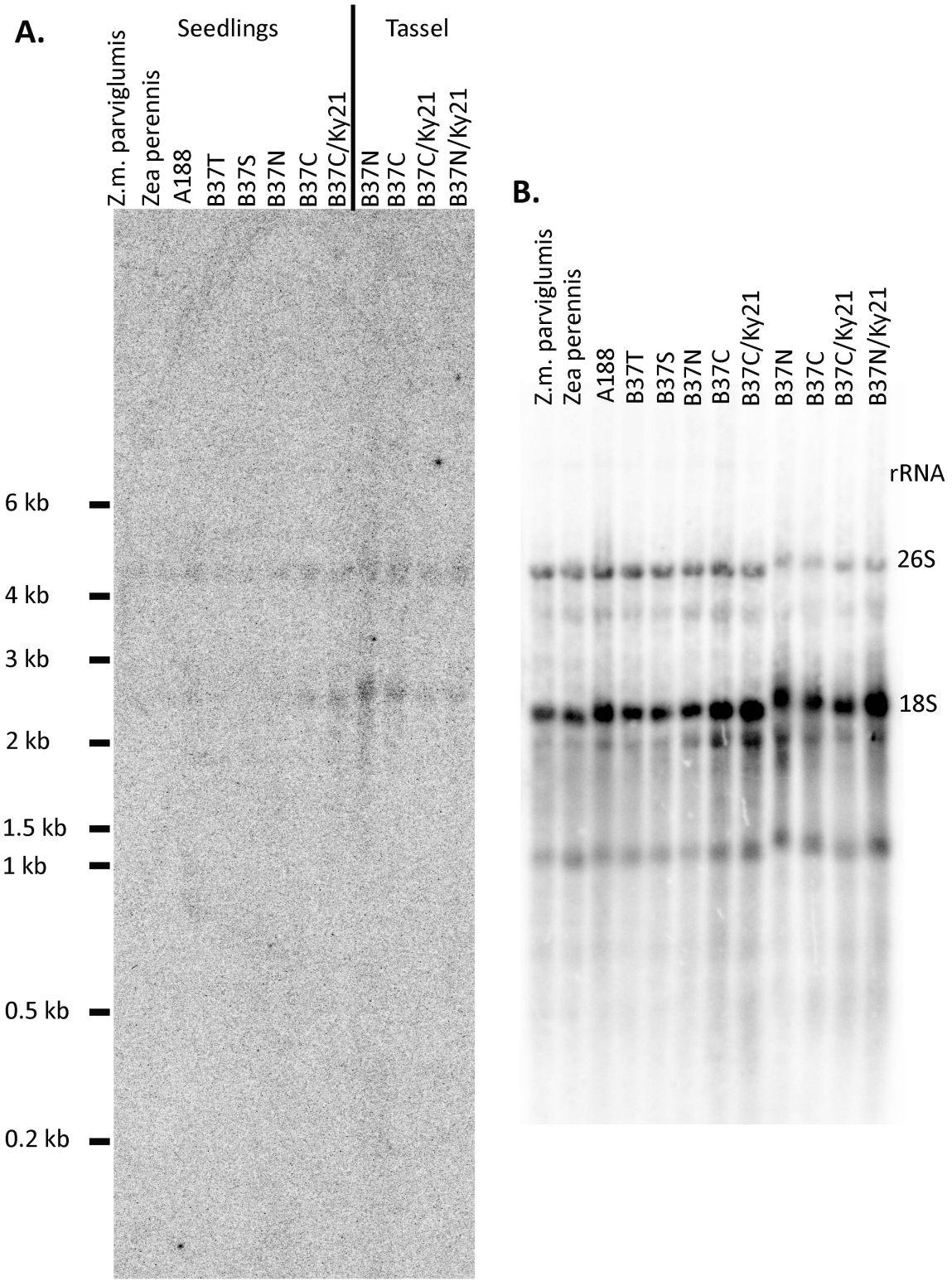


Figure 14: Transcripts hybridized with an *rps13* probe

A. Northern blot hybridized with a probe to the coding region of *rps13*. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 14: Transcripts hybridized with an *rps13* probe

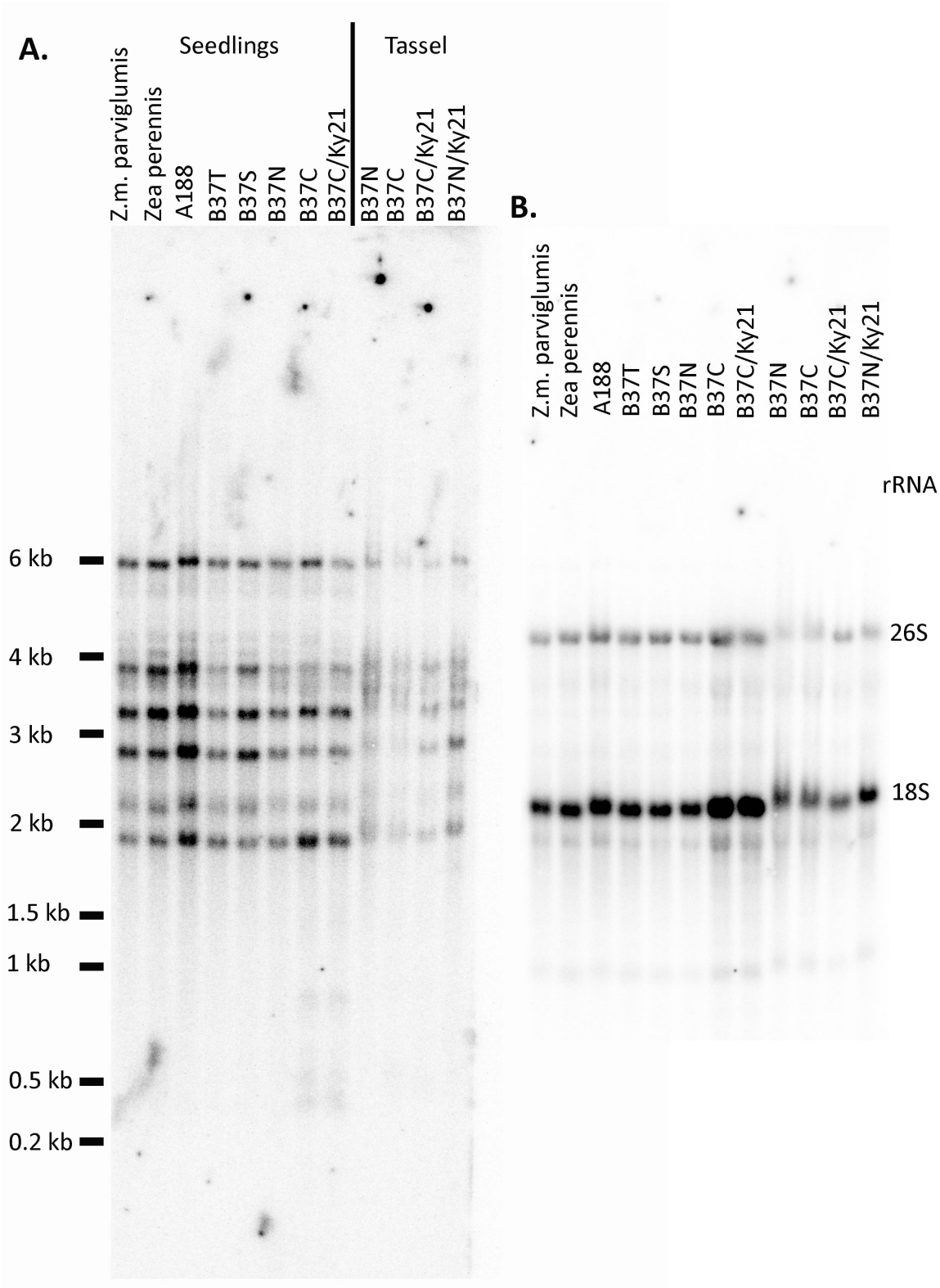


Figure 15: Transcripts hybridized with *nad7* probe

A. Northern blot hybridized with a probe to the coding region of *nad7*. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 15: Transcripts hybridized with *nad7* probe

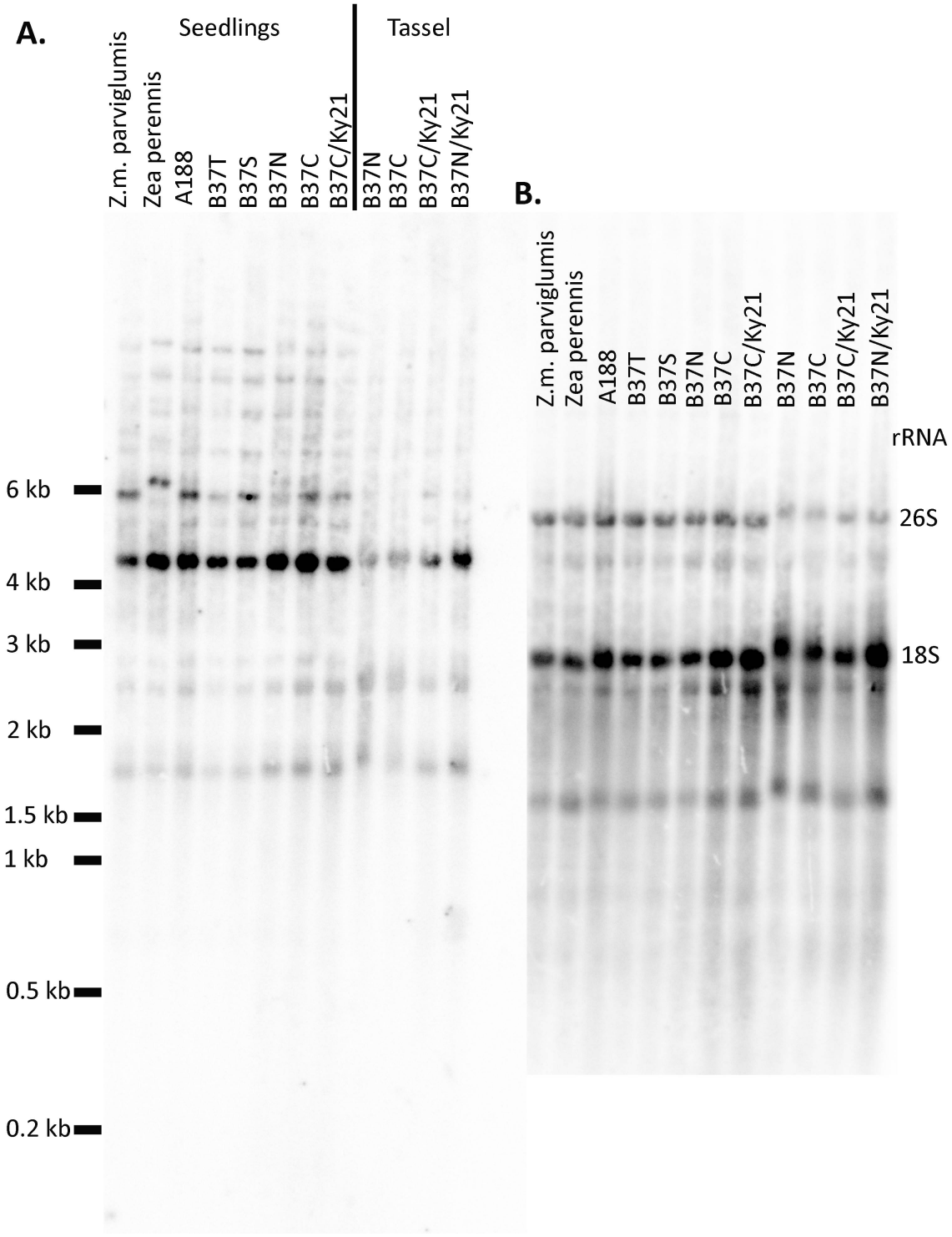


Figure 16: Transcripts hybridized with *cob* probe

A. Northern blot hybridized with a probe to the coding region of *cob*. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 16: Transcripts hybridized with *cob* probe

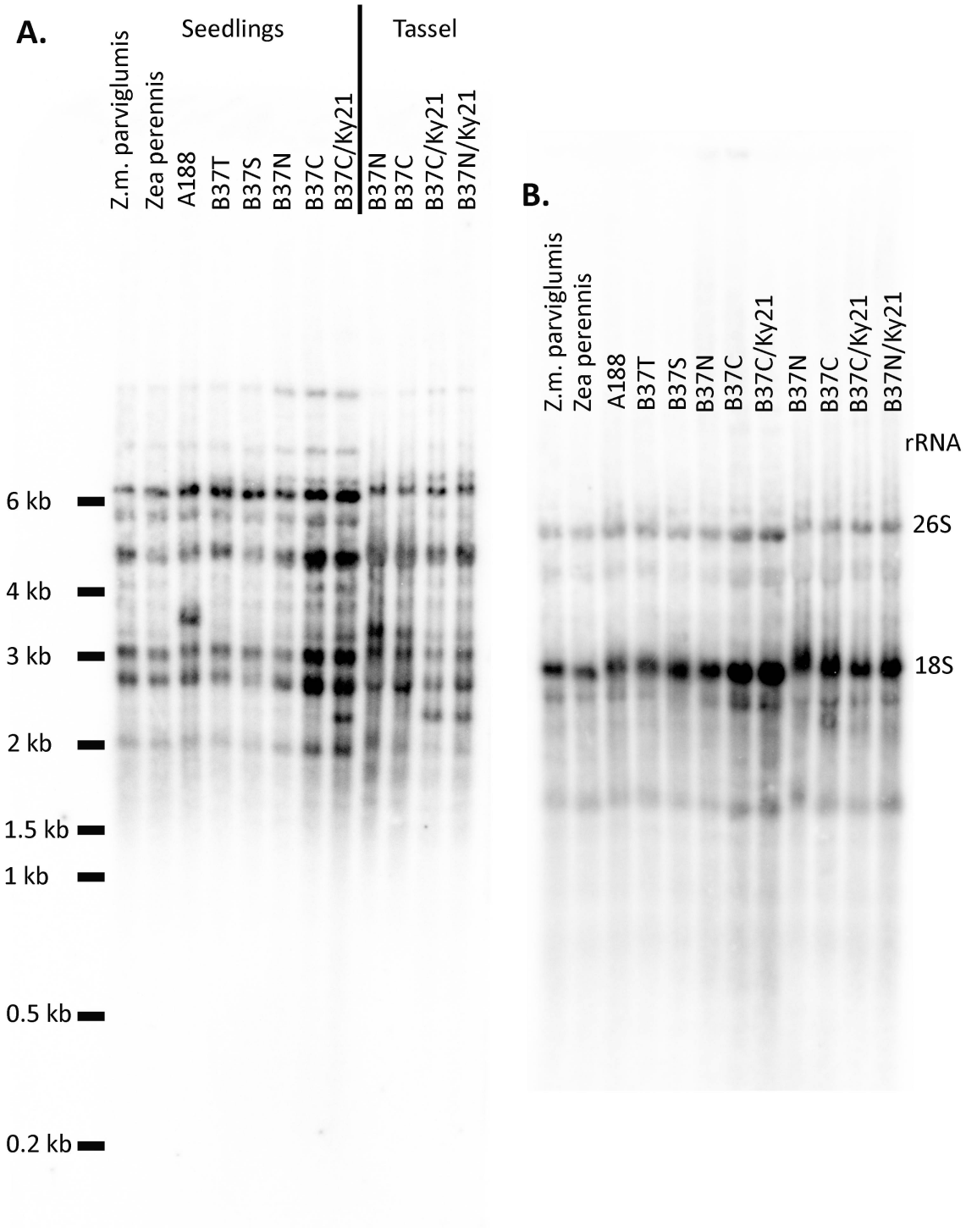


Figure 17: Transcripts hybridized with *cox2* probe

A. Diagram of the chimeric *cox2* gene in the CMS-C genome. The probe used for the northern is drawn above the conserved region of the gene; numbering is based upon the first base of the translated region. B. Northern blot hybridized with the *cox2* probe. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 17: Transcripts hybridized with *cox2* probe

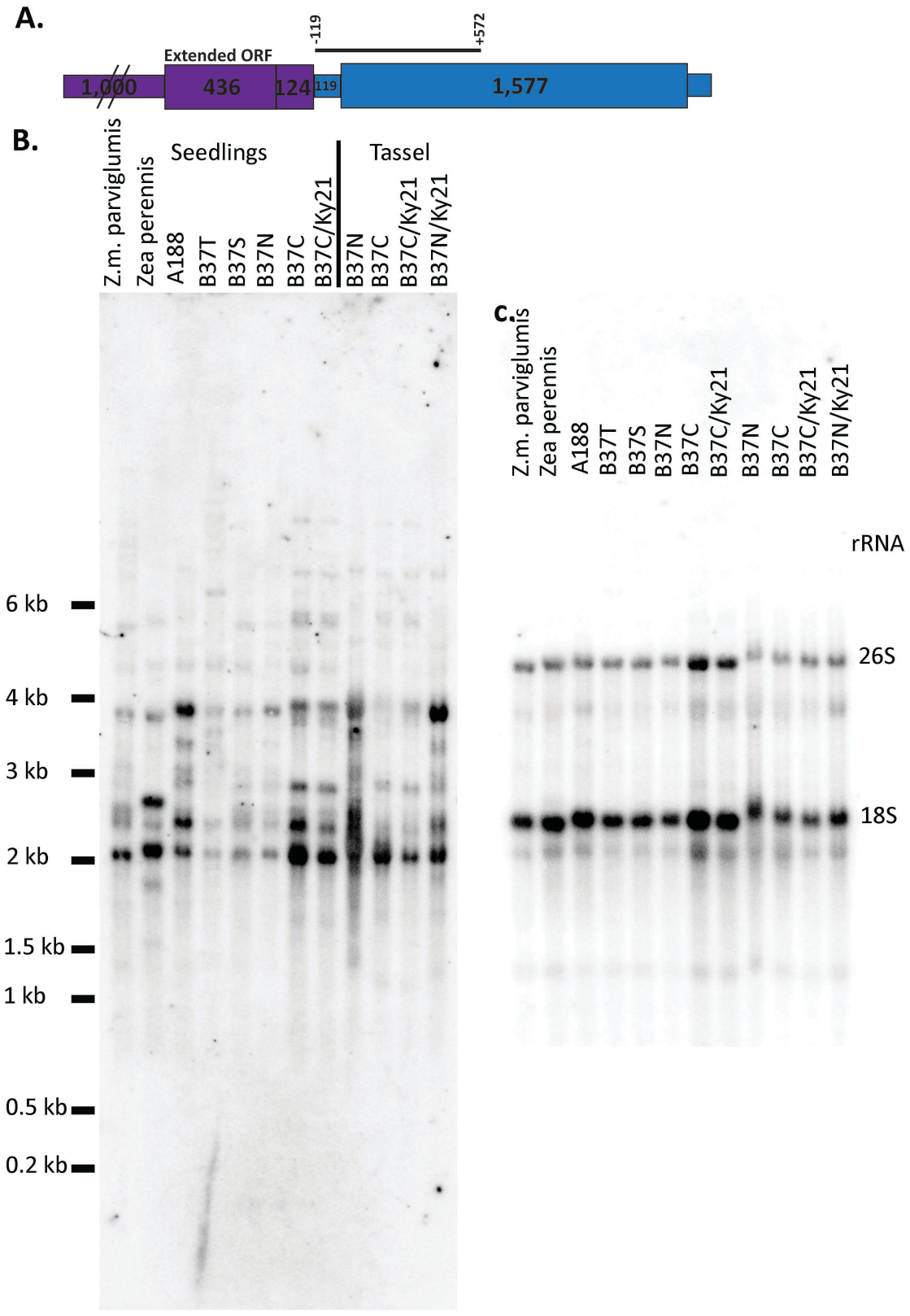


Figure 18: Transcripts hybridized with *atp6* probe

A. Diagram of the chimeric *atp6* gene in the CMS-C genome. The probe used for the northern is drawn above the conserved region of the gene; numbering is based upon the first base of the translated region. B. Northern blot hybridized with the *atp6* probe. The first eight lanes are seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes are tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 18: Transcripts hybridized with *atp6* probe

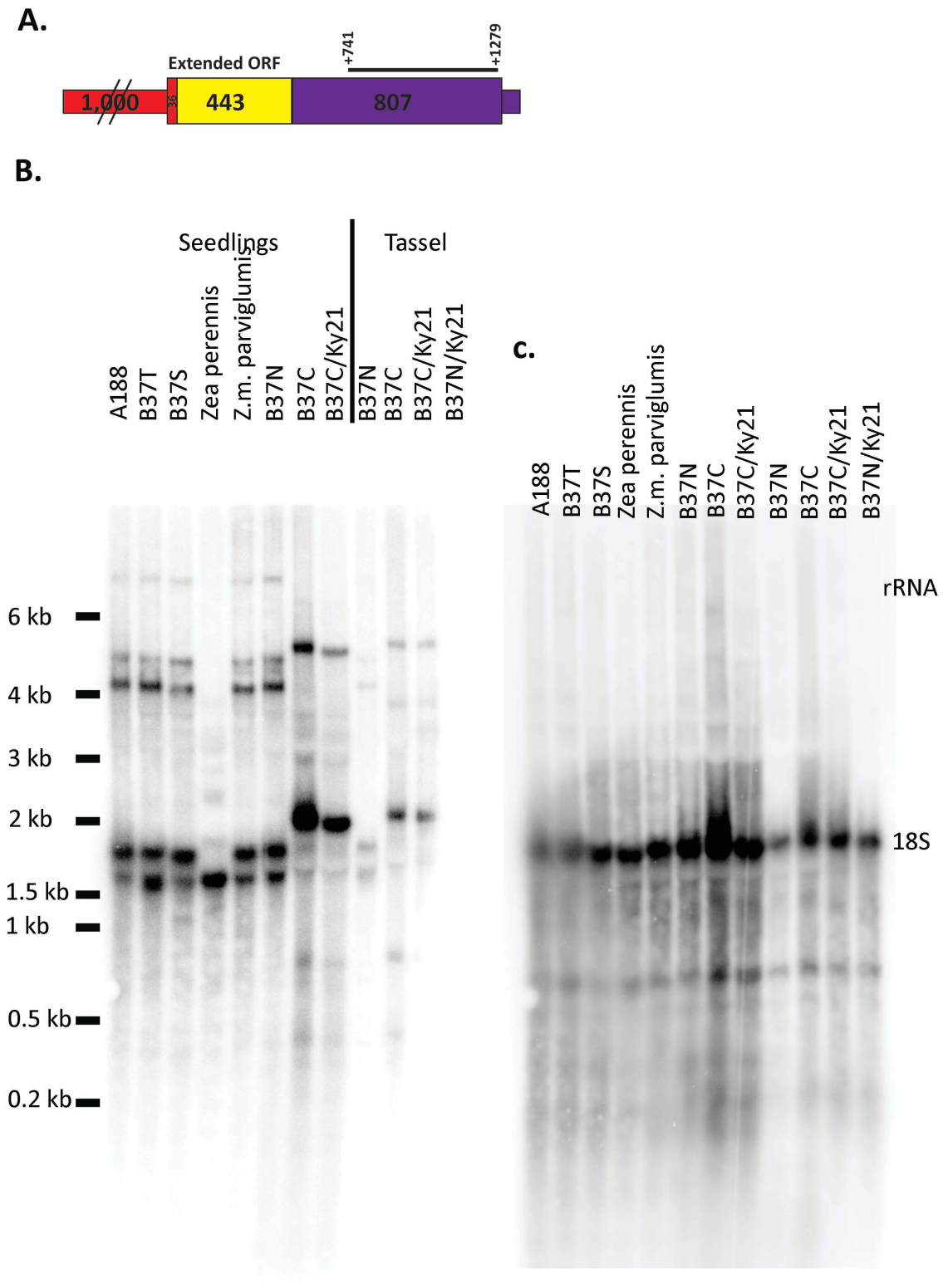


Figure 19: Transcripts hybridized with *atp9* probe

A. Diagram of the normal *atp-1* and chimeric *atp9-2* and genes in the CMS-C genome. The probe used for the hybridization is drawn above the conserved coding region of the gene; numbering is based upon the first base of the translated region. B. Northern blot hybridized with the *atp9* probe. The first eight lanes contain seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes contain tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 19: Transcripts hybridized with *atp9* probe

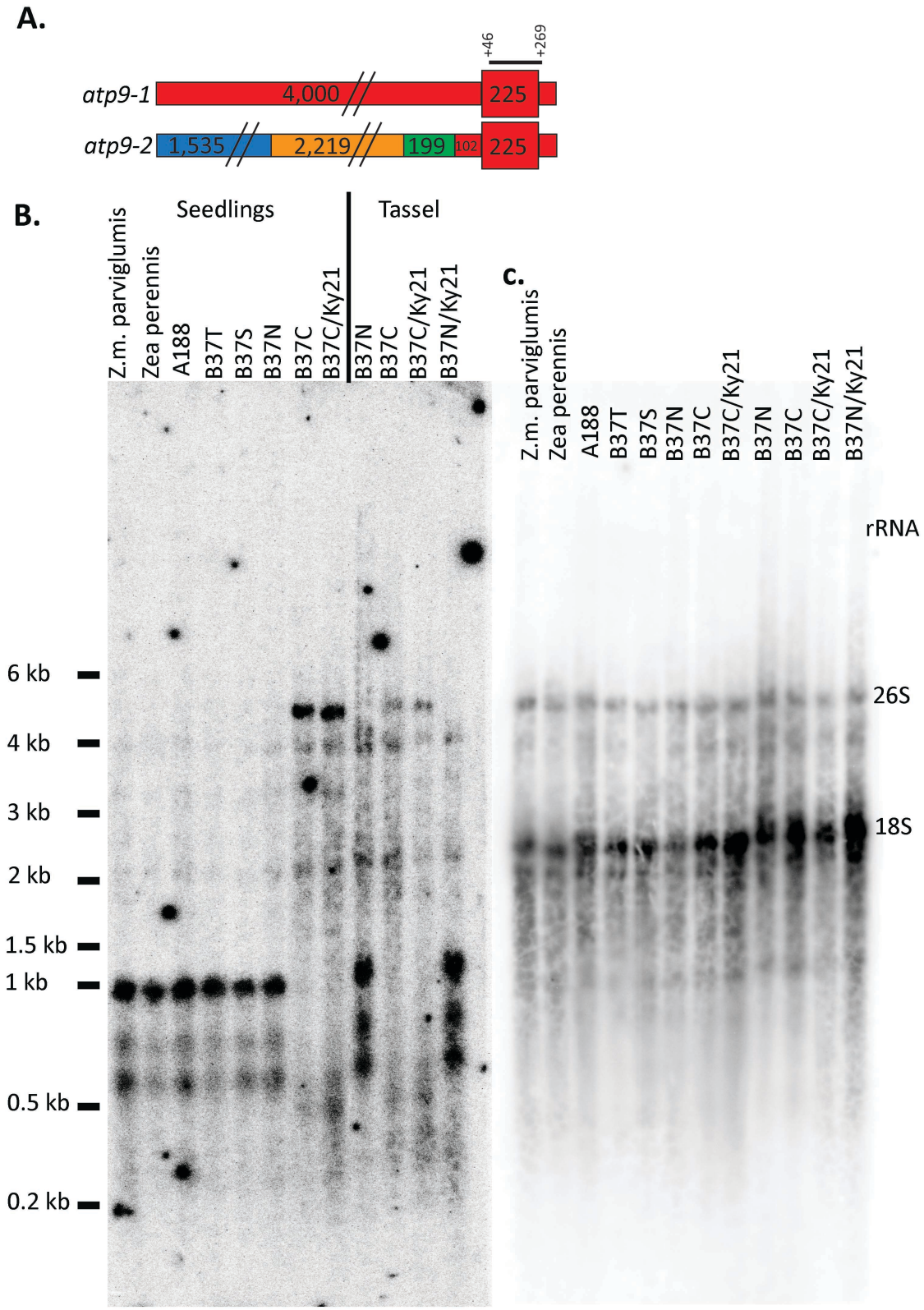
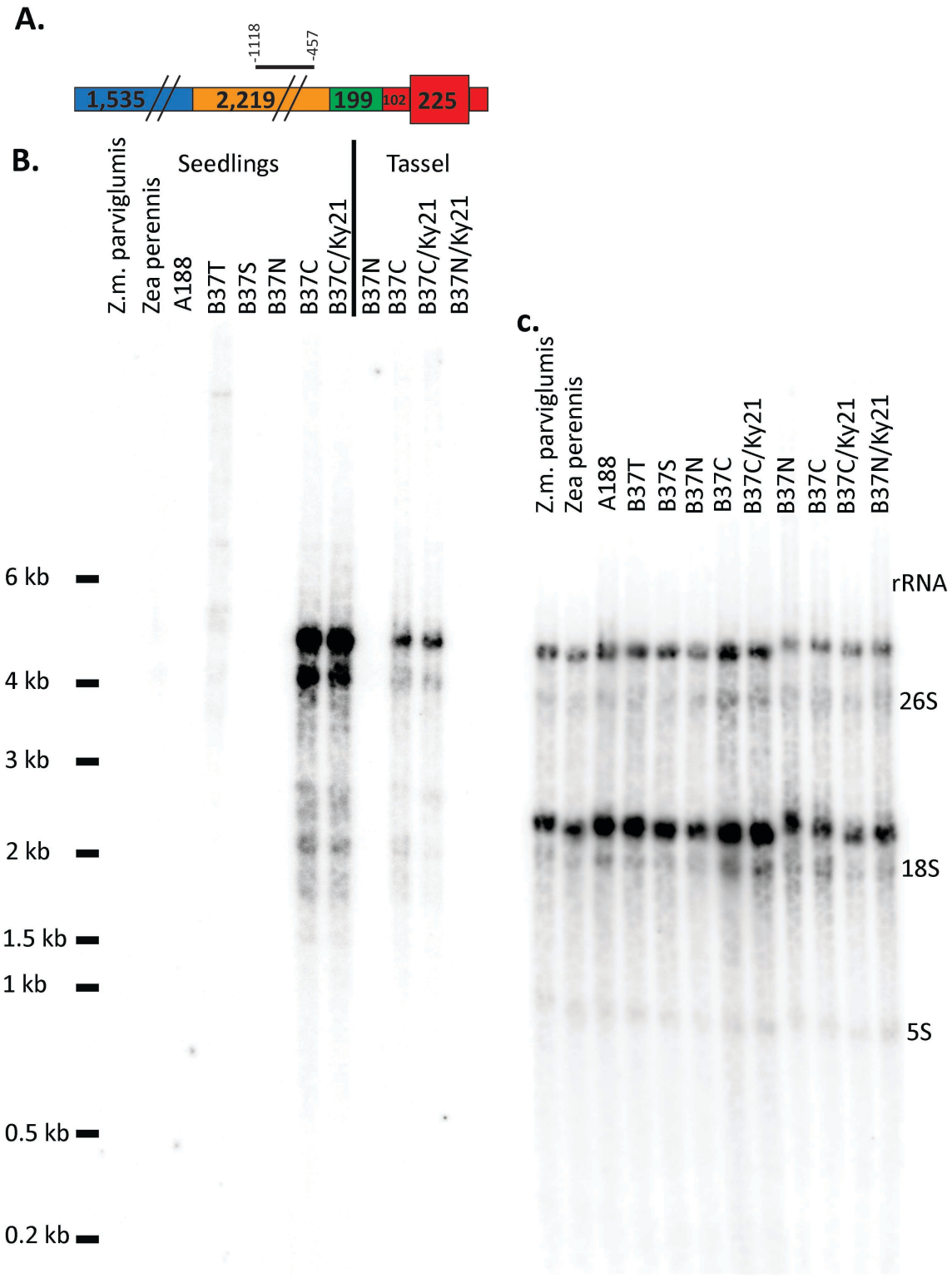


Figure 20: Transcripts hybridized with *atp9-2* probe

A. Diagram of the chimeric *atp9-2* gene in the CMS-C genome. Drawn above the gene is the region amplified for the *atp9-2* specific. Numbering is based upon the first base of the translated region. The probe lies within a region that only occurs upstream of *nad4L* in the CMS-T genome (B37T) and is not present in any other sequenced mitochondrial genome. B. Northern blot hybridized with the *atp9-2* probe. The first eight lanes contain seedling RNA from all of the sequenced *Zea* mitochondrial genomes along with the CMS-C genome in the restored background (B37C/Ky21). The last four lanes contain tassel mitochondrial RNA from the N and C cytotypes in a non-restored background (B37N and B37C respectively) and in the CMS-C restorer nuclear background (B37C/Ky21, and B37N/Ky21). The positions of the RNA ladder bands are indicated to the left of the figure. B. Loading Control: The blot from part A. hybridized with probes for the 26S and 18S rRNA.

Figure 20: Transcripts hybridized with *atp9-2* probe



3.4 Discussion

Based upon the previously identified CMS-associated genes, one would expect a novel transcript to be observed in CMS-C. This transcript would be the result of rearrangements that place an ORF near a promoter. Introduction of *Rf4* would then alter this novel transcript. However, RNA analysis of the candidate CMS causing ORFs in CMS-C did not identify any ORF matching these criteria.

3.4.1 Transcripts of Candidate ORFs

Nine of the eighteen probes used to detect potential ORF transcripts from different mitochondrial genomes detected stable RNA transcripts. However none of these transcripts were unique to C. Although these results indicate that the tested candidate ORFs are not involved in causing pollen abortion in the C genome further analysis was performed to determine if these stable RNA transcripts represented a previously unidentified gene, cross hybridization to other known genes, detection of known genes untranslated regions (UTRs), or non specific binding to rRNA transcripts.

The three chimeric ORF probes that hybridized to stable RNA transcripts did not have the expression pattern indicative of a CMS-causing gene. The orf246 probe weakly hybridized to three bands in all tissues that are the same size as nuclear and mitochondrial rRNA so are most likely not a transcript for orf246 (Fig. 5, Appendix 2). However, the probe did not show significant homology to any rRNA gene so this would

be non-specific binding and explain the weak hybridization. The fourth band that the probe detected is weak and variable among the different cytotypes; it did not change with the introduction of the restorer gene. It is most likely a large transcript from the *rps4* gene, which is located approximately 1 kb from the probe. This would place orf246 within the upstream sequence of *rps4* and allow for detection of *rps4* transcripts with a long 5'UTR.

The transcripts detected with the orf186a probe do not appear to be rRNA bands (Fig 6). The orf186a probe is only ~800 bases upstream from the start codon of the first two exons of *nad2* (*nad2*-ex1, 2). Thus it is probably detecting the *nad2* gene. The orf147b probe (Fig. 7) also appears to be detecting a gene, *cox1*, and not a unique transcript for orf147b because it is only 25 bases downstream from the *cox1* stop codon.

Twelve of the thirteen ORFs identified that were within 2 kb of rearrangement points between the C and N genomes were tested for expression. The one ORF not tested was orf117a which is identical in C and N and lies within the intron between the *nad2* exons 1 and 2. Because of its location, any probe made will detect *nad2* expression, and will not indicate if orf117a is expressed as an independent mRNA. Furthermore, because the ORF lies within the first intron of *nad2* in multiple fertile and male sterile mitochondrial genomes it is unlikely to be the cause of CMS.

Of the six ORFs at rearrangement points with detectable signals on northern blots, several could be binding non-specifically to rRNAs and not to stable transcripts from the ORF. The 4.5 and 2.9 kb bands for the orf247c/orf137a probe (Fig. 8) are migrating at

the same positions as the mitochondrial 26S and 18S rRNA respectively (Appendix 2), while the 2.7 kb band correlates with the nuclear 18s rRNA (Fig 8; Appendix 2). Another possibility for the signals with the orf247c/orf137a probe is that it is picking up transcripts for *ccmB* because the probe is ~2 kb downstream from the stop codon of *ccmB*.

The orf126a probe (Fig. 11) is hybridizes to a 2.5 kb band and is very similar to the mitochondrial 18s rRNA band (Fig. 11; Appendix 2). However, this also might be a transcript for *nad4L* because orf126a ends about 400 bases upstream of the translation start site for *nad4L*. The signals detected with the orf112b and orf105a probes (Figs. 12 & 13) are very weak and the sizes correlate with the mitochondrial 26S and 18S rRNA transcripts (Appendix 2); therefore, these probes are probably not detecting a specific transcript for orf112b or orf105a.

Orf193a is 1 kb upstream from the *atp1* start codon in all cytotypes which most likely explains the transcripts detected (Fig. 9). The *atp1* expression pattern has bands with published size estimates of 4.8 kb, 3.6 kb, and 2.4 kb (Feiler, 1986). Although the published sizes do not match the estimated sizes of the observed bands (5.8 kb, 4.2 kb, 3.5 kb, and 1.8 kb; Figure 9) the ladder used for estimating the sizes also measured the 26s rRNA band at 3.8 kb instead of the known 3.5 kb. If the rRNA was used for size estimation than the observed 5.8 kb would be closer to the 4.8 kb transcript of the published *atp1* gene. The observed 4.2 kb and 3.5 kb would be the double band seen by Feiler (1986) at 3.6 kb for *atp1*. The 2.4 kb band observed by Feiler (1986) might not be

present if the orf193a probe is to far upstream of *atp1* to detect the smaller transcript. It is interesting to note that all transcripts detected with the orf193a probe are decreased when crossed to Ky21. This may indicate that *Rf4* or another gene within the Ky21 genome alters *atp1* expression. Further analysis is needed to determine if there is a direct effect; this can be done by looking at *atp1* levels using a gene specific probe and determining if the phenotype is segregating with the *Rf4* allele.

The orf140d probe (Fig. 10) hybridized to several transcripts, all of which could be potentially explained due to the probe being ~200bp downstream of *nad5*-ex3 and ~1.2 kb downstream of *nad1*-ex5. Although according to Pereira de Souza *et al.* (1991) a *nad5*-ex3 probe did not hybridize to transcript that had not already been assembled into the mature *nad5* transcript that does not contain any 5' or 3' UTR of *nad5*-ex3. The 4.8 kb, 4 kb, and 2.2 kb bands could also be nuclear 28S, mitochondrial 26S, and nuclear 18S rRNA bands respectively (Fig 10; Appendix 2).

3.4.2 Analysis of Transcripts from Mitochondrial Genes

Genes representing each of the mitochondrial respiratory complexes were tested for expression differences in CMS-C. These included *nad7* of complex I, *cob* of complex III, *cox2* of complex IV, *atp6* and *atp9* of complex V, and the ribosomal protein *rps13.* Chimeric versions of *cox2*, *atp6* and *atp9* are found in the CMS-C mitochondrial genome (Fig. 3)

All of the non-chimeric genes tested, *nad7*, *cob* and *rps13,* showed variation in intensities and/or transcription pattern in at least one of the cytotypes tested. This indicates that *the positions of their* promoters and/or processing sites can vary. These variations are not associated with any phenotypic effects and so do not appear to affect mitochondrial function. The estimated sizes of *nad7* bands are different from the published results (major bands at 3.2 and 1.7 kb along with minor bands at 4 and 6 kb; Marienfeld and Newton, 1994). The differences in size can be explained by the differences in the markers used.

The *cob* probe gave several more bands then the major 2.9kb published by Feiler (1986), although faint bands can be seen in Feiler's image at 2.5 kb, 1.7 kb, and 1.6 kb. Assuming the Fermetas ladder is inaccurate, Feiler's 2.9 kb is most likely the 4.8 kb band based upon the rRNA and the remaining bands are not intense enough to be seen in the previous work.

When both the CMS-C (B37C) and B37N cytotypes were crossed to Ky21 (*Rf4* nuclear background) only the *cob* probe showed a change in the northern pattern (Fig. 16). In the Ky21 background a new band was seen at 2.3 kb with the *cob* probe. This was a surprising result because there are no alterations in the CMS-C mitochondrial genome involving *cob* and no indications that *cob* would be involved in CMS-C pollen abortion. The *cob* gene is probably not involved in CMS-C. The alteration of the transcript pattern may not be caused by *Rf4,* but rather by another mitochondrially targeted gene in the Ky21 background. There is a large class of RNA binding proteins that are targeted to

organelles that have been shown to alter RNA transcripts through transcriptional regulation or processing, and to be variable among lines (see chapter 7). Another possibility is that *Rf4* is affecting the *cob* transcripts coincidently with altering the expression of another gene that causes pollen abortion. Further experiments should be done to determine if the alteration of the *cob* transcription pattern is linked to the Ky21 background or to *Rf4*.

Cox2 is the most variable transcript tested, which correlates with previous work identifying *cox2* as being at a rearrangement point and having at least five functional promoters (Lupold et al., 1999a). Previous northern analysis determined that *cox2* in B37N produced 1.9, 2.3, 2.7, and 3.0 kb transcripts while CMS-C only produced 1.9 and 2.3 kb transcripts (Dewey, 1991). The results presented here have slightly different size estimations, with B37N producing 2.1, 2.4, 2.8, 3.5 and 3.8 kb transcripts and C having 2.1, 2.4, 2.8, and 3.8 kb transcripts. These differences are probably due to variations between the length of the gel ran and methods for size estimation. Dewey et al. (1991) estimated sizes based upon the migration of the 18S and 26S RNA bands while a RNA ladder was used in this research. This would mean that Dewey's 1.9 kb transcript correlates to the 2.1 and 2.4 kb transcripts, his 2.3 kb is the 2.8 kb transcript, and his 2.7 and 3.0 kb transcripts are the 3.5 and 3.8 transcripts presented here. The 3.8 kb transcript in CMS-C presented here is very weak so it might not have been detected in Dewey's analysis. It is interesting to note that even though *cox2* in CMS-C has the upstream sequence from *atp6* it still produces many of the same sized transcripts as B37N, which has the normal *cox2* upstream sequence.

Northern analysis using a probe to the coding region of *atp9* and *atp6* produced comparable results to those previously published (Dewey, 1988). Further analysis in order to distinguish between the normal *atp9-1* and chimeric *atp9-2* genes in CMS-C indicated that *atp9-2* is the major *atp9* transcript in the CMS-C genome. This is a surprising discovery since *atp9-1* appears to be completely normal in structure when compared to *atp9* in other *Zea mays* mitochondrial genomes. Further analysis is described in the next chapter to determine if *atp9-1* is being transcribed at all and why it is not exhibiting the normal transcriptional pattern and abundance. With the introduction of an abundant novel *atp9* transcript in C there is the possibility of altered translational efficiency due to the secondary structure of the large *atp9-2* transcript.

Chapter 4: REAL-TIME PCR OF CHIMERIC ATP SYNTHASE GENES

4.1 Introduction

With there being no novel transcribed ORF identified in CMS-C, further analysis was performed on the chimeric ATP synthase genes, *atp6* and *atp9-2*. Both of these genes use different promoters than the normal copy, resulting in altered RNA patterns. However, introduction of *Rf4* did not appear to alter the sizes of the transcripts for the gene. Further analysis was performed using real time PCR to determine if the change in promoters and/or introduction of the *Rf4* gene alters the stable RNA transcript levels of *atp6* or *atp9*. This would also help confirm whether *atp9-1* is transcribed at significant levels and whether *atp9-2* is the major transcript.

4.2 Materials and Methods

4.2.1 Total RNA Isolation

Tissue was harvested from 6-day-old etiolated seedling shoots, along with pre-meiotic and meiotic tassel material staged using aceto-carmine (see Section 3.2.1 for protocol). The four genotypes analyzed were B37N (NB type mitochondria), B37C (CMS-C type mitochondria), B37C/Ky21 (CMS-C mitochondria crossed to restorer line), B37N/Ky21 (NB type mitochondria crossed with restorer line). Three replicates of each tissue and genotype were prepared as biological replicates. The tissue was weighed,

surface sterilized in a 10% (v/v) sodium hypochlorite solution, and rinsed in cold DEPC treated deionized water. It was then placed in a 50 mL conical tube, frozen using liquid nitrogen and stored at -80° C until use.

Before performing the RNA prep, the sample was ground in liquid nitrogen using a mortar and pestle to a fine powder. 100 mg of frozen tissue was placed into a 1.5 mL tube with 750 µL of Invitrogen TRIzol reagent (cat no. 15596-026) and vortexed. After incubating the tissue in TRIzol at room temperature for 5 minutes 800 µL of chloroform was added and the sample was vortexed for 15 minutes. The samples were then incubated at room temperature for 10 minutes followed by centrifugation at max speed for 15 minutes.

The aqueous phase was transferred to a new tube. RNA was precipitated by adding 1.5 mL isopropyl alcohol and incubating at room temperature for 10 minutes. The RNA was pelleted by centrifuging at max speed for 10 minutes. The supernatant was discarded and the pellet washed with 3 mL of 75% ethanol. RNA was pelleted again by centrifuging at max speed for 5 minutes. The supernatant was discarded and the pellet air dried for 10 minutes. The pellets were resuspended in 20 µL DEPC treated water and stored at -80°C until use.

4.2.2 cDNA Synthesis

cDNA was synthesized for each RNA preparation using 10 μg of RNA in 22 μl of DEPC treated water. 2 μl of 10 mM dNTPs and 2 μL of 250 ng/μL random primers (Invitrogen) were added to the RNA sample and heated at 65° C for 5 minutes. The sample was placed back on ice and then 8 μL of 5x first-strand buffer, 2 μL of 0.1M DTT, 2 μL superscriptase III (Invitrogen) , and 2 μL RNaseOUT (Invitrogen) were added. The RNA sample was placed at 25° C for minutes followed by one hour incubation at 50° C for cDNA synthesis. Reactions were stopped by heating at 70° C for 15 minutes.

After cDNA synthesis RNA was removed with 2 μL RNase H incubated at 37° C for 20 minutes. cDNA was subsequently precipitated with 4 μL sodium acetate and 100 μL cold 95% ethanol incubated at -20° C for one hour. The cDNA was pelleted by centrifuging at max speed at 4° C for 20 minutes followed by washing with 70% ethanol. The pellet was air dried, resuspended in 20 μL TE, and stored at -20° C.

4.2.3 Real Time PCR

Real time PCR was performed on three tissues; six day old etiolated seedling shoots, pre-meiotic tassel, and meiotic tassel. For each tissue analyzed three separate reaction plates were used, one for each biological replicate. Each plate consisted of four technical replicates for each gene tested (*atp9*, *atp9-1*, *atp9-2*, *atp6*, and *rps13*) on four genotypes (B37N, B37C, B37C/Ky21, and B37N/Ky21). Primer information used for each gene can be seen in Appendix 1.

Master mixes were made for each genotype consisting of 20 ng cDNA, 12.5 µL SYBR green mix (Applied Biosystems cat# 4364346), and water up to 24 µL per reaction. Primer mixes were also prepared for each primer set consisting of 5 pmol/µL each of the forward and reverse primers; see appendix A for primers. To a microamp optical 96 well reaction plate (applied biosystems cat no.4306737), 1 µL of the appropriate primer mix was added to each well followed by 24 µL of the associated genotype master mix. The reactions were mixed by pippetting and the plate sealed with an optically transparent thermoseal (ISCBioExpress cat no. T-2417-8). The samples were briefly centrifuged for 2 minutes at 500xg at 4°C and then placed in an ABI 7300 real time PCR machine. The following program was used for RT-PCR: (1) - 50° C for 2 minutes, (2) - 95° C for 10 minutes, (3) – 95° C for 15 seconds, (4) – 60° C for 1 minute, repeat steps 3 and 4 40 times followed by a hold at 4° C.

4.2.4 Real Time PCR Analysis

Analysis was performed on etiolated seedling shoots, pre-meiotic tassel, and meiotic tassels independently. The ABI 7300 software was used to combine the three biological reps into one "relative quantification study". The baseline for each PCR product was manually determined in order to consist of the maximum number of cycles before amplification was observed. The threshold was also manually determined to ensure that it crossed the PCR reaction at the point of linear growth across all reactions. In order to equalize any uneven loading of the cDNA across genotypes, *rps13* was used as

an endogenous control for normalization. All relative cDNA levels were calculated using B37C/Ky21 (B37C-*Rf4*) as the reference sample. Two tailed T-tests were then performed comparing all samples for each gene and tissue tested with a 95% confidence.

4.2.5 RNA Editing Analysis

Primers flanking the *atp9-1* (appendix 1), *atp9-2* (appendix 1), and *atp6* (appendix 1) coding regions in the C genome were used to generate cDNA with the Invitrogen one-step RT-PCR kit (cat# 12574-018). 10 µM of sense and anti-sense primer for each gene were independently combined with 25 µL 2X reaction mix, 1 µL enzyme mix, 1 µg of total RNA from B37C tassels, and brought to 50 µL with DEPC treated water. The protocol was repeated with total RNA from B37N and B37C/Ky21 tassels. Reactions were placed in the thermocycler and amplified using the following program: 30 minutes at 55°C, 94°C for 2 minutes, followed by 40 cycles of 94°C for 15 seconds, 55°C for 30 seconds, and 68°C for 1 minute. A final 5 minute extension at 68°C was performed before a 4°C hold. After completion of cDNA synthesis, 5 µL of product was run on a 0.8% TAE agarose gel to confirm proper amplification. Samples were purified using a Qiagen PCR cleanup kit (cat# 28304). Samples were then submitted with the corresponding primers to the DNA Core Facility at the University of Missouri for sequencing. Sequencing results were compared to the published DNA sequence (B37N acc. # NC007982; CMS-C acc# DQ645536) and cDNA for the genes (*atp6* acc. #Z11843, *atp9* acc.#AF390542).

4.2.6 5' UTR Analysis

Five thousand bases of the 5' UTR from *atp9* in the B37N (acc. # NC007982), A (acc. # DQ490952), CMS-S (acc. # DQ645536), CMS-T (acc. # DQ490953), *Z. perennis* (acc. # DQ645538) and *Z. parviglumis* (acc. # DQ45539) mitochondrial genomes were compared with *atp9-1*, and *atp6* from the CMS-C mitochondrial genome using pairwise BLASTs from NCBI. Any differences that occurred between CMS-C (acc# DQ645536) and all other cytotypes were confirmed by sequencing. Primers representing *atp6* (Appendix 1), and *atp9-1* (Appendix 1) were used to amplify the 5' UTR from B37N and CMS-C DNA. The sequenced results were then compared to the published sequence to confirm.

4.3 Results

4.3.1 Real-Time PCR

Relative real-time PCR was performed in seedlings, and in pre-meiotic and meiotic tassels on the *atp6* (Fig. 21), total *atp9* (Fig. 22), and *atp9-1* (Fig. 23) transcripts in B37N, B37C, B37C/Ky21, and B37N/Ky21 backgrounds along with the *atp9-2* (Fig. 24) transcripts in B37C and B37C/Ky21 backgrounds. Two major differences in RNA levels were observed in the analysis. First the *atp6* gene in B37C had increased expression of ~10.9, ~1.8, and ~3.2 fold in seedling, pre-meiotic, and meiotic tassel when compared to

the B37N cytotype in the same nuclear background. The second major difference was the drastic decrease of *atp9-1* (normal version of *atp9*) transcripts in B37C ranging from a ~14 to ~23.7 fold decrease. Another interesting difference in RNA levels when comparing B37C to B37N was overall *atp9* levels in both pre-meiotic and meiotic tassel samples had a ~1.34 and ~1.38 fold increase in B37C respectively.

4.3.2 RNA Editing

Sequencing of the *atp6* cDNA identified 18 shared C to U RNA editing sites in the B37N, B37C, and B37C/Ky21 genomes when compared to their respective DNA sequences (Table 3A). The last edit site at 1243 introduces the published stop codon for the *atp6* gene. An additional edit site 20 bases from the start codon was observed for B37C and B37C/Ky21. This edit site is part of the extended ORF were homology to the *atp9* coding region exists. All of the edit sites identified confirm the previously identified C to U edit sites in *atp9*, *atp6*, and *atp6c* (Dewey, 1985).

The results of sequencing the *atp9-1* in B37N, B37C, and B37C/Ky21 and *atp9-2* in B37C and B37C/Ky21 identified seven C to U edit sites in the coding region, as compared to the published DNA sequence (Table 3B). These represent all of the known edit sites within the *atp9* coding region (Mulligan et al., 1988).

Figure 21: Relative *atp6* levels

A. Diagram of the chimeric *atp6* gene in the CMS-C mitochondrial genome. The amplification product from the real-time PCR is drawn above the conserved coding region of the gene. Numbering for the probe location is based upon the first base of the coding region. B. Relative levels of *atp6* RNA in seedling shoots. N represents B37N, C is B37C, C-*Rf4* is B37C/Ky21, and N-*Rf4* is B37N/Ky21. C. Relative levels of *atp6* RNA in pre-meiotic tassel material. D. Relative levels of *atp6* RNA in meiotic tassel material. All *atp6* RNA levels were compared relative to C-*Rf4* and were normalized based upon *rps13* levels. Relative levels of RNA for all samples were calculated using three biological replicates and four technical replicates for each biological sample.

Figure 21: Relative *atp6* levels

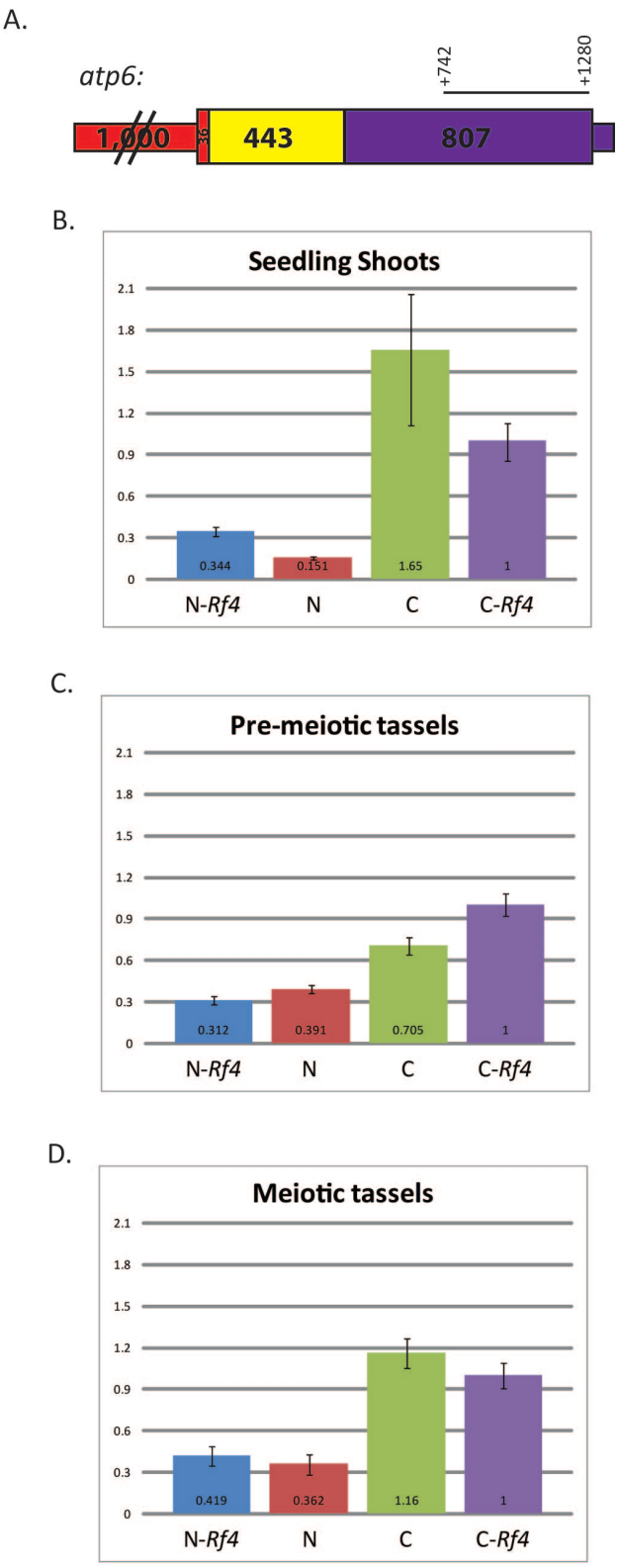


Figure 22: Relative *atp9* levels

A. Diagram of the normal *atp9-1* and chimeric *atp9-2* genes in the CMS-C mitochondrial genome. The amplification product from the real-time PCR is drawn above the conserved coding region of the gene. Numbering for the probe location is based upon the first base of the coding region. B. Relative levels of *atp9* RNA in seedling shoots. N represents B37N, C is B37C, C-*Rf4* is B37C/Ky21, and N-*Rf4* is B37N/Ky21. C. Relative levels of *atp9* RNA in pre-meiotic tassel material. D. Relative levels of *atp9* RNA in meiotic tassel material. All *atp9* levels were compared relative to C-*Rf4* and were normalized based upon *rps13* levels. Relative levels of RNA for all samples were calculated using three biological replicates and four technical replicates for each biological sample.

Figure 22: Relative *atp9* levels

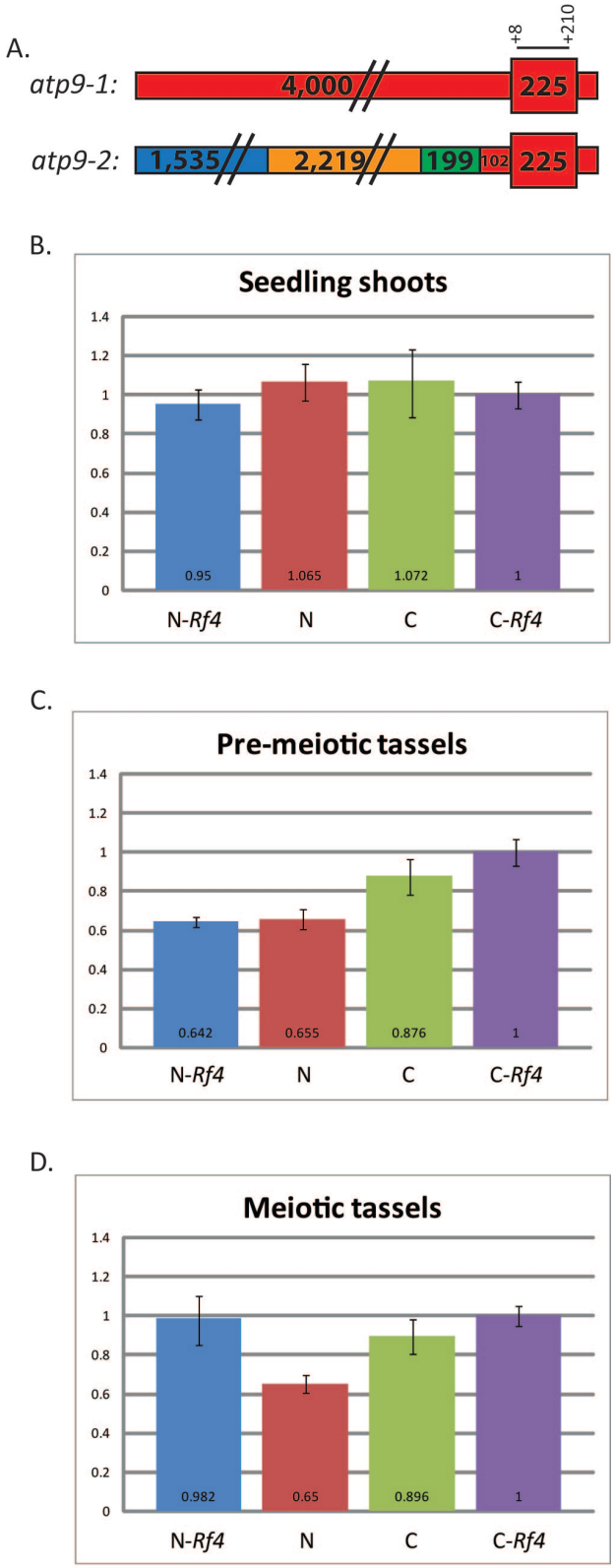


Figure 23: Relative *atp9-1* levels

A. Diagram of the normal *atp9-1* gene in the CMS-C mitochondrial genome. The amplification product from the real-time PCR is drawn above the conserved coding region of the gene. Numbering for the probe location is based upon the first base of the coding region. B. Relative levels of *atp9-1* RNA in seedling shoots. N represents B37N, C is B37C, C-*Rf4* is B37C/Ky21, and N-*Rf4* is B37N/Ky21. C. Relative levels of *atp9-1* RNA in pre-meiotic tassel material. D. Relative levels of *atp9-1* RNA in meiotic tassel material. All *atp9-1* levels were compared relative to C-*Rf4* and were normalized based upon *rps13* levels. Relative levels of RNA for all samples were calculated using three biological replicates and four technical replicates for each biological sample.

Figure 23: Relative *atp9-1* levels

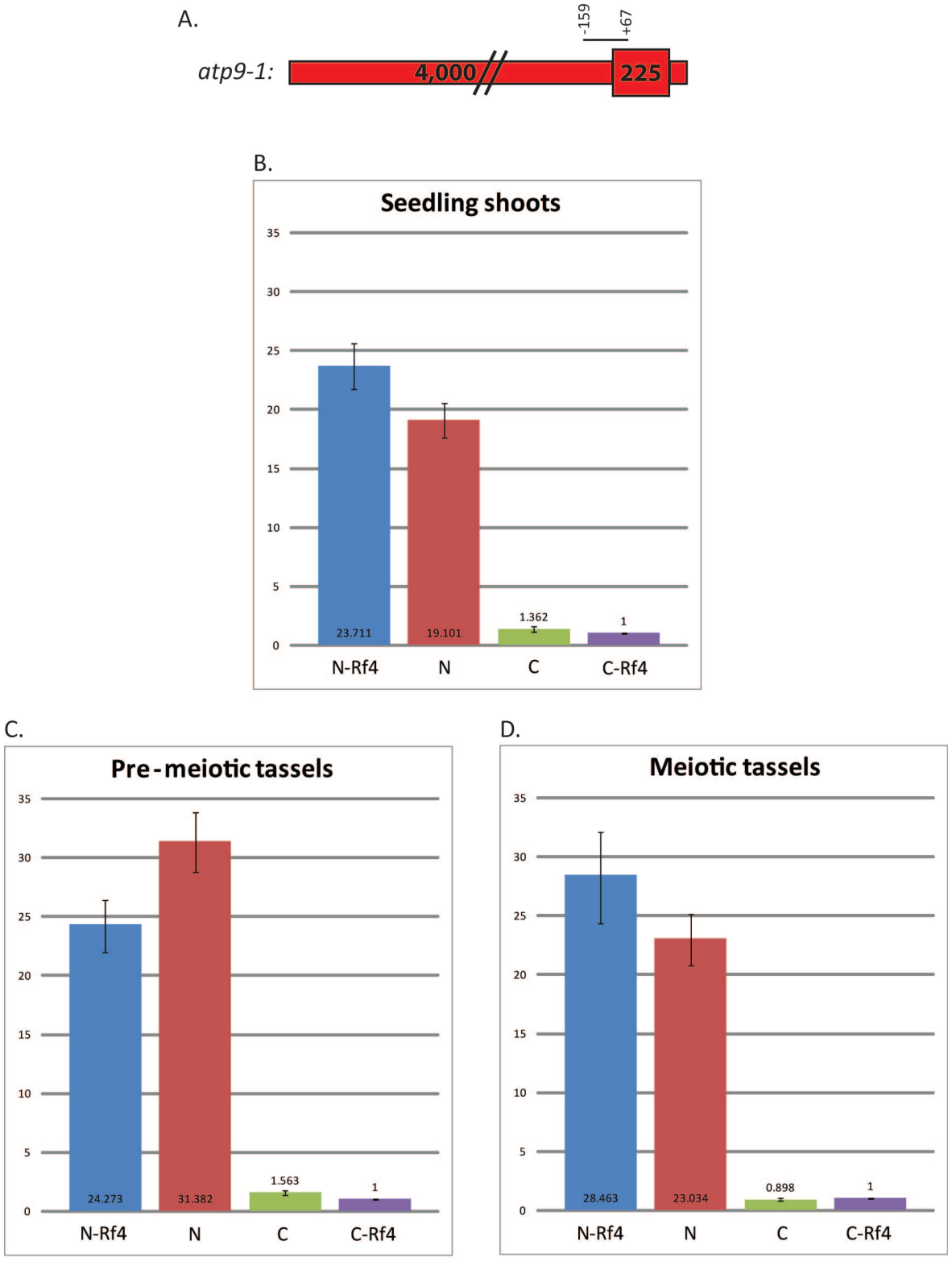
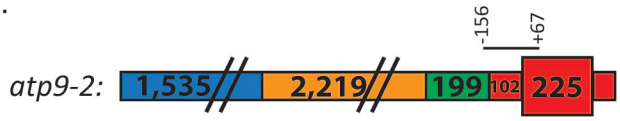


Figure 24: Relative *atp9-2* levels

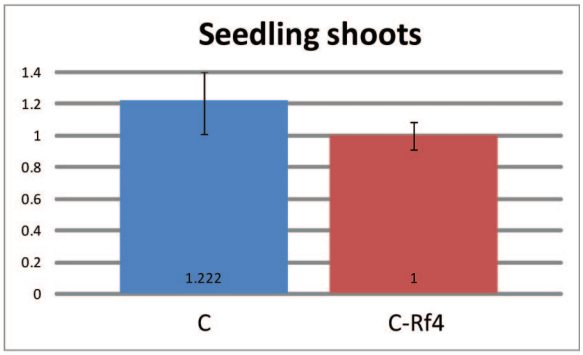
A. Diagram of the chimeric *atp9-2* gene present only in the CMS-C mitochondrial genome. The amplification product from the real-time PCR is drawn above the conserved coding region of the gene. Numbering for the probe location is based upon the first base of the coding region. B. Relative levels of *atp9-2* RNA in seedling shoots. N represents B37N, C is B37C, C-*Rf4* is B37C/Ky21, and N-*Rf4* is B37N/Ky21. C. Relative levels of *atp9-2* RNA in pre-meiotic tassel material. D. Relative levels of *atp9-2* RNA in meiotic tassel material. All *atp9-2* levels were compared relative to C-*Rf4* and were normalized based upon *rps13* levels. Relative levels of RNA for all samples were calculated using three biological replicates and four technical replicates for each biological sample.

Figure 24: Relative *atp9-2* levels

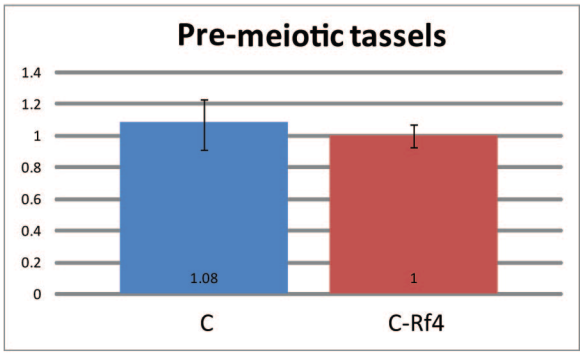
A.



B.



C.



D.

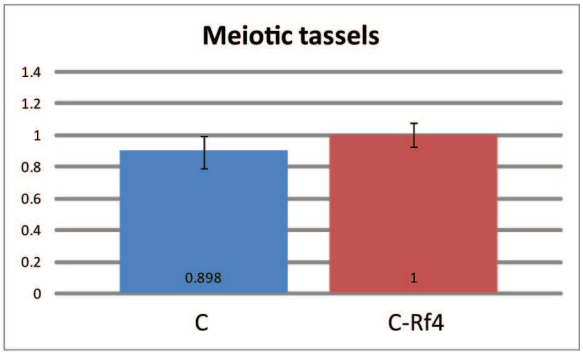


Table 3: RNA editing sites in *atp6* and *atp9* transcripts

A. *atp6* edit sites

Edit Site Location (bases from start codon in CMS-C)	572	588	641	692	761	779	787	794	795
DNA	C	C	C	C	C	C	C	C	C
B37N cDNA	T	T	T	T	T	T	T	T	T
B37C cDNA	T	T	T	T	T	T	T	T	T
B37C/Ky21 cDNA	T	T	T	T	T	T	T	T	T
Amino acid change	S→F	L→L	S→L	P→L	S→L	S→L	R→C	P→L	P→L
Edit Site Location (bases from start codon in CMS-C)	988	1010	1051	1181	1189	1196	1205	1232	1243
DNA	C	C	C	C	C	C	C	C	C
B37N cDNA	T	T	T	T	T	T	T	T	T
B37C cDNA	T	T	T	T	T	T	T	T	T
B37C/Ky21 cDNA	T	T	T	T	T	T	T	T	T
Amino acid change	H→Y	S→L	S→L	S→L	H→Y	S→F	S→L	T→I	Q→ Stop

B. *atp9* edit sites

Edit Site Location (Bases from start codon in CMS-C)	20	82	92	134	182	191	212
DNA	C	C	C	C	C	C	C
B37N cDNA	T	T	T	T	T	T	T
B37C *atp9-1* cDNA	T	T	T	T	T	T	T
B37C/Ky21 *atp9-1* cDNA	T	T	T	T	T	T	T
B37C *atp9-2* cDNA	T	T	T	T	T	T	T
B37C/Ky21 *atp9-2* cDNA	T	T	T	T	T	T	T
B37C *atp6* cDNA	T	N/A	N/A	N/A	N/A	N/A	N/A
B37C/Ky21 *atp6* cDNA	T	N/A	N/A	N/A	N/A	N/A	N/A
Amino acid change	S→L	P→L	S→L	S→L	S→L	L→F	S→L

4.3.3 *atp9* Upstream Sequence

Sequencing of the *atp9-1* cDNA sequence in B37C (Fig. 25) and B37C/Ky21 identified overlapping sequence reads starting at 66 bases before the translation initiation site, that did not occur for *atp9* in B37N (Fig. 25). The two sequences were read separately from each other in the chromatogram and were determined to be due to some cDNA sequences having two copies of the five base pair region and other sequences having three copies. This generated a five base pair shift in the sequencing read for some of the transcripts. The experiment was repeated with a second RNA isolation from new material along with sequencing of individual cDNA clones to confirm the results. Two clones were sequenced from B37C, one with two copies and another with three copies of the five base pair region (Fig. 25). Four clones were sequenced from B37C/Ky21 resulting in two sequences with two copies and two sequences with three copies of the five base pair region.

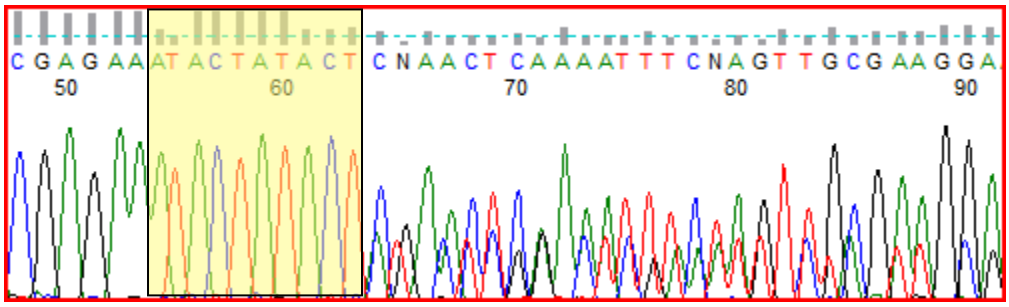
Comparisons between the upstream sequences of CMS-C *atp9-1* and *atp6* (contains the 5'UTR from *atp9*) from the published genome identified no differences between the two sequences for at least 5 kb upstream of the coding sequences. When both of these sequences were compared to the 5kb upstream sequence of *atp9* in other sequenced mitochondrial genomes three mutations were constantly observed (Table 4). CMS-C has a single nucleotide polymorphism identified at 221 bases upstream of the translation initiation site where the C genome sequences contain a C and the other genomes have a T. There was also a five base insertion (CTAAA) at -762 in the C genome that does not occur within the other genomes. The C genome also shared a mutation with *Z.*

Figure 25: Chromatograms of *atp9-1* cDNA sequencing results

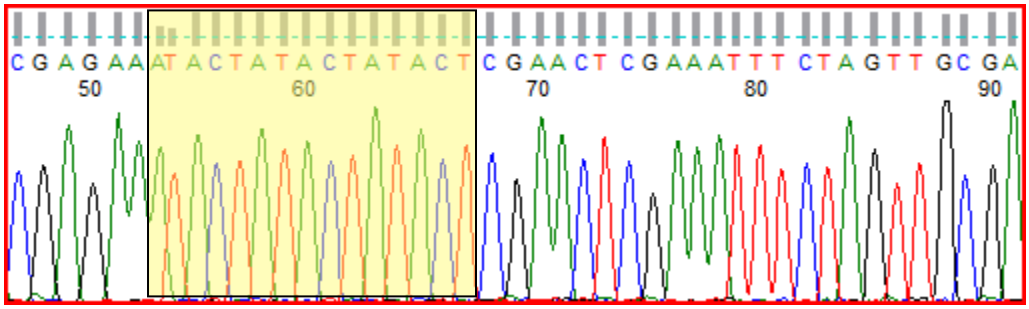
Chromatogram results from sequencing the *atp9-1* gene in CMS-C and *atp9* in B37N. Direct sequencing of *atp9-1* in CMS-C resulted in two copies of the five bp repeat ATACT (highlighted) followed by overlay of two sequences, one with a third copy of ATACT. Individual clones from the B37C *atp9-1* cDNA show the two sequences present, one with two copies of the ATACT repeat (highlighted) and the other with three copies (highlighted). The *atp9* transcripts in B37N always have three copies of the ATACT sequence (highlighted).

Figure 25: Chromatograms of *atp9-1* cDNA sequencing results

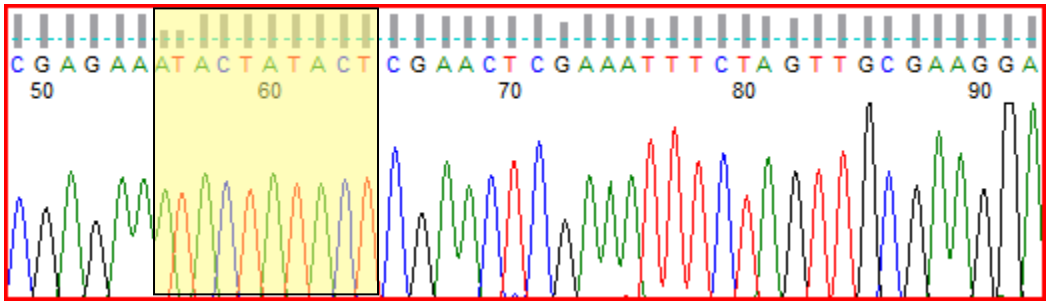
B37C: *atp9-1* cDNA



B37N: *atp9* cDNA



B37C: *atp9-1* cDNA clone1



B37C: *atp9-1* cDNA clone2

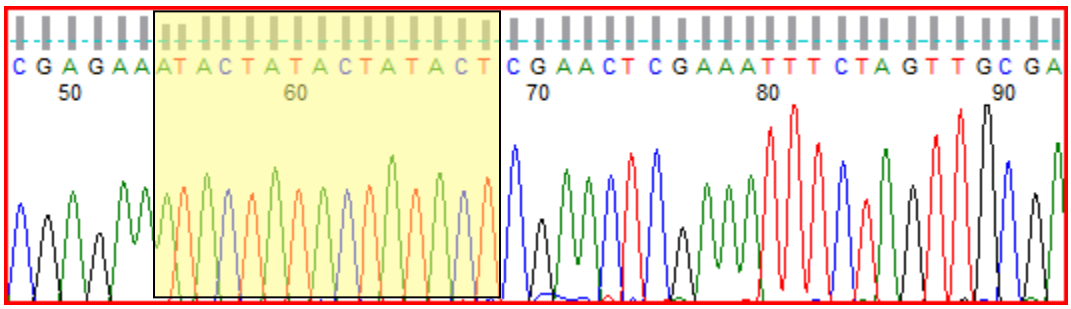


Table 4: Comparison of *atp9* upstream sequence in *Zea* genomes

The identical upstream sequence of CMS-C *atp9-1* and *atp6* were compared to the *atp9* upstream sequences in the other sequenced *Zea* mitochondrial genomes. Any sequence variations that occur as compared to CMS-C are highlighted. One Indel at -762 and a SNP at -221 bases upstream of the start codon occurs in every comparison of the CMS-C *atp9-1* to the other sequenced *Zea* mitochondrial genomes.

Bases upstream of start codon	CMS-C *atp9-1* and *atp6*	B37N *atp9*	CMS-S *atp9*	CMS-T *atp9*	A619 *atp9*	Z. perrenis *atp9*	Z. parviglumis *atp9*
5768	-----	CGGCA	-----	-----	-----	-----	-----
5735	T	T	T	T	G	T	T
3335	ACTTT	ACTTT	ACTTT	ACTTT	ACTTT	-----	ACTTT
3306	G	T	G	G	G	G	G
2853	T	T	A	T	T	T	T
2625	-----	-----	-----	-----	-----	TACTA	-----
2560	----	----	----	----	----	GATT	----
2373	-----	-----	-----	-----	-----	-----	TTGCG
2357	AAAAAACCT	AAAAAACCT	AAAAAACCT	GAAAAACCT	AAAAAACCT	-----TCAA	AAAAAACCT
1956	C	C	C	C	C	A	C
1951	CTTCGATTCT	CTTCGATTCT	CTTCGATTCT	CTTCGATTCT	CTTCGATTCT	----------	CTTCGATTCT
1909	-----	-----	-----	-----	-----	TAGTA	-----
1865	----	----	----	----	----	CAGC	----
1414	T	T	T	T	T	G	T
1401	G	G	T	G	G	G	G
1199	A	A	C	A	A	A	A
1167	GA	GA	GA	GA	GA	TT	GA
762[1]	CTAAA	-----	-----	-----	-----	-----	-----
519	------	------	TATCAG	TATCAG	------	------	------
315	GATAT	GATAT	GATAT	GATATGATAT	GATAT	-----	GATAT
221[1]	C	T	T	T	T	T	T
76	-----*	ATACT	ATACT	ATACT	ATACT	-----	ATACT
51	T*	T	T	T	T	G	T

* Sequences also present upstream of CMS-C *atp9-2* gene.

[1]Variation in sequence that occurs across all sequenced *Zea* mitochondrial genomes

perennis at -76 where both contain only two copies of a five base sequence (ATACT) and the other genomes have three copies. All other sequence differences identified in C were only seen when compared to one or two other mitochondrial genomes.

4.4 Discussion

4.4.1 Real-Time PCR

The *atp6* gene in CMS-C had a minimum increase of 1.8 fold in all tissues tested of CMS-C when compared to the B37N cytotype in the same nuclear background. This increase is most likely caused by the *atp6* gene in CMS-C using the *atp9* promoter instead of the standard *atp6* upstream sequence found in other cytotypes. This indicates that the *atp9* promoter is a stronger promoter than the *atp6* promoter. Introduction of the restored background does not have a consistent effect on the *atp6* transcript even between the tassel samples, meaning that either *atp6* is not involved in CMS-C pollen abortion or restoration is occurring after transcription and RNA processing.

Overall, the *atp9* levels do not drastically change when comparing B37C cytotype to B37N, although there is a slight increase of *atp9* levels in B37C within the tassel samples. This slight increase might be an indirect effect of pollen abortion or be caused by the presence of two *atp9* genes one of which is using a *cox2* promoter. Either way the slight change in *atp9* expression most likely has no significant biological affect. This is supported by the fact that introduction of the restorer background does not change the *atp9* levels in B37C.

The drastic decrease of *atp9-1* levels in B37C as compared to N is an interesting find that supports the previous northern data. This means that the primary transcript in CMS-C is not generated from the B37N like *atp9-1* gene but from the 5' chimeric C specific *atp9-2*. Introduction of the restorer background did not restore *atp9-1* levels back to normal in CMS-C or dramatically alter *atp9-2* levels. Thus, the *Rf4* gene may affect the translation of *atp9-2*, rather than altering RNA transcription or processing.

The introduction of the restorer background to both the CMS-C and B37N cytotypes generated a number of small tissue and stage specific changes in RNA levels of the genes tested. These alterations were not consistent between CMS-C and B37N in the restored background. Since all of these changes are minor and do not correlate with CMS-C and pollen abortion they are unlikely to have anything to do restoration of fertility. These variations might be due to heterotic effects on the mitochondria or due to specific alleles in the Ky21 background.

4.4.2 RNA Editing

RNA editing of the *atp6*, *atp9-1*, and *atp9-2* tassel transcripts appears to be completely normal in the CMS-C genome and no changes occur with the introduction of the restorer of fertility allele (Ky21; Tables 3 and 4). The CMS-C *atp6* gene has a unique extended ORF that contains the first 36 bases from the *atp9* coding region (Dewey et al., 1991). One edit site was identified within this region that also occurs in the *atp9* gene. RNA editing appears to be normal within the tassels of the CMS-C genome with no effect from *Rf4.*

4.4.3 RNA Indel variation

Interestingly, two variants of *atp9-1* transcripts were identified in both B37C and B37C/Ky21 tassels. One variant contains two copies of a five base pair tandem duplication 76 bp upstream of the coding sequence, while the other has three copies of the five base pair sequence (Fig. 25). The DNA sequence of CMS-C contains two copies of the 5 bp duplication, so the three copies in some transcripts must be due to slippage during transcription. The *atp9-1* transcripts in B37N contain only three copies of the 5 bp duplication, which is the copy number within the DNA (Fig 25). These results were repeated with a different tassel RNA samples to ensure that it was not due to cDNA, PCR, or sequencing error. The existence of two *atp9-1* RNA species in CSM-C is unlikely to have any effect on pollen abortion since this is not the major *atp9* transcript in CMS-C and introduction of *Rf4* does not affect it. The major *atp9* transcript in CMS-C, *atp9-2*, and *atp6* have sequence homology to the *atp9-1* gene but only produce transcripts containing two copies of the 5 bp sequence.

4.4.4 *atp9* 5' UTR Alterations

Using primer extension Mulligan et al. (1988) identified some of the 5' ends of the *atp9* transcripts at -218, -251, -272, -288, -419, -496, -600, -636, -1240, and -1540 in B37N cytoplasm. Mulligan et al. (1988) also determined that the 3' end of *atp9* is at ~+330 (coding region of 225 bp). If all of the mapped 5' ends were prevalent on northern gels there would be bands at around 550 bp, 580 bp, 600 bp, 615 bp, 750 bp, 830 bp, 930 bp, 960 bp, 1570 bp, and 1870 bp. This is vastly more complex than the

observed northern pattern in B37N of 600 bp, 800 bp, and 1000 bp. This means that the majority of the mapped 5' ends do not correlate with the most prevalent *atp9* transcripts. Based upon the major *atp9* transcripts, the estimated 5' ends of *atp9* would be at -270bp, -470 bp and -670 bp, which most closely correlate to the 5' ends mapped by Mulligan (1988) at -272 bp, -496 bp, and -636 bp.

In order to determine if any of the SNPs or Indels could cause the decrease in *atp9-1* transcription, they were compared to the mapped and estimated 5' ends. An insertion at -762 in CMS-C *atp9-1* and a SNP at -221 in CMS-C *atp9-1* are both conserved when compared to all sequenced *Zea* mitochondrial genomes. The insertion at -762 is too far upstream of the probable 5' end of the 1 kb transcript to be at a processing or transcriptional start site. However, the insertion at -762 could be interfering with a transcription binding site and interfere with the production of the 1000 bp transcript. Interesting the 5 bp insertion occurs immediately after the monocot conserved promoter sequence YYTA (Fey and Marechal-Drouard, 1999). The SNP at -221 is ~50 bp downstream of the closest mapped and northern-estimated 5' end. This SNP might interfere with a binding site for transcription or RNA processing and interfere with production of the 600 bp transcript. Although there is no SNP or Indel that accounts for the lack of an 800bp transcript in C *atp9-1,* it may be missing due to the fact that it is a processed form of the 1000 bp transcript. If the larger 1000 bp transcript is not present, there would be no processing to produce the 800bp band. This could also explain the lack of a 600 bp transcript if the SNP at -221 is not involved.

Chapter 5: 1D AND 2D ANALYSIS OF MITOCHONDRIAL PROTEINS

5.1 Introduction

Bioinformatics and RNA analysis did not identify a novel gene being transcribed in CMS-C. Although this reduces the potential for a toxic protein being synthesized from a novel transcript in CMS-C it does not rule it out. In order to determine if a novel protein is present in CMS-C, 2D differential in gel electrophoresis (DIGE) was used CMS-C and normal (B37N) mitochondrial proteins were labeled with different fluorescent dyes. The samples can then be combined and be separated by isoelectric focusing (IEF) followed by SDS-PAGE. This allows for labeling and comparing some of the mitochondrial proteins.

There were two chimeric genes identified, *atp6* and *atp9-2,* that have potential to cause pollen abortion. The chimeric *atp6* uses the *atp9* promoter and has a unique extended ORF. A chimeric ATP6 protein could be generated from this transcript if the extended ORF is not proteolytically cleaved off. Another possibility is that the extended orf gets cleaved off but does not get degraded; this would leave a chimeric protein with the first 12 amino acids of ATP9. The *atp9-2* transcript could also produce a chimeric ATP9 protein by using a different start codon within its chimeric 5'UTR. Alternatively, the *atp9-2* transcript could be folding in such a way as to prevent efficient translation. This would reduce the ATP9 levels present. These hypotheses can be tested by the 2D DIGE experiments and with the use of antibodies raised to the N terminus of the ATP6 and ATP9 proteins.

5.2 Materials and Methods

5.2.1 Material

Seedling material was harvested from six day old etiolated seedling shoots. Seeds were soaked in water containing Captan 50 overnight and then rolled in germination paper (Anchor Paper cat # SD3815L). The seedlings were placed at 28° C for five days before harvesting the shoots for mitochondrial isolation.

Plants for tassel material were grown in the field during the summers of 2006 and 2007. Pre-emergent tassels were harvested and staged according to the oldest male florets on the tassel. Pre-meiotic material had no anthers that had entered into meiosis. Meiotic tassels represent all stages of development up to the tetrad stage at the end of meiosis. All tassel preps contained pooled samples to represent pre-meiotic or meiotic tassel material.

5.2.2 Mitochondrial protein isolation

The protocol used to isolate mitochondria was a modification from the protocol outlined by (Stern and Newton, 1985). All solutions used had been autoclaved and all isolation steps were conducted at 4° C.

Tissue was harvested from 6-day-old etiolated seedling shoots of B37N (NB type mitochondria), B37C (CMS-C type mitochondria), and B37C/Ky21 (CMS-C mitochondria

crossed to restorer line), and B37N/Ky21 (CMS-N mitochondria crossed to restorer line), along with pre-emergent tassels from B37N, B37C, B37C/Ky21, and B37N/Ky21.

The pre-emergent tassel material was determined to have entered into meiosis by staging the older anthers on the tassel using aceto-carmine. The anthers were removed from the male floret and fixed in a 1.5 mL eppendorf tube containing 3 parts 95% ethanol to 1 part glacial acetic acid. The anthers were allowed to soak for a minimum of 10 minutes before being placed on a microscope slide. After one drop of aceto-carmine was added on top of the anthers they were smashed using rusty probes. A cover slip was placed over the smashed anthers followed by gently heating the slide over an ethanol lamp. The anthers were then examined under a 100x magnification to determine the stage of development.

The tissue was weighed; surface sterilized in a 10% (v/v) sodium hypochlorite solution, and rinsed in cold DEPC-treated deionized water. The tissue was then placed in a cooled blender jar containing cold homogenization buffer (a ratio of 4 mL buffer per 1 gm tissue). The homogenization buffer contained 0.4 M mannitol, 25 mM MOPS pH 7.8, 1 mM EGTA, 4 mM cysteine and 0.15% bovine serum albumin was added just prior to use. Pulses of two seconds each at low followed by high speed was used to homogenize the tissue in a Waring blender. The homogenate was filtered through a sterile funnel, lined with four layers of cheese cloth on top of one layer of Miracloth, into autoclaved 250 mL centrifuge bottles on ice. Nuclei, plastids, starch, and cellular debris were removed from the homogenate through two differential spins in an RC-5B

Sorvall centrifuge at 4° C, with a pre-cooled GSA rotor. First the filtrate was centrifuged at 2500 rpm (1000 x g) for 5 minutes. The supernatant was carefully decanted into sterile, cooled 250 mL bottles. The supernatant was then centrifuged again at 4000 rpm (2000 x g) for 10 minutes to remove additional plastids. The supernatant was then transferred to a sterile, cooled 250 mL centrifuge bottle and centrifuged at 8000 rpm (10,000 x g) for 15 minutes to pellet the mitochondria. The supernatant was carefully poured off and the pellet resuspended in 1 mL of homogenization buffer per 10g starting material.

Mitochondrial protein concentration was determined using the Bradford assay. 2 ul of sample or standard was added to 1 mL of Bradford assay. Four technical replicates were performed per sample to get more accurate concentration determination. The samples in Bradford assay solution were mixed, placed in cuvettes, and measured at 595 nm. Concentration was determined by applying the sample absorbance to a standard curve calculated using BSA at 1 mg/mL, 2.5 mg/mL, 5 mg/mL, and 10 mg/mL. 1mg aliquots were made of each sample followed by centrifugation at 10,000 x g for 15 minutes to pellet the mitochondria. The supernatant was removed and the pellet stored at -80° C until use.

5.2.3 Protein Purification for 2D DIGE

The protein extraction protocol was modified from the one used at the University of Missouri proteomics center (Hurkman and Tanaka, 1986). For each sample 1 mL of Tris

pH 8.8 buffered phenol was added to the pelleted mitochondria, resuspended, and transferred to a 15 mL Falcon tube. Another 1.5 mL of phenol was added, followed by 2.5 mL of extraction media (0.1 M Tris-HCl pH 8.8, 10mM EDTA, 0.4% 2-mercaptoethanol, and 0.9 M sucrose). The sample was then homogenized for one minute and incubated at 4°C with agitation for 30 minutes. After the incubation the sample was centrifuged at 4°C for 10 minutes at 5000xg. The upper phenol phase was transferred to a 50 mL falcon tube and the aqueous phase was back extracted with 2.5 mL of phenol pH 8.8. The back extraction was vortexed and centrifuged again at 5000xg for 10 minutes, the phenol phase was then combined with the previous extract. The proteins were precipitated out of the phenol by adding 5 volumes of ice cold 100% methanol containing 0.1 M ammonium acetate. The sample was vortexed and stored at -20°C overnight followed by a centrifuge at 20,000xg at 4°C for 20 minutes to pellet the proteins. The pellet was washed twice with 100% methanol containing 0.1 M ammonium acetate followed by three washes with 80% acetone containing 10 mM DTT. The pellet was completely resupsended with each wash followed by a 15 minute incubation at -20°C and pelleting at 4°C with a 20 minute spin at 20,000xg. After the final wash the pellet was dried at 37°C for 15 minutes and stored at -20°C.

5.2.4 2D DIGE

Protein pellets were resuspended in labeling buffer containing 7 M urea, 2 M thiourea, 4% CHAPS, 30 mM Tris-Cl, and a final pH of 8.5 by adding NaOH. The samples were quantified using the Bradford assay (see section 5.2.2). After quantification, 50 µg

of each sample was combined with 1 µL of either Cy3 (GE Healthcare, Piscataway, NJ cat #25-8008-83) or Cy5 (GE Healthcare, Piscataway, NJ cat #25-8008-62) at 400 pmol/µL and incubated in the dark on ice for 30 minutes. The reaction was stopped with the addition of 1 µL of 10 mM lysine and incubating on ice in the dark for 10 minutes. All samples were brought to a final concentration of 2.5 µg/µL using labeling buffer for electrophoresis.

First-dimension IEF separation and second-dimension SDS-PAGE was performed at the MU proteomics center. For the first-dimension IEF separation 25 µg each of Cy3 and Cy5 labeled proteins were combined along with 275 µg unlabeled protein from each sample. The samples were brought up to 250 µL with 2% triton X-100, 5% glycerol, 60 mM DTT, 130 mM 2-HED, 7 M urea, 2 M thiourea, and 4% CHAPS. A 3-10 IEF strip was then rehydrated with the sample at 4°C for 15 hours and subsequently electrophoresed with the following program; 250V- 250Vhrs fast ramp, 1000 V – 500Vhrs fast ramp, 8000V – 2hr. gradient ramp, 8000V – 30,000 Vhrs fast ramp, and a re-ramp to 8,000V the following day prior to the second dimension run.

After IEF separation the samples were loaded onto a 15-20% Tris-glycine gradient gel and electrophoresed at 0.5 watt per gel for 30 minutes followed by an overnight run at 7.5 mA per gel. The gels were washed for one hour in 20% methanol and scanned for Cy3/Cy5 images. Gels were then fixed in 10% methanol and 7% acetic acid for 1 hour and stained overnight with SYPRO (BioRad) followed by imaging with the FujiFLA 5000. Spots of interest were manually extracted and given to the MU proteomics center for trypsin digestion, MALDI TOF/TOF MS and MS/MS in the positive ion mode.

5.2.5 1D gel Electrophoresis

A 1 mg mitochondrial aliquot was removed from the freezer and 80 µL TE added for each sample to be run on the gel. The concentration of the samples was then determined using the Bradford assays as described in section 5.2.2. In a separate tube 10 µg of mitochondrial proteins were brought up to 9.75 mL with TE followed by the addition of 3.75 µL loading buffer (Invitrogen cat# NP0007) and 1.5 µL reducing agent (Invitrogen NP0004). The samples were mixed by pipetting and then heated with agitation at 70° C for 10 minutes. Non-solubalized material in the sample was then pelleted by centrifuging at max speed for 5 minutes. Samples were loaded onto a NuPAGE 12% Bis-Tris Gel (Invitrogen NP0343) and electrophoresed at 200V until the migration front reached the bottom of the gel using NuPAGE MES-SDS running buffer (Invitrogen cat# NP0002) with antioxidant (Invitrogen cat# NP0005).

5.2.6 Blotting

After completion of electrophoresis, the gel cassette was disassembled and the gel was soaked in transfer buffer (25 mM Tris-HCl, 192 mM Glycine, and 20% Methanol pH 8.3) for a minimum of 15 minutes. Blotting was performed using the Fisher semi-dry blotting unit (FB-SDB-2020). First, three pieces of Whatman paper slightly larger than the gel were wetted in transfer buffer and placed on the blotter one at a time rolling out air bubbles between each piece. The equilibrated gel was then placed on top of the

paper followed by a piece of 0.2 um PVDF (Millipore ISEQ00010) the exact size of the gel that had been wetted in 100% methanol followed by 1-2 minute soaks in water and a final 1-2 minute soak in transfer buffer. Three more pieces of Whatman paper slightly larger than the gel were wetted in transfer buffer and placed on top of the PVDF one at a time, rolling out air bubbles between each layer. The blotter was assembled and ran at 0.8 mA per cm^2 of gel for 2.5 hours. After blotting the apparatus was disassembled and the blot dried overnight.

5.2.7 Antibody Incubations

Before hybridization the blot was re-equilibrated by wetting in 100% methanol followed by washing in water. The blot was then placed in 25 mL blocking solution (1x TBS, 5% (w/v) nonfat dry milk, and 0.1% (w/v) Tween-20) and incubated at room temperature with gentle agitation for one hour, followed by three 5 minute washes with 1x TBS containing 0.1% Tween-20. After washing, the blot was incubated in the primary antibody solution (1x TBS, 5% (w/v) BSA, 0.1% (v/v) Tween-20) for 1 hour at room temperature with gentle agitation. A list of the antibodies used is shown in Table 5. The blot was then washed again three times for 5 minutes with 1x TBS containing 0.1% Tween-20. After washing, the blot was soaked in blocking solution containing a 1:3000 dilution of the appropriate secondary antibody for 1 hour at room temperature with gentle agitation. A final three 5-minute washes, were performed with 1x TBS containing 0.1% (v/v) Tween-20, followed by 5 minutes in Pierce SuperSignal West Pico

Chemiluminescent Substrate (cat# 34080). The blot was placed in a sheet protector and exposed to Blue Ultra Autorad Film (ISCBioExpress cat# F-9029-8x10) for imaging. The film was the scanned into the computer and imported into Image Gauge for quantification. The amount of fold change was calculated relative to the B37N sample after normalization based upon MnSOD levels. A change of 1.5 fold or higher was considered significant.

Table 5: Antibodies used for westerns

Protein Recognized	Monoclonal/Polyclonal Antibody	Dilution used	Organism antibody was raised in
ATPα[1]	Monoclonal	1:500	Mouse
E1α[1]	Monoclonal	1:500	Mouse
ATPγ[2]	Polyclonal	1:8000	Rabbit
MnSOD[3]	Polyclonal	1:1000	Rabbit
ATP6[2]	Polyclonal	1:8000	Rabbit
HSP22[1]	Monoclonal	1:500	Mouse
ATPδ'[2]	Polyclonal	1:8000	Rabbit
ATP9	Polyclonal	1:7500	Rabbit

[1]Antibody generously provided by Dr. Tom Elthon
[2]Antibody generously provided by Dr. Christine Chase
[3]Antibody generously provided by Dr. Daniel Kliebenstein

5.2 Results

5.3.1 IEF/SDS DIGE Analysis

IEF/SDS page analysis on seedlings labeled with CyDyes identified two proteins with differential expression (Fig. 26 and 27). One of these proteins (spot #1) was around 21 kDa in size and is increased 4.76 fold in B37N as compared to B37C. MALDI TOF/TOF MS analysis did not find any significant matches with this protein. The other differentially

expressed protein (spot #2) was ~16kDa in size and was 5.34 fold increased in B37C compared to B37N.

When B37N and B37C samples from pre-meiotic tassels were compared on IEF/SDS 2D DIGE, 13 proteins were decreased by at least 2 fold in B37C compared to B37N while one protein was increased (Fig. 28 and 29). The approximate size and fold change of each protein is summarized in Table 6. Protein #14 (Table 6), which has increased expression in B37C, is not a protein spot but a smear of proteins at the pole of the IPG strip. Proteins #2, 3, 4, 6, and 7 were chosen for MALDI TOF/TOF MS and MS/MS analysis. Protein #2 had MS and seven out of eight MS/MS matches to maize *Rf2* aldehyde dehydrogenase. There was no significant MS match for protein #3 however two of the eight ions analyzed with MS/MS were from the rice ATPδ precursor. Two of the eight ions analyzed for protein #4 were identified to be similar to rice ATPD although the MS results did not have any significant matches. The remaining two proteins (#6 and #7) did not have any significant matches for MS or MS/MS analysis.

As a control, comparisons of pre-meiotic B37C/Ky21 and B37N/Ky21 (B37C and B37N in the restored background) were made. This comparison only identified three 2-fold or greater protein differences (Fig. 30 and 31). Two of these proteins (2 and 3 on Fig. 30 and 31) are increased in B37C/Ky21, protein 2 by 2.12 fold and protein 3 by 2.38. Protein 1 is ~2.5 greater in abundance in B37N/Ky21 and is a streak along the edge of the IPG strip like protein #15 from the pre-meiotic B37N vs. B37C comparison. Both pre-meiotic comparisons had a blank streak in the center of the gel.

5.3.2 Western Analysis

Westerns were performed on seedling and meiotic tassel material from B37N, B37C, and B37C/Ky21 with several antibodies (summarized in table 5): the 60 kDa ATPα, 43 kDa E1α, 34 kDa ATPγ, 24 kDa MnSOD, 24 kDa ATP6, 20 kDa ATPδ', and the 9 kDa ATP9 (Fig. 31). No novel protein was identified by the ATP9 or ATP6 antibody in the seedling or meiotic tassel material (Fig. 32A and B). There were also no significant changes in seedling protein levels with the antibodies tested (Fig. 32 C). However, in meiotic tassels ATP9 showed a 2.1 fold decrease in B37C relative to B37N while no significant difference was seen between B37C/Ky21 and B37N (Fig 32D). A similar result was obtained for ATP6, which shows a 1.9 fold decrease in B37C while B37C/Ky21 levels are the same as B37N. None of the proteins tested showed a significant decrease in the meiotic samples. The meiotic B37C sample also had expression of a stress protein HSP22 and an apparent degradation product of ATPδ'.

Figure 26: Seedling 2D DIGE B37N and B37C overlay

Seedling mitochondrial proteins from B37C were labeled with Cy3 (red) and separated by IEF/SDS-PAGE along with Cy5 labeled B37N (green). Two proteins were identified as differentially expressed and are labeled as 1 and 2 within the figure. Sizes of the SigmaMarker Wide Range (Sigma #S8445) labeled with Cy3 (red) are labeled to the left of the ladder.

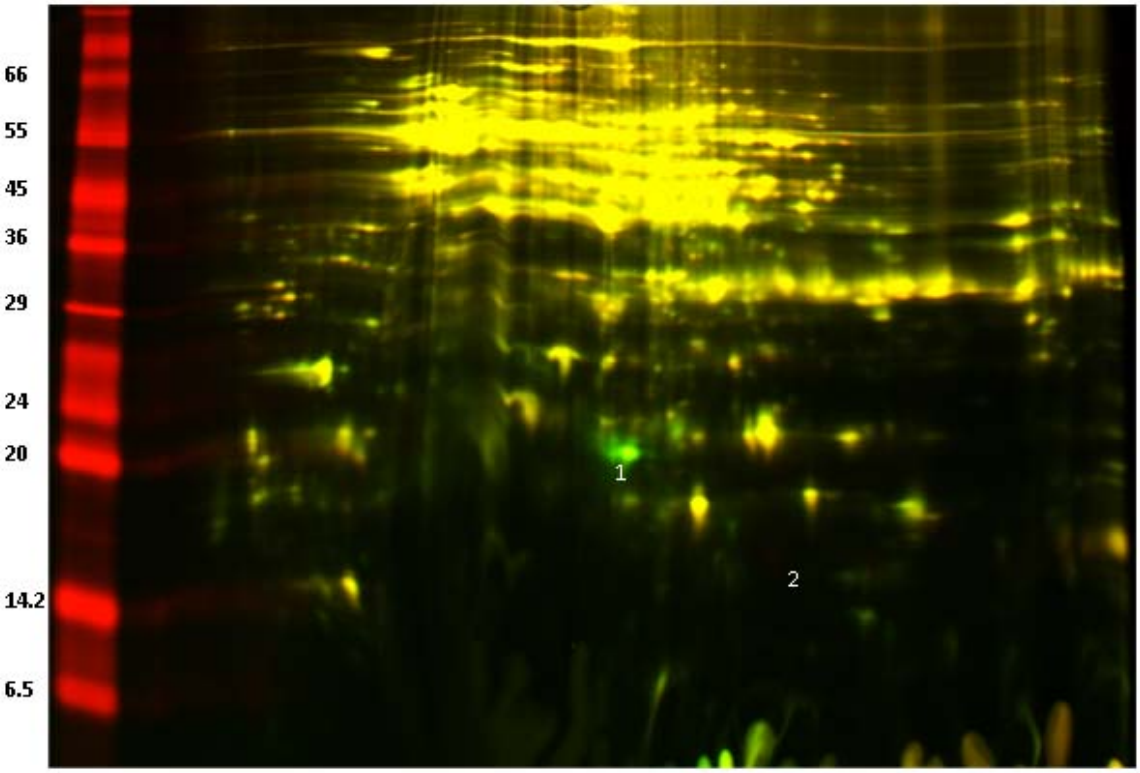


Figure 27: Seedling 2D DIGE B37C and B37N individual channels

Seedling mitochondrial proteins from B37C were labeled with Cy3 (A) and separated by IEF/SDS-PAGE along with Cy5 labeled B37N (B). Two proteins were identified as differentially expressed and are labeled as 1 and 2 within the figure.

Figure 27: Seedling 2D DIGE B37C and B37N individual channels

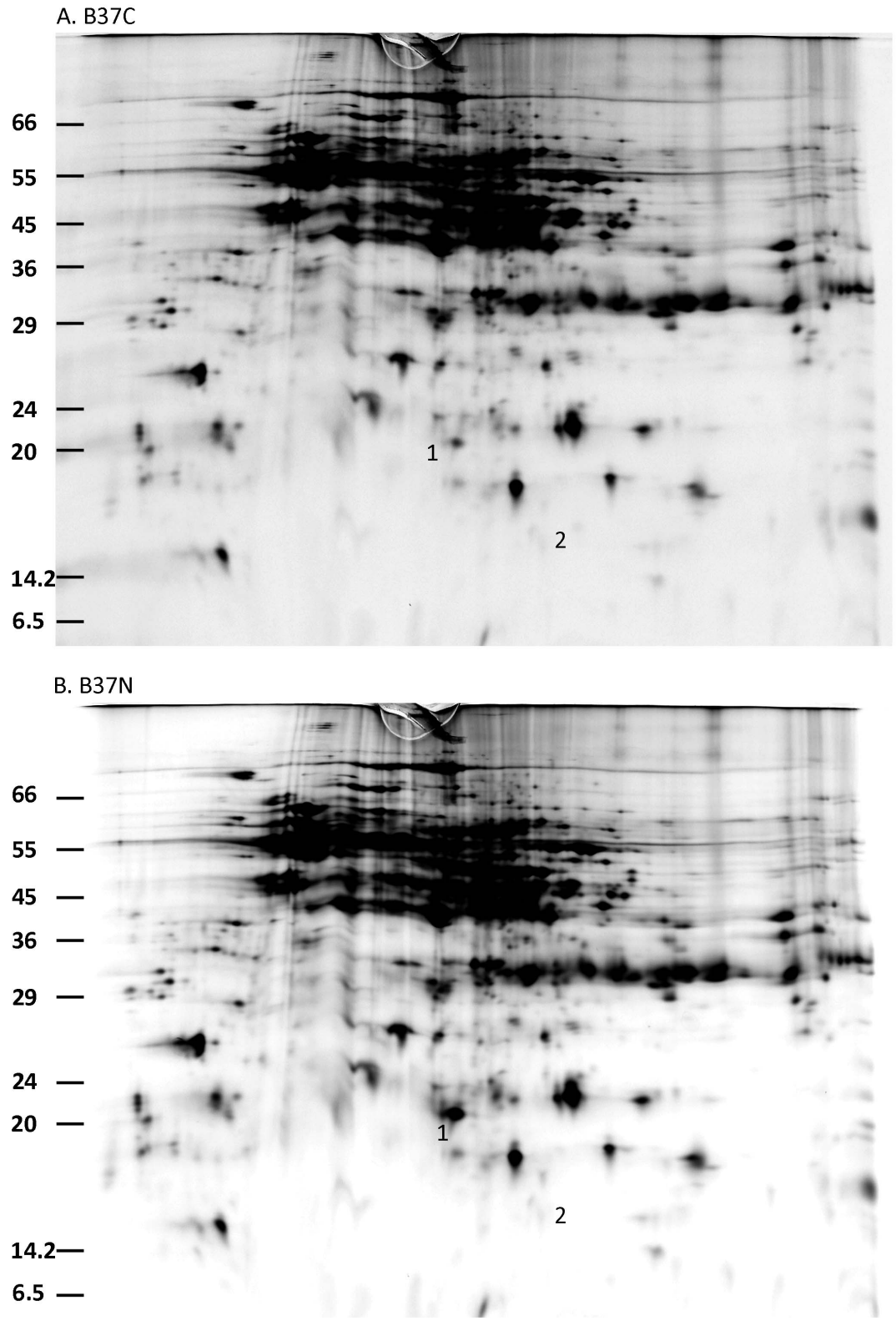


Figure 28: Pre-meiotic 2D DIGE B37C and B37N overlay

Pre-meiotic mitochondrial proteins from B37C were labeled with Cy3 (red) and separated by IEF/SDS-PAGE along with Cy5 labeled B37N (green). Thirteen proteins were identified as having higher expression in B37N (spots 1-13). There was only one spot that had higher expression in B37C (spot 14). Sizes of the SigmaMarker Wide Range (Sigma #S8445) labled with Cy3 are labeled to the left of the ladder.

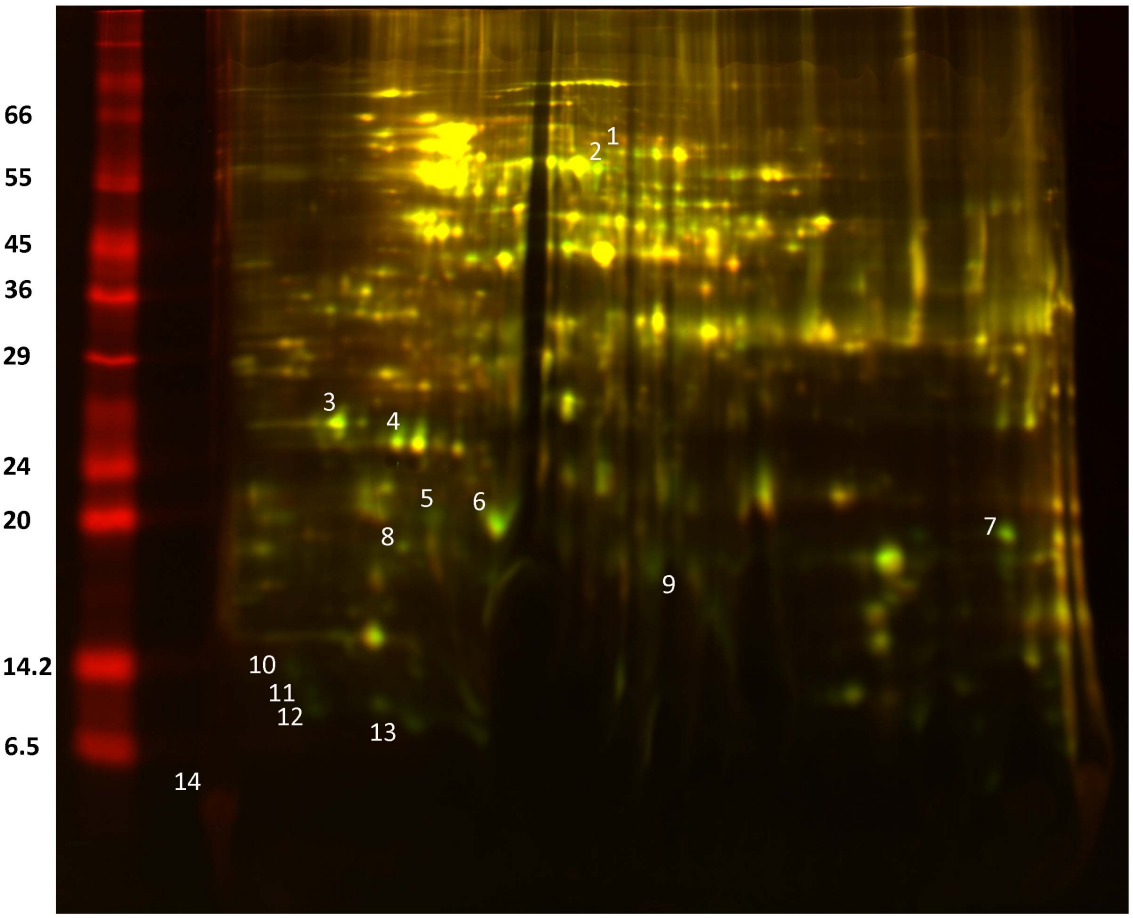


Figure 29: Pre-meiotic 2D DIGE B37C and B37N individual channels

Pre-meiotic mitochondrial proteins from B37C were labeled with Cy3 (A) and separated by IEF/SDS-PAGE along with Cy5 labeled B37N (B). Thirteen proteins were identified as having higher expression in B37N (spots 1-13). There was only one spot that had higher expression in B37C (spot 14). Images labeled the same as the overlay image (Figure 26).

Figure 29: Pre-meiotic 2D DIGE B37C and B37N individual channels

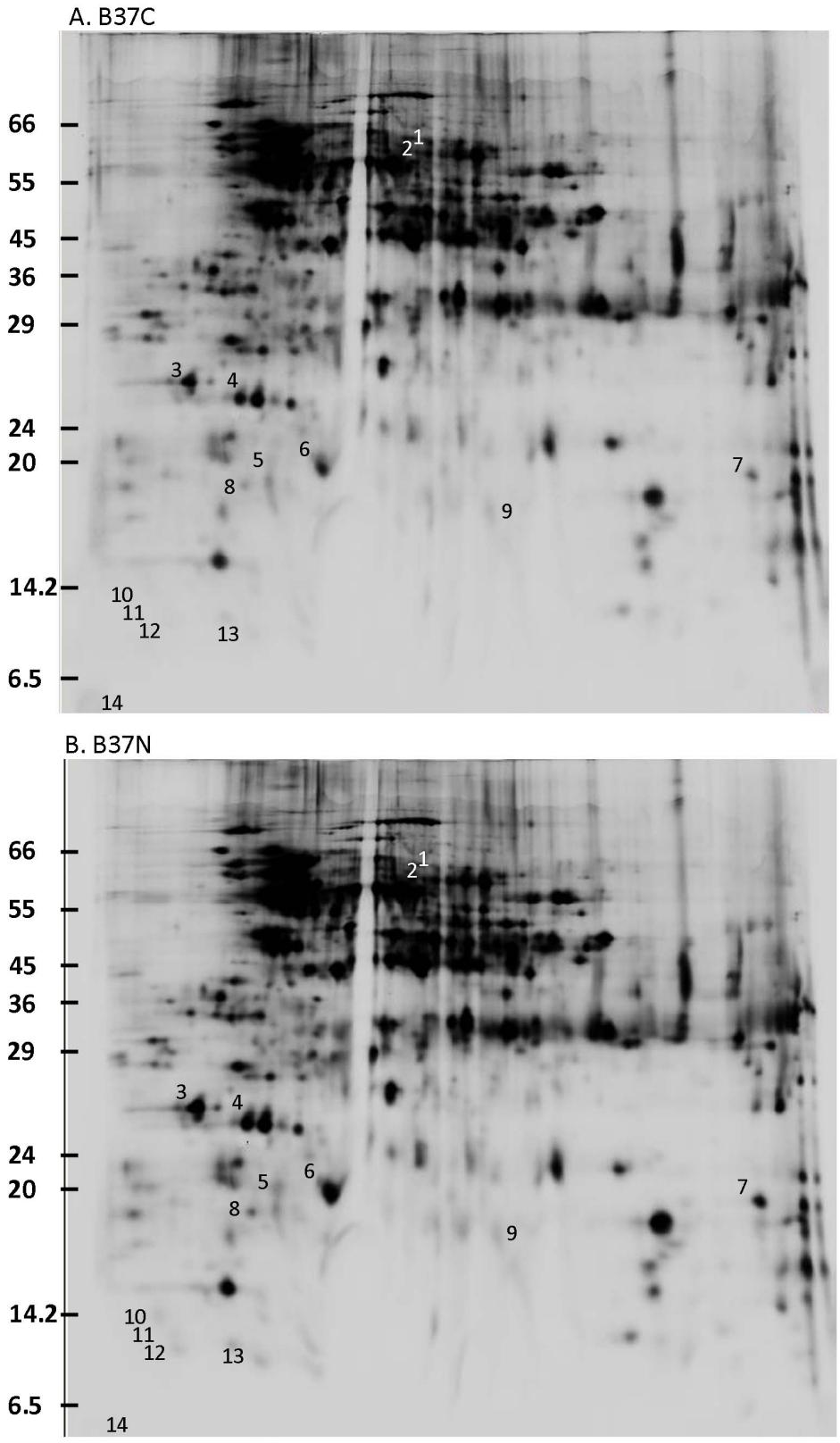


Table 6: Pre-meiotic 2D DIGE differential proteins

Protein Spot	Mass	Fold Change in B37C	MS match	MS/MS match	% Coverage of MS/MS matches
1	61 kDa	-2.01	-	-	-
2	58 kDa	-2.65	Maize Rf2	Maize Rf2	50%
3	26 kDa	-2.04	No matches	Rice putative ATPδ precursor	21%
4	25 kDa	-2.04	No matches	Rice putative ATPD	35%
5	20 kDa	-2.24	-	-	-
6	20 kDa	-2.14	No matches	No matches	No coverage
7	20 kDa	-2.51	No matches	No matches	No coverage
8	18 kDa	2.07	-	-	-
9	17 kDa	-2.11	-	-	-
10	14 kDa	-2.69	-	-	-
11	11 kDa	-2.33	-	-	-
12	9 kDa	-2.67	-	-	-
13	7 kDa	-2.23	-	-	-
14	4 kDa	~5	-	-	-

* All proteins that did not have MS and MS/MS performed on them have a – in the relative fields.

Figure 30: Pre-meiotic 2D DIGE B37N/Ky21 and B37C/Ky21 overlay

Pre-meiotic mitochondrial proteins from B37N/Ky21 were labeled with Cy3 (red) and separated by IEF/SDS-PAGE together with Cy5 labeled B37C/Ky21 (green). Three proteins were identified as differentially expressed between the two samples (spots 1-3). Sizes of the SigmaMarker Wide Range (Sigma #S8445) labled with Cy3 are labeled to the left of the ladder.

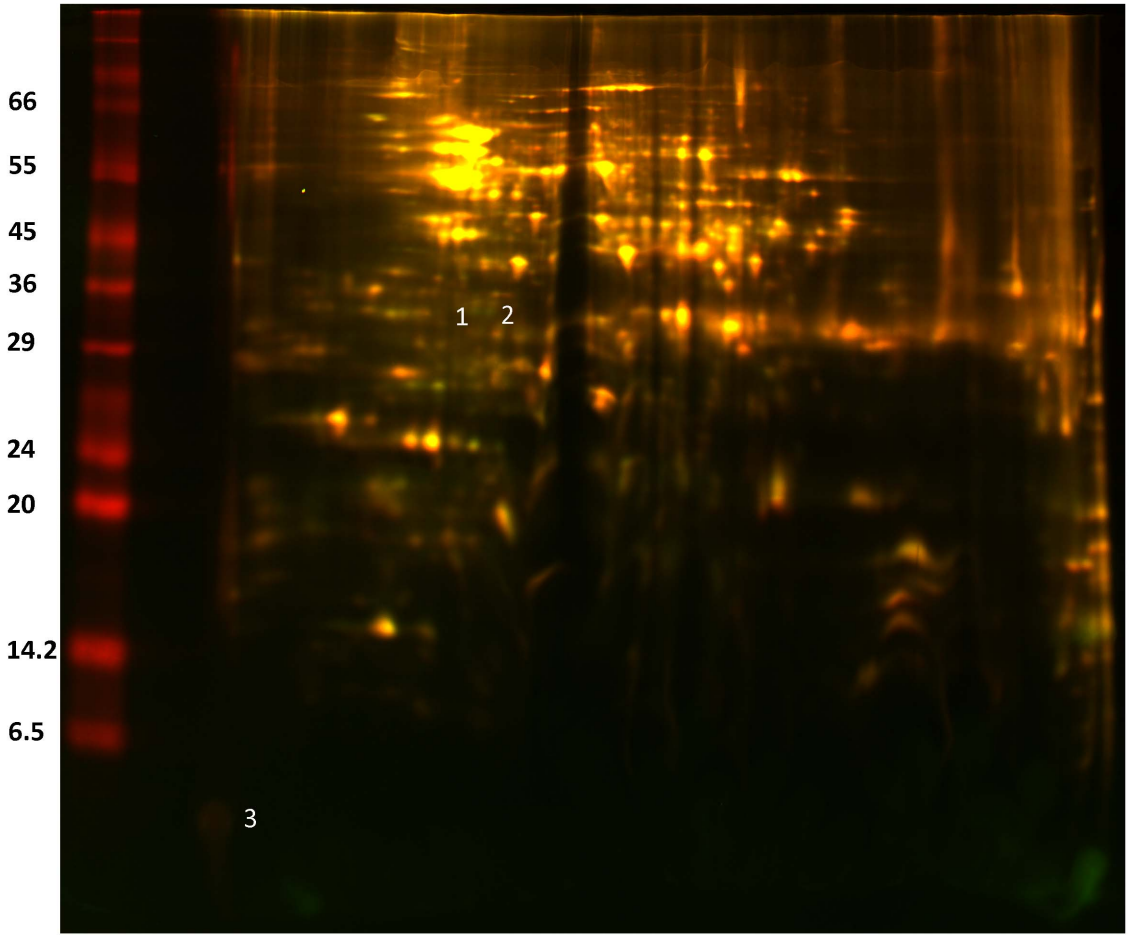


Figure 31: Pre-meiotic 2D DIGE B37N/Ky21 and B37C/Ky21 individual channels

Pre-meiotic mitochondrial proteins from B37N/Ky21 were labeled with Cy3 (A) and separated by IEF/SDS-PAGE along with Cy5 labeled B37C/Ky21 (B). Three proteins were identified as differentially expressed between the two samples (spots 1-3).

Figure 31: Pre-meiotic 2D DIGE B37N/Ky21 and B37C/Ky21 individual channels

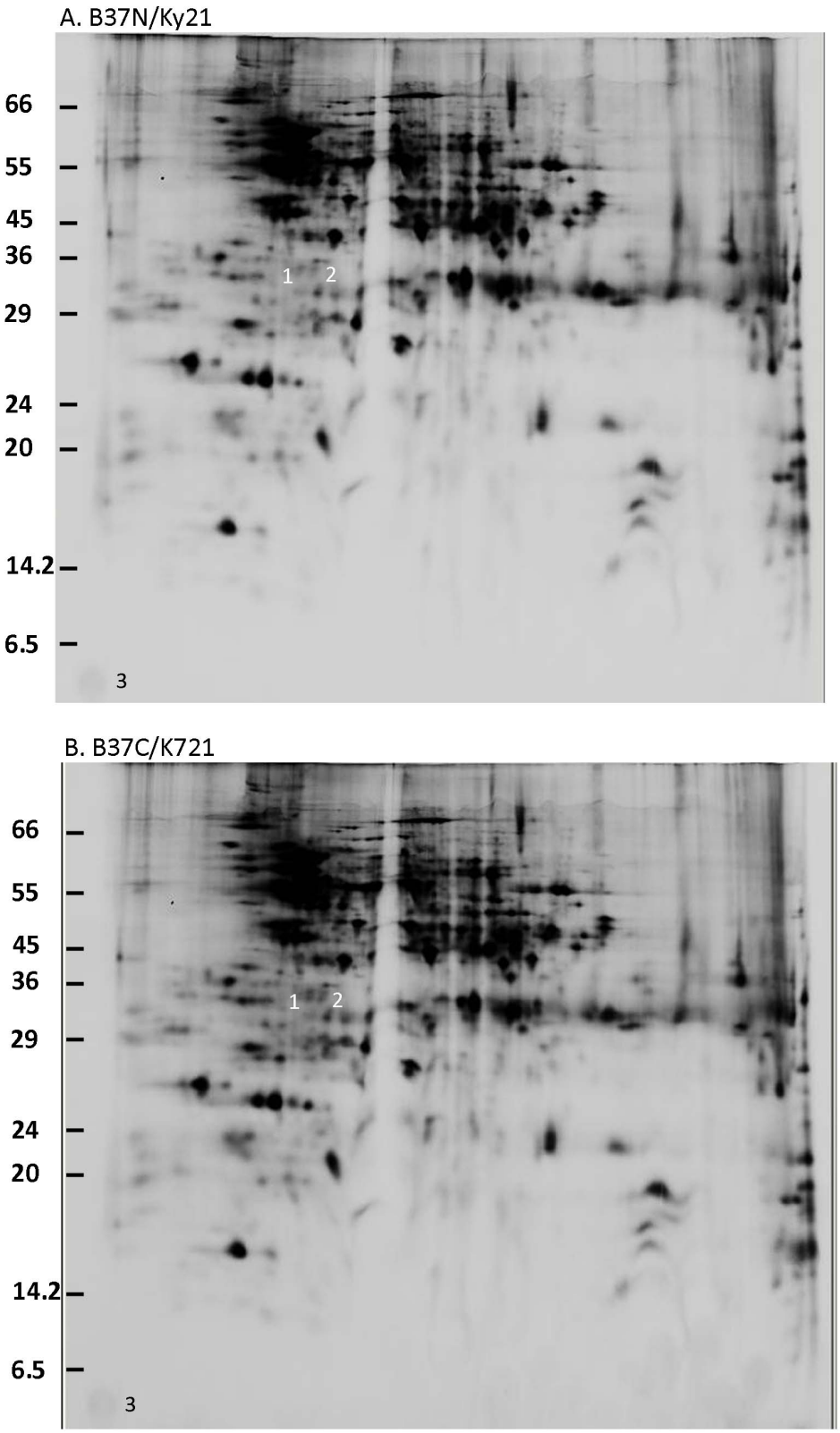
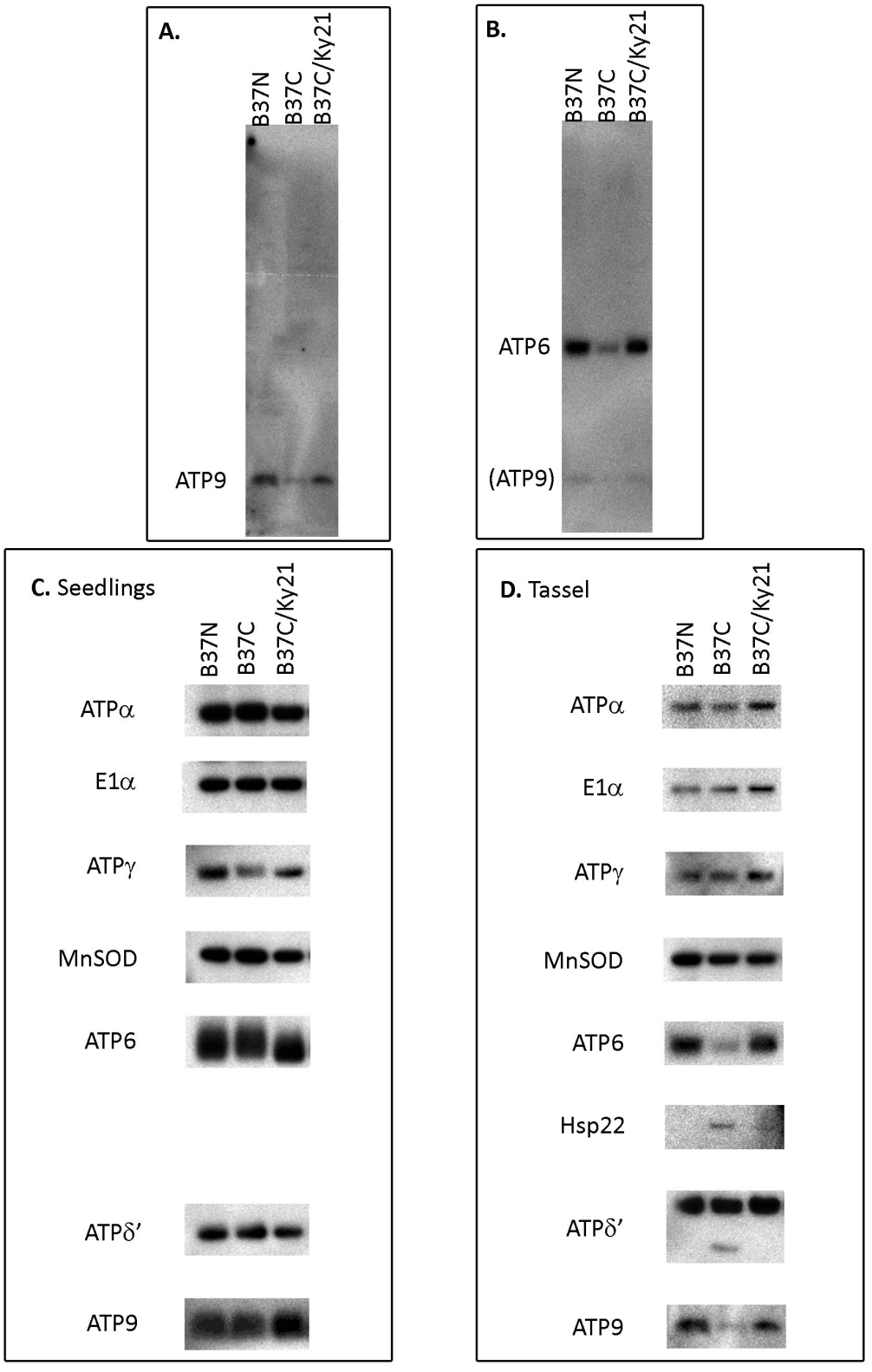


Figure 32: 1D immunoblots

Mitochondria were isolated from either etiolated seedling shoots or meiotic tassels and run on a 12% Bis-Tris gel from Invitrogen. The gels were blotted and hybridized to several antibodies. A. Full length gel blot of meiotic tassel mitochondria hybridized to the ATP9 antibody showing no other proteins being detected. B. The same blot was re-hybridized to the ATP6 antibody. Other than the remnant ATP9, no other proteins were detected. C. Seedling mitochondria hybridized to several components of the ATP synthase complex (ATPα, ATPγ, ATP6, ATPδ', and ATP9), along with a component of pyruvate dehydrogenase (E1α) and the mitochondrial superoxide dismutase (MnSOD). No proteins were differentially expressed in the seedling samples. D. Meiotic tassel mitochondria hybridized to the same antibodies as the seedling mitochondria along with a stress protein (HSP22). There was induction of HSP22 in B37C along with a decrease in ATP6 and ATP9 levels. The B37C meiotic tassel sample also had an extra band with the ATPδ' antibody. There were no significant differences between B37N and B37C in the restored background (B37C/Ky21).

Figure 32: 1D immunoblots



5.4 Discussion

5.4.1 IEF/SDS DIGE Analysis

DIGE analysis on seedlings and pre-meiotic tassels gave very different results. First, seedlings were analyzed in order to compare the results to previously published work. These samples served as controls since CMS-C has no apparent defects at this stage of development (Forde and Leaver, 1980). Pre-meiotic tassel material was also analyzed to quantify protein levels right before the physical signs of abortion are seen in meiosis. Ideally using DIGE and pre-abortion tassels would 1) increase the probability that the protein causing pollen abortion will be visualized 2) minimize the number of secondary effects caused by pollen abortion.

Only two major differences were seen in the seedling material. One protein increased in the B37N sample and another increased in the B37C sample. Neither of these proteins were identified based upon MS analysis or appeared differentially expressed in the pre-meiotic tassel material. In the pre-meiotic tassel 15 differentially expressed proteins were observed (Fig. 28 and 29, Table 6). All of these proteins except one had higher expression in B37N. The five differentially expressed proteins that had the highest levels were selected for MS analysis. Three of these proteins could be assigned with confidence from the MS^2 spectra.

One of these proteins (#2 on Table 6) was identified as an aldehyde dehdrogenase encoded for by the *restorer of fertility gene 2* (*Rf2*). *Rf2* is a mitochondrially localized aldehyde dehydrogenase gene that is involved in the restoration of fertility of CMS-T in

maize. It is interesting to note that, while CMS-T requires the expression of a particular aldehyde dehydrogenase to restore fertility, CMS-C decreases the expression of an aldehyde dehydrogenase. *Rf2* has been shown to be necessary for normal pollen development in N mitochondria as well as CMS-T (Liu et al., 2001). It is thought that Rf2 may play a role in detoxifying acetaldehyde produced during microspore development (Liu et al., 2001; Tadege et al., 1999).

Two other proteins that had significant matches from MALDI-TOF-TOF analyses were components of the ATP synthase complex. Protein #3 had two ions matching the precursor of the ATP synthase δ subunit (ATPδ), which is the same protein as ATPδ'. There were also two ion matches for Protein #4 to the ATP synthase D subunit (ATPD). Although neither of these proteins are mitochondrially encoded, this finding does imply that CMS-C abortion is affecting the ATP synthase complex. This would suggest that the remaining unidentified differentially expressed proteins could be components of the ATP synthase complex. There might be a reduction in ATP synthase components due to a chimeric protein interfering with the proper formation of the ATP synthase complex. Alternatively, reduced synthesis of a mitochondrially encoded ATP synthase subunit may also result in lower levels of all ATP synthase components.

The only protein that was detected as increased in the pre-meiotic B37C sample consisted of a smear along the length of the gel followed by a large spot at the bottom localizing at the IEF pole. This could represent a membrane protein being synthesized uniquely to CMS-C. Because of the non-polar nature of membrane proteins they do not

usually separate well on IEF strips and tend to cluster at the end of the strip. However, this might also represent unsolubilized proteins in the B37C sample and have nothing to do with the protein composition of the samples. Interestingly, this was not seen on the seedling IEF/SDS DIGE gel even though the same sample preparation method was used.

When the pre-meiotic B37N and B37C cytotypes were compared to each other in the restored background (B37N/Ky21 versus B37C/Ky21), all of the differences disappeared except for the streak at the IEF pole (Fig. 30 and 31). This streak was still present to a lesser degree however it was now showing an increase in the B37N/Ky21 sample instead of the B37C/Ky21. This may indicate restoration of fertility affecting a protein in this streak. It may also mean that there is a dye bias since both the B37C and B37N/Ky21 samples were labeled with Cy3.

There were two other proteins in the B37N/Ky21 v B37C/Ky21 comparison that had differential expression. They were increased in the B37C/Ky21 sample and migrated right next to each other; thus, they may be isoforms or modified forms of the same protein. These proteins do not correlate to any proteins identified as differentially expressed in the B37N verses B37C comparison (Fig 28 and 29). Unfortunately the concentrations of the proteins were too low for MS analysis.

Several contradictions arose in the performance of this analysis. One is that neither the results obtained from seedlings nor pre-meiotic tassels matched the previously published data. The previous work analyzing protein synthesis differences in mitochondria of CMS-C and normal seedlings, identified B37C as missing a 15.5 kDa

protein that is present in B37N, while gaining a 17.5 kDa protein (Forde and Leaver, 1980). These proteins were identified as mitochondrially synthesized because they were labeled *in organelle* with 35S-methionine. In contrast, the analysis presented here was performed on total mitochondria proteins. Any differentially expressed proteins identified by Forde and Leaver (1980) may be present at levels too low to be detected when looking at all of the mitochondria proteins. Another possibility is that, in the current study, some differences are being masked on the IEF-SDS gels due to overlapping proteins so that the previously identified variations are not seen. A third explanation is that these proteins are rapidly turned over and can only be detected with *in organello* protein synthesis approaches.

It is possible that the proteins identified by Forde and Leaver (1980) are the extended ORFs of the *atp6* gene. Both the extended ORFs in B37N and CMS-C *atp6* are predicted to be 16.3 kDa proteins. Although this is not the size of either of the identified proteins, proteins on SDS gels do not always migrate at their actual size due it variations in SDS binding (Rath et al., 2009). This may be the case for the extended ORFs especially since the N extended ORF has a theoretical isoelectrical point at 6.14 while the CMS-C extended ORF's is at 4.4.

It was not surprising to see several proteins having differential expression in tassels and not in seedlings since this is the tissue in which pollen abortion occurs. However, it was unexpected to see two proteins with large differences in levels within seedlings and having none in the pre-meiotic tassel samples. It is possible that these proteins have

nothing to do with the CMS-C phenotype. Since CMS-C and B37N have the same nuclear background, any differences should be attributable to the mitochondrial genotype. Furthermore, CMS-C has no known effects outside pollen development, so any changes due to CMS-C mitochondrial expression should be in all tissues or be tassel specific; differences should not occur just in the seedlings.

In order to draw more solid conclusions from the analysis, more biological replicates and dye swaps would need to be performed. However, since these preliminary results indicate that ATP synthase is being affected by pollen abortion in CMS-C and that a membrane protein may be involved, alternative methods were used for further investigation.

5.4.2 Western Blot Analysis

Two candidate genes were identified for causing CMS, the chimeric *atp6* and *atp9-2* genes (Fig. 3); both are components of the ATP synthase complex. It is possible that the *atp6* extended orf is producing a chimeric protein with an ATP9 N-terminus. Alternatively, the extended orf is not being proteolytically cleaved, and so a chimeric ATP6 is synthesized. *Atp9-2* could also produce a chimeric ATP9 protein if it uses a different translation start site. The 4kb *atp9-2* transcript in CMS-C could affect the availability of the start site and inhibit the efficient translation of the gene.

No chimeric protein was identified in the 2D gels making it less likely that there is a chimeric protein synthesized although the protein could have been masked by other

proteins on the gel. Also, if the protein was hydrophobic it would not separate on the IEF and this would prevent identification if the protein. To test whether there is a chimeric ATP6 or ATP9 protein, westerns were performed on blots of SDS gels with antibodies raised to the N terminus of the ATP6 and ATP9 proteins.

The antibody to ATP6 only detected only the mature ATP6 protein (Figure 30B). This means that the extended orf is being proteolytically cleaved after translation, so a chimeric ATP6 protein is likely not involved in CMS-C. When tested, the ATP9 antibody did not detect a chimeric protein on the blots (Fig. 32A). This means that the *atp9-2* gene is not synthesizing a stable chimeric ATP9 protein. It is also likely that the C-*atp6* extended ORF, which contains the first 12 codons of ATP9, is being degraded. However, since the antibody was raised against the first 13 amino acids of ATP9 and only the first 12 amino acids are shared with the C-*atp6* extended ORF, there is a one amino acid difference. This difference could cause the antibody not to bind to the C-atp6 extended ORF; therefore, it cannot be positively determined that this protein is not produced.

Because no chimeric protein was identified in CMS-C, ATP synthase proteins were examined to see if these proteins are affected during pollen abortion. If ATP9 synthesis is decreased, this could affect other components of the ATP synthase complex, especially those proteins within the F_0 complex (ATP9 and ATP6, Fig. 2).

Two different proteins were used as loading controls for the analysis, MnSOD which is involved in removal of superoxide radicals (Zhu and Scandalios, 1992), and the E1α subunit of the mitochondrial pyruvate dehydrogenase (Grof et al., 1995). Both of these proteins did not show significant differences in seedlings or tassels among the tested

genotypes. A third protein, HSP22, was used as a control for stress response in the CMS-C sample. HSP22 was not expressed in the N tassel, but it did show expression in the CMS-C tassel. Interestingly, lower levels of HSP22, were still detectable when *Rf4* was introduced.

Two F_1 ATP synthase proteins, ATPα and ATPγ, do not appear to be altered in seedlings or meiotic tassels. A third F_1 component tested, ATPδ', showed a second band in the CMS-C meiotic tassels. This could represent a protein cleavage product but would need to be tested further to confirm this hypothesis. In comparison, both of the F_0 subunits tested, ATP6 and ATP9, were not altered in seedlings but their levels were significantly decreased within the CMS-C meiotic tassels. The introduction of Rf4 restored ATP9 and ATP6 levels.

Previous studies have shown that the F_0 and F_1 subunits can assemble independently of each other (Tzagoloff et al., 2004). This would explain why ATP9 and ATP6 are decreased but ATPα and ATPγ are unaffected. Furthermore it has also been shown that ATPδ' binds to ATP9 and assembly of the ATP synthase complex can occur independently of ATPδ' (Duvezin-Caubet et al., 2003). If the chimeric *atp9-2* is being translated less efficiently, resulting in a decrease in ATP9 levels when there is a high demand, then fewer F_0 subunits would be formed. The other components of the F_0 complex could then be targeted for degradation resulting in the observed decrease of ATP6. The F_1 could still assemble, but fewer complete ATP synthase complexes would form, causing less ATP to be synthesized. Since ATPδ' is not essential for F_1 formation and binds to ATP9 for normal function, the protein could be destabilized and partially

degraded when ATP9 levels are decreased. If this is what is causing pollen abortion, then in CMS-C meiotic tassels, one would expect to see a decrease in the ATP synthase complex and an increase in the free ATP synthase F_1 subunit. Restoration of fertility would restore the ATP synthase complex to normal levels by increasing ATP9 levels.

Chapter 6: 2D BLUE NATIVE ELECTROPHORESIS

6.1 Introduction

Because decreases of ATP6 and ATP9 are observed in CMS-C meiotic tassel material, the levels of mitochondrial ATP synthase complex could be affected during pollen abortion. This would be especially the case for the F_0 subunit of ATP synthase, because ATP6 and ATP9 are two of the three proteins that make up the F_0 subunit (Fig. 2). To determine if the ATP synthase complex is altered in CMS-C mitochondria, 2D blue-native polyacrylamide gel electrophoresis (BN-PAGE) can be used.

BN-PAGE uses mild detergents to isolate intact respiratory chain complexes from the mitochondria (Eubel et al., 2005; Schagger, 1991). The complexes are then separated on a non-denaturing polyacrylamide gel using Commassie Blue G-250 to positively charge the proteins (Eubel et al., 2005; Schagger, 1991). This methodology allows for separation of protein complexes from 3-4 MDa to 100 kDa in size (Eubel et al., 2005). A lane of a first-dimension blue native gel can then be run on a second denaturing, SDS-polyacrylamide gel to visualize the individual proteins of each complex. Combining this technique with CyDye labeling of the proteins allows for highly sensitive identification of proteins within the mitochondrial respiratory complexes.

If the decrease in ATP6 and ATP9 is causing pollen abortion, the CMS-C mitochondria would be expected to show a decrease of the ATP synthase complex, with no other

complexes being affected. The alteration of ATP synthase levels would then be restored to normal levels with the introduction of the fertility restoring *Rf4* allele.

6.2 Materials and Methods

6.2.1 Materials

Tissue was harvested from pre-emergent tassels of B37N (NB type mitochondria), B37C (CMS-C type mitochondria), and B37C/Ky21 (CMS-C mitochondria crossed by restorer line). Two separate isolations of each genotype were performed from pooled tassel material in summer of 2007 and an isolation of each was performed from pooled material grown in the summer of 2006. This gave three biological replicates for each genotype.

The pre-emergent tassel material was determined to have entered into meiosis by staging the older anthers on the tassel using aceto-carmine (Freeling and Walbot, 1994). The anthers were removed from the male floret and fixed in a 1.5mL eppendorf tube containing 3 parts 95% ethanol to 1 part glacial acetic acid. The anthers were allowed to soak for a minimum of 10 minutes before being placed on a microscope slide. After one drop of aceto-carmine was added on top of the anthers they were smashed using rusty probes. A cover slip was placed over the smashed anthers followed by gently heating the slide over an ethanol lamp. The anthers were then examined microscopically using a 100x magnification to determine the stage of development.

6.2.2 Mitochondrial Protein Isolation

The protocol used to isolate mitochondria was a modification from the protocol outlined by (Stern and Newton, 1985). All solutions used had been autoclaved and all isolation steps were conducted at 4° C.

The tissue was weighed, surface sterilized in a 10% sodium hypochlorite solution, and rinsed in cold DEPC treated de-ionized water. The tissue was then placed in a cooled blender jar containing cold homogenization buffer (a ratio of 4 mL buffer per 1 gm tissue). The homogenization buffer contained 0.4 M mannitol, 25 mM MOPS pH 7.8, 1 mM EGTA, 4 mM cysteine and 0.15% Bovine serum albumin was added just prior to use. Pulses of two seconds each at low followed by high speed was used to homogenize the tissue in a Waring blender. The homogenate was filtered through a sterile funnel, lined with four layers of cheese cloth on top of one layer of Miracloth, into autoclaved 250 mL centrifuge bottles on ice. Nuclei, plastids, starch, and cellular debris were removed from the homogenate through two differential spins in an RC-5B Sorvall centrifuge at 4° C, with a pre-cooled GSA rotor. First the filtrate was centrifuged at 2500 rpm (1000 x g) for 5 minutes. The supernatant was carefully decanted into sterile, cooled 250 mL bottles. The supernatant was then centrifuged again at 4000 rpm (2000 x g) for 10 minutes to remove additional plastids. The supernatant was then transferred to a sterile, cooled 250 mL centrifuge bottle and centrifuged at 8000 rpm (10,000 x g) for 15 minutes to pellet the mitochondria. The supernatant was carefully poured off and the pellet resuspended in 1 mL of homogenization buffer per 10g starting material.

Mitochondrial protein concentration was determined using the Bradford assay. 2 µL of sample or standard was added to 1 mL of Bradford assay. Four technical replicates were used per sample in order to get more accurate concentration determination. The samples in Bradford assay solution were mixed, placed in cuvettes, and absorbance determined at 595nM. Concentration was determined by applying the sample absorbance to a standard curve calculated from BSA at 1 mg/mL, 2.5 mg/mL, 5 mg/mL, and 10 mg/mL. 1mg aliquots were made of each sample followed by centrifugation at 10,000 x g for 15 minutes to pellet the mitochondria. The supernatant was removed and the pellet stored at -80°C until use.

6.2.3 Isolation of Mitochondrial Complexes

All blue native protocols used were modified from Schagger *et al.* (1991) and Eubel *et al.* (2005).

A 1 mg aliquot of mitochondria was resuspended in 100 µL blue native lysis buffer (30 mM HEPES pH 7.4, 150 mM potassium acetate, and 10% (v/v) glycerol) containing 10 mg of digitonin (10 g digitonin added to blue native lysis buffer, heated to 90°C and divided into 100 µL aliquots and stored at -20° C) and incubated on ice for 20 minutes. During the 20 minute incubation the samples were pipeted every 5 minutes to mix. The samples were then centrifuged for 30 minutes at 100,000 x g and 4° C using a TLA 100.2 rotor. After centrifugation the supernatant was removed and the concentration determined using the Bradford assay as previously described. For western analysis a

250 µg aliquot was transferred to a 0.5 mL tube and brought up to 80 µL using blue native lysis buffer and 4 µL of blue native loading buffer (5% commossie G-250 and 750 mM aminocarpoic acid) was added.

6.2.4 DIGE Labeling of Mitochondrial Proteins

Pre-meiotic tassel mitochondria protein complexes from B37N, B37C, and B37C/Ky21 were labeled with Cy5, Cy3, and Cy2 dyes respectively for three biological replicates. One dye swap was performed labeling B37N with Cy3 and B37C with Cy5, (B37C/Ky21 was not tested in the dye swap experiment). 100 µg of protein complexes was transferred to a 0.5 mL tube after isolation (Section B) and brought up to 22 µL using blue native lysis buffer. To each tube 1µL of the appropriate CyDye at a concentration of 400 pmol was added, mixed, briefly spun down, and incubated on ice in the dark for 30 minutes. The labeling reaction was then terminated by adding 2 µL of 10 mM lysine, mixed by pipetting, and then incubated in the dark for 10 minutes on ice. The labeled samples were then combined and 3.75 µL of blue native loading buffer was added.

6.2.5 1D Blue Native Electrophoresis (BN-PAGE)

The mitochondrial complexes were separated using a 28 cm long, 0.75 cm thick gel. First a 7.6 cm long 12% acrylamide spacer was poured consisting of 2.3 mL water, 3.3 mL 3x blue native buffer (150 mM Bis Tris and 1.5 M aminocarpoic acid pH 7.0), 2 mL 50%

Glycerol, 2.4 mL 49.5% acrylamide (49.5T/3C), 35 µL 10% ammonium persulfate, and 3.5 µL TEMED. The gel was overlayed with isopropanol until polymerization.

The isopropanol was then rinsed off with water and the inside of the gel cassette dried using chromatography paper. A 4-12% separation gradient was then poured from the top using a gradient pourer connected to a peristaltic pump. The 4% gel solution was made by combining 4.51 mL water, 3.05 mL 3x blue native buffer, 0.925 mL 50% Glycerol, and 0.74 mL 49.5% acrylamide (49.5T/3C). The 12% gel solution consisted of 2.1 mL water, 3.05 mL 3x blue native buffer, 1.85 mL 50% Glycerol, and 2.22 mL 49.5% acrylamide (49.5T/3C). The solutions were added to the gradient pourer at which point 25µL 10% ammonium persulfate and 2.5µL TEMED was added just prior to turning on the pump and allowing the two chambers to mix. After completion of pouring the gel was overlayed with isopropanol and allowed to polymerize.

After the gel polymerized the isopropanol was rinsed off with water and the inside of the gel cassette dried out using chromatography paper. A stacker was then poured on top of the separating gel to the top of the cassette consisting of 2.89 mL water, 1.67 mL 3x blue native buffer, 0.4 mL 49.5% acrylamide, 40 µL 10% ammonium persulfate, and 4 µL TEMED. The gel comb was then inserted into the stacker. After polymerization the gel was placed at 4°C to cool until use.

The gel was placed in a C.B.S. Scientific gel rig filling the cathode chamber with a buffer consisting of 50 mM Tricine, 15 mM BisTris pH 7.0, and 0.02% coomassie blue G-250; and the anode chamber with 50 mM BisTris pH 7.0. The comb was removed and the wells rinsed out to remove any unpolymerized acrylamide. The gel was then loaded

and ran at 200V for 24 hours in the dark until the migration front reached the bottom spacer.

6.2.6 DIGE 2D BN-PAGE

After the 1D gel electrophoresis was completed the lanes were cut out and placed in a 15 mL concial tube containing 3 mL of a 1% SDS and 1% β-mercaptoethanol. The tubes were wrapped in aluminum foil and incubated with gentle agitation for a minimum of 30 minutes. The gel strip was washed briefly with water and then placed on top of a 1 cm thick, 22 cm 15-22% Tris-Glycine gel with a 1 cm stacker. Once the gel strip is flush with the stacking layer it was embedded with a 1.2% agarose solution contacting methylene blue. Before running the gel was scanned using a GE Ettan DIGE Imager. The gel then ran overnight at 150V until the migration front reached the bottom of the gel. The gel was then scanned again using with a GE Ettan DIGE Imager.

6.2.7 DIGE 2D BN-PAGE Analysis

The scanned gel images were analyzed using General Electric DeCyder 2D software v6.5. A protein was identified as differentially expressed if it had a 1.5 or larger fold change.

6.2.8 2D Gel Electrophoresis for Western Analysis

After the 1D gel electrophoresis was completed the top 10cm of each lane was cut out and placed in a 15 mL conical tube containing 3 mL of a 1% SDS and 1% B-mercaptoethanol. The tubes were incubated with gentle agitation for a minimum of 30 minutes. The gel strip was washed briefly with water and then placed on top of a 1cm thick, 20 cm long 15-22% Tris-Glycine gel with a 1 cm stacker. Once the gel strip was flush with the stacking layer it was embedded with a 1.2% agarose solution contacting methylene blue. The gel then ran overnight at 150V until the migration front reached the bottom of the gel.

6.2.9 Western Blotting

Blotting was performed according to standard methods (Maniatis et al., 1989). After completion of the gel run the gel cassette was disassembled and the gel cut just above the migration front followed by an equilibration in transfer buffer (25 mM Tris-HCl, 192 mM Glycine, and 20% Methanol pH8.3) for a minimum of 15 minutes. Blotting was performed using the Fisher semi-dry blotting unit (FB-SDB-2020). First three pieces of Whatman paper (Fisher cat# 05-714-4) slightly larger than the gel were wetted in transfer buffer and placed on the blotter one at a time rolling out air bubbles between each piece. The equilibrated gel was then placed on top of the paper followed by a piece of 0.2 µM PVDF (Millipore cat no.# ISEQ00010) the exact size of the gel that had been wetted in 100% methanol followed by 1-2 minute soaks in water and a final 1-2 minute soak in transfer buffer. Three more pieces of whatman paper slightly larger than

the gel were wetted in transfer buffer and placed on top of the PVDF one at a time, rolling out air bubbles between each layer. The blotter was assembled and ran at 0.8mA per cm^2 of gel for 2.5 hours. After the blotting was completed the apparatus was disassembled and the gel left out to dry overnight.

6.2.10 Immunoblot Experiments

Antibody incubations with the western blots were performed according to standard methods (Maniatis et al., 1989). Before hybridization the blot was re-equilibrated by wetting in 100% methanol followed by washing in water. The blot was then placed in 25 mL blocking solution (1x TBS, 5% nonfat dry milk, and 0.1% Tween-20) and incubated at room temperature with gentle agitation for one hour, followed by three 5 minute washes with 1x TBS containing 0.1% Tween-20. After washing the blot was incubated in the primary antibody (ATPα, ATPγ, ATPδ', ATP6, ATP9, or COX2) solution containing, 1x TBS, 5% BSA, and 0.1% Tween-20, for 1 hour at room temperature with gentle agitation. For information about each antibody used see table 3. The blot was then washed again three times for 5 minutes with 1x TBS containing 0.1% Tween-20. After washing the blot was soaked in blocking solution containing a 1:3000 dilution of the appropriate secondary antibody (table 3) for 1 hour at room temperature with gentle agitation. A final three, 5 minute washes, were performed with 1x TBS containing 0.1% Tween-20, followed by 5 minutes in Pierce SuperSignal West Pico Chemiluminescent Substrate

(34080). The blot was placed in a sheet protector and exposed to Blue Ultra Autorad Film (ISCBioExpress F-9029-8x10) for imaging.

6.2.11 Analysis of BN-PAGE Immunoblots

The film was scanned and the images loaded into Image Gauge for quantification. Boxes were drawn around each spot along with a box over an adjoining blank space for background measurement. The B37N, B37C, and B37C/Ky21 blot were normalized to each other according to COX2 levels. Comparisons were then made for every spot to the B37C sample. A comparison was considered to be significant if the fold change was 2 fold or higher.

6.3 Results

6.3.1 DIGE Blue Natives

Eight different complexes or super complexes can be visualized on the blue native gel based upon the complex size and the individual protein composition (Fig 33). The largest super complex is a dimer of one complex I (NADH dehydrogenase) and two complex III (Cytochrome C reductase) subunits ($(I+III_2)_2$). This is followed by a monomer form of the same super complex ($I+III_2$). Subsequently, there is a faint dimer form of complex V (V_2, ATP synthase) that is slightly larger than the following complex I. Two different forms of a monomer complex V can be seen next, V_1-1 and V_1-2, V_1-2 lacks

protein #11 as compared to V_1-1. The last two complexes that can be identified are a dimer of complex III (III_2) followed by the F_1 (matrix) portion of complex V (V-F_1). The F_0 portion of complex V is too small to be distinguished on blue native gels (~60kDa).

Analysis was first performed on three biological replicates without comparing to the dye swap experiment. There were 31 proteins identified to be differentially expressed between B37N and B37C within at least two of the three biological replicates performed (Fig33B, 34A and B; summarized in table 7). These proteins were assigned to a complex and given a putative identity based upon the migration pattern compared to previously published work, the results are summarized in table 7 (Eubel et al., 2003; Gieige, 2003; Jansch et al., 1996). Three are components of complex I (proteins# 2, 8, and 9) representing two unique proteins (2 and 9 are the same protein in I+III_2 and complex I respectively). There was one component of complex III differentially expressed and represented by two proteins (proteins #1 in I+III_2 and #17 in III_2). Another protein (#3) is observed in I+III_2, but was not seen when complex I and III exist independently of each other.

There are another 13 proteins (# 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, and 19) that represent seven unique components of complex V. Proteins 4, 10, and 16 are putatively identified as ATPδ' in V_2, V_1-1, and V_1-2 respectively. Proteins 5 and 6 are part of V_2 and have no putative identification; they are also seen as proteins 11 and 12 in V_1-1. One other protein (#6) has an unknown Identity and is only seen differentially expressed in V_2. Protein 7 is thought to be ATP9 and only appears to be differentially expressed in V_2.

ATPε is the only ATP synthase protein that is smaller than ATP9 so is thought to be protein 13 in V_1-1. The last ATP synthase proteins identified as differentially expressed, proteins 14, 15, 18, and 19, are thought to be ATPα in V_1-2 (#14) and V-F_1 (#18), and ATPγ in V_1-2 (#15) and V-F_1 (#19).

The remaining 12 differentially expressed proteins are not part of an identified complex. Comparisons were not made to the B37C/Ky21 (labeled with Cy2) sample because it had too high of a background for accurate quantification (results can be seen in fig 34C).

When the biological replicate results were compared to the dye swap several of the 31 proteins were not confirmed as differentially expressed (Fig 35 and 36; summarized in table 7). These proteins fell into three categories: some had higher expression in the cy3 channel in the biological replicates and the dye swap (proteins #1, 14, 15, 17, 18, and 19), other proteins were differentially expressed within the three biological replicates but not the dye swap (proteins #2, 4, 8, 9, 10, 12, 16, and 23), and proteins that could not be seen on the dye swap experiment (proteins # 3, 13, 21, 22, 25, 26, 27, 28, 29, 30).

There were seven proteins confirmed to be differentially expressed by the dye swap experiment. Three of these proteins are part of the complex V dimer (proteins #5, 6, and 7) and have higher expression in B37N. Another protein (#11) is part of complex V_1-1, has higher expression in B37N, and appears to be the same protein as #5. The

remaining three confirmed proteins (proteins # 20, 24, and 31) also have higher expression in B37N but are not part of any complexes.

6.3.2 Blue Native Immunoblots

The four forms of the ATP synthase complex observed in the DIGE blue natives were also observed with the westerns; a dimer form (V_2), two monomer forms V_1-1 and V_1-2 and the V-F_1 subunit (Fig. 37; table 8). Each blot (B37N, B37C, and B37C/Ky21) was normalized according to COX2 levels and then compared to the B37C blot. All comparisons are summarized in table 8, a significant change is considered to be 2 fold or higher.

Comparing B37N to B37C identified no significant changes with the ATPα protein in any ATP synthase complex. ATPγ also did not show any differences between the B37N and B37C samples for the V_2, V_1-1, and V_1-2 complexes. However, it did have significantly lower levels (3.13 fold) in the B37N sample within the F_1 subunit. The last component of the ATP synthase F_1 subunit tested, ATPδ', also had significantly higher levels in B37C within the F_1 subunit (2.77 fold), along with significantly lower B37C levels for the V_2 (-9.24 fold) and V_1-1 (2.21 fold) complexes. The two components of the F_0 subunit tested, ATP6 and ATP9, had equal levels in the V_1-1 and V_1-2 complexes between B37N and B37C. However, both ATP6 and ATP9 had significantly lower B37C levels in the V_2 complex, 2.47 and 4.75 fold decrease respectively.

When the proteins in B37C/Ky21 were compared to B37C, all of the F_1 proteins (ATPα, ATPγ, and ATPδ') had a significant decrease in the B37C/Ky21 F_1 subunit levels (8.38, 44.01, and 34.4 fold respectively). The only other difference identified for these proteins was a 2.01 fold increase of ATPδ' in the V_1-2 complex of B37C/Ky21. No comparison was made for ATPδ' in complex V_2 because there was no detectable spot in B37C/Ky21. ATP6 levels decreased 4.22 and 2.22 fold respectively within the V_2 and V_1-1 complexes of B37C/Ky21 when compared to B37C. Although ATP9 did not completely separate from the migration front and had tailing at the bottom, quantification was still performed. A 7.69 fold decrease was observed of ATP9 in B37C/Ky21 for the V_1-2 complex. No other significant differences were identified.

Figure 33: Blue native 2D DIGE B37C and B37N overlay

Isolated mitochondrial complexes from B37C and B37N were labeled with cy3 (green) and cy5 (red) respectively. The samples were run on a blue native first dimension and imaged (A). The identifiable complexes and supercomplexes of the mitochondrial respritory chain are labeled. The first dimension gel slice was then run into a second tris-glycine SDS gel dimension and imaged again (B). The proteins with a 1.5 fold or higher change are numbered and correlate to the protein numbers in table 5.

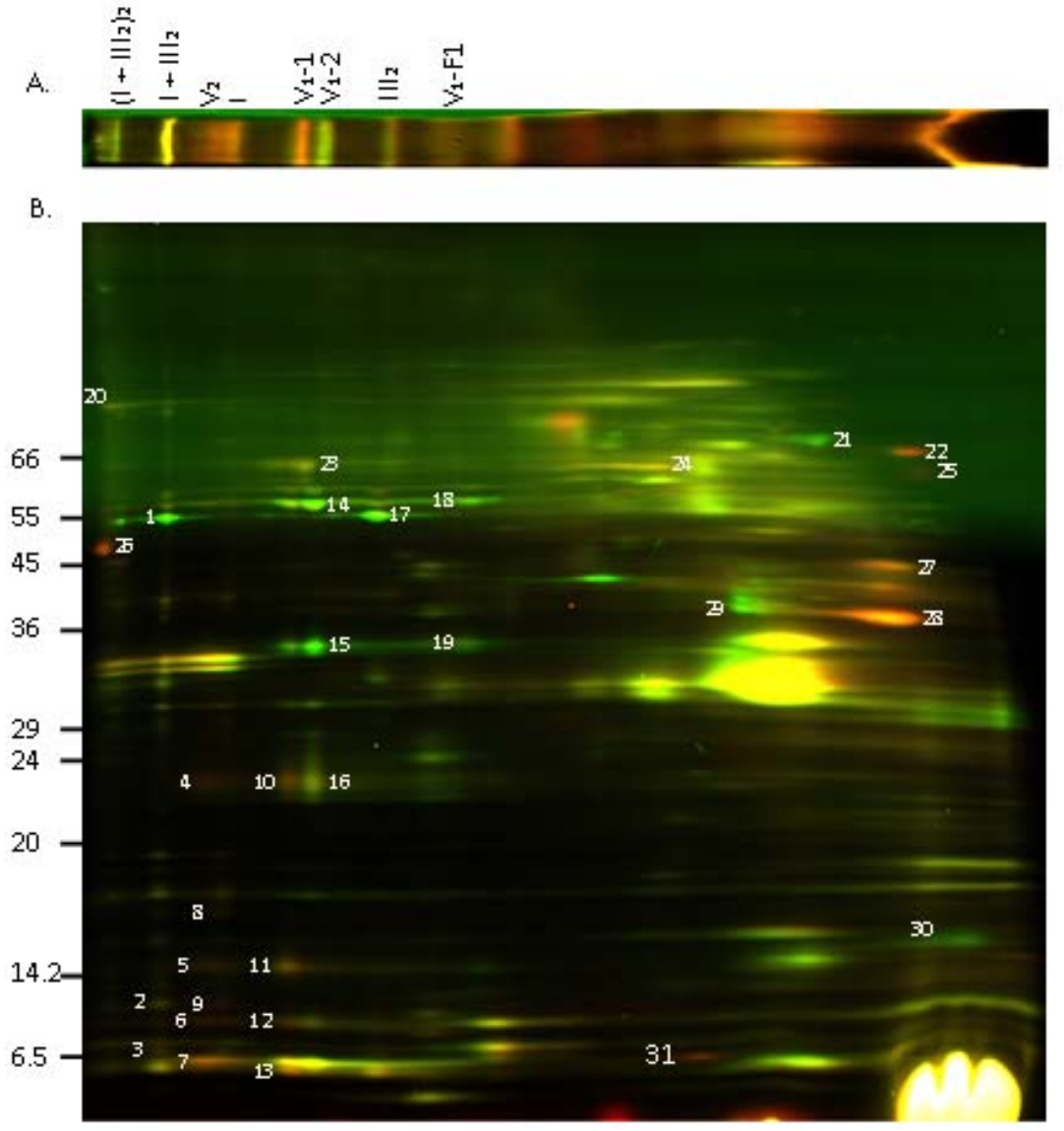


Figure 34: Blue native 2D DIGE individual channels

Isolated mitochondrial complexes from B37C (A), B37N (B), and B37C/Ky21 (C) were labeled with cy3, cy5, and cy2 respectively (overlay of A and B in Fig. 31). The samples were run on a 2D blue native PAGE and imaged. The identifiable complexes and super complexes of the mitochondrial respiratory chain are labeled. All proteins with a 1.5 fold or higher change are numbered and correlate to the protein numbers in table 5 and figure 31.

Figure 34: Blue native 2D DIGE individual channels

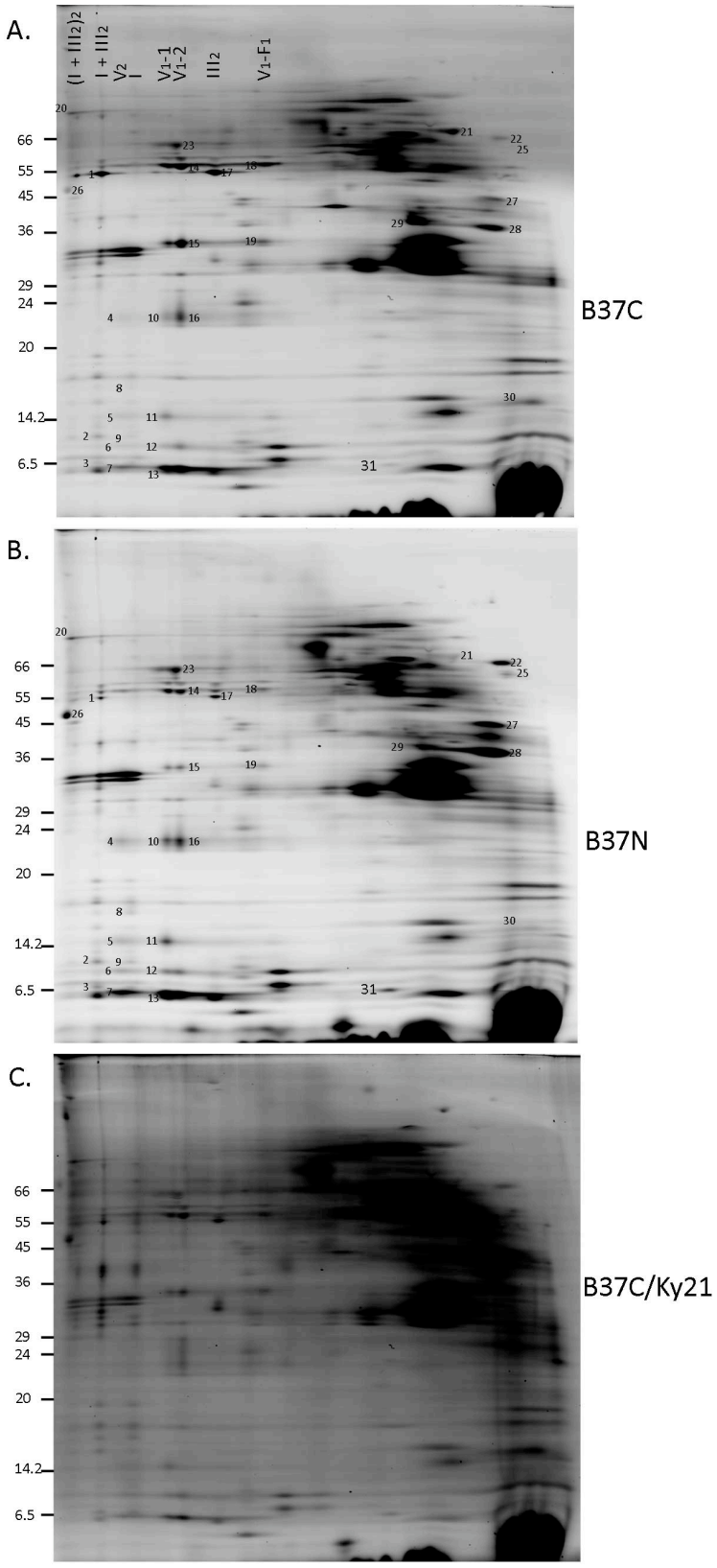


Table 7: 2D Blue native DIGE differential proteins

A summary of all proteins identified to be expressed differentially (1.5 fold or greater) in at least two of the three blue native DIGE biological replicates. This is followed by the corresponding fold change in the dye swap experiment. A positive value in the average fold change column correlates to higher levels in B37N while a negative value means B37C has higher expression. The reciprocal is true for the dye swap fold change column. Hypothetical protein identification is based upon previously published work. Size estimation was calculated according to the Sigma broad range protein standard.

Table 7: 2D blue native DIGE differential proteins

Protein Spot #	Complex	Probable Protein ID	Size (kDa)	Avg. Fold Change B37C v B37N	Dye Swap Fold Change B37N v B37C
1	$I + III_2$	Mito. processing peptidase	55	-4.75 ± 1.19	-11
2	$I + III_2$		13	1.78 ± 0.22	1.34
3	$I + III_2$		7	1.94 ± 0.08	
4	V_2	ATPδ'	23	3.67 ± 1.17	-1.19
5	V_2		15	2.13 ± 0.14	-1.73
6	V_2		10	2.4 ± 0.49	-1.58
7	V_2	ATP9	6	1.92 ± 0.23	-1.61
8	I		16	2.34 ± 0.24	1.06
9	I	Same as #2	13	2.68 ± 1.07	1.34
10	V_1 - 1	ATPδ'	23	2.19 ± 0.55	1.23
11	V_1 - 1	same as #5	15	1.73 ± 0.21	-1.43
12	V_1 - 1	same as #6	10	1.89 ± 0.38	-1.18
13	V_1 - 1	ATPε	5	1.85 ± 0.29	
14	V_1 - 2	ATPα	58	-1.99 ± 0.67	-2.26
15	V_1 - 2	ATPγ	35	-2.66 ± 1.54	-2.44
16	V_1 - 2	ATPδ'	23	2.14 ± 0.46	1.12
17	III_2	Mito. processing peptidase	55	-6.68 ± 2.49	-11
18	V - F_1	ATPα	58	-2.65 ± 0.4	-2.69
19	V - F_1	ATPγ	35	-2.33 ± 0.2	-1.89
20	-		85	1.57 ± 0.05	-2.11
21	-		70	-2.16 ± 0.46	
22	-		66	5 ± 1.67	
23	-		64	1.76 ± 0.19	1.47
24	-		62	2.83 ± 1	-2.47
25	-		60	5.4 ± 2.46	
26	-		47	2.19 ± 0.19	
27	-		45	2.21 ± 0.26	
28	-		38	3.21 ± 0.52	
29	-		38	-2.15 ± 0.3	
30	-		15	-2.19 ± 0.5	
31	-		9	1.92 ± 0.33	-2.11

Figure 35: Blue native 2D DIGE B37N and B37C dye swap overlay

Isolated mitochondrial complexes from B37N and B37C were labeled with cy3 (green) and cy5 (red) respectively. The samples were run on a 2D blue native/SDS gel, imaged, and overlayed; the identifiable complexes and super complexes of the mitochondrial respiratory chain are labeled. The proteins with a 1.5 fold or higher change are numbered and correlate to the protein numbers in table 5 and figures 31 and 32.

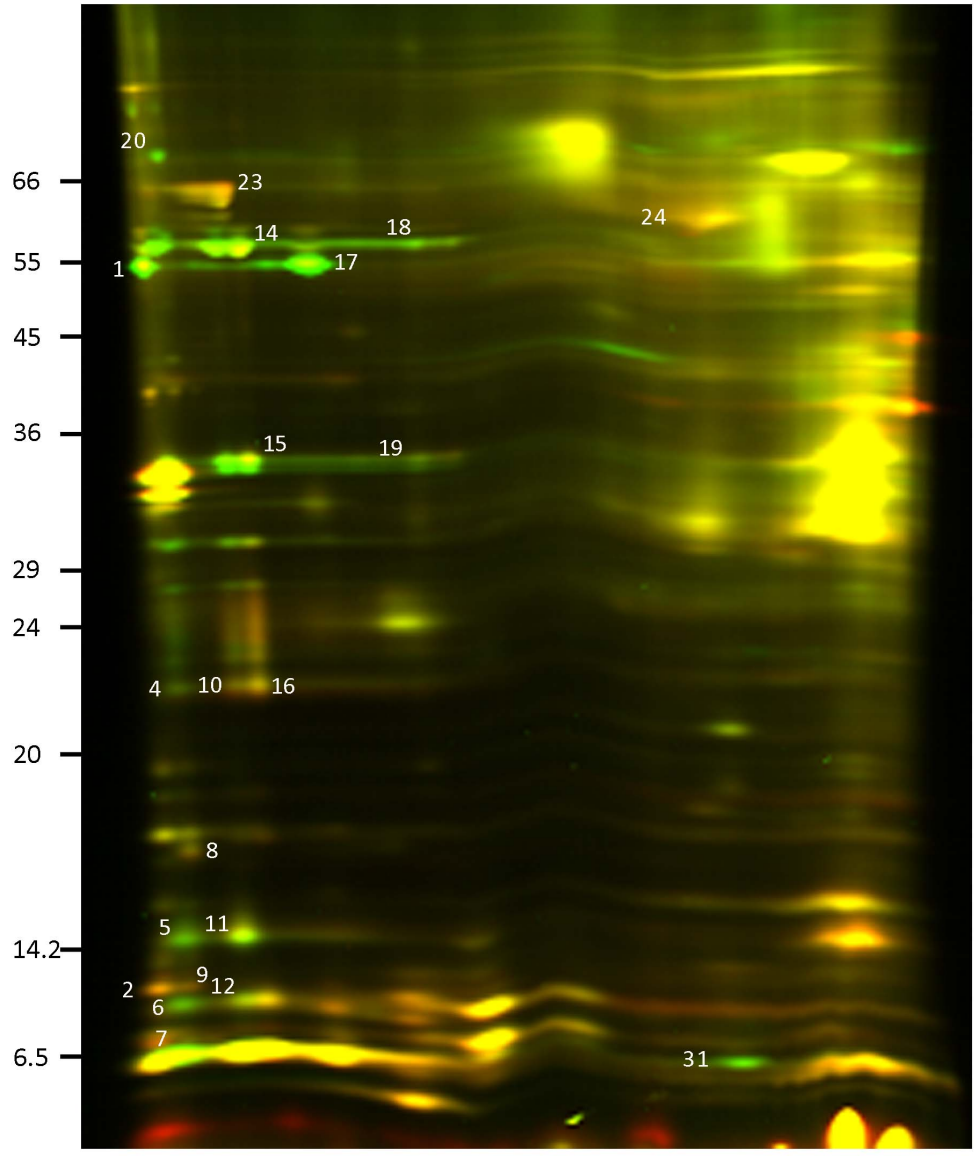


Figure 36: Blue native 2D DIGE B37N and B37C dye swap individual channels

Isolated mitochondrial complexes from B37N (A) and B37C (B) were labeled with cy3 and cy5 respectively. The samples were run on a 2D blue native PAGE and imaged. The identifiable complexes and super complexes of the mitochondrial respiratory chain are labeled. All proteins with a 1.5 fold or higher change are numbered and correlate to the protein numbers in table 5 and figures 31-33.

Figure 36: Blue native 2D DIGE B37N and B37C dye swap individual channels

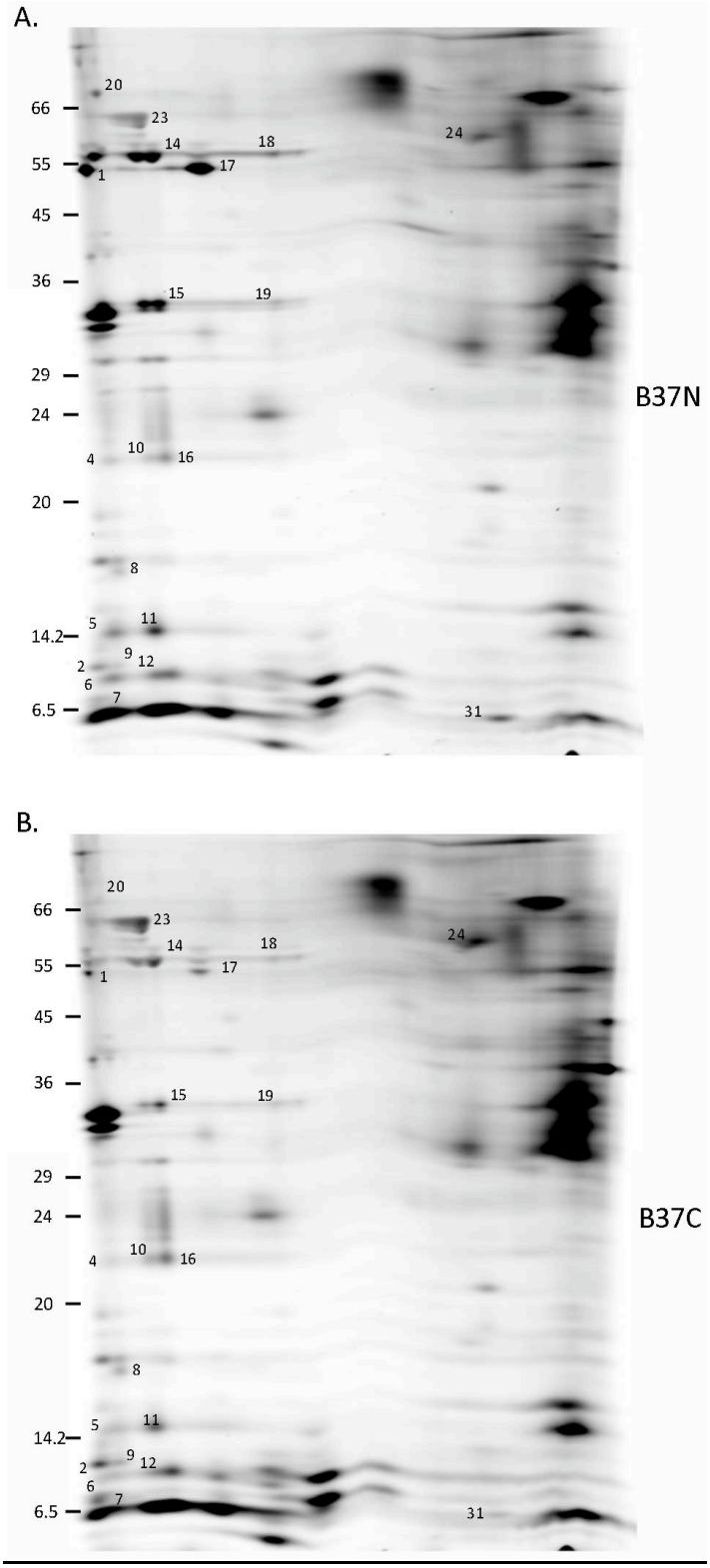


Figure 37: 2D BN-PAGE immunoblots

Mitochondrial respiratory chain complexes were isolated and separated on a 2D blue native gel for B37N (A), B37C (B), and B37C/Ky21 (C) meiotic tassel mitochondria. The gels were blotted and hybridized to ATPα, ATPγ, ATP6, ATPδ', ATP9, and COX2 antibodies. The four forms of the ATP synthase complex observed (V_2, V_1-1, V_1-2, and V-F_1) are labeled at the top of the image.

Figure 37: 2D BN-PAGE immunoblots

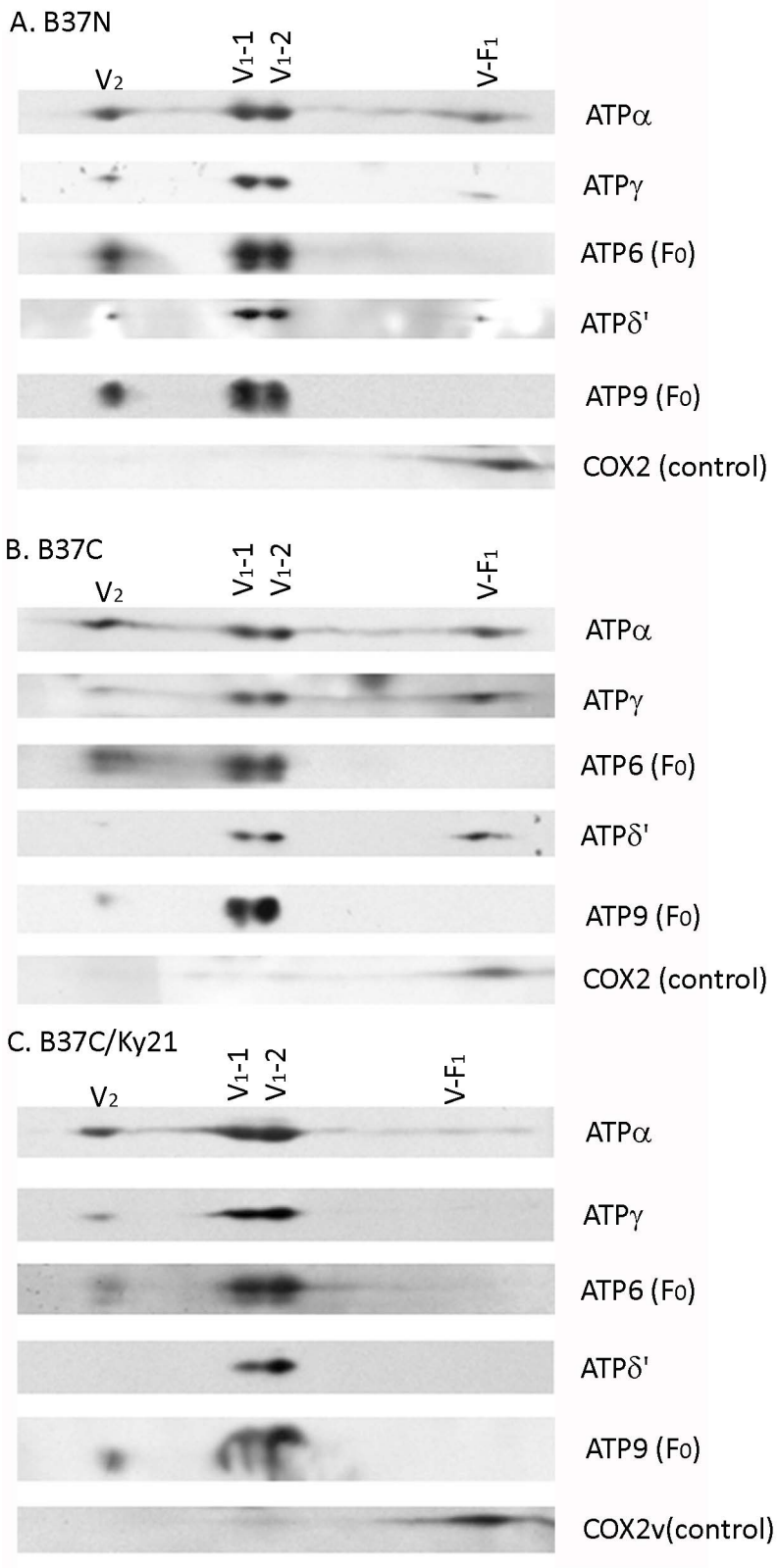


Table 8: 2D BN-PAGE immunoblot analysis

The levels of five ATP synthase proteins were measured in three samples B37N, B37C, and B37C/Ky21. All levels were compared to B37C to calculate the fold change. A 2 fold or larger change was considered to be significant. The proteins with significantly lower expression than B37C are highlighted in green and the proteins with significantly higher expression than B37C are highlighted in blue.

Fold Change compared to C					
Sample	**subunit**	**V_2**	**V_1-1**	**V_1-2**	**V-F_1**
B37N	ATPα	-1.03	1.3	1.53	-1.25
B37C	ATPα	1	1	1	1
B37C/Ky21	ATPα	-1.8	-1.05	1.22	-8.38
B37N	ATPγ	-1.27	1.39	1.12	-3.13
B37C	ATPγ	1	1	1	1
B37C/Ky21	ATPγ	-1.56	1.05	1.14	-44.01
B37N	ATPδ'	9.24	2.21	1.51	-2.77
B37C	ATPδ'	1	1	1	1
B37C/Ky21	ATPδ'		1.29	2.01	-34.4
B37N	ATP6	2.47	-1.17	1.12	
B37C	ATP6	1	1	1	
B37C/Ky21	ATP6	-4.22	-2.22	1.03	
B37N	ATP9	4.75	1.13	-1.86	
B37C	ATP9	1	1	1	
B37C/Ky21	ATP9	1.37	-1.51	-7.69	

6.4 Discussion

There were no proteins identified by the DIGE blue natives that were novel to C, making it improbable that a chimeric protein is being produced and inhibiting the normal function of a respiratory chain complex. However, blue natives only allow for the separation of large protein complexes (100 kDa or larger), so if a smaller complex or an individual protein was involved in causing pollen abortion it would not be seen in this analysis.

Several proteins were seen to be differentially expressed with the DIGE blue natives. There are three proteins differentially expressed in the super complex I +III_2 according to the biological replicates, the largest of these proteins (protein #1) is 55 kDa and is putatively identified as the complex III mitochondrial processing peptidase. This protein was identified to have an average 4.75 fold increase in B37C (Cy3 channel) but was also identified to be 11 fold increased in B37N (Cy3 channel) in the dye swap. This indicates that there is a Cy3 dye preference for this protein and would result in an average Cy3 bias of 7.88 fold and an actual 3.12 fold increase of this protein in B37N mitochondria. The putative mitochondrial processing peptidase is also seen differentially expressed when it is part of complex III_2 (protein #17). A Cy3 dye bias is still observed for this protein (average 8.84 fold), resulting in an actual 2.16 fold increase in B37N. These results need to be confirmed due to the Cy3 bias along with the high standard error within the biological replicates.

The remaining two proteins that were identified as differentially expressed within complex I+III_2 migrate at 13 kDa (#2) and 7kDa (#3), and have slightly higher expression

in N, 1.78 and 1.94 respectively. Neither of these proteins could be confirmed to be differentially expressed with the dye swap, protein #2 was not differentially expressed and protein #3 did not separate out on the gel due to its small size.

The remaining 13 proteins of the 19 proteins identified to be components of a complex are in the ATP synthase complex (complex V). Four of them (#4-7) are part of the complex V dimer (V_2) and show higher levels in B37N. Only protein #4, putatively ATPδ', was not confirmed to be differentially expressed by the dye swap. In the dye swap this protein had no change in expression. There may be a dye bias, due to preferential binding of Cy3 to the protein, which would result in a 1.74 fold increase in B37N. Proteins 4-6 are also identified as having higher expression in N for the complete form of complex V_1 (V_1-1), proteins 10-12 respectively. The fold change was not as significant for these proteins when part of V_1-1 when compared to V_2 which may be the reason for the dye swap not confirming the change in expression. More dye swap replicates should be performed to determine if a difference can be identified. The putative ATPδ' protein had indications of being biased to the Cy3 label, as was seen previously. Although the putative ATP9 protein did not appear to be differentially expressed in V_1-1 another small protein (#13) putatively identified as ATPε was identified as having higher levels in B37N. This could not be confirmed in the dye swap because the protein did not separate from ATP9 on the gel. A second form of the monomer complex V was seen on the blue natives that lacked protein #11 (V_1-2). Within V_1-2 ATPδ', protein 16, was differentially expressed similar to its levels in V_1-1, including the dye bias. However, only two other proteins were identified as altered,

proteins #14 and 15, putatively identified as ATPα and ATPγ respectively. Both of these proteins have significantly higher expression in B37C, which contradicts the slightly higher expression in N seen for ATPδ'. The dye swap experiment indicates that the expression differences for ATPα and ATPγ can be attributed purely to a Cy3 dye bias and not to an actual expression difference.

The last form of complex V observed is the matrix portion V-F_1. The putative ATP alpha and gamma (proteins #18 and 19) once again have higher levels in C that appears to be due to a dye bias according to the dye swap, with no actual change in expression.

There remain 12 proteins (#20-31) which are differentially expressed, that do not appear to be part of a large complex. A majority of these proteins cannot be confirmed in the dye swap experiment because they are not observed either because they do not occur (26) or are not present in the gel separation (3, 13, 21, 22, 25, 27, 28, 29, and 30). Since no identity can be assigned to these proteins no conclusions can be drawn about their role in pollen abortion. Most likely these proteins represent stress response in CMS-C.

Overall it appears that complex V is the only complex significantly affected by CMS-C. It appears to form fewer V_2 and V_1-1 complexes in CMS-C mitochondria. However, this is difficult to confirm because several proteins have a dye bias which complicates the end result.

In order to confirm the differences seen in B37N and B37C for the ATP synthase complex in the BN-DIGE results, BN westerns were performed using antibodies raised to

different components of the ATP synthase complex (Fig. 37). A higher cut-off for significance, two-fold or higher, was used for the westerns since no biological replicates were performed. Blue native westerns were also performed on B37C/Ky21 samples in order to determine if the ATP synthase complex changes with the introduction of *Rf4*.

The same four ATP synthase complexes as observed in the BN-DIGE experiment (V_2, V_1-1, V_1-2, and F_1) were observed in the BN westerns. Four components of the dimer complex V (V_2) were identified to have higher expression in B37N according to the BN-DIGE results. Two of those proteins were putatively identified as ATPδ' and ATP9 (Table 5), with the remaining two being unknown. When the BN gel blots were probed with antibodies these results were confirmed. Decreases of ATPδ' and ATP9 levels in B37C were observed when compared to B37N (table 8). ATPα and ATPγ had no significant changes in expression between B37N and B37C according to the BN westerns which also confirms the results from the BN-DIGE experiments.

In some cases, a direct correlation between the BN-DIGE and those from the immunoblots could not be made, because a significant difference for ATPα and ATPγ could be masked due to preferential binding of one CyDye over another. The BN westerns also identified a significant increase in B37N ATP6 (at ~30 kDa) levels for complex V_2. This result was not observed in the BN-DIGE, although the two proteins observed at ~30 kDa are very weak. The unidentified 10 and 15 kDa proteins differentially expressed in the BN-DIGE (Fig 33 and 34) could not be confirmed because no antibodies were available for those proteins.

It would be expected that all ATP synthase subunits should show a decrease if the V_2 complex was reduced in B37C (not just ATPδ', ATP6, and ATP9). This is especially true since it has been shown that V_2 formation is due to the binding of the F_0 subunits which include ATP6 and ATP9, along with ATP4 (Dienhart et al., 2002).

Of the four V_1-1 proteins identified as having higher levels in B37N as compared with B37C according to the BN-DIGE results, only ATPδ' could be tested with the BN westerns. Both the BN-DIGE and westerns showed ~2.2 fold higher levels of ATPδ' in the B37N sample relative to B37C. The remaining antibodies tested on the western blots showed no significant difference, confirming the DIGE results. It has been shown that the ATP synthase complex can assemble without the presence of ATP δ', although this can cause a proton leak and mitochondrial degradation (Mueller, 2000).

No differences were seen for the V_1-2 complex in the BN westerns. The ATPα and ATPγ had significant differences in the BN-DIGE results but that is probably due to a dye bias and not an actual difference in the samples. ATPδ' was also differentially expressed in the BN-DIGE results however this expression change was not confirmed by the western results.

A dye bias appears to account for the only proteins identified as differentially expressed for F_1 in the BN-DIGE results, ATPα and ATPγ. However, the other components were difficult to identify. The western BN showed that both ATPγ and ATPδ' were higher in B37C, by 3.13 and 2.77 fold respectively. The other F_1 subunit tested on the western blue natives, ATPα, showed no significant difference between

B37N and B37C. This contradiction in results may just be a result of the F_1 subunit in the BN-DIGE experiments producing too weak of a signal for accurate quantification in combination with the dye bias observed for some of the proteins. The increased levels of the F_1 subunit in B37C may indicate a lack of enough F_0 subunit for formation of a complete ATP synthase complex, which correlates with the decrease observed in V_2 formation.

If the decrease in the V_2 complex and the increase in F_1 levels in B37C is the result of the overall decrease in ATP6 and ATP9 levels as observed in the 1D westerns, then it would be expected that introduction of *Rf4* in B37C/Ky21 samples would restore the V_2 and F_1 levels to that of B37N. Interestingly the only difference observed for V_2 when B37C was compared to B37C/Ky21 was a decrease in B37C/Ky21 ATP6 levels, comparable results were observed for V_1-1. V_1-2 had similar levels of ATP6 between B37C and B37C/Ky21 but had a decrease of ATP9 for B37C/Ky21 and an increase in ATP delta'. The apparent decrease in ATP9 could be due to it not separating from the migration front and not an actual difference. Although B37C/Ky21 did not give the anticipated results for V_2, it did show a dramatic decrease of F_1 levels for all antibodies.

It appears that the introduction of *Rf4* does not have an overall effect on the V_2:V_1-1:V_1-2 ratio in the CMS-C cytotype but has a dramatic effect on the amount of free F_1 present. The other affect seen is the increase of ATPδ' in the V_1-2 complex. If ATP9 levels are down then it could cause a decrease of F_0 formation and therefore an increase in the amount of free F_1. Furthermore, if ATP9 is reduced there could be formation of

incomplete F_0 subunits, comprised of less than 10-14 ATP9 subunits. These reduced F_0 subunits might still be able to form into an ATP synthase complex but cause ATPδ' to bind less efficiently generating a proton leak. The increase of ATP9 seen for B37C/Ky21 in the 1D western is then allowing for more F_0 and ATP synthase to be formed without changing the ratio of the different forms of the complex. To determine if this is the case F_0 should be isolated and run out on a gel to determine if there is variation in the sizes of the ATP9 oligomers, indicative of incomplete F_0 subunits.

Another possibility is that the decrease in F_1 is the result of hybrid vigor since the CMS-C cytotype tested was an inbred line (B37C) and the CMS-C/*Rf4* is a hybrid (B37C/Ky21). To determine if *Rf4* or hybrid vigor is the cause of decreasing F_1 levels, the CMS-C cytotype should be tested as a hybrid that does not contain *Rf4* such as B37C/Mo17.

Chapter 7: IDENTIFICATION OF CANDIDATE *RF4* GENES

7.1 Introduction

Previous work on *Rf4* mapped it to the short arm of chromosome 8, 1.5±1 cM distal to the marker npi114a and 4.5±1.8 cM distal from npi220a (Sisco, 1991). The map location of *Rf4* can be used with the recently published B73 maize genome sequence to identify candidate *Rf4* genes between the markers.

Further filtering of the candidate genes can be performed using criteria from other identified *Rf* genes. First, in order for the *Rf4* gene to function it must be targeted to the mitochondria. Mitochondrial targeting sequences have been identified for several genes, which allowed the development of algorithms to predict mitochondrial targeting from the peptide sequence (Claros and Vincens, 1996; Emanuelsson et al., 2000; Small et al., 2004). Furthermore, the *Rf* genes in other CMS types have been identified, and in nearly all cases, they contain an RNA-binding PPR motifs (Saha et al., 2007). PPR proteins are a large class of nuclear encoded proteins found within plants. Most PPR proteins are targeted to either the mitochondria or chloroplasts and play a role in RNA processing, stability, and translation (reviewed by (Saha et al., 2007; Woodson and Chory, 2008).

7.2 Methods

7.2.1 Identification of Candidate *Rf4* Genes

All annotated genes lying near the end of chromosome 8 (Bin 8.00; 1-2082500 bp) were extracted from the maize sequence database (www.maizesequence.org; release 3a.50). The translated genes were then analyzed for mitochondrial targeting sequences using three different programs; MitoPROT (http://ihg2.helmholtz-muenchen.de/ihg/mitoprot.html; Claros and Vincens, 1996), Predotar (http://urgi.versailles.inra.fr/predotar/predotar.html ; Small et al., 2004), and TargetP (http://www.cbs.dtu.dk/services/TargetP; Emanuelsson et al., 2000). The genes that were predicted to contain a mitochondrial targeting sequence by at least two of the three programs with a 0.5 probability or higher, were then analyzed for PPR motifs using TPRpred (http://toolkit.tuebingen.mpg.de/tprpred; Karpenahalli et al., 2007).

7.2.2 Sequencing of a Candidate *Rf4* Gene

PCR Amplification of the gene was performed on three non-restoring *rf4* (Mo17, B37, B73) and three restoring *Rf4* (Oh43, A619, Ky21) lines using primers designed to flank the coding region of the candidate gene (Appendix 1). Each PCR product was then gel purified using the Promega Wizard SV gel and PCR clean-up system (cat# A9282), and sequenced at the MU DNA core using the primers from PCR amplification along with internal primers every ~500 bases (Appendix1). The sequencing products for each genotype were than combined and compared to each other for SNPs using CLUSTALW

(http://align.genome.jp/). Further analysis was performed on the sequences by combining the exons, translating them, and comparing them again using CLUSTALW.

7.3 Results

Analysis of all annotated genes in Bin 8.00 at maizesequence.org identified ten genes that code for proteins which contain mitochondrial targeting motifs (Table 9). Eight PPR motifs were identified within one of the mitochondrially targeted proteins in Bin 8.00, protein AC205241.2_FG021 (Fig 38). There were no PPR motifs within the remaining nine mitochondrially targeted proteins within Bin 8.00.

Sequencing of AC205241.2_FG021 identified nucleotide changes in all tested genomes. Most polymorphisms occurred within the introns. However, as seen in figure 37 there were several polymorphisms, as compared to B73, within exons resulting in amino acid changes. There were three amino acid codons altered within the last PPR motif of B37. Mo17 had six amino acid changes, three tandem changes towards the N terminus before the first PPR motif and three within the last PPR motif. A619 had two amino acids that could not be determined due to poor sequencing results and another five amino acids that are altered as compared to B73. Four of A619 amino acid changes are towards the N terminus and one is within the last PPR motif. Ky21 had three polymorphic amino acids all within the last PPR motif. Oh43 had one polymorphic amino acid before the first PPR motif and two unknown amino acids due to poor sequencing results. As can be observed in figure 37 and 38, there were no

polymorphisms that were conserved in all the *Rf4* lines (A619, Ky21, Oh43) that varied from the *rf4* lines (B73, B37, and Mo17).

Figure 38: Comparison of AC205241.2_FG021 protein sequence

Multiple alignment of the predicted protein sequence corresponding to AC205241.2_FG021 was performed on seven sequences: the published sequence from B73, three sequencing results from *rf4* lines B73, B37, and Mo17, along with three sequencing results from *Rf4* lines Ky21, A619, and Oh43. Above the amino acid sequence the identified PPR motifs are denoted with asterisks. All amino acids that vary from the published sequence are highlighted.

Figure 38: Comparison of AC205241.2_FG021 protein sequence

```
PPR Motif
published-rf4      MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ
B73-rf4            MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ
B37-rf4            MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ
Mo17-rf4           MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ
Ky21-Rf4           MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ
A619-Rf4           MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ
Oh43-Rf4           MWALRRAGNPLFRAQKVASAHVCASLEVLLAADAKNVEEHHEADCQKMCCCHKPRSPAFQ

PPR Motif
published-rf4      LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPAAVQNDDGLASEDSSDEDAA
B73-rf4            LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPAAVQNDDGLASEDSSDEDAA
B37-rf4            LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPAAVQNDDGLASEDSSDEDAA
Mo17-rf4           LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPAAVQNDDGLASEDSSDEDAA
Ky21-Rf4           LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPAAVQNDDGLASEDSSDEDAA
A619-Rf4           LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPPAVQNDDGLASEDSSDEDAA
Oh43-Rf4           LSFSSGQFACSRGFSSQPGANSGDKADELEDGFSDLEVPPAAVQNDDGLASEDSSDEDAA

PPR Motif
published-rf4      DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN
B73-rf4            DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN
B37-rf4            DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN
Mo17-rf4           DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN
Ky21-Rf4           DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN
A619-Rf4           DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN
Oh43-Rf4           DEFGLPDVDADAKPEKEHTRRLSDSILLKTLSDSILLKTVLEAPRHQVTSALEKWAKDGN

PPR Motif
published-rf4      AFDRGELYYVLLNLRKRHWYSKALELVEWVQKSQLFEFVERDYAAHLDLTAKVYGLHKAE
B73-rf4            AFDRGELYYVLLNLRKRHWYSKALELVEWVQKSQLFEFVERDYAAHLDLTAKVYGLHKAE
B37-rf4            AFDRGELYYVLLNLRKRHWYSKALELVEWVQKSQLFEFVERDYAAHLDLTAKVYGLHKAE
Mo17-rf4           AFDRGELYYVLLNLRKRHWYSKALELVEWVKKSQLFELVERDYAAHLDLTAKVVWSHKAE
Ky21-Rf4           AFDRGELYYVLLNLRKRHWYSKALELVEWVQKSQLFEFVERDYAAHLDLTAKVYGLHKAE
A619-Rf4           AFDRGELYYVLLNLRKRHWYSKALELVEWVQKSQLFEFVERDYAAHLDLTAKVLWSHKAE
Oh43-Rf4           AFDRGELYYVLLNLRKRHWYSKALELVEWVQKSQLFEFVERDYAAHLDLTAKVYGRHKAE

PPR Motif                        ************************************
published-rf4      QTIEKIPASCRGEIVYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR
B73-rf4            QTIEKIPASCRGEIVYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR
B37-rf4            QTIEKIPASCRGEIVYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR
Mo17-rf4           QTIEKIPASCRGEIVYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR
Ky21-Rf4           QTIEKIPASCRGEIVYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR
A619-Rf4           QTIEKIPASCRGEIVYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR
Oh43-Rf4           QTIEKIPASCRGEIXYRTLLANCVSASNIKKSEQVFNRMKDLGFPVTQFSFNQLLLLYKR

PPR Motif                                ************************************
published-rf4      LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
B73-rf4            LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
B37-rf4            LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
Mo17-rf4           LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
Ky21-Rf4           LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
A619-Rf4           LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
Oh43-Rf4           LDRKKIADVLAMMEKEDVKPSLFTYKILVDAKGSVGDIEAMEKVIESMEKDSIEPDLTFN
```

```
PPR Motif          *********************************
published-rf4      ATIARHYIFYGQREKAEALLKAMEGDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED
B73-rf4            ATIARHYIFYGQREKAEALLKAMEGDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED
B37-rf4            ATIARHYIFYGQREKAEALLKAMEGDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED
Mo17-rf4           ATIARHYIFYGQREKAEALLKAMEGDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED
Ky21-Rf4           ATIARHYIFYGQREKAEALLKAMEGDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED
A619-Rf4           ATIARHYIFYGQREKAEALLKAMXXDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED
Oh43-Rf4           ATIARHYIFYGQREKAEALLKAMEGDDIQKNRAACKTLLPLHAFLGNSDAVERIWKACED

PPR Motif              ******************************************************
published-rf4      NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE
B73-rf4            NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE
B37-rf4            NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE
Mo17-rf4           NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE
Ky21-Rf4           NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE
A619-Rf4           NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE
Oh43-Rf4           NTRVAESVSAIEAFGKLGDVEKAEKVFEDMLVHWKTLSSKFYNALLKVYADHNLLDKGKE

PPR Motif          ********************************************** **********
published-rf4      LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY
B73-rf4            LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY
B37-rf4            LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY
Mo17-rf4           LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY
Ky21-Rf4           LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY
A619-Rf4           LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY
Oh43-Rf4           LAKRMDENRVRFGAPTLDALVKLYVEAGEVEKAESLVHKLSIQNHIKPIYNTYMTLLDSY

PPR Motif          ******************************************************
published-rf4      SKKGDVRNSEKVFNKLRQSGYTGRIRMYQVLLHAYVRAKAPAYGFRERMKADNIFPNSVV
B73-rf4            SKKGDVRNSEKVFNKLRQSGYTGRIRMYQVLLHAYVRAKAPAYGFRERMKADNIFPNSVV
B37-rf4            SKKGDVRNSEKVFNKLRQSGYTGRIRMYQSLLHAYVHAKAPAYGFRERMKADNIFPNSAV
Mo17-rf4           SKKGDVRNSEKVFNKLRQSGYTGRIRMYQSLLRAYVHAKAPAYGFRERMKADNIFPNSVV
Ky21-Rf4           SKKGDVRNSEKVFNKLRQSGYTGRIRMYQSLLHAYVHAKAPAYGFRERMKADNIFPNSAV
A619-Rf4           SKKGDVRNSEKVFNKLRQSGYTGRIRMYQVLLHAYVRAKAPAYGFRERMKADNIFPNSAV
Oh43-Rf4           SKKGDVRNSEKVFNKLRQSGYTGRIRMYQVLLHAYVXAKAPAYGFRERMKADNIFPNSVV

PPR Motif
published-rf4      STLIAATDPFVKKKSISDLLD
B73-rf4            STLIAATDPFVKKKSISDLLD
B37-rf4            STLIAATDPFVKKKSISDLLD
Mo17-rf4           STLIAATDPFVKKKSISDLLD
Ky21-Rf4           STLIAATDPFVKKKSISDLLD
A619-Rf4           STLIAATDPFVKKKSISDLLD
Oh43-Rf4           STLIAAXDPFVKKKSISDLLD
```

7.4 Discussion

Sisco (1991) previously mapped *Rf4* to the short arm of chromosome 8, 1.5±1 cM distal to the marker npi114a and 4.5±1.8 cM distal from npi220a. Furthermore, the IBM genetic map places *Rf4* proximal to the marker si486079a04b. The maize sequence database (maizesequence.org) does not have the marker npi114a placed on the physical map however npi220a is said to be located at 2,07,7600 to 2,082,500 bp and si486079a04b at 1,925,700 to 1,930,600 bp.

In order to ensure that any potential candidate gene was not missed, all of the annotated genes from the tip of the short arm of chromosome 8 to npi220a (2,082,500 bp) were analyzed for mitochondrial targeting sequences. This identified 10 putative mitochondrially targeted genes (Table 9), two of which (AC193967.2_FG003, and AC193967_FG011) are located between the markers si486079a04 and npi220a.

Since all previously identified restorer genes are PPR proteins except for maize Rf2, which is an aldehyde dehydrogenase and has roles in pollen development in non-CMS plants, the mitochondrially targeted genes identified were analyzed for a PPR motif using TPRpred. Only one gene was identified to be a PPR protein, AC205241.2_FG021. Although this gene is located proximal to si486079a04, it was considered to be the best candidate for *Rf4* due to the PPR motifs.

This gene was further analyzed for any SNP's or indels within its coding region by comparing three restorer lines (A619, Ky21, and Oh43) and three non-restorer lines

(B37, B73, and Mo17). Sequencing of AC205241.2_FG021 in B73 had 100% identity to the published sequence indicating that the primers were amplifying the correct gene. When the sequences were compared to each other no conserved polymorphisms were observed among the *Rf4* lines (Appendix 3). The majority of the polymorphisms present were located within the introns, although every genome did have polymorphisms present within the exons. To determine if the polymorphisms altered the predicted amino acid sequence, the gene was translated for each genome and compared (Fig. 38). Although a few amino acids do change, they do not seem to be conserved among the *Rf4* lines as compared to the *rf4* lines. This means that most likely AC205241.2_FG021 is not *Rf4*. *Rf4* is possibly one of the other mitochondrially targeted proteins, a gene not annotated at maizesequence.org, or it may not be present in the B73 genome.

Chapter 8: CONCLUSIONS

Two previous studies have investigated the cause of CMS-C pollen abortion. The first study tried to identify any protein expression differences in the CMS-C genome as compared to B37N using *in organello* protein synthesis (Forde and Leaver, 1980). Forde's study showed that CMS-C mitochondria lacked the expression of a 15.5kDa protein found in B37N. Conversely, a 17.5kDa protein was present in CMS-C but not found in B37N. However, introduction of *Rf4* to the CMS-C cytotype did not alter its protein profile. This makes it unlikely that the proteins are involved in pollen abortion within CMS-C.

Subsequent research was performed by Dewey *et al.* (1991) to determine if any rearrangements occur within several mitochondrial genes in CMS-C. These studies identified three genes with an altered RNA banding pattern in CMS-C (*atp9*, *atp6*, and *cox2*). The rearrangements occurred in the upstream sequence of all three genes and in the extended ORFs of *atp6* and *cox2* (Fig. 3). However, restoration of fertility by *Rf4* did not appear to alter these transcripts in seedling shoots. Further analysis in the Dewey *et al.* (1991) paper determined that COX2 protein levels were unaffected by the chimeric transcript. Furthermore, no chimeric COX2 protein was found to be synthesized in CMS-C.

The research presented in this paper utilized the recently sequenced CMS-C mitochondrial genome and the development of newer methodologies in an attempt to

elucidate the molecular basis of CMS-C pollen abortion. Based upon the characteristics of CMS-associated genes in other organisms, 28 ORFs were identified as potential candidates for causing pollen sterility in CMS-C (Fig 4; Tables 1 and 2). Twenty of these ORFs were tested for expression on RNA blots. None of the tested ORFs had expression unique to CMS-C or that was altered in the presence of Rf4, which would be expected of a gene causing CMS. The eight ORFs that were not tested were either intronic to a mitochondrial gene in CMS-C and B37N or were not within 2kb of a known promoter so they are unlikely to be transcribed. This means that it is unlikely that rearrangements have caused the expression of a new gene in CMS-C, unless the gene encodes a protein smaller than 50 amino acids or it does not have any of the properties of other known CMS causing genes.

Bioinformatics analysis of the CMS-C genome also found the three functional chimeric genes (*atp9; atp9-2* in this research*, atp6*, and *cox2*) that were previously identified by Dewey *et al.* (1991; Fig 3). Interestingly a second copy of *atp9* (denoted *atp9-1* in the research) was found in the CMS-C genome that had no rearrangements. RNA blot analysis of *atp9-2*, *atp6*, and *cox2* found that, there were no tassel-specific modifications of the CMS-C transcripts in the presence of *Rf4* (Fig. 17-20). Relative real time PCR also showed that *Rf4* did not significantly alter the RNA levels for the *atp9* or *atp6* transcripts (Figs. 21-24).

Interestingly both the RNA blot analysis and real time PCR confirm that CMS-C preferentially uses the chimeric *atp9-2* gene over the non-rearranged *atp9-1*. This

change is potentially due to a 5 bp insertion identified 762 nucleotides upstream of the start codon, or a SNP 221 bases upstream of the start codon (Table 4). Although PCR was able to amplify edited transcripts of *atp9-1,* their levels are so low that they were not observed on RNA blots and real-time analysis determines them to have at least ~20 fold lower levels in CMS-C as compared to B37N.

With no novel transcript identified as a probable cause of CMS-C, total mitochondrial proteins in CMS-C were analyzed in order to determine if there is any novel protein present. Preliminary experiments run using 2D IEF/SDS DIGE did not identify any novel protein in CMS-C mitochondria. However, it was discovered that a few proteins had decreased expression within CMS-C tassels as compared to B37N. MS data was obtained for some of the differentially expressed proteins and they were discovered to be nuclear encoded components of the ATP synthase complex. This indicates that whatever is causing pollen abortion may be affecting the ATP synthase complex. Although no novel protein was observed on the 2-D DIGE gels, one could still be present. The CMS-causing protein could be masked by another protein, or it could be hydrophobic and not resolved on the first dimension IEF strip.

With no novel transcript or protein observed in CMS-C, further analysis was performed on the chimeric C-*atp6* and C-*atp9-2* genes. The C-*atp6* gene could be producing a chimeric protein from the extended ORF containing the N terminus of ATP9. The C-*atp6* extended ORF could also be retained after translation, generating a chimeric ATP6 protein. There is also the possibility that the 4kb transcript from C-*atp9-2* is not

translated as efficiently as the N-atp9 transcripts. This could reduce the levels of ATP9 protein in the plant. Hybridizations with antibodies raised against the N-terminus of ATP9 or ATP6 did not identify any novel protein, decreasing the probability of a chimeric protein being produced from these transcripts. However a tassel-specific decrease in ATP9 and ATP6 levels was observed.

In order to determine how the ATP synthase complex is being affected, and if a novel protein is associating with the complex, 2D Blue Native/SDS gels were performed. The results did not identify any novel protein associating with ATP synthase in CMS-C but there were alterations with the observed forms of the ATP synthase complex. The most notable result found in CMS-C was an increase within the tassel of free F_1, a subcomplex of ATP synthase, that is then brought back down below B37N levels with the introduction of *Rf4*.

CMS-C appears to be a unique system for pollen abortion. No CMS-associated transcript is observed which is characteristic of other CMS systems. Instead there appears to be a tassel-specific decrease in CMS-C of ATP6 and ATP9 according to SDS gels and an increase in free F_1 based upon blue native gels. With the introduction of *Rf4* ATP6 and ATP9 levels go up to normal levels and the amount of free F_1 decreases. This may be due to a secondary effect of a yet unidentified protein. However, no novel protein was observed on the IEF/SDS DIGE or BN-DIGE experiments.

In the absence of evidence for expression of a “toxic” novel chimeric protein, it is more likely that the cause of CMS-C is that the large 4kb *atp9-2* transcript is not

translated efficiently. In non-tassel tissue, mitochondria are able to compensate, but in tassels (especially the tapetum layer) demand increases and mitochondria cannot compensate. ATP9 levels decrease, which causes a reduced assembly of the F_0 subunit, thereby reducing ATP6 and other F_0 component levels. This in turn would increase free F_1 present because there is not enough F_0 for assembly of the whole ATP synthase complex. The mitochondria would not produce enough ATP which would eventually lead to premature cell death in the tapetum. Further work will need to be done to determine if the decrease in ATP9 and ATP6 levels in tassels cause the pollen abortion or are an effect of the pollen abortion.

Appendix 1: Primers used for all experiments

Chapter3: Analysis of Transcripts from Candidate CMS-C Causing Genes

Gene/ORF	Forward	Tm	Reverse	Tm	Product
410	agaagcccacatgcccactc	61	ggtcttcaccttgaaacggcgag	61	271
246a/62	agtgccagtagcaccagttcatcg	62	agaaggggatggctggttcacag	62	207
191c	agacttggattttggtcggcgttg	60	ccatttcttcgagtgtgggtgtcc	60	96
186a	tgcccaggttccacaggtg	61	tatggccgcgttgcccag	62	229
163b	tatcagaggggttctcggcgtg	61	aagtagagcctgatgcgctcc	60	380
147b	aaagaagcccacttacccactc	58	ggaggcttgatgtttccgttcgtg	61	197
59b	atacgttcgtagcagtagctg	55	ttctcagagctgaggtcaacc	57	114
307a	ggcattcctaaggatgaggtcagc	61	gcgttcaaggatgtcagacgactc	61	528
264a	acctttcagggtggagcgtttgca	64	gcacgttagtggtaataggtgcga	60	299
247c/137a	actacggggcgtttacctagccaa	62	tctagaaagccgtatgctttgga	57	342
227a	ccaaacctcgtgtggggctaatcg	63	caacttccgcattcgagggca	61	559
193a	gcgagattggaagaatcgcagagc	61	atacgcatccctcccgtattggtg	61	532
140d	gcgctagttgctcaatccgttgc	62	gccctccaaaggtgcctagag	61	244
126a	ttggatgggcgagagggtgttcca	65	aggctatcagtgcactcttctgg	59	345
120a	aaagcaggtggcattgggatctg	61	gttctgggcaataagtcgacacca	60	235
112b	tggctgaacgaatccattcccac	61	agtcactactgtgtccatgaa	54	234
110a	acgacgactggaacaccctctc	61	cattaaggcgggggcgtatgtgtc	63	204
105a	aaagcgatgctgcccactctg	61	ccaaaccaacaactgagctttcga	59	211
atp6	ttttcccttgcatctcggtca	65	ttggctcctcgtttttatgcaa	64	538
atp9	ctttagcgggagctgctgtcggta	63	tttccttcccaggaagacggagaa	60	223
atp9-1	agtaagcgggcaccgaagtcag	62	gccttgaaaggattggttccca	59	662
atp9-2	ttagtgggcgctgacagttcg	60	gccactggaaagtccactgaacc	61	661
cob	gagatgttgaaggggctggt	65	tccttggtgaatcggtcgaaa	65	763
cox2ex1	ccataggctcctatgctgggag	64	ctacttcattgcccaacaaggcag	63	739
nad7ex2	gcggagtcactgaatgaagtc	60	agaaagcagccctagcctct	60	247
rps13	tggatggaattggaccgaaaa	65	cttgcgagcagtccttgcatt	65	271

Chapter4: Real Time PCR of Chimeric ATP Synthase Genes: Real-Time PCR

Gene/ORF	Forward	Tm	Reverse	Tm	Product
atp6	ttttcccttgcatctcggtca	65	ttggctcctcgtttttatgcaa	64	538
atp9	aaatcaataggtgctggagctg	61	cgaatgagatcagaaaggccatc	64	201
atp9-1	ccagtttcattgaagcaactcc	56	cgacagcagctcccgctaaagc	64	226
atp9-2	tggccatccatcaactgaggca	62	cgacagcagctcccgctaaagc	64	223
rps13	tggatggaattggaccgaaaa	65	cttgcgagcagtccttgcatt	65	271

Chapter 4: Real-Time PCR of Chimeric ATP Synthase Genes: RNA editing

Gene/ORF	Forward	Tm	Reverse	Tm	Product
atp9-1	ccagtttcattgaagcaactcc	56	tttccttcccaggaagacggagaa	60	430
atp9-2	tggccatccatcaactgaggca	62	tttccttcccaggaagacggagaa	60	427
atp6-c	ccagtttcattgaagcaactcc	56	ctggaacggcaactgcctacat	61	1224
atp6-n	tcgagattgtgtgggtgttcag	58	ctggaacggcaactgcctacat	61	1421

Chapter4: Real-Time PCR of Chimeric ATP Synthase Genes: 5’ UTR

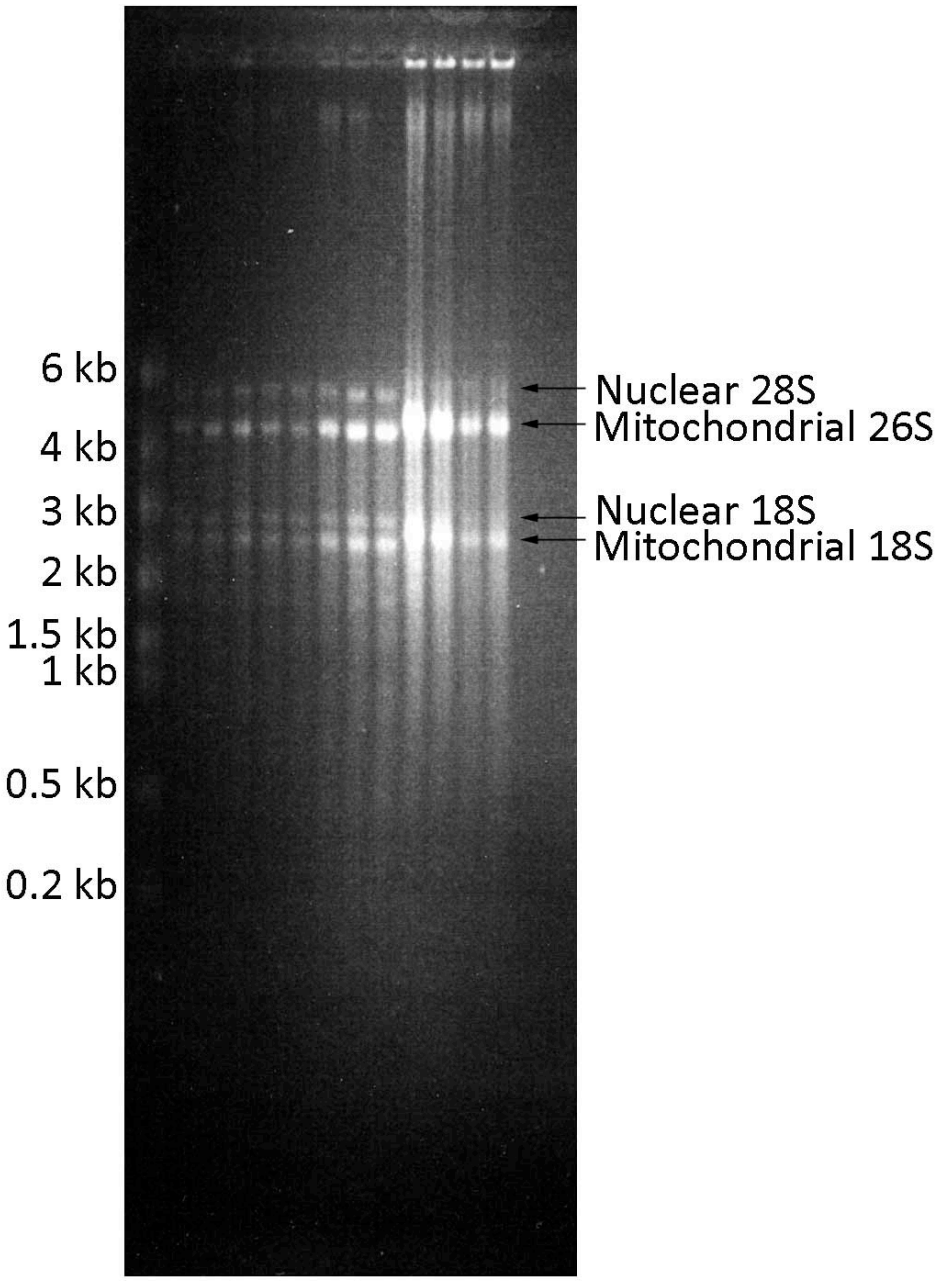
Gene/ORF	Forward	Tm	Reverse	Tm	Product
atp9-1	agcggataaccctgtgctattcc	60	cgacagcagctcccgctaaagc	64	1039
atp6	agcggataaccctgtgctattcc	60	ccggaatcggaatcggaacaccat	62	1078

Chapter 7: Identification of Candidate *Rf4* Genes

Name	Primer Sequence	Tm	Location
AC205241FGP021R4	tcctccacgtctcctcaacagc	57	-40bp from start codon
AC205241G21F605	gcatgctcaggatcgcccttcta	62	+605bp from start codon
AC205241G21F1028	cgaatttggcttgcctgatgtgga	61	+1028bp from start codon
AC205241G21F1512	tgtgaccctggaaacctgctaagg	62	+1512bp from start codon
AC205241G21R2511	tctcacccctacaagatgcaggaa	60	+2511bp from start codon
AC205241G21F2902	agaggcacttctgaaggcaatgga	61	+2902bp from start codon
AC205241G21R3048	cctctatagcagacacggactcag	59	+3048bp from start codon
AC205241FGP021F7	cacggaaacatgatggatctcac	57	+3676bp from start codon

Appendix 2: Ethidium Bromide Stained RNA Gel Image

Image of one of the RNA gels used for blotting and hybridizations stained with ethidium bromide. The sizes of the RNA ladder are labeld to the right and the species of rRNA to the left.



Appendix 3: DNA Sequence Comparison of AC205241.2_FG021

Multiple alignment of the DNA sequences corresponding to AC205241.2_FG021 was performed on seven sequences: the published sequence from B73, three sequencing results from *rf4* lines B73, B37, and Mo17, along with three sequencing results from *Rf4* lines Ky21, A619, and Oh43. The gene exons are highlighted. Nucleotides that are in common are denoted with an asterisk bellow the DNA sequence.

Appendix 3: DNA Sequence Comparison of AC205241.2_FG021

```
published-rf4   AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
B73-rf4         AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
B37-rf4         AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
Mo17-rf4        AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
Ky21-Rf4        AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
A619-Rf4        AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
Oh43-Rf4        AGCGACGCCATGTGGGCTCTTCGCCGAGCTGGCAACCCCCTCAGGTAGTGATGCTCCCGG
                **********************************************************

published-rf4   CCGCTGGTTATCCGACTAGGTCTCTTATTGTCA-CCCCCCCTGCATCCGTCGTTTTAACG
B73-rf4         CCGCTGGTTATCCGACTAGGTCTCTTATTGTCA-CCCCCCCTGCATCCGTCGTTTTAACG
B37-rf4         CCGCTGGTTATCCGACTAGGTCTCTTATTGTCA-CCCCCCCTGCATCCGTCGTTTTAACG
Mo17-rf4        CCGCTGGTTATC-GACTAGGTCTCTTATTGTC--CCCCCCCTGCATCCGTCGTTTTAACG
Ky21-Rf4        CCGCTGGTTATCCGACTAGGTCTCTTATTGTCA-CCCCCCCTGCATCCGTCGTTTTAACG
A619-Rf4        CCGCTGGTTATCCGACTAGGTCTCTTATTGTCACCCCCCCCCTGCATCCGTCGTTTTAACG
Oh43-Rf4        CCGCTGGTTATCCGACTAGGTCTCTTATTGTCA-CCCCCCCTGCATCCGTCGTTTTAACG
                *********** ******************  *************************

published-rf4   GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
B73-rf4         GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
B37-rf4         GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
Mo17-rf4        GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
Ky21-Rf4        GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
A619-Rf4        GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
Oh43-Rf4        GCGGCCTATTTGATCCGCATGGGGATGCTTATTCGCTATGCTTGTTGCGTCGCAGCGCCC
                ************************************************************

published-rf4   TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
B73-rf4         TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
B37-rf4         TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
Mo17-rf4        TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
Ky21-Rf4        TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
A619-Rf4        TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
Oh43-Rf4        TCTGCACCCCGAAATGTTTGATCCCCTGTTAGGCTGTTAGTTGGTCGCGATTGCTTGGAA
                ************************************************************

published-rf4   CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAAATATTTTGTTCAATTGTTCATTAT
B73-rf4         CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAAATATTTTGTTCAATTGTTCATTAT
B37-rf4         CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAAATATTTTGTTCAATTGTTCATTAT
Mo17-rf4        CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAAATATTTTGTTCAATTGTTCATTAT
Ky21-Rf4        CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAAATATTTTGTTCAATTGTTCATTAT
A619-Rf4        CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAAATATTTTGTTCAATTGTTCATTAT
Oh43-Rf4        CACACGGGTATAGCATTTTTTAGGGGTTAAAGTTAATTATTTTGTTCAATTGTTCATTAT
                ********************************** ***********************

published-rf4   CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
B73-rf4         CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
B37-rf4         CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
Mo17-rf4        CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
Ky21-Rf4        CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
A619-Rf4        CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
Oh43-Rf4        CATGTTGAGATTATTGGTATCACTTGACACTATGTTGGATGTATTGTGTTTCTGCGCCGA
                ************************************************************
```

```
published-rf4   TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGTATCTGTT
B73-rf4         TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGTATCTGTT
B37-rf4         TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGCATCTGTT
Mo17-rf4        TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGTATCTGTT
Ky21-Rf4        TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGCATCTGTT
A619-Rf4        TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGTATCTGTT
Oh43-Rf4        TTTAAGGATAGCTATATCTATGTAATTGCCAAGACATAGAGTGAATCATCTGTATCTGTT
                ************************************************** *******

published-rf4   ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTATATATTGTGGGAATTTGCTACTAC
B73-rf4         ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTATATATTGTGGGAATTTGCTACTAC
B37-rf4         ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTACATATTGTGGGAATTTGCTACTAC
Mo17-rf4        ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTATATATTGTGGGAATTTGCTACTAC
Ky21-Rf4        ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTACATATTGTGGGAATTTGCTACTAC
A619-Rf4        ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTATATATTGTGGGAATTTGATACTAC
Oh43-Rf4        ATGATTTCTTCTATTTTATTGTCTATAGTTTGTTTATATATTGTGGGAATTTGCTACTAC
                ********************************** *************** ******

published-rf4   GAGTGCGTCAGATGTTTTTAATGAGAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
B73-rf4         GAGTGCGTCAGATGTTTTTAATGAGAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
B37-rf4         GAGTGCGTCAGATGTTTTTAATGANAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
Mo17-rf4        GAGCGCGTCANATGTTTTTAATGANAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
Ky21-Rf4        GAGTGCGTCAGATGTTTTTAATGAGAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
A619-Rf4        GAGTGCGTCAGATGTTTTTAATGAGAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
Oh43-Rf4        GAGTGCGTCAGATGTTTTTAATGAGAAGAGCAATTTCTGATACTTGTATGTTCAACTTCA
                *** ****** ************* ************************************

pusblished-rf4  CAGTTAATTGCCTGTTCATCTTTCATGTTTTTTTATTAGCATGTGAATGCATGTTTTTCT
B73-rf4         CAGTTAATTGCCTGTTCATCTTTCATGTTTTTTTATTAGCATGTGAATGCATGTTTTTCT
B37-rf4         CAGTTAATTGCCNGTTCATCTTTCATGTTTTTTTATTAGCATGTGAATGCATGTTTTT-T
Mo17-rf4        CAGTTAATTGCCTGTTCATCTTTCATGTTTTTTTATTANCATGTGAATGCATGTTTTTCT
Ky21-Rf4        CAGTTAATTGCCTGTTCATCTTTCATGTTTTTTTATTAGCATGTGAATGCATGTTTTT-T
A619-Rf4        CAGTTAATTGCCTGTTCATCTTTCATGTTTTTTTATTAGCATGTGAATGCATGTTTTTCT
Oh43-Rf4        CAGTTAATTGCCTGTTCATCTTTCATGTTTTTTTATTAGCATGTGAATGCATGTTTTT--
                ************ ************************ ******************

published-rf4   AGCATCATGTGAATGCATGCTCAGGATCGCCCTTCTATTGTATGTTGTGCATCCAATATA
B73-rf4         AGCATCATGTGAATGCATGCTCAGGATCGCCCTTCTATTGTATGTTGTGCATCCAATATA
B37-rf4         ANCATCATGTGAATGCATGCTCAGGATCGCCCTTCTATTGTATGTTGTGCATCCNATATA
Mo17-rf4        ACCATCATGTGAATGCNTGCTCAAGATCNCCCTTCTATTGTATGTTGTGCATCCAATATA
Ky21-Rf4        AGCATCATGTGAATGCATGCTCAGGATCGCCCTTCTATTGTATGTTGTGCATCCAATATA
A619-Rf4        AGCATCATGTGAATGCATGCTCAGGATCGCCCTTCTATTGTATGTTGTGCATCCAATATA
Oh43-Rf4        ----------------------------------------------------CTCATATA
                                                                    *  *****

published-rf4   TTCTAATAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
B73-rf4         TTCTAATAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
B37-rf4         TTCTA-TAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
Mo17-rf4        TTCTAATAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
Ky21-Rf4        TTCTAATAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
A619-Rf4        TTCTAATAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
Oh43-Rf4        TTCTAGTAATTGTCTGCAGATTTGGCTAACTCTTGGTTCTGCTTCAGGTTCCGTGCCCAA
                ***** ******************************************************
```

```
published-rf4   AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
B73-rf4         AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
B37-rf4         AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
Mo17-rf4        AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
Ky21-Rf4        AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
A619-Rf4        AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
Oh43-Rf4        AAAGTTGCAAGTGCCCATGTCTGTGCGAGTCTTGAGGTGCTGCTAGCTGCTGATGCAAAG
                ************************************************************

published-rf4   AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
B73-rf4         AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
B37-rf4         AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
Mo17-rf4        AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
Ky21-Rf4        AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
A619-Rf4        AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
Oh43-Rf4        AATGTAGAAGAGCACCATGAGGCAGACTGCCAGAAAATGTGCTGCTGTCATAAACCAAGG
                ************************************************************

published-rf4   TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
B73-rf4         TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
B37-rf4         TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
Mo17-rf4        TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
Ky21-Rf4        TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
A619-Rf4        TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
Oh43-Rf4        TCACCAGCCTTTCAATTGTCATTTTCATCTGGCCAGTTTGCTTGCAGTAGGGGTTTCTCC
                ************************************************************

published-rf4   TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
B73-rf4         TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
B37-rf4         TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
Mo17-rf4        TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
Ky21-Rf4        TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
A619-Rf4        TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
Oh43-Rf4        TCTCAGCCAGGTGCAAATTCTGGTGACAAAGCTGATGAGTTAGAGGACGGATTTTCTGAT
                ************************************************************

published-rf4   CTCGAGGTTCCACCAGCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
B73-rf4         CTCGAGGTTCCACCAGCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
B37-rf4         CTCGAGGTTCCACCAGCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
Mo17-rf4        CTCGAGGTTCCACCAGCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
Ky21-Rf4        CTCGAGGTTCCACCAGCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
A619-Rf4        CTCGAGGTTCCACCACCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
Oh43-Rf4        CTCGAGGTTCCACCAGCAGCTGTTCAAAACGATGACGGGTTGGCATCTGAAGATAGTTCT
                *************** ************************************************

published-rf4   GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
B73-rf4         GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
B37-rf4         GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
Mo17-rf4        GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
Ky21-Rf4        GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
A619-Rf4        GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
Oh43-Rf4        GATGAGGATGCTGCTGACGAATTTGGCTTGCCTGATGTGGATGCTGATGCAAAACCTGAG
                ************************************************************
```

```
published-rf4   AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
B73-rf4         AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
B37-rf4         AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
Mo17-rf4        AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
Ky21-Rf4        AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
A619-Rf4        AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
Oh43-Rf4        AAGGAGCACACTAGAAGATTATCTGATTCTATTCTCCTCAAGACATTATCTGATTCTATT
                ************************************************************

published-rf4   CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
B73-rf4         CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
B37-rf4         CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
Mo17-rf4        CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
Ky21-Rf4        CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
A619-Rf4        CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
Oh43-Rf4        CTCCTCAAGACTGTGTTGGAAGCTCCGAGGCATCAAGTCACTTCAGCACTTGAGAAATGG
                ************************************************************

published-rf4   GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
B73-rf4         GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
B37-rf4         GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
Mo17-rf4        GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
Ky21-Rf4        GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
A619-Rf4        GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
Oh43-Rf4        GCAAAGGATGGCAATGCATTTGATAGGGGTGAACTCTATTATGTTCTTCTTAATCTTCGA
                ************************************************************

published-rf4   AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
B73-rf4         AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
B37-rf4         AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
Mo17-rf4        AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
Ky21-Rf4        AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
A619-Rf4        AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
Oh43-Rf4        AAGCGGCATTGGTACTCCAAAGCCTTGGAGGTATTTAACATATAATTTACTGAGAGGTAG
                ******************************************************************

published-rf4   TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
B73-rf4         TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
B37-rf4         TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
Mo17-rf4        TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
Ky21-Rf4        TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
A619-Rf4        TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
Oh43-Rf4        TTTTATACTTAAATTGGGTGCAAATATTTAGTATGTGGGATTGGAGGTGCTTAACCAAGA
                ************************************************************

published-rf4   GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
B73-rf4         GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
B37-rf4         GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
Mo17-rf4        GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
Ky21-Rf4        GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
A619-Rf4        GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
Oh43-Rf4        GGTGGATCTATATATTTTTTAAGAACACTTAGTATGCTGGATAATATGTGAAACAAACAA
                ************************************************************
```

```
published-rf4   TATGCATTACTCTTTAGCTTGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGC
B73-rf4         TATGCATTACTCTTTAGCTTGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGC
B37-rf4         TATGCATTACTCTTTAGCTCGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGA
Mo17-rf4        TATGCATTACTCTTTAGCTTGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGC
Ky21-Rf4        TATGCATTACTCTTTAGCTCGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGA
A619-Rf4        TATGCATTACTCTTTAGCTTGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGC
Oh43-Rf4        TATGCATTACTCTTTAGCTTGTAATAGCCTGCTCAATATAAAAATAAGTATTTAGCCAGA
                ******************* ****************************************

published-rf4   AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
B73-rf4         AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
B37-rf4         AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
Mo17-rf4        AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
Ky21-Rf4        AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
A619-Rf4        AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
Oh43-Rf4        AGCTTAGAAGCATGTTTGCTTTGTGACCCTGGAAACCTGCTAAGGCACAGGCAGGTGCTC
                ************************************************************

published-rf4   CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
B73-rf4         CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
B37-rf4         CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
Mo17-rf4        CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
Ky21-Rf4        CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
A619-Rf4        CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
Oh43-Rf4        CCAGGCACATGATTTGCTGAGCCTTTGGGGTCTGAGTTACTGCCTTGCTGCTCACCTGCC
                ************************************************************

published-rf4   TGGCAGGCAGCATCCAAACAGGCCCCTAATGCTGAAATGTAGTTTTTAAGGTAAAAACAG
B73-rf4         TGGCAGGCAGCATCCAAACAGGCCCCTAATGCTGAAATGTAGTTTTTAAGGTAAAAACAG
B37-rf4         TGGCAGGCAGCATCCAAACGGGCCCCTAATGCTGAAATGTAGTTTTTAAGGTAAAAACAG
Mo17-rf4        TGGCAGGCTGCATCCAAACAGGCCCCTAATGCTGAAATGTAGTTTTTAAGGTAAAAACAG
Ky21-Rf4        TGGCAGGCAGCATCCAAACGGGCCCCTAATGCTGAAATGTAGTTTTTAAGGTAAAAACAG
A619-Rf4        TGGCAGGCAGCATCCAAACAGGCCCCTAATGCTGAAACGTAGTTTTTAAGGTAAAAACAG
Oh43-Rf4        TGGCAGGCAGCATCCAAACAGGCCCCTAATGCTGAAATGTAGTTTTTAAGGTAAAAACAG
                ******** ********** ***************** **********************

published-rf4   GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
B73-rf4         GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
B37-rf4         GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
Mo17-rf4        GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
Ky21-Rf4        GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
A619-Rf4        GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
Oh43-Rf4        GTAAATTACTTAACATGCAGATTTTTTCCCTGTACTTATTTTTAAACATCTTACAGAGGG
                ************************************************************

published-rf4   TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAAGGGGTTGGCTAAAATG
B73-rf4         TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAAGGGGTTGGCTAAAATG
B37-rf4         TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAANGGGTTGGCTAAAATG
Mo17-rf4        TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAAGGGGTTGGCTAAAATG
Ky21-Rf4        TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAAGGGGTTGGCTAAAATG
A619-Rf4        TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAAGGGGTTGGCTAAAATG
Oh43-Rf4        TAAAGACAAGAAAGGTGCAAAGATACTTAAGTAAAAGCTTAGAAGGGGTTGGCTAAAATG
                ****************************************** ***************
```

```
published-rf4   TTGCATTATGTGAAAGCATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
B73-rf4         TTGCATTATGTGAAAGCATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
B37-rf4         TTGCATTATGTGAAAGCATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
Mo17-rf4        TTGCATTATGTGAAAGCATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
Ky21-Rf4        TTGCATTATGTGAAAGCATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
A619-Rf4        TTGCATTATGTGAAAGCATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
Oh43-Rf4        TTGCATTATGTGAAAGNATAGCATGATAATCCCAGATCATATGAATGATATGTAAACTGC
                *************** ****************************************

published-rf4   ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
B73-rf4         ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
B37-rf4         ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
Mo17-rf4        ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
Ky21-Rf4        ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
A619-Rf4        ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
Oh43-Rf4        ATCTTTTCTGGTGTATCTGTGATTCTGTATAGTTCTAAGTAATGCTTTGGTCTTTTGTTT
                ************************************************************

published-rf4   TGTTT----------ATTGCAAAAGGCTATTTGTTGCTTTCTGCTCAGTATAACTTTGTT
B73-rf4         TGTTT----------ATTGCAAAAGGCTATTTGTTGCTTTCTGCTCAGTATAACTTTGTT
B37-rf4         TGTTT----------ATTGCAAAAGGCTATTTGTTGCTTTCTGCTCAGTATAACTTTGTT
Mo17-rf4        TGTTT----------ATTGCAAAAGGCTATTTGTTGCTTTCTGCTCAGTATGACTTTGTT
Ky21-Rf4        TGTTT----------ATTGCAAAAGGCTATTTGTTGCTTTCTGCTCAGTATAACTTTGTT
A619-Rf4        TGTTT----------ATTGCAAAAGGCTATTTGTTGCTTTCTGCTCAGTATAACTTTGTT
Oh43-Rf4        TNTTTTGTTTTNNNTNNNNNAANANNNTATTTGTTGCTTTNTGCTCANNANAANTTTGAT
                * ***              ** *   ************* ******  *  * **** *

published-rf4   AACTCATTTTTCTAACTATTTTTTTTCCTTTATACTTAATTATGTAGTCTTTCTTTGCAT
B73-rf4         AACTCATTTTTCTAACTATTTTTTTTCCTTTATACTTAATTATGTAGTCTTTCTTTGCAT
B37-rf4         AACTCATTTTTCTAACTAT-TTTTTTCCTTTATACTTAATTATGTAGTCTTTCTTTGCAT
Mo17-rf4        AACTCTTTTTTCTAACTAGGTTTTTTCCTTTATACTTAATTATGTAGTCTTTCTTTGCAT
Ky21-Rf4        AACTCATTTTTCTAACTAT-TTTTTTCCTTTATACTTAATTATGTAGTCTTTCTTTGCAT
A619-Rf4        AACTCATTTTTCTAACTAT-TTTTTTCCTTTATACTTAATTATGTAGTCTTTCTTTGCAT
Oh43-Rf4        AACTCATTTTTNNAANTATTTTTTTTCCTTTATAATTAATTANGTAGTNNTTCTCNNCAT
                ***** *****  ** **  ************** ******* *****  ****   ***

published-rf4   TTCTAATCTACCTCTATGTATTACCAGTACATTTATCCTATGTACGCATCGCAGTAGTTG
B73-rf4         TTCTAATCTACCTCTATGTATTACCAGTACATTTATCCTATGTACGCATCGCAGTAGTTG
B37-rf4         TTCTAATCTACCTCTATGTATTACCAGTACATTTATCCTATGTACGCATCGCAGTAGTCG
Mo17-rf4        TTCCAATCAACCTCTATGTATTACCAGTACATTTATCCAATGGACGCATCGCAGTAGTCG
Ky21-Rf4        TTCTAATCTACCTCTATGTATTACCAGTACATTTATCCTATGTACGCATCGCAGTAGTCG
A619-Rf4        TTCTAATCTACCTCTATGTATTACCAGTACATTTATCCTATGTACGCATCGCAGTAGTCG
Oh43-Rf4        TTATAATCTACCTNNANGTATCANCAGNNCATTTATCNTANGTNNGCNTNGCAGTAGTNG
                **  **** ****  * **** * ***  ********  * *   ** * ******** *

published-rf4   TACAATACCACCAAGGAAAGCAACACAGCTAACGATGGAGGTGCACACTTATAGTTGGTT
B73-rf4         TACAATACCACCAAGGAAAGCAACACAGCTAACGATGGAGGTGCACACTTATAGTTGGTT
B37-rf4         TACAATACCACCAAGGAAAGCAACACAGCTAACGATGGAGGTGCACACTTATAGTTGGTT
Mo17-rf4        TACAATACCACCAAGGAAAGCAACACAGCTAACGATGGAGGTGCACACTTATAGTTGGTT
Ky21-Rf4        TACAATACCACCAAGGAAAGCAACACAGCTAACGATGGAGGTGCACACTTATAGTTGGTT
A619-Rf4        TACAATACCACCAAGGAAAGCAACACAGCTAACGATGGAGGTGCACACTTATAGTTGGTT
Oh43-Rf4        NACAANACCACCAAGGAAANCAACNCAGCTAANGANGGAGGNGCACANTNATAGTTGGTN
                 **** ************* **** ******* ** ***** ***** * *********
```

```
published-rf4   TGAAACATAATGATAACAACCTTCTTGGTAAATGCATGTGTTCAACTTCATGCTGCCACT
B73-rf4         TGAAACATAATGATAACAACCTTCTTGGTAAATGCATGTGTTCAACTTCATGCTGCCACT
B37-rf4         TGAAACATAATGATAACAACCTTCTTGGTAAATGCATGTGTTCAACTTCATGCTGCCACT
Mo17-rf4        TGAAACATAATGATAACAACCTTCTTGGTAAATGCATGTGTTCAACTTCATGCTGCCACT
Ky21-Rf4        TGAAACATAATGATAACAACCTTCTTGGTAAATGCATGTGTTCAACTTCATGCTGCCACT
A619-Rf4        TGAAACATAATGATAACAACCTTCTTGGTAAATGCATGTGTTCAACTTCATGCTGCCACT
Oh43-Rf4        TGAAACATAATGATAACANCCTTNNTGGTAAANGCATGTGTTCAACTTCATGCCNCCACN
                ****************** ****  ******* ********************  ****

published-rf4   TCCCAATTTCCAGATAGGATCTCTATTATTTGATAATATGTAATCAGAAATGATCACTTG
B73-rf4         TCCCAATTTCCAGATAGGATCTCTATTATTTGATAATATGTAATCAGAAATGATCACTTG
B37-rf4         TCCCAATTTCCAGATAGGATCTCTATTATTTGATAATATGTAATCAGAAATGATCACTTG
Mo17-rf4        TCCCAATTTCCAGATAGGATCTCTATTATTTGATAATATGTAATCAGAAATGATCACTTG
Ky21-Rf4        TCCCAATTTCCAGATAGGATCTCTATTATTTGATAATATGTAATCAGAAATGATCACTTG
A619-Rf4        TCCCAATTTCCAGATAGGATCTCTATTATTTGATAATATGTAATCAGAAATGATCACTCG
Oh43-Rf4        NCNCNNNNNCCAGATAGGATNTNTATTATTTGATAATANGTAATCAGAAATNATCANNTG
                 * *     *********** * *************** ************ ****   *

published-rf4   ATTAAACATTCAACTATGCCTGTATGATATGAAGAGATAACTGATGTTGTATCAGGGTGA
B73-rf4         ATTAAACATTCAACTATGCCTGTATGATATGAAGAGATAACTGATGTTGTATCAGGGTGA
B37-rf4         ATTAAACATTCAACTATGCCTGTATGATATGAAGAGA--ACTGATGTTGTATCAGGGTGA
Mo17-rf4        ATTAAACATTCAACTATGCCTGTATGATATGAAGAGATAACTGATGTTGTATCAGGGTGA
Ky21-Rf4        ATTAAACATTCAACTATGCCTGTATGATATGAAGAGA--ACTGATGTTGTATCAGGGTGA
A619-Rf4        ATTAAACATTCAACTATGCCTGTATGATACGAAGAGATAACTGATGTTGTATCAGGGTGA
Oh43-Rf4        ANNCANCANTCAANTNNNCCTGTATGATANGAAGAGATAANTGATGNNGNATCAGGGTGA
                *   * ** **** *   *********** *******  * *****  * **********

published-rf4   TTGTGTACCTTCATATCAT---ATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACT
B73-rf4         TTGTGTACCTTCATATCAT---ATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACT
B37-rf4         TTGTGTACCTTCATATCATCATATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACT
Mo17-rf4        TTGTGTACCTTCATATCATCATATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACG
Ky21-Rf4        TTGTGTACCTTCATATCATCATATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACT
A619-Rf4        TTGTGTACCTTCATATCATCATATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACT
Oh43-Rf4        TNGNNNNCCTTCNNATCAT---ATGTGTGCTGTTCTGACTTGTGAAATAATTTTGTCACT
                * *    *****  *****   ************************************

published-rf4   ATTTCTCTGCAGCTTGTTGAGTGGGTTCAAAAATCACAGCTATTCGAATTTGTAGAGCGT
B73-rf4         ATTTCTCTGCAGCTTGTTGAGTGGGTTCAAAAATCACAGCTATTCGAATTTGTAGAGCGT
B37-rf4         ATTTCTCTGCAGCTTGTTGAGTGGGTTCAAAAATCACAGCTATTCGAATTTGTAGAGCGT
Mo17-rf4        GTTTCTCTGCAGCTTGTTGAGTGGGTTAAAAAATCACAGCTATTCGAACTTGTAGAGCGT
Ky21-Rf4        ATTTCTCTGCAGCTTGTTGAGTGGGTTCAAAAATCACAGCTATTCGAATTTGTAGAGCGT
A619-Rf4        ATTTCTCTGCAGCTTGTTGAGTGGGTTCAAAAATCACAGCTATTCGAATTTGTAGAGCGT
Oh43-Rf4        ATTTCTCTGCAGCTTGTTGAGTGGGTTCAAAAATCACAGCTATTCGAATTTGTAGAGCGT
                 ************************** ******************** ***********

published-rf4   GATTATGCTGCACATCTTGATTTAACAGCTAAGGTC-TATGGTCTTCATAAAGCAGAACA
B73-rf4         GATTATGCTGCACATCTTGATTTAACAGCTAAGGTN-TATGGTTTTCATAAAGCAGAACA
B37-rf4         GATTATGCTGCACATCTTGATTTAACAGCTAAGGTC-TATGGTCTTCATAAAGCAGAACA
Mo17-rf4        GATTATGCTGCACATCTTGATTTAACAGCTAAGGTCGTATGGTCT-CATAAAGCAGAACA
Ky21-Rf4        GATTATGCTGCACATCTTGATTTAACAGCTAAGGTC-TATGGTCTTCATAAAGCAGAACA
A619-Rf4        GATTATGCTGCACATCTTGATTTAACAGCTAAGGTCTTATGGTCT-CATAAAGCAGAACA
Oh43-Rf4        GATTATGCTGCACATCTTGATTTAACAGCTAAGGTC-TATGGTCGTCATAAAGCAGAACA
                ********************************  ******   **************
```

```
published-rf4   GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTTTATAGAACACTTTTGGC
B73-rf4         GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTTTATAGAACACTTTTGNC
B37-rf4         GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTTTATAGAACACTTTTGGC
Mo17-rf4        GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTGTATAGAACACTTTTGGC
Ky21-Rf4        GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTTTATAGAACACTTTTGGC
A619-Rf4        GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTGTATAGAACACTTTTGGC
Oh43-Rf4        GACTATTGAGAAAATTCCTGCATCTTGTAGGGGTGAGATTGTNTATAGAACACTTTTGGC
                ****************************************** *************** *

published-rf4   CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
B73-rf4         CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
B37-rf4         CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
Mo17-rf4        CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
Ky21-Rf4        CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
A619-Rf4        CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
Oh43-Rf4        CAATTGTGTTTCTGCATCCAATATAAAAAAATCGGAGCAAGTTTTTAATAGGATGAAGGA
                ************************************************************

published-rf4   TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
B73-rf4         TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
B37-rf4         TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
Mo17-rf4        TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
Ky21-Rf4        TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
A619-Rf4        TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
Oh43-Rf4        TCTTGGATTTCCGGTTACACAATTTTCTTTCAATCAGCTTCTGTTACTCTACAAAAGGCT
                ************************************************************

published-rf4   GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
B73-rf4         GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
B37-rf4         GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
Mo17-rf4        GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
Ky21-Rf4        GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
A619-Rf4        GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
Oh43-Rf4        GGACAGGAAGAAGATTGCTGATGTTCTTGCAATGATGGAAAAGGAGGATGTGAAACCATC
                ************************************************************

published-rf4   TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
B73-rf4         TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
B37-rf4         TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
Mo17-rf4        TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
Ky21-Rf4        TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
A619-Rf4        TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
Oh43-Rf4        TCTCTTTACTTACAAGATTCTTGTAGATGCCAAGGGTTCAGTTGGAGATATAGAAGCTAT
                ************************************************************

published-rf4   GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
B73-rf4         GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
B37-rf4         GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
Mo17-rf4        GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
Ky21-Rf4        GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
A619-Rf4        GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
Oh43-Rf4        GGAGAAAGTGATTGAATCAATGGAGAAAGATAGCATTGAGCCAGATCTCACGTTTAATGC
                ************************************************************
```

```
published-rf4   TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
B73-rf4         TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
B37-rf4         TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
Mo17-rf4        TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
Ky21            TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
A619            TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
Oh43-Rf4        TACAATTGCCAGACACTACATATTTTATGGCCAGCGTGAGAAAGCAGAGGCACTTCTGAA
                ************************************************************

published-rf4   GGCAATGGAGGGTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTACCTTT
B73-rf4         GGCAATGGAGGGTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTACCTTT
B37-rf4         GGCAATGGAGGGTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTACCTTT
Mo17-rf4        GGCAATGGAGGGTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTACCTTT
Ky21-Rf4        GGCAATGGAGGGTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTACCTTT
A619-Rf4        GGCAATGGNNNNTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTGCCTTT
Oh43-Rf4        GGCAATGGAGGGTGATGATATCCAGAAAAATCGTGCTGCTTGCAAGACTTTATTACCTTT
                ********    ***************************************** *****

published-rf4   ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCGTGTGAGGATAA
B73-rf4         ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCGTGTGAGGATAA
B37-rf4         ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCGTGTGAGGATAA
Mo17-rf4        ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCGTGTGAGGATAA
Ky21-Rf4        ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCGTGTGAGGATAA
A619-Rf4        ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCTTGTGAGGATAA
Oh43-Rf4        ACATGCTTTTCTGGGCAATAGTGATGCTGTGGAAAGAATTTGGAAGGCGTGTGAGGATAA
                ********************************************** ***********

published-rf4   TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
B73-rf4         TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
B37-rf4         TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
Mo17-rf4        TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
Ky21-Rf4        TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
A619-Rf4        TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
Oh43-Rf4        TACTCGCGTAGCTGAGTCCGTGTCTGCTATAGAGGCCTTTGGTAAGCTTGGTGATGTTGA
                ************************************************************

published-rf4   AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
B73-rf4         AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
B37-rf4         AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
Mo17-rf4        AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
Ky21-Rf4        AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
A619-Rf4        AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
Oh43-Rf4        AAAGGCAGAGAAGGTTTTTGAGGATATGCTTGTGCATTGGAAAACACTCTCTTCCAAATT
                ************************************************************

published-rf4   CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
B73-rf4         CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
B37-rf4         CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
Mo17-rf4        CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
Ky21-Rf4        CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
A619-Rf4        CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
Oh43-Rf4        CTACAACGCGTTGCTAAAGGTGTATGCTGATCACAACCTCTTAGACAAGGGTAAGGAACT
                ************************************************************
```

```
published-rf4   AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
B73-rf4         AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
B37-rf4         AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
Mo17-rf4        AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
Ky21-Rf4        AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
A619-Rf4        AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
Oh43-Rf4        AGCAAAGCGGATGGATGAGAACCGTGTCAGGTTTGGGGCCCCAACTCTGGATGCACTTGT
                ************************************************************

published-rf4   GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
B73-rf4         GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
B37-rf4         GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
Mo17-rf4        GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
Ky21-Rf4        GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
A619-Rf4        GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
Oh43-Rf4        GAAGCTATACGTTGAGGCTGGAGAGGTGGAGAAAGCTGAGTCCCTAGTGCACAAGCTATC
                ************************************************************

published-rf4   TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
B73-rf4         TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
B37-rf4         TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
Mo17-rf4        TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
Ky21-Rf4        TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
A619-Rf4        TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
Oh43-Rf4        TATACAGAACCACATCAAGCCTATATACAACACATACATGACGCTGCTTGACAGCTACTC
                ************************************************************

published-rf4   AAAGAAAGGAGACGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
B73-rf4         AAAGAAAGGAGACGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
B37-rf4         AAAGAAAGGAGACGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
Mo17-rf4        AAAGAAAGGAGATGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
Ky21-Rf4        AAAGAAAGGAGACGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
A619-Rf4        AAAGAAAGGAGACGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
Oh43-Rf4        AAAGAAAGGAGACGTCCGCAACTCAGAGAAAGTGTTCAACAAGTTGCGGCAGTCCGGCTA
                ************ ************************************************

published-rf4   CACTGGGAGGATCAGGATGTACCAGGTGCTGCTCCATGCTTATGTACGTGCTAAGGCCCC
B73-rf4         CACTGGGAGGATCAGGATGTACCAGGTGCTGCTCCATGCTTATGTACGTGCTAAGGCCCC
B37-rf4         CACTGGGAGGATCAGGATGTACCAGTCGCTGCTCCATGCTTATGTACATGCTAAGGCCCC
Mo17-rf4        CACTGGGAGGATCAGGATGTACCAGTCGCTGCTCCGTGCTTATGTACATGCTAAGGCCCC
Ky21-Rf4        CACTGGGAGGATCAGGATGTACCAGTCGCTGCTCCATGCTTATGTACATGCTAAGGCCCC
A619-Rf4        CACTGGGAGGATCAGGATGTACCAGGTGCTGCTCCATGCTTATGTACGTGCTAAGGCCCC
Oh43-Rf4        CACTGGGAGGATCAGGATGTACCAGGTGCTGCTCCATGCTTATGTACRTGCTAAGGCCCC
                ************************  ******** *********** ************

published-rf4   TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGTTGTGTC
B73-rf4         TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGTTGTGTC
B37-rf4         TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGCTGTGTC
Mo17-rf4        TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGTTGTGTC
Ky21-Rf4        TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGCTGTGTC
A619-Rf4        TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGCTGTGTC
Oh43-Rf4        TGCTTATGGATTCAGAGAGAGGATGAAGGCTGACAACATTTTCCCCAACAGTGTTGTGTC
                ************************************************** ******
```

```
published-rf4   AACTTTGATTGCCGCTACCGATCCGTTTGTCAAGAAGAAATCGATATCTGACTTGCTCGA
B73-rf4         AACTTTGATTGCCGCTACCGATCCGTTTGTCAAGAAGAAATCGATATCTGACTTGCTCGA
B37-rf4         AACTTTGATTGCCGCTACTGATCCGTTTGTCAAGAAGAAATCGATATCTGACTTGCTCGA
Mo17-rf4        AACTTTGATTGCCGCTACCGATCCGTTTGTCAAGAAGAAATCAATATCTGACTTGCTCGA
Ky21-Rf4        AACTTTGATTGCCGCTACTGATCCGTTTGTCAAGAAGAAATCGATATCTGACTTGCTCGA
A619-Rf4        AACTTTGATTGCCGCTACTGATCCGTTTGTCAAGAAGAAATCGATATCTGACTTGCTCGA
Oh43-Rf4        AACTTTGATTGCCGCTACYGATCCGTTTGTCAAGAAGAAATCGATATCTGACTTGCTCGA
                ********************* ************************ *****************

published-rf4   TTAGGTTATGGAGTAGTGTGAGATC
B73-rf4         TTAGGTTATGGAGTAGTGTGAGATC
B37-rf4         TTAGGTTATTGAGTAGTGTGAGATC
Mo17-rf4        TTAGGTTATGGAGTAGTGTGAGATC
Ky21-Rf4        TTAGGTTATTGAGTAGTGTGAGATC
A619-Rf4        TTAGGTTATGGAGTAGTGTGAGATC
Oh43-Rf4        TTAGGTTATGGAGTAGTGTGAGATC
                ********** ***************
```

References

Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J., and Palmer, J.D. (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature, 408(6810), 354-357.

Akagi, H., Sakamoto, M., Shinjyo, C., Shimada, H., and Fujimura, T. (1994) A unique sequence located downstream from the rice mitochondrial atp6 may cause male sterility. Curr Genet, 25(1), 52-58.

Allen, J.O., Fauron, C.M., Minx, P., Roark, L., Oddiraju, S., Lin, G.N., Meyer, L., Sun, H., Kim, K., Wang, C., Du, F., Xu, D., Gibson, M., Cifrese, J., Clifton, S.W., and Newton, K.J. (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics, 177(2), 1173-1192.

Barnabas, B. (1985) Effects of water loss on germination ability of maize (*zea mays*) pollen. Ann. Bot., 48, 861-864.

Bateson, W., Gairdner, E. (1921) Male-sterility in flax. J. Genet., 11, 269-275.

Beckett, J.B. (1971) Classification of male-sterile cytoplasms in maize (*Zea mays* L.). crop science, 11(sept.-oct.), 724-727.

Bedinger, P. (1992) The remarkable biology of pollen. Plant Cell, 4(8), 879-887.

Begu, D., Graves, P.V., Araya, A., and Litvak, S. (1988). In Proc. 3rd Int. Workshop on the mitochondrial genomeof higher plants (F. Quetier, and Lejeune, B., ed, Centre National de la Recherche Scientifique, France, pp 2.

Bentolila, S., Alfonso, A.A., and Hanson, M.R. (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A, 99(16), 10887-10892.

Binder, S., Marchfelder, A., and Brennicke, A. (1996) Regulation of gene expression in plant mitochondria. Plant Mol Biol, 32(1-2), 303-314.

Boeshore, M.L., Hanson, M.R., and Izhar, S. (1985) A variant mitochondrial-DNA arrangment specific to petunia stable sterile somatic hybrids. Plant Mol. Biol., 4, 125-132.

Brown, G.G. (1999) Unique Aspects of Cytoplasmic Male Sterility and Fertility Restoration in Brassica napus. J Hered., 90, 351-356.

Brown, G.G., Formanova, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W.Y., and Landry, B.S. (2003) The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J, 35(2), 262-272.

Carpousis, A.J., Vanzo, N.F., and Raynal, L.C. (1999) mRNA degradation. A tale of poly(A) and multiprotein machines. Trends Genet, 15(1), 24-28.

Claros, M.G., and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem, 241, 779-786.

Clifton, S.W., Minx, P., Fauron, C.M., Gibson, M., Allen, J.O., Sun, H., Thompson, M., Barbazuk, W.B., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R.K., and Newton, K.J. (2004)

Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol, 136(3), 3486-3503.
Crofts, A.R. (2004) The cytochrome bc1 complex: function in the context of structure. Annu Rev Physiol, 66, 689-733.
Crofts, A.R., Shinkarev, V.P., Kolling, D.R., and Hong, S. (2003) The modified Q-cycle explains the apparent mismatch between the kinetics of reduction of cytochromes c1 and bH in the bc1 complex. J Biol Chem, 278(38), 36191-36201.
Cui, X., Wise, R.P., and Schnable, P.S. (1996) The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. Science, 272(5266), 1334-1336.
Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Cattolico, L., Falentin, C., Giancola, S., Renard, M., Budar, F., Small, I., Caboche, M., Delourme, R., and Bendahmane, A. (2003) Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep, 4(6), 588-594.
Devos, K.M., Beales, J., Nagamura, Y., and Sasaki, T. (1999) Arabidopsis-rice: will colinearity allow gene prediction across the eudicot-monocot divide? Genome Res, 9(9), 825-829.
Dewey, R.E., Levings, C.S., 3rd, and Timothy, D.H. (1986) Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm. Cell, 44(3), 439-449.
Dewey, R.E., Levings III, Charles S., and Timothy, D.H. (1985) Nucleotide sequence of ATPase subunit 6 gene of maize mitochondria. Plant Physiol., 79, 914-919.
Dewey, R.E., Timothy, D.H., and Levings, C.S., 3rd. (1991) Chimeric mitochondrial genes expressed in the C male-sterile cytoplasm of maize. Curr Genet, 20(6), 475-482.
Dewey, R.E.a.L.I., C.S. (1988) Molecular studies of cytoplasmic male sterility in maize. Phil. Trans. R. Soc. Lond., 319, 177-185.
Dienhart, M., Pfeiffer, K., Schagger, H., and Stuart, R.A. (2002) Formation of the yeast F1F0-ATP synthase dimeric complex does not require the ATPase inhibitor protein, Inh1. J Biol Chem, 277(42), 39289-39295.
Dombrowski, S., Brennicke, A., and Binder, S. (1997) 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. Embo J, 16(16), 5069-5076.
Duvezin-Caubet, S., Caron, M., Giraud, M.F., Velours, J., and di Rago, J.P. (2003) The two rotor components of yeast mitochondrial ATP synthase are mechanically coupled by subunit delta. Proc Natl Acad Sci U S A, 100(23), 13235-13240.
Duvick, D.N. (1959) The use of cytoplasmic male-sterility in hybrid seed production. Econ. Bot., 13, 167-195.
Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol, 300(4), 1005-1016.
Eubel, H., Braun, H.P., and Millar, A.H. (2005) Blue-native PAGE in plants: a tool in analysis of protein-protein interactions. Plant Methods, 1(1), 11.
Eubel, H., Jansch, L., and Braun, H.P. (2003) New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. Plant Physiol, 133(1), 274-286.
Fauron, C., Casper, M., Gao, Y., and Moore, B. (1995) The maize mitochondrial genome: dynamic, yet functional. Trends Genet, 11(6), 228-235.
Feiler, H.S. (1986) Altered mitochondrial gene expression in the nonchromosomal stripe mutants of maize. University of Missouri-Columbia, Columbia.

Fey, J., and Marechal-Drouard, L. (1999) Compilation and analysis of plant mitochondrial promoter sequences: An illustration of a divergent evolution between monocot and dicot mitochondria. Biochem Biophys Res Commun, 256(2), 409-414.
Figueroa, P., Leon, G., Elorza, A., Holuigue, L., Araya, A., and Jordana, X. (2002) The four subunits of mitochondrial respiratory complex II are encoded by multiple nuclear genes and targeted to mitochondria in Arabidopsis thaliana. Plant Mol Biol, 50(4-5), 725-734.
Fillingame, R.H., Angevine, C.M., and Dmitriev, O.Y. (2003) Mechanics of coupling proton movements to c-ring rotation in ATP synthase. FEBS Lett, 555(1), 29-34.
Folkerts, O., and Hanson, M.R. (1991) The male sterility-associated pcf gene and the normal atp9-1 gene in Petunia are located on different mitochondrial DNA molecules. Genetics, 129(3), 885-895.
Forde, B.G., and Leaver, C.J. (1980) Nuclear and cytoplasmic genes controlling synthesis of variant mitochondrial polypeptides in male-sterile maize. Proc Natl Acad Sci U S A, 77(1), 418-422.
Freeling, M., and Walbot, V. (1994) The Maize Handbook: Springer-Verlag.
Gabay-Laughnan, S., Chase, C.D., Ortega, V.M., and Zhao, L. (2004) Molecular-genetic characterization of CMS-S restorer-of-fertility alleles identified in Mexican maize and teosinte. Genetics, 166(2), 959-970.
Gagliardi, D., and Leaver, C.J. (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. Embo J, 18(13), 3757-3766.
Gagliardi, D., Perrin, R., Marechal-Drouard, L., Grienenberger, J.M., and Leaver, C.J. (2001) Plant mitochondrial polyadenylated mRNAs are degraded by a 3'- to 5'-exoribonuclease activity, which proceeds unimpeded by stable secondary structures. J Biol Chem, 276(47), 43541-43547.
Gieige, P., Sweetlove, Lee J., and Leaver, Christopher J. (2003) Identification of mitochondrial protein complexes in *Arabidopsis* using two-dimensional blue-native polyacrylamide gel electrophoresis. Plant Mol Biol Reporter, 21, 133-144.
Grof, C.P., Winning, B.M., Scaysbrook, T.P., Hill, S.A., and Leaver, C.J. (1995) Mitochondrial pyruvate dehydrogenase. Molecular cloning of the E1 alpha subunit and expression analysis. Plant Physiol, 108(4), 1623-1629.
Gromiha, M.M., and Suwa, M. (2005) A simple statistical method for discriminating outer membrane proteins with better accuracy. Bioinformatics, 21(7), 961-968.
Gu, J., Dempsey, S., and Newton, K.J. (1994) Rescue of a maize mitochondrial cytochrome oxidase mutant by tissue culture. Plant J, 6(6), 787-794.
Handa, H. (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. Nucleic Acids Res, 31(20), 5907-5916.
Hanson, M., and Folkerts, O. (1992) Structure and function of the higher plant mitochondrial genome. INT Rev Cytol, 141, 129-172.
Hanson, M.R. (1984) Stability, variation, and recombination of plant mitochondrial genomes via cell and tissue culture. Oxf. Surv. Plant Mol. Cell Biol., 1, 33-52.
Hanson, M.R., and Bentolila, S. (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell, 16 Suppl, S154-169.
Hayes, R., Kudla, J., and Gruissem, W. (1999) Degrading chloroplast mRNA: the role of polyadenylation. Trends Biochem Sci, 24(5), 199-202.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics, 14(4), 378-379.
Hunt, M.D., and Newton, K.J. (1991) The NCS3 mutation: genetic evidence for the expression of ribosomal protein genes in Zea mays mitochondria. EMBO J, 10(5), 1045-1052.
Hurkman, W.J., and Tanaka, C.K. (1986) Solubilization of Plant Membrane Proteins for Analysis by Two-Dimensional Gel Electrophoresis. Plant Physiol, 81(3), 802-806.
Iwabuchi, M., Koizuka, N., Fujimoto, H., Sakai, T., and Imamura, J. (1999) Identification and expression of the kosena radish (Raphanus sativus cv. Kosena) homologue of the ogura radish CMS-associated gene, orf138. Plant Mol Biol, 39(1), 183-188.
Jansch, L., Kruft, V., Schmitz, U.K., and Braun, H.P. (1996) New insights into the composition, molecular mass and stoichiometry of the protein complexes of plant mitochondria. Plant J, 9(3), 357-368.
Josephson, L.M., Morgan, T.E., and Arnold, J.M. (1978) Genetics and inheritance of fertility restoration of male-sterile cytoplasms in corn. Proc. 33rd Ann. Corn Sorghum Res. Conf., 33, 48-59.
Kaleikau, E.K., Andre, C.P., and Walbot, V. (1992) Structure and expression of the rice mitochondrial apocytochrome b gene (cob-1) and pseudogene (cob-2). Curr Genet, 22(6), 463-470.
Kamps, T.L., McCarty, D.R., and Chase, C.D. (1996) Gametophyte genetics in Zea mays L.: dominance of a restoration-of-fertility allele (Rf3) in diploid pollen. Genetics, 142(3), 1001-1007.
Karpenahalli, M.R., Lupas, A.N., and Soding, J. (2007) TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. BMC Bioinformatics, 8, 2.
Karpova, O.V., and Newton, K.J. (1999) A partially assembled complex I in NAD4-deficient mitochondria of maize. The Plant Journal, 17(5), 511-521.
Kazama, T., and Toriyama, K. (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice. FEBS Lett, 544(1-3), 99-102.
Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S., and Imamura, J. (2003) Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J, 34(4), 407-415.
Komori, T., Ohta, S., Murai, N., Takakura, Y., Kuraya, Y., Suzuki, S., Hiei, Y., Imaseki, H., and Nitta, N. (2004) Map-based cloning of a fertility restorer gene, Rf-1, in rice (Oryza sativa L.). Plant J, 37(3), 315-325.
Kudla, J., Hayes, R., and Gruissem, W. (1996) Polyadenylation accelerates degradation of chloroplast mRNA. Embo J, 15(24), 7137-7146.
Kuhn, J., Tengler, U., and Binder, S. (2001) Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. Mol Cell Biol, 21(3), 731-742.
Landgren, M., Zetterstrand, M., Sundberg, E., and Glimelius, K. (1996) Alloplasmic male-sterile Brassica lines containing B. tournefortii mitochondria express an ORF 3' of the atp6 gene and a 32 kDa protein. off. Plant Mol Biol, 32(5), 879-890.
Laser, K.D., Lersten, N.R. (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterility angiosperms. Bot. Rev., 38, 425-454.
Lauer, M., Knudsen, C., Newton, K.J., Gabay-Laughnan, S., and Laughnan, J.R. (1990) A partially deleted mitochondrial cytochrome oxidase gene in the NCS6 abnormal growth mutant of maize. New Biol, 2(2), 179-186.

Laughnan, J.R., Gabay-Laughnan, S. (1983) Cytoplasmic male sterility in maize. Ann. Rev. Genet., 17, 27-48.
Lee, S.J., Gracen, V.E., and Earle, E.D. (1979) The cytology of pollen abortion in C-cytoplasmic male-sterile corn anthers. Amer. J. Bot., 66(6), 656-667.
Levings, C.S., 3rd, and Siedow, J.N. (1992) Molecular basis of disease susceptibility in the Texas cytoplasm of maize. Plant Mol Biol, 19(1), 135-147.
Lisitsky, I., Klaff, P., and Schuster, G. (1996) Addition of destabilizing poly (A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA. Proc Natl Acad Sci U S A, 93(23), 13398-13403.
Liu, F., Cui, X., Horner, H.T., Weiner, H., and Schnable, P.S. (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. Plant Cell, 13(5), 1063-1078.
Lough, A.N., Roark, L.M., Kato, A., Ream, T.S., Lamb, J.C., Birchler, J.A., and Newton, K.J. (2008) Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. Genetics, 178(1), 47-55.
Lupold, D.S., Caoile, A.G., and Stern, D.B. (1999a) Genomic context influences the activity of maize mitochondrial cox2 promoters. Proc Natl Acad Sci U S A, 96(20), 11670-11675.
Lupold, D.S., Caoile, A.G., and Stern, D.B. (1999b) The maize mitochondrial cox2 gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units. J Biol Chem, 274(6), 3897-3903.
Lupold, D.S., Caoile, A.G., and Stern, D.B. (1999c) Polyadenylation occurs at multiple sites in maize mitochondrial cox2 mRNA and is independent of editing status. Plant Cell, 11(8), 1565-1578.
Makaroff, C.A., Apel, I.J., and Palmer, J.D. (1989) The atp6 coding region has been disrupted and a novel reading frame generated in the mitochondrial genome of cytoplasmic male-sterile radish. J Biol Chem, 264(20), 11706-11713.
Makaroff, C.A., and Palmer, J.D. (1988) Mitochondrial DNA rearrangements and transcriptional alterations in the male-sterile cytoplasm of Ogura radish. Mol Cell Biol, 8(4), 1474-1480.
Maniatis, T., Sambrook, J., and Fritsch, E.F. (1989) Molecular Cloning A Laboratory Manual: Cold Spring Harbor Laboratory Press.
Marechal-Drouard, L., Weil, J., and Dietrich, A. (1993) Transfer RNAs and transfer RNA genes in plants. Ann Rev Plant Physiol Plant Mol Biol, 44, 13-32.
Marienfeld, J.R., and Newton, K.J. (1994) The maize NCS2 abnormal growth mutant has a chimeric nad4-nad7 mitochondrial gene and is associated with reduced complex I function. Genetics, 138(3), 855-863.
Marres, C.A., de Vries, S., and Grivell, L.A. (1991) Isolation and inactivation of the nuclear gene encoding the rotenone-insensitive internal NADH: ubiquinone oxidoreductase of mitochondria from Saccharomyces cerevisiae. Eur J Biochem, 195(3), 857-862.
Meuter-Gerhards, A., Riegart, S. and Wiermann, R. (1999) Studies on sporopollenin biosynthesis in Curcurbita maxima (DUCH)-II: the involvment of aliphatic metabolism. J. Plant Physiol., 154, 431-436.
Millar, A.H., Eubel, H., Jansch, L., Kruft, V., Heazlewood, J.L., and Braun, H.P. (2004) Mitochondrial cytochrome c oxidase and succinate dehydrogenase complexes contain plant specific subunits. Plant Mol Biol, 56(1), 77-90.
Mitchell, P. (1976) Possible molecular mechanisms of the protonmotive function of cytochrome systems. J Theor Biol, 62(2), 327-367.

Moller, I.M. (2001) PLANT MITOCHONDRIA AND OXIDATIVE STRESS: Electron Transport, NADPH Turnover, and Metabolism of Reactive Oxygen Species. Annu Rev Plant Physiol Plant Mol Biol, 52, 561-591.
Moller, I.M., Rasmusson, A.G., and Fredlund, K.M. (1993) NAD(P)H-ubiquinone oxidoreductases in plant mitochondria. J Bioenerg Biomembr, 25(4), 377-384.
Moneger, F., Smart, C., and Leaver, C. (1994) Nuclear restoration of cytoplasmic male sterility in sunflower is associated with the tissue-specific regulation of a novel mitochondrial gene. Embo J, 13, 8-17.
Mueller, D.M. (2000) Partial assembly of the yeast mitochondrial ATP synthase. J Bioenerg Biomembr, 32(4), 391-400.
Mulligan, R.M., Lau, G.T., and Walbot, V. (1988) Numerous transcription initiation sites exist for the maize mitochondrial genes for subunit 9 of the ATP synthase and subunit 3 of cytochrome oxidase. Proc Natl Acad Sci U S A, 85(21), 7998-8002.
Mulligan, R.M., Leon, P., and Walbot, V. (1991) Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression. Mol Cell Biol, 11(1), 533-543.
Namslauer, A., and Brzezinski, P. (2004) Structural elements involved in electron-coupled proton transfer in cytochrome c oxidase. FEBS Lett, 567(1), 103-110.
Newton, K.J., Knudsen, C., Gabay-Laughnan, S., and Laughnan, J.R. (1990) An abnormal growth mutant in maize has a defective mitochondrial cytochrome oxidase gene. Plant Cell, 2(2), 107-113.
Newton, K.J., Mariano, J.M., Gibson, C.M., Kuzmin, E., and Gabay-Laughnan, S. (1996) Involvement of S2 episomal sequences in the generation of NCS4 deletion mutation in maize mitochondria. Dev Genet, 19(3), 277-286.
Newton, K.J., Winberg, B., Yamato, K., Lupold, S., and Stern, D.B. (1995) Evidence for a novel mitochondrial promoter preceding the cox2 gene of perennial teosintes. Embo J, 14(3), 585-593.
Nivison, H.T., and Hanson, M.R. (1989) Identification of a mitochondrial protein associated with cytoplasmic male sterility in petunia. Plant Cell, 1(11), 1121-1130.
Nivison, H.T., Sutton, C.A., Wilson, R.K., and Hanson, M.R. (1994) Sequencing, processing, and localization of the petunia CMS-associated mitochondrial protein. Plant J, 5(5), 613-623.
Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. (2002) The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics, 268(4), 434-445.
Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and et al. (1992) Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol, 223(1), 1-7.
Pacini, E., Franchi, G.G., and Hesse, M. (1985) The tapetum: its form, function and possible phylogeny in Embryophyta. Pl. Syst. Evol, 149, 155-185.
Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L., and Song, K. (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A, 97(13), 6960-6966.
Palmer, J.D., and Herbon, L.A. (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol, 28(1-2), 87-97.
Papa, S., Capitanio, N., Capitanio, G., and Palese, L.L. (2004) Protonmotive cooperativity in cytochrome c oxidase. Biochim Biophys Acta, 1658(1-2), 95-105.

Pecina, P., Houstkova, H., Hansikova, H., Zeman, J., and Houstek, J. (2004) Genetic defects of cytochrome c oxidase assembly. Physiol Res, 53 Suppl 1, S213-223.
Pereira de Souza, A., Jubier, M.F., Delcher, E., Lancelin, D., and Lejeune, B. (1991) A trans-splicing model for the expression of the tripartite nad5 gene in wheat and maize mitochondria. Plant Cell, 3(12), 1363-1378.
Pruitt, K.D., and Hanson, M.R. (1991) Transcription of the Petunia mitochondrial CMS-associated Pcf locus in male sterile and fertility-restored lines. Mol Gen Genet, 227(3), 348-355.
Rapp, W.D., and Stern, D.B. (1992) A conserved 11 nucleotide sequence contains an essential promoter element of the maize mitochondrial atp1 gene. Embo J, 11(3), 1065-1073.
Rasmusson, A.G., Heiser, V., Irrgang, K.D., Brennicke, A., and Grohmann, L. (1998a) Molecular characterisation of the 76 kDa iron-sulphur protein subunit of potato mitochondrial complex I. Plant Cell Physiol, 39(4), 373-381.
Rasmusson, A.G., Heiser, V.V., Zabaleta, E., Brennicke, A., and Grohmann, L. (1998b) Physiological, biochemical and molecular aspects of mitochondrial complex I in plants. Biochim Biophys Acta, 1364(2), 101-111.
Rasmusson, A.G., Soole, K.L., and Elthon, T.E. (2004) Alternative NAD(P)H dehydrogenases of plant mitochondria. Annu Rev Plant Biol, 55, 23-39.
Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G., and Deber, C.M. (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci U S A, 106(6), 1760-1765.
Regnier, P., and Arraiano, C.M. (2000) Degradation of mRNA in bacteria: emergence of ubiquitous features. Bioessays, 22(3), 235-244.
Rhoads, D.M., Levings, C.S., III, and Siedow, J.N. (1995) URF13, a ligand-gated, pore-forming receptor for T-toxin in the inner membrane of CMS-T mitochondria. J. Bioenerg. Biomembr., 27, 437-445.
Rogers, J.S., and Edwardson, J.R. (1952) The utilization of cytolasmic male-sterile inbreds in the production of corn hybrids. Agron. J., 44, 8-13.
Saalaoui, E., Litvak, S., and Araya, A. (1990) The apocytochrome b from an alloplasmic line of wheat (*T. aestivum,* cytoplasm-*T. Timopheevi*) exists in two differently expressed forms. Plant Science, 66, 237-246.
Sabar, M., Gagliardi, D., Balk, J., and Leaver, C.J. (2003) ORFB is a subunit of F1F(O)-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. EMBO Rep, 4(4), 381-386.
Saha, D., Prasad, A.M., and Srinivasan, R. (2007) Pentatricopeptide repeat proteins and their emerging roles in plants. Plant Physiol Biochem, 45(8), 521-534.
Schagger, H.a.J.G.v. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Analytical biochemistry, 199, 223-231.
Schuster, G., Lisitsky, I., and Klaff, P. (1999) Polyadenylation and degradation of mRNA in the chloroplast. Plant Physiol, 120(4), 937-944.
Scott, R.J. (1994) Pollen exine: the sporopollenin enigma and the physics of pattern. In Molecular and Cellular Aspects of Plant Reproduction (R.J.S.a.M.A. Stead, ed, Cambridge Cambridge University Press, pp 49-81.
Sisco, P.H. (1991) Duplications complicate genetic mapping of Rf4, a restorer gene for cms-C cytoplasmic male sterility in corn. Crop Sci., 31, 1263-1266.
Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics, 4(6), 1581-1590.
Song, J., and Hedgcoth, C. (1994) A chimeric gene (orf256) is expressed as protein only in cytoplasmic male-sterile lines of wheat. Plant Mol Biol, 26(1), 535-539.

Steiglitz, H. (1977) Role of β-1,3-glucanase in postmeiotic microspore release. Dev. Biol., 57, 87-97.
Stern, D.B., and Newton, K.J. (1985) Mitochondrial gene expression in Cucurbitaceae: conserved and variable features. Curr Genet, 9(5), 395-404.
Tadege, M., Dupuis, I.I., and Kuhlemeier, C. (1999) Ethanolic fermentation: new functions for an old pathway. Trends Plant Sci, 4(8), 320-325.
Taiz, L., Zeiger, E. (2002) Plant Physiology. Sunderland, Mass.: Sinauer Associates Inc.
Tang, H.V., Pring, D.R., Shaw, L.C., Salazar, R.A., Muza, F.R., Yan, B., and Schertz, K.F. (1996) Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. Plant J, 10(1), 123-133.
Tzagoloff, A., Barrientos, A., Neupert, W., and Herrmann, J.M. (2004) Atp10p assists assembly of Atp6p into the F0 unit of the yeast mitochondrial ATPase. J Biol Chem, 279(19), 19775-19780.
Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat Genet, 15(1), 57-61.
Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y., and Liu, Y.G. (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell, 18(3), 676-687.
Weber, J., and Senior, A.E. (2003) ATP synthesis driven by proton transport in F1F0-ATP synthase. FEBS Lett, 545(1), 61-70.
Wen, L., and Chase, C.D. (1999) Pleiotropic effects of a nuclear restorer-of-fertility locus on mitochondrial transcripts in male-fertile and S male-sterile maize. Curr Genet, 35(5), 521-526.
Wise, R.P., Pring, D.R., and Gengenbach, B.G. (1987) Mutation to male fertility and toxin insensitivity in Texas (T)-cytoplasm maize is associated with a frameshift in a mitochondrial open reading frame. Proc. Natl. Acad. Sci. USA, 84, 2858-2862.
Woloszynska, M., Kmiec, B., Mackiewicz, P., and Janska, H. (2006) Copy number of bean mitochondrial genes estimated by real-time PCR does not correlate with the number of gene loci and transcript levels. Plant Mol Biol, 61(1-2), 1-12.
Woodson, J.D., and Chory, J. (2008) Coordination of gene expression between organellar and nuclear genomes. Nat Rev Genet, 9(5), 383-395.
Yagi, T. (1991) Bacterial NADH-quinone oxidoreductases. J Bioenerg Biomembr, 23(2), 211-225.
Yamamoto, M.P., Kubo, T., and Mikami, T. (2005) The 5'-leader sequence of sugar beet mitochondrial atp6 encodes a novel polypeptide that is characteristic of Owen cytoplasmic male sterility. Mol Genet Genomics, 273(4), 342-349.
Young, E.G., and Hanson, M.R. (1987) A fused mitochondrial gene associated with cytoplasmic male sterility is developmentally regulated. Cell, 50(1), 41-49.
Zabala, G., Gabay-Laughnan, S., and Laughnan, J.R. (1997) The nuclear gene Rf3 affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize. Genetics, 147(2), 847-860.
Zhu, D., and Scandalios, J.G. (1992) Expression of the maize MnSod (Sod3) gene in MnSOD-deficient yeast rescues the mutant yeast under oxidative stress. Genetics, 131(4), 803-809.

VITA

Louis Meyer was born in St. Louis, Missouri on August 19, 1978. He received his Bachelors of Science and Bachelors of Arts in Biology, along with minors in Chemistry and Philosophy at Truman State University Kirksville, Missouri. He then went on to obtain his Masters in Biology at the University of Missouri in Columbia, Missouri. Louis Meyer has currently accepted a post doctorial research position performing analytical biochemistry, working towards his goal of perusing a job in scientific research.