Molecular Characterization of *empty pericarp5* in Maize

LIU, Yujun

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Biology

The Chinese University of Hong Kong
June 2013

Thesis/Assessment Committee

Professor Honming LAM (Chair)

Professor Baocai TAN (Thesis Supervisor)

Professor Weichang YU (Committee Member)

Professor Yiji. XIA (External Examiner)

Statement

All the experiments reported in this thesis were performed by the author, unless specially stated otherwise in the content.

_____

LIU Yujun

Acknowledgements

I would like to express my deepest and sincerest gratitude to my supervisor, Dr. B.C. Tan, for offering me a chance to study at The Chinese University of Hong Kong and to carry out such a meaningful project. Thank you for all your help, guidance and encouragement throughout my graduate study. Your excellent view and great passion for science have sparked my interest in plant science and encouraged my decision to pursue an academic career in Biological Sciences.

I also would like to acknowledge Dr. H.M. Lam, Dr. W.C. Yu and Dr. Y.J. Xia for being on my committee and giving insightful advices on my project. Besides, my appreciations shall also go to Dr. Kenneth Cline (University of Florida, USA), Dr. Gregorio Hueros (Universidad de Alcalá, Spain), and Dr. Tsuyoshi Nakagawa (Shimane University, Japan) for their contributions to this work.

Thanks to all my friends and labmates in SC LG90. Most of you give me emotional and technical supports, and some training on my EQ. Discussions and debates with you give me much inspiration beyond the project. Special thanks go to Dr. Chunhui Xu who gave me suggestions in rice transformation, also to Dr. Qianfeng Li, who gave me a lot of help in these four years.

Many thanks go to my parents, for your love and support on my whole journey of growth. Most special thanks go to my dear husband, Liping Kuang. Without your love, encouragement and support, this could never have happened. I am so lucky to have you always stand beside me. Last thanks to my baby daughter, the coming of you is my greatest happiness in these years.

A million thanks go to all the people who helped me in accomplishing my Ph.D study.

Abstract

Seed development is a critical stage in the life cycle of flowering plants. Understanding the mechanism governing this process is both a fundamental question in plant biology and also an important task in agriculture application as seeds are staple food and seed quality and size are controlled by the genes governing seed development process. Maize as a typical monocot plant, is also an excellent model system for monocot seed development research.

In flowering plants, RNA editing is a post-transcriptional mechanism that converts specific cytidines to uridines in both mitochondrial and plastidial transcripts, altering the genetic information encoded by these genes. It is important for posttranscriptional regulation and in some cases critical to the functions of the encoded proteins. For example, editing can restore a conserved amino acid codon, create an initiation or stop codon, or remove a stop codon that leads to a functional larger protein. Therefore, deficiency in editing may result in a compromised or complete loss of function for the encoded proteins, leading to a severe consequence in plant growth and development.

In this study, we report the molecular characterization of the *empty pericarp 5* (*emp5*) mutant in maize (*Zea mays*). Null mutation of *Emp5* results in abortion of embryo and endosperm development at early stages. *Emp5* encodes a mitochondrion targeted DYW-subgroup PPR protein. Analysis of the mitochondrial transcripts reveals that loss of the EMP5 function abolishes the C-to-U editing of *rpl16*-458 (100% edited in the wildtype), decreases the editing at nine sites in *nad9, cox3* and *rps12*, and surprisingly increases the editing at five sites of *atp6, nad1, cob* and *rpl16*. EMP5 lacking the E+ and DYW domain still retains the substrate specificity and editing function, only at reduced

efficiency. This suggests that the E+ and DYW domains of EMP5 are not essential to the EMP5 editing function, but necessary for efficiency. Analysis of the ortholog in rice indicates that OsEMP5 has a conserved function in C-to-U editing of the rice mitochondrial *rpl16*-458 site. Knock-down expression of *OsEmp5* results in slow growth seedlings and defective seeds. These results demonstrate that EMP5 encodes a PPR-DYW protein that is required for the editing of multiple transcripts in mitochondria and the editing events, particularly the C-to-U editing at the *rpl16*-458 site, are critical to the mitochondrial functions and hence to seed development in maize.

摘要

种子发育是开花植物生命周期中的一个关键阶段。理解控制这一过程的机理既是植物生物学的重要基础研究课题，也是农业应用上的重要任务。因为种子是我们的主食，而种子质量和大小正是主要由那些控制种子发育过程的基因控制着。玉米作为一种典型的单子叶植物，是很好的研究单子叶植物种子发育的模式植物。

在开花植物中，RNA编辑是发生在线粒体和质体中的一种转录后机制，主要将特定的胞嘧啶转换为尿嘧啶。这一过程很多时候将改变基因组的遗传信息。这是一种非常重要的转录后调节机制，许多案例表明它对编码蛋白的功能十分关键。例如，RNA编辑能修复一个保守的氨基酸密码，创造一个起始或终止密码子，或移除一个终止密码子而编码出一个更大的功能蛋白。因此，RNA编辑不能正常进行可能损害或完全失去编码蛋白的功能，对植物的生长和发育造成严重后果。

本研究中，我们报道了玉米种子突变体*emp5(empty pericarp 5)*的分子特征研究。*Emp5*基因的无效突变导致玉米种子胚和胚乳发育在早期发育阶段停止。*Emp5*基因编码一个定位在线粒体的PPR-DYW蛋白。对*emp5*突变体线粒体转录组的分析表明，EMP5蛋白功能失效阻碍了*rpl16*-458（野生型中100%被编辑）这一位点的C-to-U RNA编辑，降低了*nad9, cox3*和*rps12*这三个转录本中总共9个位点的C-to-U RNA编辑水平。令人意外的是，同时也增加了*atp6, nad1, cob* 和 *rpl16* 4个转录本中共5个位点的RNA编辑水平。EMP5蛋白缺失E+和DYW结构域仍然保留了底物的特异性和RNA编辑功能，只是编辑效率有所降低。这表明EMP5蛋白的E+和DYW结构域对其编辑功能不关键，但对编辑效率是必需的。对EMP5在水稻中的同源蛋白的分析表明，OsEMP5在水稻线粒体*rpl16-458*位点的编辑功能是保守的。OsEMP5的基因沉默表达导致水稻植株生长缓慢及种子缺陷。这些结果表明*Emp5*基因编码的这一

PPR-DYW蛋白对多个线粒体转录本的RNA编辑是必需的。尤其是*rpl16*-458这一位点的编辑对线粒体的功能十分重要，因而对玉米种子发育非常关键。

Table of Contents

Thesis/Assessment Committee i

Statement ii

Acknowledgements iii

Abstract iv

摘要 vi

Table of Contents viii

List of Tables xii

List of Figures xiii

List of Abbreviations xv

Chapter 1. General Introduction 1

Chapter 2. Literature Review 6

2.1. Seed development in flowering plants 7

2.1.1 Seed morphogenesis of flowering plants 7

2.1.2 Molecular mechanisms of seed development 8

2.2 PPR protein family 11

2.2.1 Definition of PPR protein 11

2.2.2 Subgroups of PPR protein 12

2.2.3 PPR protein distribution and evolution 13

2.2.4 Functions of PPR proteins in plants 14

2.3 PPR proteins involving in seed development 15

2.4 PPR proteins studied in maize 17

2.5 RNA processing in plant mitochondria 19

2.5.1 RNA splicing 20

2.5.2 5’ and 3’ processing of plant mitochondrial transcrips 23

2.5.3 RNA stabilization 24

2.6 RNA editing 25

2.7 Status of C to U RNA editing mechanism 28

Chapter 3. Materials and methods 31

3.1 Plant Materials 32

3.2 Light Microscopy of Cytological Sections 32

3.3 Immunohistochemistry Analysis Materials 32

3.4 Isolation of genomic DNA Materials 33

3.5 Southern Analysis Materials 34

3.6 Inverse PCR cloning Materials 35

3.7 RNA Extraction and RT-PCR Materials 35

3.8 Subcellular Localization of EMP5 Protein Materials 36

3.9 Analysis of Mitochondrial RNA Editing Materials 37

3.10 Rice Transformation Materials 37

3.10.1 *OsEmp5* RNAi vector construction Materials 37

3.10.2 Callus induction from mature rice seeds Materials 38

3.10.3 Callus subculture Materials 38

3.10.4 Preparation of *Agrobacterium tumefaciens* Materials 38

3.10.5 Co-cultivation Materials 39

3.10.6 Callus washing and selection Materials 39

3.10.7 Regeneration Materials 40

3.10.8 Screening of transgenic plants Materials 40

Chapter 4. Results 42

4.1 Phenotypic and Genetic Characterization of *emp5-1* 43

4.2 Cloning of *Emp5* 45

4.3 *Emp5* Encodes a Mitochondrion-Targeted PPR-DYW Subclass Protein 48

4.4 Expression of *Emp5* 51

4.5 EMP5 is Required for Mitochondrial RNA Editing 52

4.6 Molecular Characterization of *emp5-4* Allele 54

4.7 Functional Analysis of the Rice *OsEmp5* Gene 58

Chapter 5. Discussion 62

5.1 Abortion of *emp5-1* Mutant Seed Development is Caused by Defective Mitochondrial RNA Editing 63

5.2 Increased Editing in the *emp5* Mutant 65

5.3 Substrate Specifying Sequence of EMP5 Are Not Conserved 66

5.4 The E+ and DYW Motif of EMP5 Is not Essential for This Protein Function ..67

Chapter 6. Conclusion 75

References 78

List of Tables

Table 1. Discovery of RNA editing in different genetic systems.....26

Table 2. Primers used in this study.....72

List of Figures

Figure 1. Subfamily and motif structure of PPR protein 13

Figure 2. Distribution of PPR genes among eukaryotes.....14

Figure 3. RNA processing in plant mitochondria.....20

Figure 4. Cis- and trans-splicing of *nad5* transcripts 21

Figure 5. Group II intron splicing mechanism 22

Figure 6. C-U, U-C and A-I RNA editing by conversion.. 27

Figure 7 Mutant *emp5-1* kernels abort early in seed development 43

Figure 8. Developmental comparisons of wild-type and *emp5-1* kernels at 8 and 13 DAP. 45

Figure 9. Effects of *emp5-1* on basal transfer layer.....46

Figure 10. *Emp5* gene cloning 47

Figure 11. EMP5 protein structure 49

Figure 12. Alignment of EMP5 protein with rice (*Oryza sativa*) OsEMP5 and a Sorghum (*Sorghum bicolor)* ortholog SbEMP5 protein 50

Figure 13. EMP5 protein localization 51

Figure 14. Expression pattern of *Emp5* in maize organs indicated.. 52

Figure 15. RNA editing defects of mitochondria genes in *emp5-1* mutant 56

Figure 16. *emp5-4* is a weak allele 57

Figure 17. OsEMP5 is a mitochondrion localized PPR-DYW protein 59

Figure 18. RNAi knock-down of *OsEmp5* expression in transgenic rice 60

Figure 19. Analysis of mitochondrial *rpl16* editing in *OsEmp5*-RNAi transgenic lines. 61

Figure 20. Alignment of -40 to +20 sequence of 15 RNA editing sites in 7 mitochondrial gene transcripts changed in *emp5-1*. 70

Figure 21. The hypothetical proteins as predicted based on the two transcripts detected in the *emp5-4* mutant seeds. 71

List of Abbreviations

cDNA	Complementary deoxyribonucleic acid
DAF	Day(s) after flowering
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
Dek	Defective kernel
dGTP	Deoxyguanine triphosphate
DNA	Deoxyribonucleic acid
dTTP	Deoxythymine triphosphate
E.coli	*Escherichia coli*
EB	Ethidium bromide
EDTA	Ethylenediaminetetraacetic acid
Emp	empty pericarp
EST	Expressed sequence tag
FAO	Food and Agricultural Organization of the United Nations
GST	Glutathione S-transferase
GUS	β-Glucuronidase reporter gene
HPT	*hygromycin phosphotransferase*
LB	Luria Broth
MORF	Multiple organellar RNA editing factor
NEB	New England BioLabs, Inc
OD_{600}	Optical density measured at 600nm
PCR	Polymerase chain reaction
PPR	Pentatricopeptide repeat
RNA	Ribonucleic acid
RNAi	RNA interference
RT-PCR	Reverse transcription polymerase chain reaction
TPR	Tetratricopeptide repeat
UTR	Untranslated region

Chapter I.
General Introduction

Seed development is a critical stage in the life cycle of flowering plants. Understanding the mechanism governing this process is both a fundamental question in plant biology and also an important task in agriculture application. This importance is due to the fact that seeds are staple food, and major seed traits are controlled by genes that define the seed developmental process.

As a staple food and feed stalk, maize is an important crop around the world. According to FAO (Food and Agriculture Organization of the United Nations) statistics in 2011, world maize production is ~840 million tons, ranking at No. 1 among all cereal crops. Maize as a typical monocot, is also an excellent model system for monocot seed development research. First, comparing with other crops such as rice and wheat, seed phenotype observation and cross in maize are convenient. Second, as a traditional genetic research material, maize has richer genetic resources than other monocots. In addition, the maize genome is sequenced (Schnable et al., 2009), which also provides genome information for maize molecular biology study. Maize transformation is feasible and becoming easier as the technique advances rapidly in this area.

In recent years, much genetic, molecular and biochemical evidence suggests that pentatricopeptide repeat (PPR) protein family may have essential functions in seed development. PPR proteins are defined by the presence of a canonical 35-amino-acid motif repeated up to 30 times in tandem (Small and Peeters, 2000). PPRs belong to a large family of genes that are particularly prevalent in terrestrial plants. Studies have showed that these proteins mainly localize in plastids and mitochondria and may have a range of essential functions in posttranscriptional processes in these organelles, including RNA stabilization, editing, splicing, and translation (Schmitz-Linneweber and Small, 2008).

Loss of function mutations in PPR genes result in defective seed phenotypes, implying that PPR genes play essential roles in plant seed development. Mutations in the *Arabidopsis PPR2* gene causes defects in cell proliferation during embryogenesis (Lu et al., 2011). Arabidopsis PPR gene *GRP23* is essential for early embryo

development through interaction with RNA Polymerase II Subunit III (Ding et al., 2006). PPR gene mutation is a common cause for defective embryogenesis in Arabidopsis. Among 250 Arabidopsis embryo defective mutants (*emb*), seventeen (6.8%) have been shown to be impaired in putative PPR genes (Tzafrir et al., 2004; Cushing et al., 2005). In rice, PPR gene *OGR1* is also involved in seed development (Kim et al., 2009). OGR1 functions in the editing of five mitochondrial transcripts. Mutant *ogr1* seeds develop opaque endosperm and are slightly smaller in width and thickness than the WT. In maize, three PPR genes are shown to be essential for seed development, all of which are localized in mitochondria. *Empty pericarp4* (*Emp4*) is necessary for early seed development and is shown to be required for the correct expression of a small subset of mitochondrial genes (Gutierrez-Marcos et al., 2007). PPR2263 is the only reported PPR protein thus far that is required for RNA editing of mitochondrial transcripts in maize. Loss of PPR2263 function reduces both the embryo and endosperm size, hence the kernel size (Sosso et al., 2012b). A recently reported mitochondrial targeted PPR protein, MPPR6, was shown to be essential for 5' maturation and translation initiation of *rps3* mRNA. Mutation in MPPR6 coincides with abnormal formation of transfer cells, delay of embryo development, and reduction of starch accumulation (Manavski et al., 2012). These multiple lines of evidence suggest that PPR genes are implicated in functions that are essential to seed development in flowering plants.

Most of the PPR proteins affecting seed development are localized in mitochondria or chloroplasts and function through RNA processing. RNA editing is one of the most important RNA processing processes, the last step to modify genetic information after transcription by alteration of RNA sequences via insertion, deletion and conversion of nucleotides. RNA editing in plants was discovered by three independent groups reporting C-to-U changes in mitochondrial mRNAs of wheat (*Triticum aestivum*) and evening primrose *(Oenothera berteriana*） at the same time in 1989 (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989). Plant RNA editing includes C-to-U, U-to-C and A-to-I editing. Rare A-to-I editing occurs

on cytosolic tRNAs, and the majority of editing in plants occurs in mitochondrial and plastid transcripts. Compared with fewer than 50 editing sites in the chloroplast transcripts, 300–500 editing sites have been detected in the mitochondrial transcripts in flowering plants (Chateigner-Boutin and Small, 2010).

Although RNA editing is an important step in post-transcriptional control of organelle gene expression (Shikanai, 2006; Grennan, 2011), little is known about the mechanisms of C to U RNA editing in plants until 2005. The breakthrough concerning the nature of editing factors was made by Dr. T. Shikanai's group in the analysis of chlororespiratory mutant *CHLORORESPIRATORY REDUCTION4* (*crr4*). *Crr4* encodes a PPR protein that is required for editing of a single site within the chloroplast transcript *ndhD* (Kotera et al., 2005). Since then, more genes involved in editing have been identified, such as *OGR1* in rice, *MEF1, MEF9, CRR21, CLB19* and *YS1* in *Arabidopsis* (Okuda et al., 2007; Chateigner-Boutin et al., 2008; Kim et al., 2009; Zehrmann et al., 2009; Zhou et al., 2009; Takenaka, 2010). Mutations in all these genes impair the editing of one or a few editing sites. All of these genes encode PPR proteins, strongly indicating that PPR proteins are important specificity editing factors. The sequences surrounding individual editing sites have been tested both in vivo and in vitro for both mitochondria and chloroplasts. These experiments revealed that, in most cases, the cis-element required for editing was contained in the region -25/+10 compared to the editing sites (Chaudhuri and Maliga, 1996; Okuda et al., 2006). Recently, Takenaka's group identified another type of proteins essential for organelle RNA editing, named multiple organellar RNA editing factor (MORF) (Takenaka et al., 2012). Loss of MORF function abolishes or lowers editing at multiple sites. This feature distinguishes MORFs from PPR proteins which typically affect only one or a few editing sites.

Although accumulating evidence suggests that PPR genes may have essential functions in seed development, many important questions remain to be answered. First, since PPR belongs to a large family of RNA-binding proteins, how these PPR proteins recognize such a vast different RNA targets with such specificity? Which

motifs are essential and which one is not important for its function? Also, no protein structure information is known about PPR proteins. The little detail of PPR proteins involved in the regulation of seed development is known. So, more studies are needed to further understand the detailed molecular mechanisms of this process. In this PhD thesis, I focused on the molecular characterization of an empty pericarp mutant (*emp5*) in maize. The *emp5* mutant arrests both the embryo and the endosperm development. Cloning by transposon tagging indicates that *Emp5* encodes a DYW subgroup PPR protein that is localized in mitochondria. Functional analysis indicates that EMP5 is involved in the C-to-U editing of multiple sites in several transcripts. Absence of EMP5 causes a complete loss of editing in the *rpl16*-458 site, causing a change from leucine to proline in the protein in the mutant. Further analysis of the transposon insertional allele suggests that the DYW domain of EMP5 is not essential for the editing function but required for efficiency. The EMP5 function is conserved in rice as knock-down expression of rice *Emp5* causes similar seed developmental defects and reduced editing on rice *rpl16*-458. This work defines that function and mechanism of a key protein involved in seed development in maize and shed lights on the function of DYW domain that is associated with editing function in all land plants.

Chapter II.

Literature review

2.1 Seed development in flowering plants

2.1.1 Seed morphogenesis of flowering plants

Seed development is a complex and multistage process, which is initiated by double fertilization in all flowering plants. Double fertilization involves two events, fusion of an egg cell with one sperm cell resulting in a diploid zygote, and fusion of two-nucleus central cell with the other sperm cell, leading to the production of the triploid endosperm. In Arabidopsis and many other dicots, embryo development is characterized by four stages, i.e. globular stage, heart stage, linear cotyledon stage and mature green stage embryo (Schulz and Jensen, 1971). The endosperm development involves three stages. In the first phase of endosperm development, the fertilized central cell undergoes rapid proliferation and expansion without cell division to generate a large and multinucleated cell defined as the syncytial phase and results in a large increase in the size or volume of a seed cavity (Olsen, 2001). Through cellularization process, this syncytium is then divided into individual cells. However, in the second phase, when embryo reaches its heart stage, the embryo grows to fill the cavity and replaces most of the endosperm volume at its maturity, because the endosperm is consumed to provide the nutrients required for embryo development. At mature stage, the endosperm only contains a single layer cells in Arabidopsis. The maternal integument ultimately becomes the seed coat (Olsen, 2001; Berger, 2003).

For maize and most of monocots, early endosperm development is similar with Arabidopsis, also undergoes coenocytic and cellularization stages. However, distinguishing from Arabidopsis, after cellularization stage, the maize endosperm continues to differentiate to four parts that include starch endosperm, embryo surrounding region, aleurone cell layer and transfer cell. Finally, through maturation stages, endosperm represents the main storage component of the seed and persists at maturity. Endosperm serves the function of reservoir providing nutrients to the young seedling upon germination. The embryo development in maize is characterized by three main stages, transition, coleoptilar and late embryogenesis stage. During the first

phase the basal-apical asymmetry of the embryo is established and the embryo is regionalized into suspensor and embryo proper; during the second phase radial asymmetry appears and the embryonic axis and meristem are established; and during the third stage vegetative structures are elaborated (Sheridan and Clark, 1993).

2.1.2 Molecular mechanisms of seed development

After many years of research on seed morphogenesis, recent studies via either forward or reverse genetics start to focus on many genes that are involved in the molecular regulation of seed development. Recent techniques such as global transcript analysis facilitate the identification of regulatory networks that are important for programming seed development (Bemer et al., 2010; Le et al., 2010). A global analysis of gene activity with Affymetrix GeneChips during Arabidopsis seed development identified 289 seed-specific genes, including 48 transcription factors (Le et al., 2010). MADS-box protein family defines a fairly large group of transcriptional factors that are involved in Arabidopsis flower and fruit development. Transgenic plants with translational the type I MADS-box genes fusion with GFP and GUS were generated. This analysis yielded expression profiles for 38 genes. All these genes expressed in the female gametophyte and developing seed. These results indicate that the entire type I subfamily is involved in seed development in Arabidopsis (Bemer et al., 2010).

A number of mutants that impair seed development also have been studied. The mutation of *transparent testa glabra*2 (*ttg2*) resulted in destroyed endosperm and seed growth through prevention of cell elongation in the integument (Garcia et al., 2005). Similarly, except for the reduction of integument cell elongation, two *haiku* (*iku*) mutations also cause early endosperm cellularization, reduced proliferation of endosperm, and reduced embryo development (Garcia et al., 2003). *HAIKU2 (IKU2)* is expressed only in the endosperm but not in the embryo. It encodes a leucine-rich repeat (LRR) receptor kinase. Because *ttg2* showed defective seed integument and iku mutations showed reduction of endosperm size, these genes were considered as the

primary regulators of seed size. The mutant of *MINISEED3 (MINI3)*, a WRKY family gene, has similar small seed phenotypes and similar patterns of altered seed development as two *haiku* mutants (*iku1* and *iku2*), producing embryos and integument cell dimensions smaller than wild type (Luo et al., 2005). These three genes, *IKU1, IKU2, and MINI3*, are implicated in the same pathway in controlling seed development (Luo et al., 2005). The homeobox-leucine zipper gene *PHABULOSA (PHB)*, putative transcriptional regulator *INNER NO OUTER (INO)*, and gene *WUSCHEL (WUS)* involve in pattern formation during early ovule development in *Arabidopsis thaliana* (Sieber et al., 2004). Mutations in these genes result in defective seeds due to their deleterious impact on integument initiation.

Several genes disrupting embryo pattern formation have also been described. The *GURKE* gene is required for normal organization of the apical region in the embryo and may also play a role during postembryonic development. Mutations in the *GURKE* gene give rise to seedlings with highly reduced or no cotyledons (Torres-Ruiz et al., 1996). *SIN1* (*SHORT INTEGUMENTS1*) gene is also essential for embryo pattern formation and viability in Arabidopsis. The postembryonic activity of this gene is required for flowering time control and ovule morphogenesis process in reproductive development. The *SIN1* gene (At1g01040), identical to previously identified *CAF* (*CARPEL FACTORY*) gene, is important for normal flower morphogenesis. The *SUS1* (*SUSPENSOR1*) gene is essential for embryogenesis. Probably through specific mRNA posttranscriptional regulation, the maternal *SIN1/SUS1/CAF* gene functions in early Arabidopsis development. (Golden et al., 2002).

A large group of polycomb (PcG) proteins are involved in epigenetic regulation of endosperm development, which includes FERTILIZATION INDEPENDENT SEED 2 (FIS2), FERTILIZATION-INDEPENDENT ENDOSPERM (FIE/FIS3), MEDEA (MEA/FIS1), MULTICOPY SUPRESSOR OF IRA (MSI1), and SWINGER (SWN) (Kohler and Makarevich, 2006; Pien and Grossniklaus, 2007). These five proteins form polycomb repressive complexes that suppress gene expression through

DNA and histone methylation. Loss of their function results in defective seed formation in the mutants. For example, the mutation of FIE gene specifically affects the central cell, allowing the replication of the central cell nucleus and endosperm development without fertilization but lacks endosperm cellularization (Kohler and Makarevich, 2006). After fertilization, embryo development of the corresponding mutants is arrested and the mutant embryo displays various defects in cell proliferation and morphogenesis (Pien and Grossniklaus, 2007).

All of the above seed development research is performed in dicot model plant Arabidopsis. In monocot model plant maize, several mutations arresting seed development at defined stages have been described, such as the defective kernel (*dek*) and empty pericarp (*emp*) mutants which affect development of both embryo and endosperm (Neuffer and Sheridan, 1980), and the endosperm or embryo specific (*emb*) mutants, in which development of the endosperm or embryo is profoundly altered without disrupting the other part (Sheridan and Clark, 1993; McCarty et al., 2005). However, due to the lethality and difficulties in gene cloning of these maize mutants, only a few of these mutants have been characterized at molecular level. Mutant *empty pericarp2* (*emp2*) arrests embryo development at the early stage of embryogenesis, revealed a heat shock response mechanism in the seed (Fu et al., 2002). EMP2 is an essential negative regulator of the heat shock transcriptional response. Loss of EMP2 function results in an unattenuated heat shock response that causes embryo lethality in the maize kernel. The *defective kernel 1* (*dek1*) gene is required for aleurone cell development in the maize endosperm. *Dek1* encodes a membrane protein of the calpain gene superfamily that is presumably involved in protein cleavage or degradation (Becraft et al., 2002; Lid et al., 2002; Wang et al., 2003). The *ZmPRPL35* gene encodes a plastid ribosomal protein, and a lesion in ZmPRPL35 causes the *emb* phenotype. It suggests a link between plastids and embryo-specific signaling events (Magnard et al., 2004). Another *emb* mutant, *lethal embryo 1* (*lem1*), is a result of an Ac insertion in *lem1* gene which encodes a plastid ribosomal protein PRPS9 (Ma and Dooner, 2004). Recently, a new emb mutant, *emb12*, was reported (Shen et al., 2013).

This *Emb12* encoded the plastid IF3 is essential to embryogenesis, but not to endosperm development in maize.

Three PPR proteins, EMP4, PPR2263 and MPPR6 were reported to be involved in maize seed development through different molecular mechanism (Gutierrez-Marcos et al., 2007; Manavski et al., 2012; Sosso et al., 2012a). For example, PPR2263 is required for RNA editing in maize mitochondria, and mutant ppr2263 seeds showed reduction in kernel size involved in both the embryo and endosperm (Sosso et al., 2012b). The maize PPR protein MPPR6 is reported to be directly involved in the 5' maturation and translation initiation of *rps3* mRNA. Both the embryo and endosperm development are arrested in the *mppr6* muant (Manavski et al., 2012). Because of the large family of PPR proteins and the large class of *emp* mutants (McCarty et al., 2005), further research is strongly needed to reveal the mechanism by which the seed development is regulated in maize and other crops. As seeds are directly related to cereal crops, the information potentially shed lights on bioengineering of crops with increased yield in agriculture.

2.2 PPR protein family

2.2.1 Definition of PPR protein

In 2000, Dr. Ian Small's group discovered a large family of proteins that has not been realized before by analyzing the Arabidopsis Genome. The key feature of these proteins is the presence of a tandem array of degenerate 35-amino-acid repeats (Small and Peeters, 2000). Most of these proteins in Arabidopsis are predicted to be targeted to either mitochondria or chloroplasts. This protein family was also identified independently on the basis of other criteria by Aubourg group in the same year, and named AtPCMP (Aubourg et al., 2000). Initially, the repeat structure of these proteins were overlooked, and considered as TPR (tetratricopeptide) motifs (Fisk et al., 1999) which is caused largely by the known presence of TPR in animals. In fact, although

the 35-amino-acid repeats do resemble TPR motifs, there is almost no overlap in the sets of sequences returned by MOTIFSEARCH using the TPR and this new type profiles. The protein secondary structure analysis also revealed that they have significant and characteristic differences. To distinguish them from TPR motifs, this new type 35-amino-acid repeats are called PPR (pentatricopeptide) motifs (Small and Peeters, 2000).

2.2.2 Subgroups of PPR protein

According to the number of amino acids in a PPR motif and whether there is gap between PPR motifs, PPR motifs can be divided into classic PPR and PPR-like motifs, which includes PPR-like S (for short) and PPR-like L (for long) motifs (Lurin et al., 2004). Many PPR proteins contain sequences unrelated to PPR motifs either before or after the tandem array of PPR motifs. The sequences before are mostly comprised of organelle targeting sequences. However, the sequences following the PPR motifs show considerable sequence conservation in many cases. The polypeptide sequences C-terminal of the last PPR-related motif from all of the Arabidopsis PPR proteins were aligned, three C-terminal motifs (E, E+, and DYW) were identified. Interestingly, these C-terminal motifs were only found in the subfamily defined by PPR-Like (PLS) repeats. The identification of these motifs allows classification of PPR proteins into five subclasses: (1) All classic PPR (P) protein without C-terminal motif. (2) PPR-like (PLS) proteins that do not have any of the three C-terminal motifs, (3) PPR-like proteins with the E motif alone, (4) PPR-like proteins with E and E+ motifs, and (5) PPR-like proteins with the E, E+, and DYW motifs (Figure 1), named the P, PLS, E, E+, and DYW subgroups, respectively (Lurin et al., 2004).

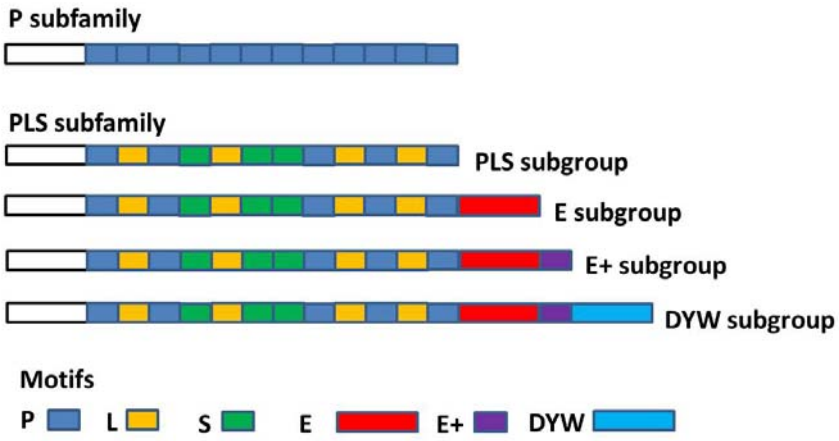

Figure 1. Subfamily and motif structure of PPR protein. The classification of proteins and nomenclature of motifs are based on Lurin et al. (Lurin et al., 2004).

2.2.3 PPR protein distribution and evolution

The PPR protein family is most prevalent in terrestrial plants. The Arabidopsis thaliana genome encodes more than 450 and vitis about 600 PPR proteins, whereas *Drosophila* and human genome contains only two and 6 PPR proteins respectively (Figure 2) (Lurin et al., 2004; Schmitz-Linneweber and Small, 2008). Analysis of expressed sequence tag (EST) data from a range of angiosperms reveals that PPR genes are very popular in terrestrial plants. In gymnosperms and bryophytes, PPR genes are also abundant. By contrast, the algal, yeast and animal genomes sequenced so far contain relatively few PPR genes, even no PPR genes. (Schmitz-Linneweber and Small, 2008).

Bioinformatics analysis as well as experimental studies on a set of PPR proteins have showed that PPR proteins are mainly targeted to plastids and mitochondria. This finding would suggest a mitochondrial or plastid origin of the PPR motif. However, no organelle genomes are found to encode a PPR protein, nor the bacterial genomes that most closely resemble the ancestral endosymbiont (Schmitz-Linneweber and Small, 2008). Hence, a symbiotic origin of PPR motifs is unlikely. Analysis of different genomes revealed a strong correlation between the number of RNA

processing events and the number of PPR genes present in the genomes (Fujii and Small, 2011). This correlation implies a co-evolutional relationship, *i.e*. an increasing number of PPR genes in early land plants evolves with increasing RNA processing events. But how the PPR genes evolve so rapidly is an enigma that remains to be solved. To achieve this goal, large amounts of new sequences from key species and comparative genomics are required to improve our understanding of PPR protein distribution and evolution.

2.2.4 Functions of PPR proteins in plants

The RNA-binding activity has been proved in several PPR proteins through in vivo and in vitro studies, such as in the maize plastid PPR proteins CRP1, PPR4 and PPR5 (Schmitz-Linneweber et al., 2005; Schmitz-Linneweber et al., 2006; Beick et al., 2008), the Arabidopsis plastid HCF152, CRR4 (Nakamura et al., 2003; Okuda et al., 2006), and rice mitochondrial Rf1 (Kazama et al., 2008).

Rapid progress has been made recently in understanding the functions of PPR proteins at the molecular level. Increasing number of studies have showed that PPR

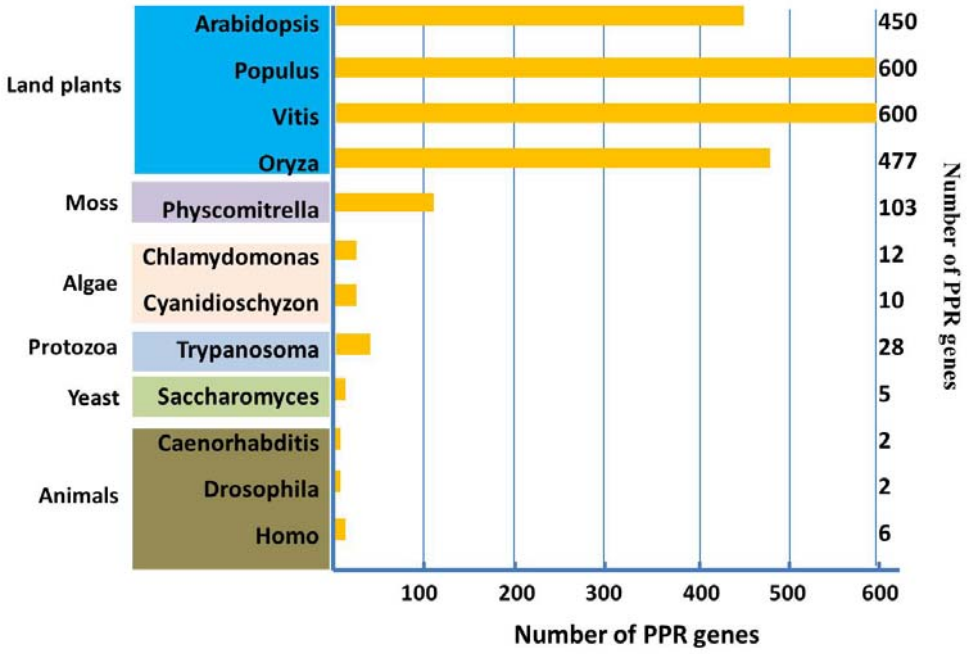

Figure 2. Distribution of PPR genes among eukaryotes (Schmitz-Linneweber and Small, 2008).The total number of PPR genes in each organism is given on the right.

proteins have a range of essential functions in mitochondria and plastid posttranscriptional processes, including RNA stabilization, editing, splicing, and translation (Schmitz-Linneweber et al., 2008). In land plants, chloroplast and mitochondrial gene expression is both independent and dependent. PPR proteins might be one of the major families of nuclear-encoded proteins involved in crucial functions of organelle gene expression processes. One line of the earliest evidence of PPR proteins regulating chloroplast mRNA is exhibited in the maize nuclear gene *CRP1*, which is required for the RNA stabilization and translation of *petD* and *petA* mRNAs from a polycistronic precursor (Barkan et al., 1994; Fisk et al., 1999). PPR10, a nuclear encoded chloroplast PPR protein, can bind to the *atpI–atpH* intergenic region, and is also involved in the RNA stabilization and translation (Prikryl et al., 2011). CRR4 is a site recognition factor in chloroplast RNA editing, mutation in which results in loss of RNA editing that is aimed for creating the translational initial codon of the plastid *ndhD* gene (Kotera et al., 2005). MEF1 is also involved in RNA editing of three specific sites of mitochondrial mRNAs in *Arabidopsis thaliana*, *rps4*-956, *nad7*-963 and *nad2*-1160 (Zehrmann et al., 2009). ZmPPR5 is also bound in vivo to the unspliced precursor of *trnG*-UCC for stabilizing the trnG-UCC precursor by directly binding and protecting an endonuclease-sensitive site (Beick et al., 2008). HCF152 is involved in the splicing of *petB* RNA and the stabilization of the spliced product (Meierhoff et al., 2003). Another protein PPR4 facilitates the trans-Splicing of the maize chloroplast *rps12* Pre-mRNA (Schmitz-Linneweber et al., 2006).

2.3 Functions of PPR proteins in seed development

At the physiological level, PPR proteins have been shown to play an important role in several developmental processes in plants. Different mutant phenotypes were observed through mutagenesis studies mainly in Arabidopsis and maize. For example, *Rf* gene mutation will result in cytoplasmic male sterility in several species. These

fertility restorer genes, *Rf* of petunia, *Rfk1* (*Rfo*) of radish, *Rf1* (*PPR13*) of sorghum and *Rf-1* of rice, all are found to code for PPR proteins (Saha et al., 2007). These findings lead to a conclusion that PPR proteins are involved in fertility restoration in CMS plants. PPR proteins are also important for chloroplast development and normal plant growth. Mutation of several chloroplast-localized PPR proteins results in yellow leave or albino leaf phenotype, such as PPR2, PPR4, PPR5 and PPR10 in maize (Williams and Barkan, 2003; Schmitz-Linneweber et al., 2006; Beick et al., 2008; Pfalz et al., 2009).

At present, a number of PPR genes have been reported that are essential for seed development. By systematic genetic screens for embryo-lethal mutations, the Arabidopsis SeedGenes project identified a number of embryo defective mutants with T-DNA insertions in PPR genes (Tzafrir et al., 2003), suggesting that PPR genes have important functions in embryogenesis. In addition, functions of PPR proteins are required for normal embryogenesis at various stages. Mutant *emb175* of *Arabidopsis* showed embryo development arrested at a very early stage, before globular to heart transition (Cushing et al., 2005). In another mutant with T-DNA insertion in At1g53330, embryo development is arrested during the transition from the globular to heart stage. This gene also codes for a PPR protein. (Kocabek, et al., 2006). The Arabidopsis PPR protein GLUTAMINE-RICH PROTEIN23 (GRP23) is involved in embryo development (Ding et al., 2006). The *grp23* mutant showed embryonic growth arrest at the 16-celled dermatogen state, an early stage of embryo development. GRP23 is the only known PPR protein that is localized in the nucleus. The GRP23 protein was found to interact physically with subunit III of RNA polymerase II through its C-terminal Gln-rich WQQ domain. This protein also can bind directly to cis-regulatory elements of DNA through its N-terminal basic domain (Ding et al., 2006). Mutations in the Arabidopsis *PPR2* gene have been found to cause defects in cell proliferation during embryogenesis (Lu et al., 2011).

A Rice PPR gene, *OGR1* is also involved in seed development, through function in RNA editing of five mitochondrial transcripts. Mutant *ogr1* seeds showed opaque

endosperm and were slightly smaller in width and thickness than the WT (Kim et al., 2009). Another mitochondria-targeted PPR protein OTP43 is required for trans-splicing of intron 1 in mitochondrial *nad1* transcript in NADH Complex I (De Longevialle et al., 2007). Arabidopsis mutant plants with a disrupted *OTP43* gene do not possess detectable mitochondrial Complex I activity and show severe defects in seed development and germination (De Longevialle et al., 2007).

Three PPR proteins in maize were shown to be essential for seed development, all of which are localized in mitochondria. Mutations in *Empty pericarp4* (*Emp4*), a maize PPR-encoding gene, confer an embryo-lethal phenotype (Gutierrez-Marcosa, et al., 2007). A loss of *Emp4* function causes a highly irregular differentiation of transfer cells in the nutrient-importing basal endosperm region. Molecular analyses suggest that EMP4 is required to regulate the correct expression of mitochondrial genes, *rps2A, rps2B, rps3/rpl16*, and *mttb* (*orfX*) in the endosperm (Gutierrez-Marcosa, et al., 2007). The other two are PPR2263 and MPPR6. PPR2263 is the only reported PPR protein required for RNA editing in maize mitochondria, and MPPR6 is directly involved in 5' maturation and translation initiation of *rps3* mRNA. Absence of any of the two genes causes reduction in the embryo and endosperm size (Manavski et al., 2012; Sosso et al., 2012a). These results indicate that PPR genes play important roles in organnele gene expression which is critial to seed development.

2.4 PPR proteins studied in maize

Most of the PPR proteins characterized thus far are in Arabidopsis, some in moss and rice. In maize, only a few PPR proteins are studied although the genome is estimated to harbor more than 450 PPR genes (Fujii and Small, 2011). The pioneering work in PPRs is on the chloroplast RNA processing1 (CRP1) in maize. CRP1 is required for the stabilization and processing of *petB* and *petD* mRNAs from its polycistronic precursor. This protein also activates the translation of another chloroplast mRNA *petA*. In *crp1* mutants, these mRNAs were absent which resulted in the loss of

cytochrome b6f complex, one of the four major multiprotein complexes in the chloroplast (Barkan et al., 1994; Fisk et al., 1999; Schmitz-Linneweber et al., 2005). PPR5 and PPR10 are also essential for RNA stability (Beick et al., 2008; Prikryl et al., 2011). PPR5 protein can stabilize trnG-UCC tRNA precursor in maize chloroplasts. RIP-Chip and coimmunoprecipitation assays identified *atpH* and *psaJ* RNA associated with PPR10 protein in chloroplasts. Further studies revealed that PPR10 can bind with specificity to the consensus sequence in the *atpH* 5-UTR and *psaJ* 3-UTR in vivo and in vitro to stabilize these two RNAs. PPR4 is required for *rps12* trans-splicing in chloroplasts (Schmitz-Linneweber et al., 2006). RIP-Chip analysis revealed that PPR4 is associated with the first intron of the plastid *rps12* pre-mRNA in vivo, and poisoned primer extension analysis demonstrated a loss of trans-spliced *rps12* RNA in the ppr4 mutants. PPR2 is involved in plastid ribosome accumulation (Williams and Barkan, 2003). *ppr2* mutants develop albino leaves and are deficient in plastid rRNA and translation products. The population of plastid transcripts is similar to other maize mutants lacking plastid ribosomes. This result suggests that PPR2 functions in the synthesis or assembly of one or more components of the plastid translation machinery. For RNA processing, THA8 is required for the splicing of *ycf3*-2 and *trnA* group II introns in angiosperm chloroplasts (Khrouchtchova et al., 2012). THA8 is a short PPR protein with only four PPR motifs. ATP4, belonging to a subclass of PPR proteins that contain a small MutS-related (SMR) domain, is necessary for the translation of the chloroplast *atpB* open reading frame (Zoschke et al., 2012). ATP4 associates *in vivo* with the sequences near the 5' end of the unusually long 5'UTR of *atpB/E* mRNA, where it facilitates ribosome association, This mechanism is required for the efficient translation and accumulation of the chloroplast ATP synthase in planta.

All the above characterized maize PPR proteins are targeted to chloroplasts. Only three PPR proteins that have been characterized so far are localized in maize mitochondria. The first one is EMP4, which is necessary for normal endosperm development and shown to be required for the correct expression of a small group of

mitochondrial genes (Gutierrez-Marcos et al., 2007). The second one is PPR2263, which is the only reported PPR protein required for RNA editing in maize mitochondria, at sites *nad5*-1550 and *cob*-908 (Sosso et al., 2012b). MPPR6 was recently reported which is directly involved in the 5' maturation and translation initiation of *rps3* mRNA. Absence of MPPR6 results in malformed transfer cells, delayed embryo development and reduced starch accumulation (Manavski et al., 2012).

2.5 RNA processing in plant mitochondria

Not only in higher plants, mitochondria are indispensable for all eukaryotic organisms. Mitochondria are the essential compartments for many metabolic pathways and most importantly the energy production by oxidative phosphorylation. Most plant mitochondrial genomes contain 54-60 known genes, encoding a set of about 30-35 proteins, 15-20 tRNAs and 3 rRNAs. The set differs slightly among species, but the encoded products are exclusively involved in ATP production, either directly or indirectly. The genes are usually spread over the circular genome, giving rise to mono- or polycistronic primary transcripts. After transcription, the primary transcripts undergo a series of processing steps including RNA editing, splicing, and maturation of secondary 5' and 3' ends, until they obtain their final mature form (Figure 3). RNA stabilization also seems to influence gene expression, bringing RNA degradation processes into the realm of gene expression regulation. To understand the mechanisms of these processes, important progress has been made in recent years. In the next section, I will summarize the recent advances in these processes, except RNA editing for which I will discuss extensively in later sections.

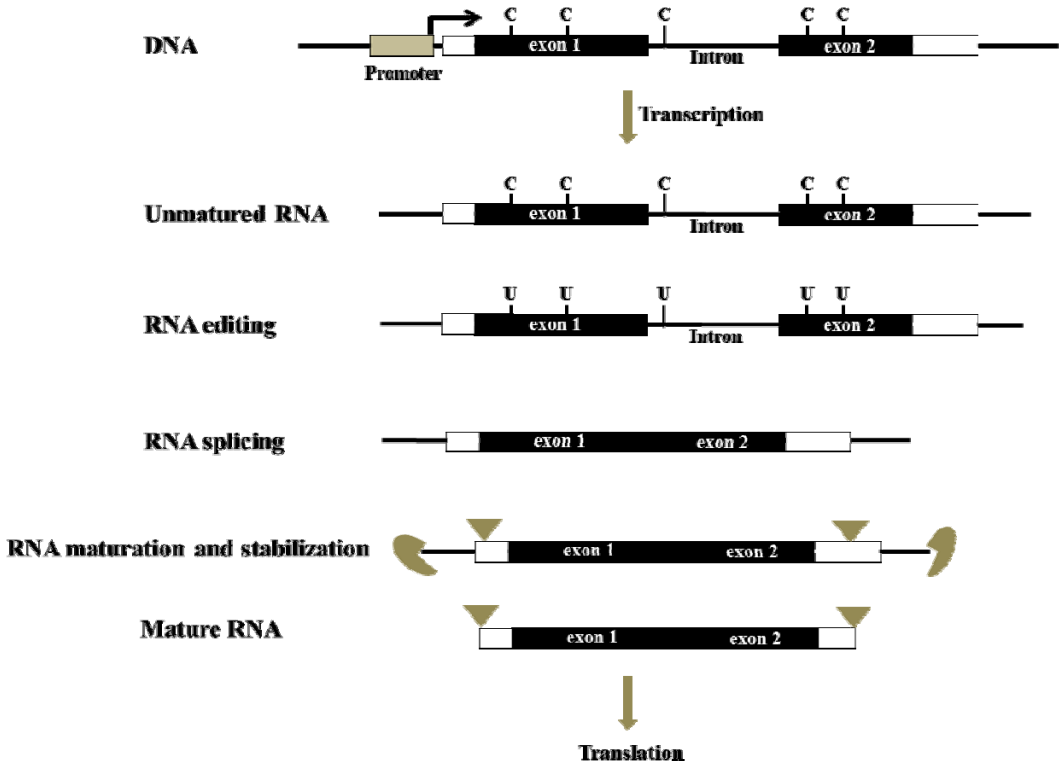

Figure 3. RNA processing in plant mitochondria.

2.5.1 RNA splicing

Most plant mitochondrial introns are arranged in a cis configuration, which means the intron and flanking exons are encoded as a single RNA molecule. However, trans-splicing also exists in plant mitochondrial introns splicing. For example, the *nad1, nad2, nad5* and *rps3* genes are fragmented in the maize mitochondrial genomes. The resulting RNA molecules, transcribed from more than one primary transcript, are joined by *cis*- and *trans*-splicing (Figure 4).

Almost all the mitochondrial introns in seed plants fall in the group II category, although several genera of land plants also contain a group I intron located in the *cox1* gene (Cho et al., 1998). As such, the splicing mechanism mainly involves group II intron splicing mechanism, which involves two transesterification steps. First, the 2' hydroxyl of an internal bulged adenosine near 3' end of the intron attacks the 5' splice site and forms a 2'-5' phosphodiester bond at the branch point. Second, a nucleophile

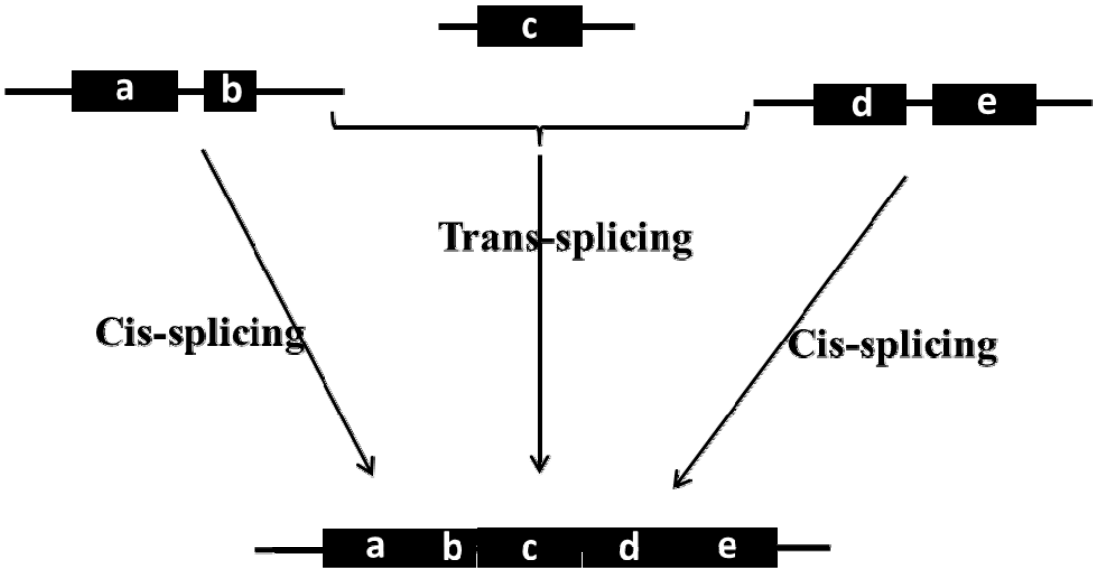

Figure 4. Cis- and trans-splicing of *nad5* transcripts.

Mature *nad5* mRNAs are generated from three precursor molecules via two cis- and two trans-splicing events.

attack from the liberated 3' hydroxyl group of the first exon occurs on the 3' splice site. The intron is thus released as a lariat and two exons are ligated (Figure 5a). In some plant mitochondrial introns, the bulged adenosine is missing, and therefore an alternative splicing pathway will operate. In this pathway, the first step is that water acts as the first attacking nucleophile rather than a transesterification, the second step remaining unchanged. The intron is thus excised as a linear form in this pathway (Figure 5b).

Additional protein factors have also been demonstrated to be involved in the splicing process in plant mitochondrial introns. MatR, which encoded by intron 4 ORF of *nad1* gene, is highly likely involved in the splicing as a maturase (Wahleithner et al., 1990; Lambowitz and Zimmerly, 2004). Other protein factors are encoded in the nucleus of higher plants. MS1 is the first proof of a nuclear-encoded factor involving in a splicing event in plants (Brangeon et al., 2000). Four other nuclear genes, which are conserved between Arabidopsis and rice, encode maturase-related proteins predicted to be imported into mitochondria

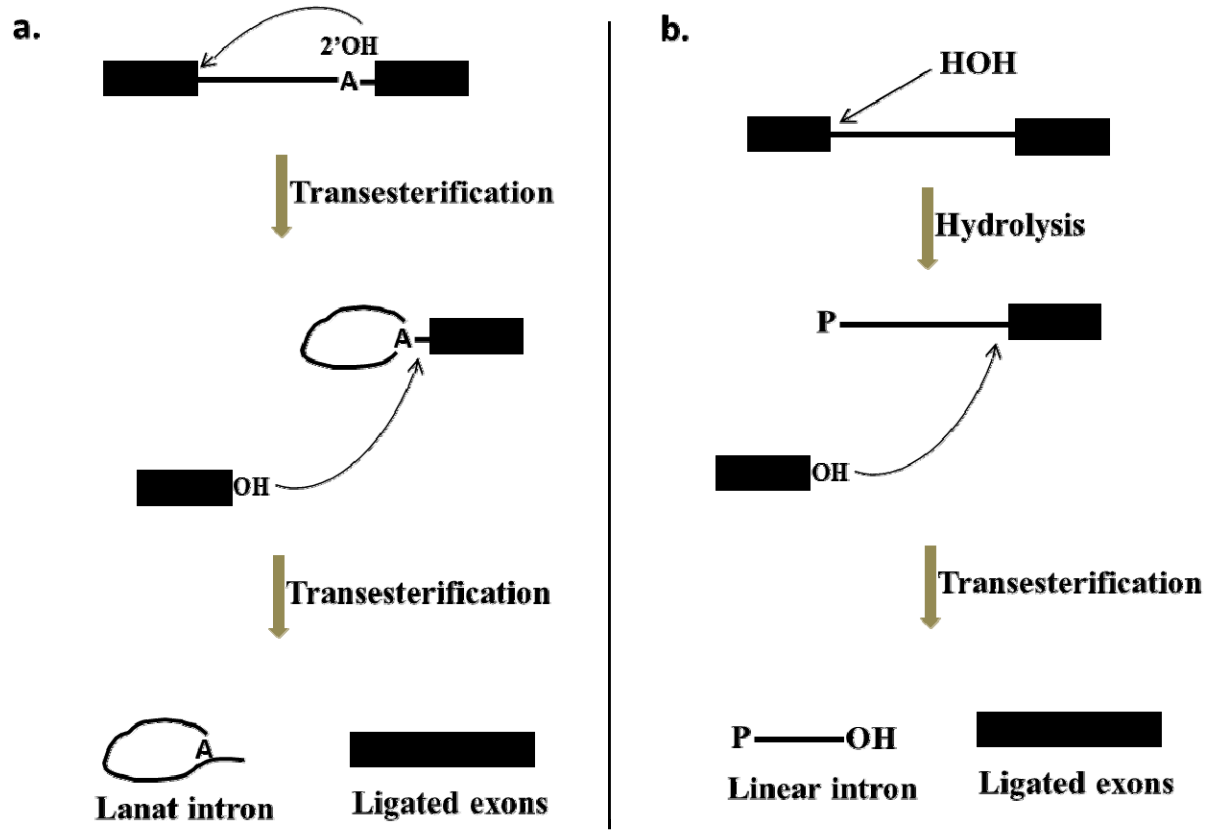

Figure 5. Group II intron splicing mechanism.

(a) In the classical pathway, the intron is spliced by two transesterification reactions and is released as a lariat. (b) In the alternative pathway, the attack on the first exon-intro boundary is hydrolytic; the second step is the same as the classical pathway. The intron is released as a linear molecule.

(Mohr and Lambowitz, 2003). Among the four factors involved in intron splicing in maize chloroplasts, three of them (CRS1, CAF1 and CAF2) have orthologs in the *Arabidopsis* nuclear genome and are predicted to be imported into mitochondria (Ostheimer et al., 2003). Recently, nuclear-encoded PPR proteins have been found involving in mitochondrial splicing, such as THA8, OTP70, OTP51 and PPR4 (Schmitz-Linneweber et al., 2006; de Longevialle et al., 2008; Chateigner-Boutin et al., 2011; Khrouchtchova et al., 2012).

2.5.2 5’ and 3’ processing of plant mitochondrial transcripts

All mitochondrial RNAs, whether tRNAs, rRNAs or mRNAs, have to be processed from longer precursor molecules. Theoretically, three types of activities could be responsible for generating mature 5’ and/or 3’ extremities from larger precursor RNAs: 5’ to 3’ exoribonucleases, endoribonucleases and 3’ to 5’ exoribonucleases. However, up to now, there is no experimental evidence demonstrating that a 5’ to 3’ exoribonuclease exists in plant mitochondria, and also no plant mitochondria endoribonuclease has been identified. The existence of mitochondrial 3’ to 5’ exoribonucleases has only been inferred from in vitro processing or degradation experiments. But recently, two genes encoding exonucleases PNPase and RNase II have been characterized in Arabidopsis. While mtPNPase is localized exclusively in mitochondria, RNase II is targeted to both mitochondria and chloroplasts (Perrin et al., 2004b; Bollenbach et al., 2005). RNase II is able to degrade unstructured RNA and plays a role in the final trimming of 3’ extremities (Perrin et al., 2004b; Bollenbach et al., 2005). PNPase appears to be a key player in plant mitochondrial RNA maturation and degradation (Perrin et al., 2004a; Perrin et al., 2004b). Although substantial progresses have been made in recent years, further studies are needed to reveal the mechanism of 5’ and 3’ processing events.

Loss of certain 5’ processing might be the results of an inactivation of a factor required for efficient cleavage of the substrate. In recent years, a number of PPR proteins have been reported involving in RNA maturation, e.g. RNA PROCESSING FACTOR1 (RPF1), RPF2, RPF3, and MPPR6. RPF1, 2 and 3. These proteins are highly similar to RFs in amino acid sequences and all participate in 5’ processing of the major transcripts of mitochondrial genes, such as subunits 4 and 9 of NADH dehydrogenase (*nad4* and *nad9*), subunit 3 of cytochrome coxidase (*cox3*), and cytochrome c maturation protein C (*ccmC*) in Arabidopsis mitochondria (Holzle et al., 2011; Jonietz et al., 2011). MPPR6 is associated in vivo with the 5’ untranslated region (UTR) of mitochondrial (*rps3*) mRNA, and the mapping of transcript termini

showed specifically extended *rps3* 5’ ends in the *mppr6* mutant, thus suggesting that MPPR6 is directly involved in 5’ maturation and translation initiation of *rps3* mRNA (Manavski et al., 2012). Besides a potential role for translation initiation, 5’ processing is potentially important for RNA stability. However, no experimental data are present to support this hypothesis yet.

2.5.3 RNA stabilization

In plant mitochondria, steady-state levels of functional RNAs are defined by equilibrium between transcription and degradation. RNA stability is crucial in determining the steady-state level of each specific transcript. Up to now, some cis- and trans- factors have been found that are required to specify stability and turnover of transcripts.

Secondary structures, forming single or double-stem loops at the 3’ ends, play a role as processing signals and can also influence transcript stability. This phenomenon has been confirmed in rapeseed *orf138* mRNA (Bellaoui et al., 1997), rice and wheat *cob* mRNA (Saalaoui et al., 1990; Kaleikau et al., 1992), and pea *atp9* mRNA (Dombrowski et al., 1997; Kuhn et al., 2001). It is likely that secondary structure present at the 3’ extremities prevents the progression of 3’ to 5’ exoribonucleases such as RNaseII and PNPase. Several nuclear-encoded PPR proteins have been reported involving in the stability of mitochondrial transcripts as trans-factors, such as PPR5, CRP1, and PPR10. PPR5 can stabilize the *trnG*-UCC precursor by directly binding and protecting an endonuclease-sensitive site (Beick et al., 2008). Mutation of *CRP1* results in loss of functional cytochrome f/b6 complex, due to absence of *petD* mRNA and *petA* translation. CRP1 is involved in chloroplast transcript *petD* stabilization and *petA* translation (Barkan et al., 1994). PPR10 can serve as a barrier to RNA decay from either the 5' or 3' direction and a bound protein provides an alternative to an RNA hairpin as a barrier to 3' exonucleases, implying that protein 'caps' at both 5' and 3' ends can define the termini of chloroplast mRNA segments

(Pfalz et al., 2009; Prikryl et al., 2011).

2.6 RNA editing

RNA editing is defined by the alteration of RNA sequences via insertion, deletion and conversion of nucleotides. In 1986, RNA editing was discovered when four uridine nucleotides were found inserted into the specific sites of trypanosome mitochondrial *cox2* mRNAs (encoding cytochrome oxidase subunit 2) which reconstitute the correct reading frame (Benne et al., 1986). Followed this discovery, RNA sequence alterations as a result of different types of RNA editing were subsequently identified in organisms separated over wide phylogenetic distances. A timeline for these discoveries is given in Table 1.

There are two major types of RNA editing, 1) RNA editing by insertion or deletion, and 2) RNA editing by conversion: C-U, U-C and A-I editing. In 1986, RNA editing through the insertion and deletion of uracil was first found in kinetoplasts in mitochondria of *Trypanosoma brucei* (Benne et al., 1986; Blum et al., 1990; Benne, 1994). The mechanism of this type of RNA editing has been studied. It starts with the base-pairing of the unedited primary transcript with a guide RNA (gRNA). This type of gRNA contains complementary sequences to the regions around the insertion/deletion points. The newly formed double stranded region is then enveloped by an editosome. The editosome is a large multi-protein complex, which catalyzes the editing (Arts and Benne, 1996; Alfonzo et al., 1997). The editosome opens the transcript at the first mismatched nucleotide and starts inserting uridines. The inserted uridines will base-pair with the gRNA and insertion will continue as long as A or G is present in the gRNA. Finally, the insertion will stop when a C or U is existed (Blum et al., 1990; Kable et al., 1997).

RNA editing by conversion includes C-U, U-C and A-I editing (Figure 6). U to C and A to I editing is rare. A to I editing occurs in regions of double stranded RNA (dsRNA). This type of editing can be either specific (a single adenosine is edited within the stretch of dsRNA), or promiscuous (up to 50% of the adenosines are

Date	Organisms	Type of RNA editing	Affecting	In genetic system
1986	Kinetoplastids	U insertions, U deletions	Diverse mRNAs	Mitochondrial
1987	Mammals	C to U	Some mRNAs and viral RNAs	Nuclear
1988	Paramyxoviruses	G and A insertions	P (phosphoprotein) and glycoprotein mRNAs	Viral
1989	Plants	C ↔ U	mRNAs, tRNAs (rRNAs)	Mitochondrial
1991	Plants	C ↔ U	mRNAs, tRNAs (rRNAs)	Chloroplast
1991	Myxomycota	C insertions, U insertions, dinucleotide insertions, C to U	mRNAs, rRNAs, tRNAs	Mitochondrial
1991	Metazoa	A to I (G)	Many mRNAs, tRNAs, miRNAs, and viral RNAs	Nuclear
1993	*Acanthamoeba* and chytridiomycete fungi	N to N′	tRNA 5′ acceptor stem	Mitochondrial
1993	Marsupials	C to U	tRNA anticodon	Mitochondrial
2002	Dinoflagellate	A ↔ G, C ↔ U, U to R, G to C	mRNA	Mitochondrial
2004	Dinoflagellate	A ↔ G, C ↔ U, R to C, U to G	mRNA, rRNA	Chloroplast
2009	Archaeae	C to U	tRNAs	Bacterial
2009	Placozoa	U to C	mRNA	Mitochondrial

Table 1. Discovery of RNA editing in different genetic systems (Knoop, 2011).

edited). Specific editing occurs within short duplexes (e.g. those formed in an mRNA where intronic sequence base pairs with a complementary exonic sequence); whereas promiscuous editing occurs within longer regions of duplex (e.g. pre- or pri-miRNAs, duplexes arising from transgene or viral expression, duplexes arising from paired repetitive elements). A to I editing can result in many effects. These effects include alteration of coding capacity, altered miRNA or siRNA target populations, heterochromatin formation, nuclear sequestration, cytoplasmic sequestration, endonucleolytic cleavage by Tudor-SN, inhibition of miRNA and siRNA processing and altered splicing.

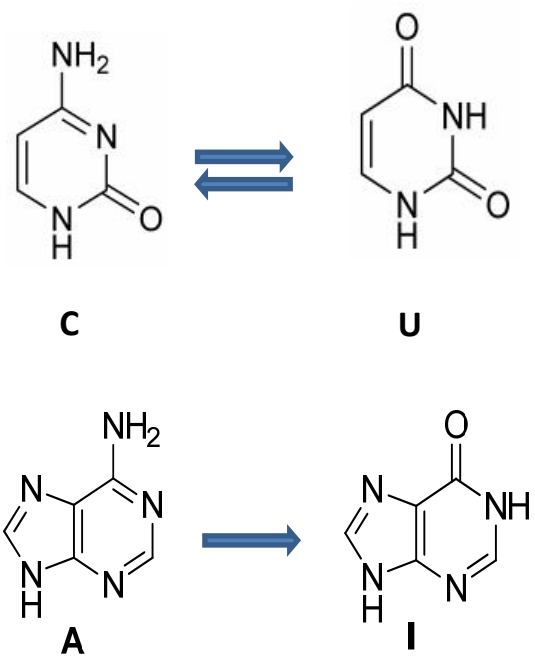

Figure 6. C-U, U-C and A-I RNA editing by conversion.

The most frequent editing events are C to U changes in messenger RNAs. The C-to-U editing enzyme is still unknown, but several hypotheses were proposed for the mechanism of C-to-U editing, including (1) RNA cleavage, cytidine release and uridine insertion then ligation, (2) cytidine deamination, (3) transamination, (4) transglycosylation (Chateigner-Boutin and Small, 2010).

Plant RNA editing was first reported in 1989 by three different groups reporting on cytidine (C) to uridine (U) exchanges in mitochondrial mRNAs independently (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989). In flowering plants, post-transcriptional modification of transcripts includes C-to-U, U-to-C and A-to-I editing. The majority of editing in plants occurs in mitochondrial and plastid transcripts, however, A-to-I editing also occurs in cytosolic tRNAs. The most frequent RNA editing results in conversions of C-to-U in mRNA (Chateigner-Boutin and Small, 2010). Compared with fewer than 40 editing sites in the chloroplast transcripts, RNA editing alters 350–500 sites in the mitochondrial transcripts (Chateigner-Boutin and Small, 2010).

The RNA editing events are important posttranscriptional regulation process in

mitochondria. They usually can restore a conserved amino acid codon, create an initiation or stop codon, alternatively, or remove a stop codon (Shikanai, 2006; Chateigner-Boutin and Small, 2010). Loss of these nucleotide changes usually results in a translated protein with compromised or completely lost function, which affects essential functions of mitochondria. This observation is supported by a number of mitochondrial RNA editing defective mutants that show severe phenotypes, such as seed defects, growth and development delays (Kim et al., 2009; Sung et al., 2010; Hammani et al., 2011; Sosso et al., 2012b; Yuan and Liu, 2012b). Editing can also occur in introns and untranslated regions, potentially playing a role in intron splicing and transcript stability, but also possibly being non-specific (Chateigner-Boutin and Small, 2010).

2.7 Status of C to U RNA editing mechanism

Although RNA editing is such an important step in the post-transcriptional control of organelle gene expression (Shikanai, 2006; Grennan, 2011), little is known about the mechanisms of C-to-U RNA editing in plants. The first breakthrough came with the identification of CRR4, a pentatricopeptide repeat (PPR) protein, which is required for the C to U editing of a single site within the chloroplast transcript *ndhD* in maize (Kotera et al., 2005). Since then, more PPR proteins have been identified as the trans-acting factors responsible for RNA editing in plastids and mitochondria (Fujii and Small, 2011). All of these PPR proteins are localized in either mitochondria or chloroplasts. Analysis of their domain structure indicates that they belong to E or DYW subclasses of the PPR protein family (Okuda et al., 2007; Schmitz-Linneweber and Small, 2008; Fujii and Small, 2011). The DYW domain shows a significant level of similarity to deaminase, raising the possibility of being a catalytic domain to convert C-to-U (Salone et al., 2007). Although this notion is not supported by the finding that truncated proteins lacking the DYW motifs can completely restore RNA editing *in vivo* (Okuda et al., 2009), a recently reported protein DYW1, which contains no identifiable PPR motifs but does contain a clear DYW domain, is

confirmed that this protein and a E subclass PPR protein CRR4 act together to edit the *ndhD-1* site (Boussardon et al., 2012). Recently, a new type proteins essential for organelle RNA editing have been identified, protein family multiple organellar RNA editing factor (MORF), including RIP1(=MORF8), loss of which alters editing efficiency at multiple sites, distinguished from PPR protein only affecting one or several sites (Bentolila et al., 2012; Takenaka et al., 2012). RIP1 is localized in both chloroplasts and mitochondria, affecting numerous editing sites in the two organelles. Either MORF2 or MORF9 are required for almost all sites of RNA editing in the chloroplast, and mutations of MORF1 and MORF3 affects numerous editing sites in mitochondria. These MORF proteins can interact selectively with PPR proteins to establish a complex editosome, and they also can connect to form hetero- and homodimers to involve in RNA editing (Bentolila et al., 2012; Takenaka et al., 2012).

It is now commonly accepted that cis-acting elements, trans-acting factors and editing enzymes are required for the specificity of the C-to-U editing. The sequences surrounding individual editing sites have been analysed in vivo and in vitro. In most cases, the region -25/+10 compared to the editing sites is required for PPR protein binding (Chaudhuri and Maliga, 1996; Okuda et al., 2006). At present, PPR and MORF protein family are the only reported trans-acting factors in C-to-U RNA editing. The PPR repeats are assumed to recognize the target RNA sequence. The second PPR repeat of MEF11 is proven crucial for the specific editing events targeted by this protein (Verbitskiy et al., 2010). The DYW domain of PPR proteins, whose sequence is similar with cytidine deaminases (Salone et al., 2007), is correlated in plant evolution with the presence of RNA editing (Fujii and Small, 2011). Therefore, this domain is deemed to have catalytic editing activity. However, all the PPR proteins involved in editing contain an E domain, but some PPR proteins with absence of DYW domain also function in RNA editing, such as CRR4, CLB19, MEF9 and SLO1 (Kotera et al., 2005; Chateigner-Boutin et al., 2008; Sung et al., 2010; Takenaka, 2010). In addition, the DYW domains of the plastid-located CRR22, CRR28, and OTP28, and the mitochondrial factor MEF11 are not essential for RNA editing in vivo

(Okuda et al., 2009; Okuda et al., 2010; Verbitskiy et al., 2010). E domain lacks any obvious catalytic site, but it is indispensable for target site editing, therefore it is assumed to recruit the editing activity from another domain of the same protein or another editing enzyme (Okuda et al., 2009).

Chapter III.

Materials and methods

3.1 Plant Materials

The *emp5-1* allele was isolated from the UniformMu population by introgressing *Mu* active lines into inbred W22 genetic background (McCarty et al, 2005). The wild type plants were either siblings of the mutant or W22. The *emp5-2, emp5-3* and *emp5-4* alleles were isolated from the Pioneer Hi-Bred International Trait Utility System for Corn (TUSC) population by PCR screening with *Emp5* specific primers and *Mu* primers. The maize plants were grown in the experimental field at the Chinese University of Hong Kong under natural conditions. *Oryza sativa* L.ssp. *Japonica* (cv. Nipponbare) was used as the plant material for *Agrobacteria*-mediated rice transformation.

3.2 Light Microscopy of Cytological Sections

To enable a precise comparison, wildtype and *emp5-1* mutant kernels were harvested from the same ear of a self-pollinated heterozygous plant at 8 and 13 DAP. The kernel was cut along longitudinal axis and the slice containing the embryo was fixed for 1 day at room temperature in 4% paraformaldehyde. The fixed material was dehydrated in an ethanol gradient series (50%, 70%, 85%, 95%, and 100% ethanol). After clearing with xylene and paraffin wax infiltration, the sample was embedded and sectioned at 6 to 10 μm thickness under a Leica 2035 Biocut. The sections were stained with Johansen's Safranin O and Fast Green and observed with a Nikon ECLIPSE 80i microscope.

3.3 Immunohistochemistry Analysis

The sections containing 13 DAP WT and *emp5-1* seeds were deparaffinized, then air-dried for 15 mins, rinsed with PBS, and incubated in 0.2-0.3 % Triton in PBS for 15 min. Non-specific antibody binding was blocked with 3% BSA for 2 h. The tissues

were then incubated with the primary antibody anti-BETL2 (a gift from Dr. Hueros, 1:400 diluted in 1% BSA, 0.05% triton in PBS) at 4℃ overnight. The slides were washed 3 times with PBS, 15 mins each to completely remove the primary antibody. Then the slides were incubated with Alexa Fluor-568 anti-rabbit secondary antibodies (Invirtrogen, USA) for immunofluorescent detection (diluted 1:500 in PBS, RT, 1 hr). After washing 3 times with PBS, the slides were viewed and imaged under an Olympus FluoView FV1000 confocal microscope. Control sections were incubated with PBS without the primary antibody and subsequently processed as described above.

3.4 Isolation of genomic DNA

1 gram of plant material was grinded into fine powder by mortar and pertle in liquid nitrogen, followed by addition of 5ml DNA extraction buffer [For making 400ml solution: 168g Urea, 25ml 5M Sodium chloride, 20ml 1M Tris-HCl (pH8.0), 16ml 0.5M EDTA (pH8.0), 20ml 20% Sarkosyl, water was added to make total volume into 400ml]. The mixture was allowed to thaw at room temperature with periodic mixing. After that, mixture was transferred to a 12ml centrifuge tube, followed by addition of 4ml phenol:chloroform(1:1, vol/vol). Mixing was shook on a rocking platform (Bio-Rad) for 30 minutes. The tube was then centrifuged at 8000g, 10℃, for 10 minutes. Upper layer supernatant was transferred to a new 12ml centrifuge tube, and mixed with 0.1X volume of 3M sodium acetate (pH5.2). Then, 3.8ml isopropanol was added, thread-like DNA would appear after inverting the tube five to six times. DNA pellet was then recovered after centrifuge at 10000g, 10℃, for 15 minutes. DNA pellet was washed with 1ml 70% ethanol, transferred to a 1.5ml microcentifuge tube and washed with 0.5ml 70% ethanol again. After incubation for 2 to 5 minutes at room temperature, DNA pellet was recovered again by centrifuging at 10000g for 2 minutes. Supernatant was removed, and the DNA pellet was completely dried under vacuum. To dissolve the pellet, 500µl 1X TE buffer (10mM Tris, 1mM EDTA, pH8.0)

was added followed by incubation on ice at 4℃ overnight. The dissolved DNA could be stored in -20℃ until usage.

3.5 Southern Analysis

For co-segregation analysis, maize genomic DNA was extracted from *emp5-1/+* and WT seedlings, and digested by HindIII. The hybridization probe was the ~1kb HinfI fragment of the *Mu1* element. In the southern analysis of rice transgenic lines, the *hpt* gene fragment amplified from vector *pTCK303* with primer Hpt-F1 and Hpt-R1 (Supplemental Table1) was used as a probe.

To label probes with [α-^{32}P]-dCTP, 45μl of TE buffer that contain about 50ng DNA fragment was incubated in boiling water for 5 minutes, and transferred to ice for 2 minutes immediately afterwards. Condensation was brought down by brief centrifugation. All the solution in the tube was transferred to a tube containing Ready-To-Go DNA Labeling Beads (-dCTP)(GE Healthcare). Beads were completely dissolved, followed by the addition of 5μl [α-^{32}P]-dCTP (3000 Ci/mmol, GE Healthcare). This reaction mixture was incubated at 37℃ for 30 to 60 minutes. Probe mixture was loaded into ProbeQuant^{TM} G-50 Micro Column (GE Healthcare), and centrifuged at 3,000rpm (750xg) for 2 minutes. The probe was denatured by incubation in boiling water for 10 minutes and then cooled down on ice immediately for 2 minutes. `

About 13μg genomic DNA was digested with restriction enzyme for 6 hours or overnight at 37℃ with the presence of RNase. Digested DNA was precipitated with ethanol and sodium acetate. Residual salts were washed away by 70% ethanol. Pure DNA was dried under vacuum then dissolved in 30μl loading buffer, followed by separation on EB containing 0.7% agarose/TBE gel overnight. DNA fragments were transferred to Hybond™-N membrane (GE Healthcare) by capillary action, and stabilized by UV crosslink. After pre-hybridization, denatured P32-labelled probe was hybridized to the membrane overnight by incubation at 65℃. The membrane was washed with washing buffer [0.1% SDS, 40 mM $Na^{+}.PO_4$ (pH 7.2), 1mM EDTA],

then blotted to damp dry prior to X-ray film exposure for detection.

3.6 Inverse PCR cloning

For inverse PCR cloning of the *Mu* flanking sequences in the 3.4 kb HindIII fragment, Genomic DNA of heterozygous *emp5-1* plants was digested by HindIII and separated on 0.7% agarose gel. The 3.4 kb fragment was enriched by cutting a small gel slice around that size and the DNA was purified. The DNA was self-ligated at 50 ng/μl concentration for overnight at 4℃. Then, the ligated DNA was digested with NotI. *Mu1* and *Mu2* elements contain a NotI site in the internal sequences; hence the digestion will linearize the ligated circular DNA. This DNA was used as the template in the inverse PCR amplification of the *Mu* flanking sequences. Annealing temperature for the first round PCR is 60℃, and *Mu* specific degenerated TIR6 primer (AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC) was used. The second round PCR were performed with 56℃ annealing temperature and *Mu1*-62 primer (CCCTTCCCTCTTCGTCCATAAT). The amplified fragment was cloned into pCR4-TOPO and sequenced.

3.7 RNA Extraction and RT-PCR

Approximately 100mg of fresh tissue was quickly frozen in liquid nitrogen and ground to fine powder with a mortar and pestle. Total RNA was extracted with 1 mL Trizol reagent according to the manufacture's instructions (Invitrogen, USA). After isopropanol precipitation, the RNA was resuspended in 40 ul RNase-free water and treated with RNase-free DNase. SuperScript® III One-Step RT-PCR System (Invitrogen) was used to perform RT-PCR according to the instructions of the supplier. RT-PCR for the expression pattern analysis of *Emp5* in maize organs was performed with primers Emp5-F2 and Emp5-R2, 57℃ annealing temperature, and running 30 cycles. For analysis of *OsEmp5* expression in rice transgenic lines, RT-PCR was

performed with primers OsEmp5-F and OsEmp5-R, running 28 cycles, and annealing temperature is 55℃.All primers are listed in Supplemental Table1.

3.8 Subcellular Localization of EMP5 Protein

To generate a translational protein fusion between EMP5 signal peptide and GFP, full-length *Emp5* and *Emp5*^{N469} fragment was amplified by PCR from maize inbred line W22 and cloned into pENTR/D-TOPO (Invitrogen, USA) respectively. The fusion was introduced to binary vector *pGWB5* (a gift from Dr. Tsuyoshi Nakagawa, Shimane University) by GATEWAY site-specific recombination. The OsEMP5^{N372}:GFP fusion expression construct was constructed similarly. These fusion proteins were placed under the cauliflower mosaic virus 35S promoter for constitutive expression. Then, these constructs were transformed into *Agrobacterium tumefaciens* strain EHA105. The resulting strains harboring full-length EMP5:GFP and EMP5^{N469}:GFP expression plasmid were used to transform *Arabidopsis thaliana* Columbia ecotype by flower-dip method (Clough and Bent, 1998). The transgenic *Arabidopsis* was identified with PCR amplification of *Hpt* gene in pGWB5 vector with primer Hpt-F3 and Hpt-R3. The protoplasts were isolated from the transgenic leaves by digesting with an enzyme solution (1.5% cellulose R10, 0.3% pectolyase Y23, 20mM MES pH5.7, 0.4M mannitol, 20mM KCl, 10mM $CaCl_2$, 0.1% BSA). By using established protocols (van Herpen et al., 2010), the EHA105 strain harboring the OsEMP5^{N372}:GFP fusion construct was infiltrated into *Nicotiana tabacum* leaves to transiently express the fusion protein. The infiltrated tobacco leaves, transgenic *Arabidopsis* leaf samples and protoplasts were used for GFP and MitoTracker red signals detection by an Olympus FluoView FV1000 confocal microscope. The working concentration of MitoTracker was 30nM and the samples were incubated at 37°C for 30mins.

In vitro chloroplast protein import assay was performed as described previously (Cline, 1986; Martin et al., 2009). *Emp5* cDNA was placed under the SP6 promoter in

pGem-3Z vector. RNA transcripts of this construct were produced by *in vitro* transcription with SP6 polymerase (Promega, USA). The protein was translated with a homemade wheat germ translation system in the presence of ^{3}H-leucine (Cline, 1986). Peas (*Pisum sativum* L. cv. Laxton's Progress 9 Improved) used for chloroplast isolation were grown as described (Cline, 1986). The chloroplast import was performed as described (Martin et al., 2009).

3.9 Analysis of Mitochondrial RNA Editing

For RNA editing analysis in the WT and the *emp5-1* allele, total RNAs were isolated from the immature embryos and endosperms by carefully removing the pericarp. For the *emp5-4* allele and rice RNAi transgenic lines, total RNAs were isolated from seedling leaves, all using the Trizol reagent according to the instructions of the manufacture (Invitrogen, USA). The RNA was treated with DNase I (New England Biolab) and the complete removal of DNA was checked by PCR on genomic DNA. Then the DNA-free RNAs were reverse-transcribed with random hexamers and the high fidelity reverse transcriptase SuperScript III (Invitrogen, USA). Full Sequences of total 35 protein-coded maize mitochondrial genes and rice mitochondrial *rpl16* gene were amplified by PCR. The RT-PCR products were sequenced directly. This analysis was done with three biological replicates by using seeds of different DAPs. These primers are listed in Table 2.

3.10 Rice Transformation

3.10.1 *OsEmp5* RNAi vector construction

For *OsEmp5* RNAi vector construction, 518bp *OsEmp5* fragment was PCR amplified with primers OsEmp5- KpnI and OsEmp5- BamHI, OsEmp5-SpeI and OsEmp5-SacI respectively (Supplemental Table1). The PCR products were digested with BamHI

and KpnI, SpeI and SacI respectively, then ligated into the binary vector *pTCK303*. The resulting RNAi vector, *pTCK303-OsEmp5* was introduced into the WT cultivar Nipponbare using *Agrobacterium*-mediated transformation.

3.10.2 Callus induction from mature rice seeds

Mature rice seeds (*Oryza sativa* L.ssp. *Japonica*) were dehulled by forceps, sterilized with 70% ethanol for one minute, then washed thoroughly with sterilized Elix water one time. The seeds were then sterilized with commercial bleach (sodium hypochlorite 1%) that contain one drop of Tween-20, by subjecting to continuous shaking on an orbital shaker for 30 minutes. After that, the bleached seeds were washed thoroughly with sterilized Elix water for three times. After blot drying on sterile filter paper, the seeds were transferred onto 2N6 media [30g/L Sucrose , 0.3g/L Casamino acids, 0.5g/L Proline, 0.5g/L L.Glutamin, 4.0g/L CHU(N_6) Basal salt mixture, 2mg/L 2,4-Dichlorophenoxy acetic acid, 3g/L Gelrite, pH5.8] at a density of 12-15 seeds per plate. The plates were kept in darkness at 28℃ for two to four weeks.

3.10.3 Callus subculture

Calli induced from mature seeds were separated from endosperm and radicle with a sterilized surgical blade. After isolation, calli were subcultured into pieces sized around 2.5mm, and transferred to freshly prepared 2N6 media for proliferation by keeping in dark at 28℃ 7 days before transformation.

3.10.4 Preparation of *Agrobacterium tumefaciens*

Freeze-thaw method was applied to introduce desired plasmid DNA into *Agrobacterium tumefacien*s (strain: EHA105). After confirming the identity of clones by restriction enzyme digestions, the bacteria (200ul) were subcultured in fresh liquid

YEP medium (20g/L LB broth, 5g/L Yeast extract, 50mg/L Kanamycin, 30mg/L Rifampicin), after 8~12h, 1:100 or 1:200 subcultured in 50ml liquid YEP medium without antibiotics overnight until OD600 is about 0.5.

3.10.5 Co-cultivation

The *Agrobacteria* were enriched by centrifuge (5000g), and resuspended in liquid 2N6-AS medium [30g/L Sucrose, 0.3g/L Casamino acid, 10g/L Glucose, 4.0g/L CHU(N_6) Basal salt mixture, 10ml/L N6 vitamins(100x), 2mg/L 2,4-Dichlorophenoxy acetic acid, pH5.8] that contain 100μm acetosyringone (AS at 19.62 mg/L). *Agrobacteria* were resuspended into OD_{600} about 0.5, and left to sit at room temperature for one to three hours. After that, suspension mixture was transferred into a larger flask. Calli were added, swirled gently and left to sit at room temperature for about 40 minutes (shake it every 5 minutes). After incubation, excess *Agrobacteria* were removed from the surface of calli by blot drying on sterile filter paper, then individual calli were transferred onto solid 2N6-AS medium [30g/L Sucrose, 0.3g/L Casamino acid, 10g/L Glucose, 4.0g/L CHU(N_6) Basal salt mixture, 10ml/L N6 vitamins(100x), 2mg/L 2,4-Dichlorophenoxy acetic acid, 3g/L Gelrite, pH5.8] that contain 100μm acetosyringone. Rice calli were co-cultivated with the *Agrobacteria* in darkness at 22℃ for two to three days.

3.10.6 Callus washing and selection

After co-cultivation, calli were removed from the plate and transferred to a flask containing sterilized Elix water. Calli were washed by shaking gently on an orbital shaker for 20 to 30 minutes. Washing process was repeated twice with fresh sterilized Elix water.

Calli were blot dried on sterile filter paper and placed onto solid 2N6-TCH medium (30g/L Sucrose, 0.3g/L Casamino acid, 4.0g/L CHU(N_6) Basal salt mixture,

10ml/L N6 vitamins(100x), 2mg/L 2,4-Dichlorophenoxy acetic acid, 3g/L Gelrite, pH5.8) which contain 25mg/L Hygromycin B (for selection of calli transformed with desired plasmid), and 400mg/L Timentin (for killing excess *Agrobacteria*). These were used to select resistant calli at 28℃ for two weeks in darkness. After that, the calli were transferred onto solid 2N6-TCH medium that contain 50mg/L Hygromycin B and 200mg/L Timentin. To allow proliferation, calli were kept in darkness at 28℃ for another two weeks.

3.10.7 Regeneration

After two weeks or beyond, proliferated calli were transferred onto solid regeneration medium, RGH6 [30g/L Sucrose, 0.5g/L L.Glutamine, 0.5g/L Proline, 0.3g/L Casamino acid, 4.0g/L CHU(N_6) Basal salt mixture, 10ml/L N6 vitamins(100x), 3mg/L 6-Benzylaminopurine, 0.5mg/L Naphthalene acetic acid, 50mg/L Hygromycin, 3g/L Gelrite, pH5.8] then incubated in darkness at 28℃ for one week. After that, the calli were transferred to light at 28℃ for one to two weeks. After part of the calli changed from creamy yellow into green, which indicates there were differentiation of chloroplast, those calli were transferred to rooting medium [30g/L Sucrose, 4.0g/L CHU(N_6) Basal salt mixture, 5ml/L N6 vitamins(100x), 50mg/L Hygromycin, 3g/L Gelrite, pH5.8] and incubated at 28℃ in 16 hours/ 8 hours light dark cycle for two weeks. After differentiation of shoots and roots, the transgenic plantlets could then be screened and planted into soil for growth.

3.10.8 Screening of transgenic plants

To screen for transgenic plants, small portion of root (~3mm long) was removed from the plant, and histochemcial staining was performed using β-Glucuronidase Reporter Gene Staining Kit (Sigma) according to the manufacturer's protocol. For successful transformants, their roots would change from creamy white to blue. PCR with primer HygF2 (5'-ATTTCGGCTCCAACAATGTC-3') and HygR2 (5'-AATTAATTCGGGG

GATCTGG-3') was also performed to confirm the GUS staining results. Southern analysis of the three *OsEmp5* RNAi transgenic lines (line 19, 23 and 33), using the *hygromycin phosphotransferase (hpt)* gene as a probe was also performed to confirmed that they are independent transgenic lines.

Accession Numbers

Sequences data for *Emp5* gDNA, cDNA, allele *emp5-1, emp5-2, emp5-3* and *emp5-4*, can be found in the GenBank database under accession numbers JX308938, JX308939, JX308940, JX308941, JX308942 and JX308943, respectively.

Chapter IV.

Results

4.1 Phenotypic and Genetic Characterization of *emp5-1*

The *emp5-1* mutant was isolated from the UniformMu population where the active *Mu* lines were introgressed into inbred W22 (McCarty et al., 2005). When this mutant was isolated, six backcrosses to W22 had been carried out. The isolated *emp5-1* allele was back-crossed to W22 twice afterward to reduce active *Mu* copy numbers. Therefore, the mutant was considered in nearly isogenic W22 background (99.6%). The selfed progeny of the *emp5-1* heterozygotes segregated *emp* kernels in a 3:1 ratio (WT:*emp*, 455:167, $p>0.95$), indicating that *Emp5* is a monogenic, recessive and nuclear gene. In contrast to the WT, the mutant *emp5-1* kernels at maturity typically are small, containing a white pericarp that is often wrinkled (Figure 7A). Sectioning the kernels could find some residual tissues, but did not find recognizable embryo or endosperm structure within the pericarp (Figure 7B). This indicated that the *emp5* mutation blocked the development of both the embryo and the endosperm. The *emp5-1* allele is an embryo lethal mutation which is maintained in heterozygotes.

Figure 7. Mutant *emp5-1* kernels abort early in seed development.

(A)The ear segregates 3:1 for wild-type and *emp5-1* mutant kernels (arrows).

(B) Dissection of mature wild-type (left) and *emp5-1* (right) kernels.

To examine the developmental arrest in *emp5-1*, we analyzed the seed development process of the *emp5-1* mutant by light microscopy. Maize embryo development is characterized by three stages, transition, coleoptilar and late

embryogenesis. Endosperm development includes coenocytic, cellularization, differentiation, and maturation stages (Olsen, 2001). Close comparison was made by analyzing both WT and the *emp5-1* mutant in the same segregating ear. The mutant kernel can be clearly distinguished from the WT as early as 8 days after pollination (DAP), with characteristics of a smaller size and translucent appearance resulted by arrested embryo and endosperm development. And this was confirmed by PCR genotyping. Dissection of *emp5-1* mutant seeds at 8 and 13 DAP revealed that the mutant embryogenesis was arrested at the transition stage, characterized by that the radial asymmetry was introduced by the formation of an internal wedge-shape meristematic region in the upper part of the embryo. Endosperm development was blocked at the differentiation stage, characterized by the mutant endosperm contained distinguished starchy endosperm and formation of aleurone layer (Figure 8C, 8D and 8F). In contrast, the WT embryo at 8DAP has already differentiated into scutellum and shoot apical meristem, and the endosperm size was much larger than the mutant (Figure 8A and 8E).

Besides the dramatic difference in embryo and endosperm, we noticed that *emp5-1* appeared lacking basal transfer cells. Basal transfer cell layer is responsible for the uptake of solutes, which is critical to seed development (Pate and Gunning, 1972). To investigate the development of basal transfer cell layer in the *emp5-1* mutant seeds, we performed the immunohistochemistry analysis with BETL-2 antibody. BETL-2 is a basal endosperm transfer layer specific protein, expressed in early and mid-term endosperm development (Hueros et al., 1999). Therefore, it is a marker of basal transfer cell formation. The WT kernels at 13 DAP differentiated more than 3 layers of transfer cells, and BETL-2 proteins were expressed in the whole basal transfer layer region (Figure 9A). In contrast, the *emp5-1* mutant kernels from the same ear only formed a single layer of transfer cells and BETL-2 was detected in these cells (Figure 9B). Subsequent microscopic analyses did not find the formation of multiple cell layers of transfer cells in the mutant. This result indicated that the basal transfer cell development in the *emp5-1* kernels is arrested.

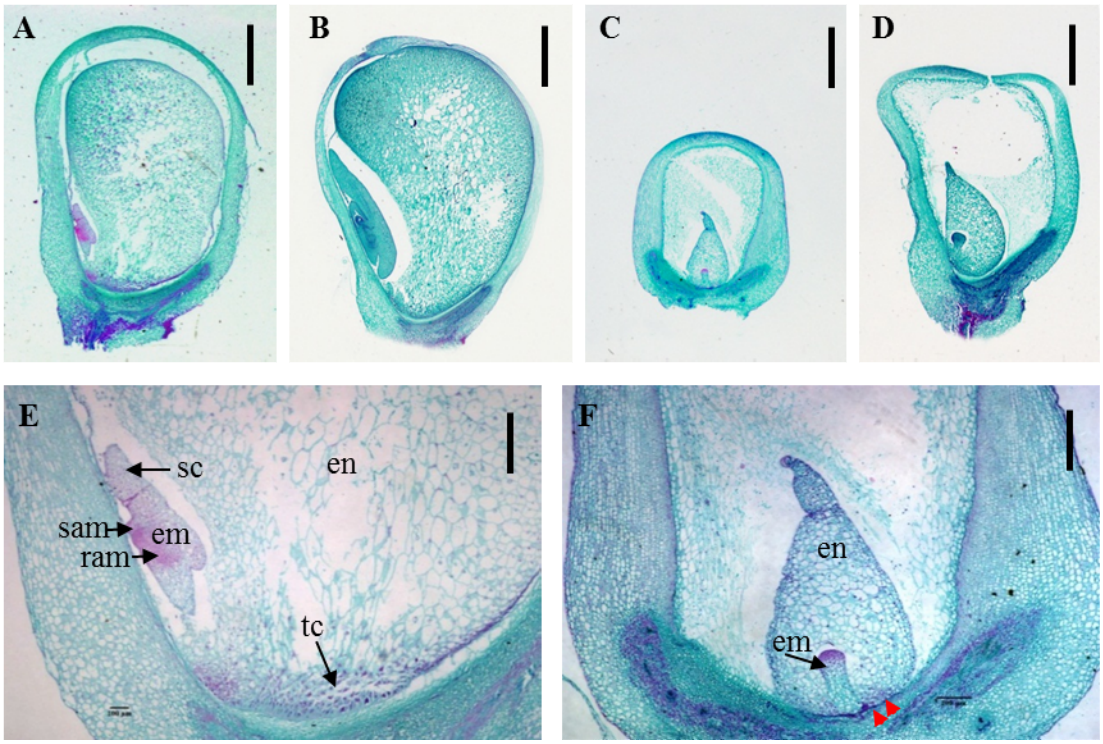

Figure 8. Mutant *emp5-1* kernels abort early in seed development.

Developmental comparisons of wild-type and *emp5-1* kernels at 8 and 13 DAP. Wild-type kernels at 8 DAP (A, E) and 13 DAP (B); *emp5-1* kernels at 8DAP (C, F) and 13 DAP (D). en:endosperm, em: embryo, tc: transfer cells, sc: scutellum, sam: shoot apical meristem, ram: root apical meristem. Red arrows point to positions where transfer cells were not clear formed in the mutant. Bars=1 mm in (A, B, C, D), 500 µm in (E, F).

4.2 Cloning of *Emp5*

Because the *emp5-1* allele was potentially tagged by *Mu* transposons, we performed Southern blot analysis to identify whether a *Mu* insertion was linked to the mutation. A segregating F2 population was created by self-crossing a heterozygous *emp5-1/Emp5* plant. Genomic DNA was isolated from the individual plants of the F2 and the genotype was determined by selfing the plant and checking for *emp5-1* mutant segregation. Because homozygous *emp5-1* is not viable, only heterozygotes (S, segregating) or wildtype (N, non-segregating) were available for this analysis. The

blots were hybridized with several *Mu* elements including *Mu1, Mu2, Mu3, Mu4*, *Mu8* and *MuDR*, only the hybridization with a *Mu1/Mu2* specific probe indentified a 3.4kb Hind III fragment that co-segregates with the *emp5* mutation, i.e. present of this

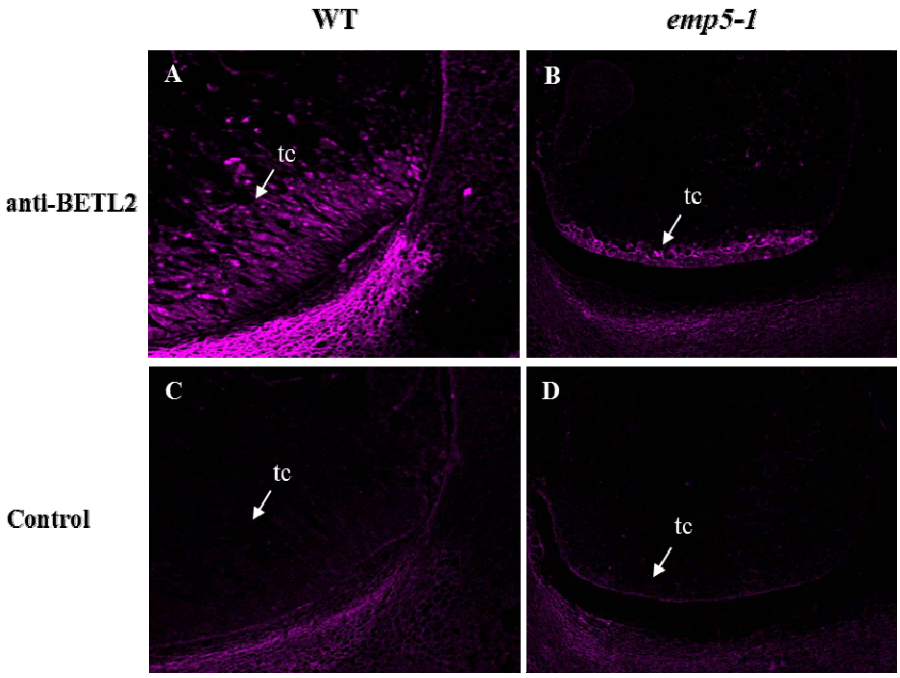

Figure 9. Effects of *emp5-1* on basal transfer layer.

Confocal visualization of BETL-2 by immunofluorescence in 13 DAP WT (A and C) and *emp5-1* mutant (B and D) kernels. Basal endosperm transfer layer specific antibody BETL-2 was used (A, B), and no BETL-2 antibody PBS buffer was used as control (C, D).

fragment in *emp5-1* heterozygote and absent in the WT (Figure 10A). No recombination was detected in the initial 22 F2 individuals tested. Increasing the population size to 120 still did not produce any recombination, suggesting a tight linkage between this *Mu* insertion and the *emp5-1* mutation. Because the *Mu1/Mu2* probe hybridizes to both *Mu1* and *Mu2*, we could not tell which one was inserted in this fragment at this stage.

The *Mu1* or *Mu2* flanking sequence in the 3.4 kb HindIII fragment was amplified by inverse PCR using *Mu* TIR primers (Refer to Methods). The size of the product was 1.7 kb with TIR sequences on both ends and a HindIII site in the middle. A 9bp target site duplication (TSD) was found flanking the *Mu* insertion. This suggests that

the *Mu* element is likely a *Mu2* (1.7kb), not a *Mu1* (1.4kb). BLAST search of the NCBI GenBank with the 1.7kb *Mu2* flanking sequence identified an expression mRNA in maize (Accession No: EU956937). This gene is located on chromosome 3. This gene appears to be a single copy gene based on the analysis of B73 whole genomic sequence draft AGPv2 (Schnable et al., 2009). Using two gene specific primers designed according to this clone, the full genomic sequence of this *Emp5* candidate gene was cloned from W22 genomic DNA. And the cDNA was amplified by RT-PCR from leaf RNAs.

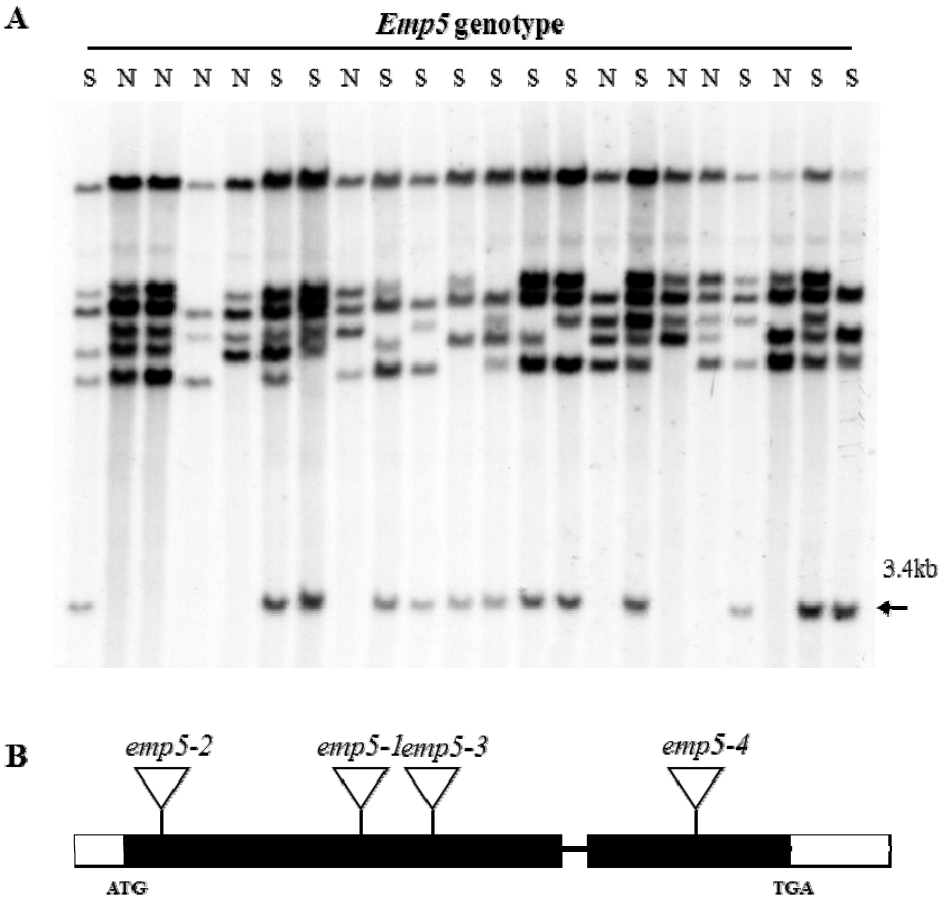

Figure 10. *Emp5* gene cloning.

(A) Cosegregation analysis of *emp5-1* segregation population using *Mu2* as a probe. *Mu2* transposon-tagged 3.4kb HindIII fragments (arrowheads) cosegregating with the *emp5-1* mutation. (B) Gene structure of *Emp5* and locations of *Mu* insertions in 4 independent alleles. Exons are filled boxes and introns are lines. *Mu* insertion sites of *emp5* alleles were marked by triangles.

To confirm that the cloned gene is the bono fide causative gene for the *emp5-1* phenotype, we isolated additional *Mu* insertional alleles from the Pioneer Hi-bred International TUSC population using *Emp5* gene specific primers in combination with *Mu*-specific primers (Bensen et al., 1995). Three independent *Mu* insertions in the *Emp5* locus were identified, named as *emp5-2, emp5-3,* and *emp5-4* (Figure 10B). The insertion sites were confirmed by sequencing the PCR products that were amplified with TIR8 primers and *Emp5* specific primers. The selfed progeny of heterozygote *emp5-2* and *emp5-3* produced *empty pericarp* kernels segregation at ratio 1:3 (*emp*:WT), but that of the *emp5-4* produced all normal kernels. Crosses between *emp5-1* heterozygotes with heterozygotes for *emp5-2* and *emp5-3* alleles produced ears segregating *empty pericarp kernels* at 1:3 ratio (*emp*: WT), whereas crossed between *emp5-1* and *emp5-4* produced all wildtype kernels. Genotyping using gene specific primers indicates that homozygous *emp5-4* seeds are viable and contain a normal embryo and endosperm. Later analysis indicated that the *emp5-4* mutation is leaky, only partially abolished the *Emp5* function (Figure 16). Because three independent alleles carried *Mu* insertions in the *Emp5* gene conditioned the empty pericarp phenotype, we concluded that the *Emp5* locus was cloned.

4.3 *Emp5* Encodes a Mitochondrion-Targeted PPR-DYW Subclass Protein

Sequence analysis indicated that the *Emp5* gene consists of 2 exons and in *emp5-1*, a *Mu2* was inserted in the first exon (Figure 10B).This *Emp5* gene encodes a 776-amino acid protein. Motif prediction analysis of EMP5 protein sequence by the algorithm TPRpred (http://tprpred.tuebingen.mpg.de/tprpred) revealed that it contains 11 PPR motifs, classifying this protein as a member of the PPR protein family (Figure 11A and 11B). The C-terminal region between residue 590 and 776 shows strong similarity to the consensus sequences of E, an E+, and a DYW domains (Lurin et al., 2004), indicating that EMP5 is a DYW subclass of the PPR protein family. Analysis of the rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*) genome identified an ortholog in each genome, named OsEMP5 and SbEMP5 respectively (Figure 12).

EMP5 showed clear divergence between monocots and dicots. Among the closely related proteins, it shares a high degree of similarity with sorghum (91%), rice (84%) and barley (83%), but a low degree of similarity with grape (69%) and Arabidopsis (64%). Functions of these proteins are not studied yet. The sequence analysis indicated that EMP5 is a typical PPR-DYW subclass protein.

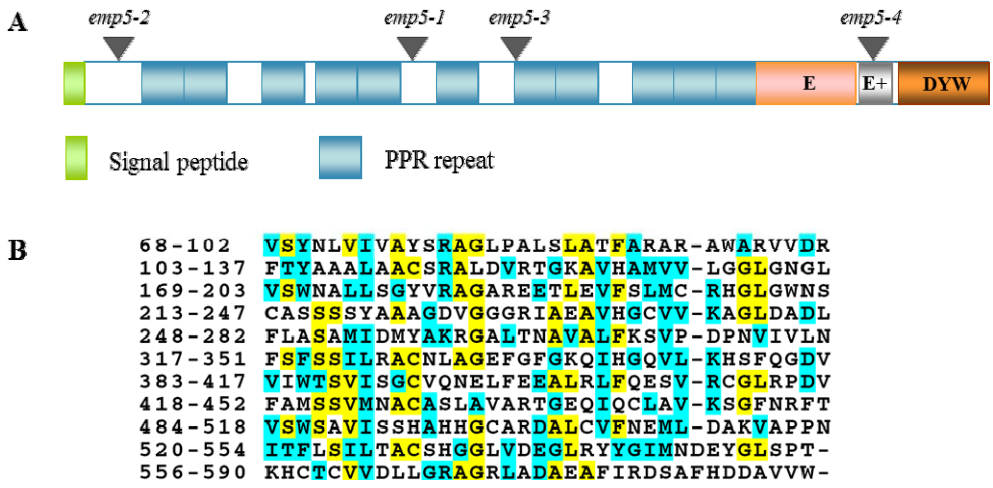

Figure 11. EMP5 protein structure.

(A) The EMP5 protein contains 11 PPR motifs, and E, E+, DYW motifs at the C-terminus. Locations of 4 *Mu* insertion alleles marked with triangles. (B) Alignment of 11 PPR motifs found in EMP5 protein. Residues identical are shaded in yellow, and similar residues are shaded in light blue.

Analysis of EMP5 with the TargetP algorithm predicted a putative mitochondrial localization, but with marginal confidence (http://www.cbs.dtu.dk/services/TargetP/). To experimentally determine the subcellular localization of EMP5, we fused full length EMP5 with green fluorescent protein (GFP) and transformed Arabidopsis. Ten lines of transgenic Arabidopsis were generated and among the transgenics GFP signals were either absent or low. We suspected that over-expression of the full length EMP5-GFP fusion may be the cause for this problem. So, the N-terminal 469 amino acid of EMP5 that contains 8 PPR repeats was fused with GFP, resulting in the

EMP5^{N469}-GFP fusion construct. Fourteen lines of transgenic Arabidopsis were produced and analyzed. Confocal laser scanning microscopy analysis of the transgenic Arabidopsis leaf samples and protoplast showed *in vivo* co-localization of

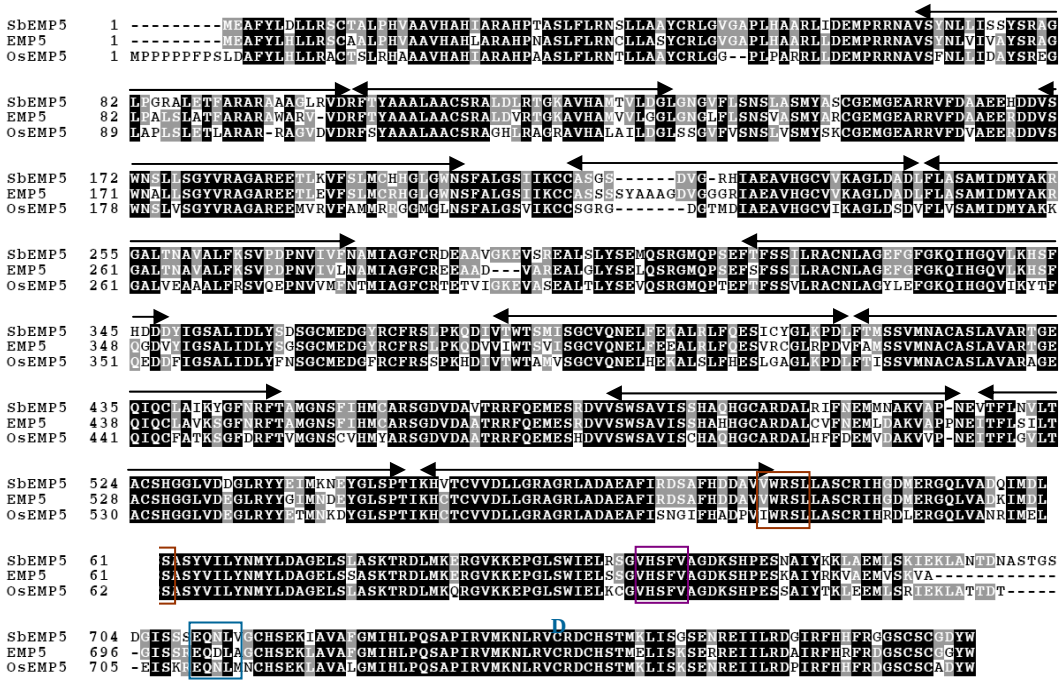

Figure 12. Alignment of EMP5 protein with rice (*Oryza sativa*) ortholog OsEMP5 and a Sorghum (*Sorghum bicolor)* ortholog SbEMP5 protein.

the GFP signal of EMP5^{N469}-GFP with MitoTracker Red in mitochondria (Figure 13A). To independently test the possibility of chloroplast localization, we also performed pea chloroplast import assay in which the EMP5 protein was labeled with ^{3}H-leucine and tested its capability in trafficking to the live chloroplasts. The result showed that EMP5 did not import into the chloroplasts as the imported chloroplasts treated with protease thermolysin eliminated all the labeled ^{3}H-EMP5 protein(Figure 13B). Thermolysin protease treatment degrades proteins unprotected by the chloroplast envelopes. Together, these results indicate that EMP5 is a PPR-DYW protein localized in maize mitochondria.

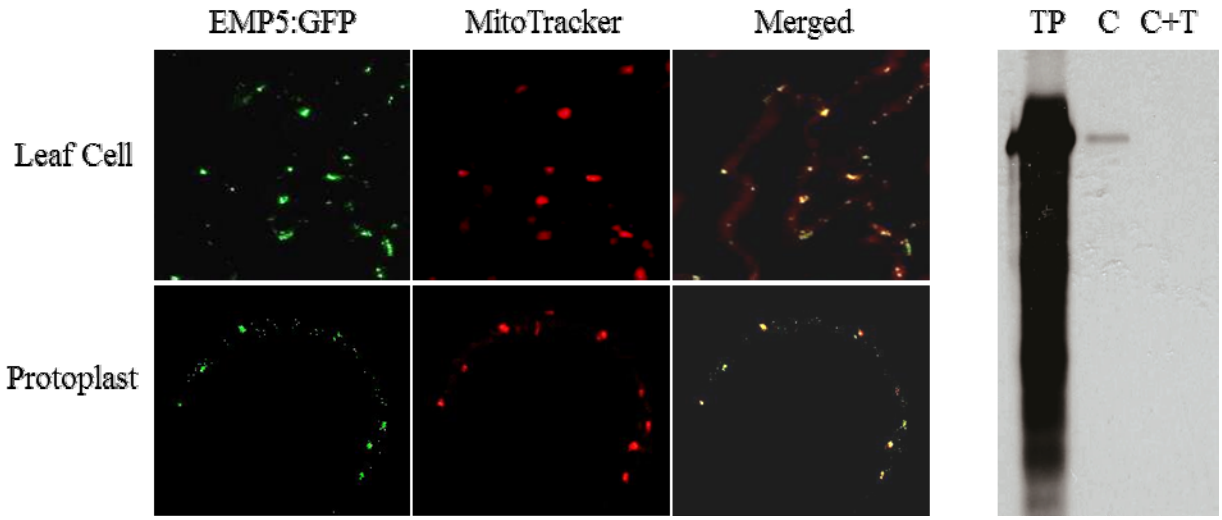

Figure 13 . EMP5 protein localization.

(A) EMP5^{N469}-GFP fusion protein that carried the N-terminus 469aa fused with GFP was expressed in transgenic Arabidopsis leaves. The leaf samples and protoplast of these leave were used to view GFP signals respectively. Fluorescence signals from EMP5^{N469}-GFP (green) and Mito Tracker stained mitochondria (red) were detected by confocal microscopy.

(B) Chloroplast import assay of full length EMP5 protein. TP: EMP5 translation precursor labeled with ^{3}H-Leu; C: chloroplasts after imported with ^{3}H-EMP5; C+T: ^{3}H-EMP5 imported chloroplasts treated with protease (thermolysin) to remove surface adhered proteins.

4.4 Expression of *Emp5*

BLAST analysis of EMP5 identified an EST from maize cDNA library of mix tissues, indicating that *Emp5* may be expressed in multiple tissues. Analysis of OsEMP5, the orthologous protein in rice, also identified two ESTs from rice panicle and callus. However, the *Emp5* mRNA could not be detected by conventional RNA gel blot analysis, suggesting that it may be expressed at low levels. Similarly, the maize mitochondrial PPR gene *Emp4* is also reported having a low expression level that could not be detected by RNA gel blot analysis (Gutierrez-Marcos et al., 2007). The *Emp5* expression can be detected by RT-PCR and the results confirmed that *Emp5* is

expressed in all vegetative and reproductive tissues tested (Figure 14). Relative high mRNA expression was in stem, leaf, root and ear, and weak expression in tassel and kernels at different developmental stages. This suggests that the EMP5 function may not be limited to embryo and endosperm development. Rather, it may have functions in other vegetative tissues during plant growth and development.

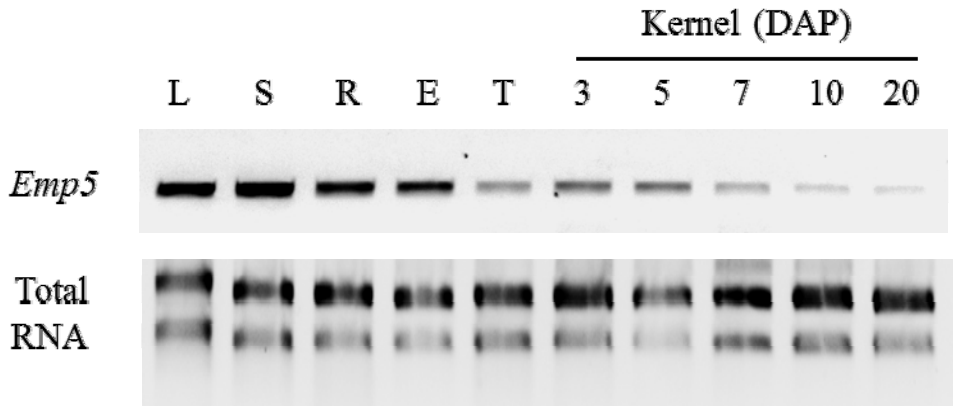

Figure 14. Expression pattern of *Emp5* in maize organs indicated.

Primer Emp5-F2 and Emp5-R2 were used in RT-PCR. L, leaf; S, stem; R, root; E, ear; T, tassel; kernel at 3, 5, 7, 10, and 20 DAP.

4.5 EMP5 is Required for Mitochondrial RNA Editing

Up to now, DYW1,MORF and PPR proteins are the only identified plant RNA editing trans-factors, and most of these PPR proteins belong to DYW-subclass (Fujii and Small, 2011) . EMP5 is a typical mitochondria-targeted DYW-subclass PPR protein. Thus, it is likely that EMP5 functions in RNA editing in maize mitochondria.

A direct comparison of the mitochondrial transcripts between the *emp5* mutant and the WT was performed by amplifying the transcripts and analyzing their sequences. Based on the maize mitochondrial genome (Clifton et al., 2004), 35 sets of primers were designed to cover the predicted 35 mitochondrial protein-coding genes. These 35 protein-coding mitochondrial genes includes 22 genes of the electron transport chain, 11 ribosomal proteins genes, a maturase gene (*mat-r*) and a transporter gene (*mttB*). The *emp5*-1 allele was chosen for its W22 background. To

eliminate possible contamination of mitochondrial DNA (most mitochondrial genes are intronless), RNAs were treated by RNase-free DNase and confirmed of DNA-free by PCR amplification without reverse transcription. The RNA was isolated from the 13 DAP kernels as the mutant became distinct. The pericarp was carefully removed to prevent contamination by maternal tissue. RT-PCR was performed using proof-reading DNA polymerase Phusion (New England Biolab). The PCR products were purified from the gel and sequenced without cloning into a vector. This allows detection of both edited and unedited sites in one sequencing reaction, and also eliminates cloning bias and random DNA polymerase errors. We amplified the mitochondrial genes in the *emp5-1* and the sequences indicated that the W22 contains a NB type mitochondrial genome (Clifton et al., 2004). As shown in Figure 15A, the C-to-U editing of *rpl16*-458 was completely abolished in the *emp5-1* allele, whereas it was completely edited in the WT. The unedited sequence codes for a proline, whereas the edited sequence codes for a leucine. The lack of C-to-U editing in *rpl16*-458 resulted in the amino acid changing from Leu to Pro in *emp5-1*. In addition, 100% C-to-U editing were found at *nad9*-190, *nad9*-356, *cox3*-245 and *cox3*-257 sites in the WT; whereas dramatically diminished in *emp5-1*, especially for *nad9*-190 and *nad9*-356 sites (Figure 15A). Five sites of the *rps12* transcript (*rps12*-71, *rps12*-196, *rps12*-221, *rps12*-269, *rps12*-284) were also reduced in the *emp5-1* mutant (Figure 15A). All these editings led to a change of the encoded amino acid as indicated in the Figure 15A.

Interestingly, we also found that RNA editing of some sites was increased in the *emp5-1* mutant. Editing of *rpl16*-444, *atp6*-953, *nad1*-536, *nad1*-832 and *cob*-1098 were dramatically increased in the *emp5-1* mutant, comparing with weaker editing of these sites in the WT (Figure 15B). However, these editings except *atp6*-953 did not change the encoded amino acids. Increased editing was also reported in the *reme1* mutants encoding a PPR protein and the *rip1* T-DNA insertional mutants (Bentolila et al., 2010; Bentolila et al., 2012). The underlying mechanism is not known.

4.6 Molecular Characterization of the *emp5-4* Allele

As described previously, homozygous *emp5-4* produced viable seeds. Genotyping with primers Emp5-F4 and Emp5-R4 and sequencing of the amplicon identified that the insertion of a 1.4kb *Mu1* element in the middle of E+ motif, thereby disrupting the E+ and DYW domain (Figure 16A). However, homozygous *emp5-4* seedlings and adult plants were macroscopically indistinguishable from the WT (Figure16B). In contrast to the use of transgenes, this allele provided an excellent genetic material for studying the effect of disrupted E+ and DYW domains on the function of EMP5. First, we examined the expression by using different *Emp5* primers anchored on different regions of the *emp5-4* allele (Figure 16A). As indicated in Figure 16C, RT-PCR results of the seedling leaf transcripts indicated that the region 5' of the *Mu1* insertion in *emp5-4* was expressed at roughly the similar level as the WT. The region including *Mu1* insertion was only expressed in *emp5-4*, as indicated in the RT-PCR using *Mu* primer TIR8 and Emp5-R1 in RT-PCR (Figure 16C). However, PCR with primers Emp5-F6/Emp5-R5 as well as Emp5-F4/Emp5-R4 across the *Mu1* insertion did not produce any products in *emp5-4* (Figure 16C). To address whether this was caused by difficulty in PCR amplification across the *Mu1* element, we used WT and homozygous *emp5-4* genomic DNA as template. PCR amplification with primers Emp5-F4 and Emp5-R4 reliably produced a 1.7kb fragment in *emp5-4* (predicted 1736bp) and a 350 bp fragment in the WT (predicted 349bp). Sequencing confirmed that the 1.7kb amplicon contains the *Mu1* insertion in the *emp5-4* allele (Figure 16C), indicating the primer set worked well in amplifying across the *Mu1* insertion. We did not detect any other alternatively spliced transcripts as some *Mu* inserted alleles showed splicing of the element. The spliced transcripts should be favored in amplification since they are short. The RT-PCR analysis indicates that the *emp5-4* allele does not produce a detectable level of *Emp5* transcripts that is likely to be translated into a WT protein, although the gene including the *Mu1* element can produce two transcripts, one contains the 5' region of the *Emp5* and ends in the *Mu1*

insertion, and the second starts from somewhere inside the *Mu1* and ends probably where WT *Emp5* ends.

To determine the impact of this mutation on editing, we amplified and sequenced the mitochondrial transcripts using the same strategy as in *emp5-1*. The templates were the seedling RNAs from the WT and the mutants from the same ear. The genotype was confirmed by PCR analysis. The results revealed that the *emp5-4* homozygotes showed similar levels of editing events in most transcripts targeted by EMP5 in comparison to the WT. These include *nad9*-356, *cox3*-245, *cox3*-257, *cob*-1098, *atp6*-953, *nad1*-536, *nad1*-832, and five editing sites in *rps12* (*rps12*-71, *rps12*-196, *rps12*-221, *rps12*-269, *rps12*-284). However, the *rpl16*-458 editing was diminished but clearly detectable in the *emp5-4* allele. This editing converts a Pro to Leu residue which is edited completely in the WT, but the editing was completely abolished in *emp5-1*. Similarly, editing of *nad9*-190 site was increased in *emp5-4* in comparison to *emp5-1*. This site is also edited completely in the WT which converts a His to Tyr residue (Figure 15A and 16D). Even in the same transcript, the other EMP5 targeted site (*nad9*-356) showed the same editing level as in the WT. Similar to *emp5-1*, the editing of *rpl16*-444 was also increased in *emp5-4*, but this editing does not change in coded amino acid (Figure 16D).These results indicated that the mutation in *emp5-4* does not affect the editing of most EMP5 targeted sites, but does decrease the editing in *rpl16*-458 and *nad9*-190.

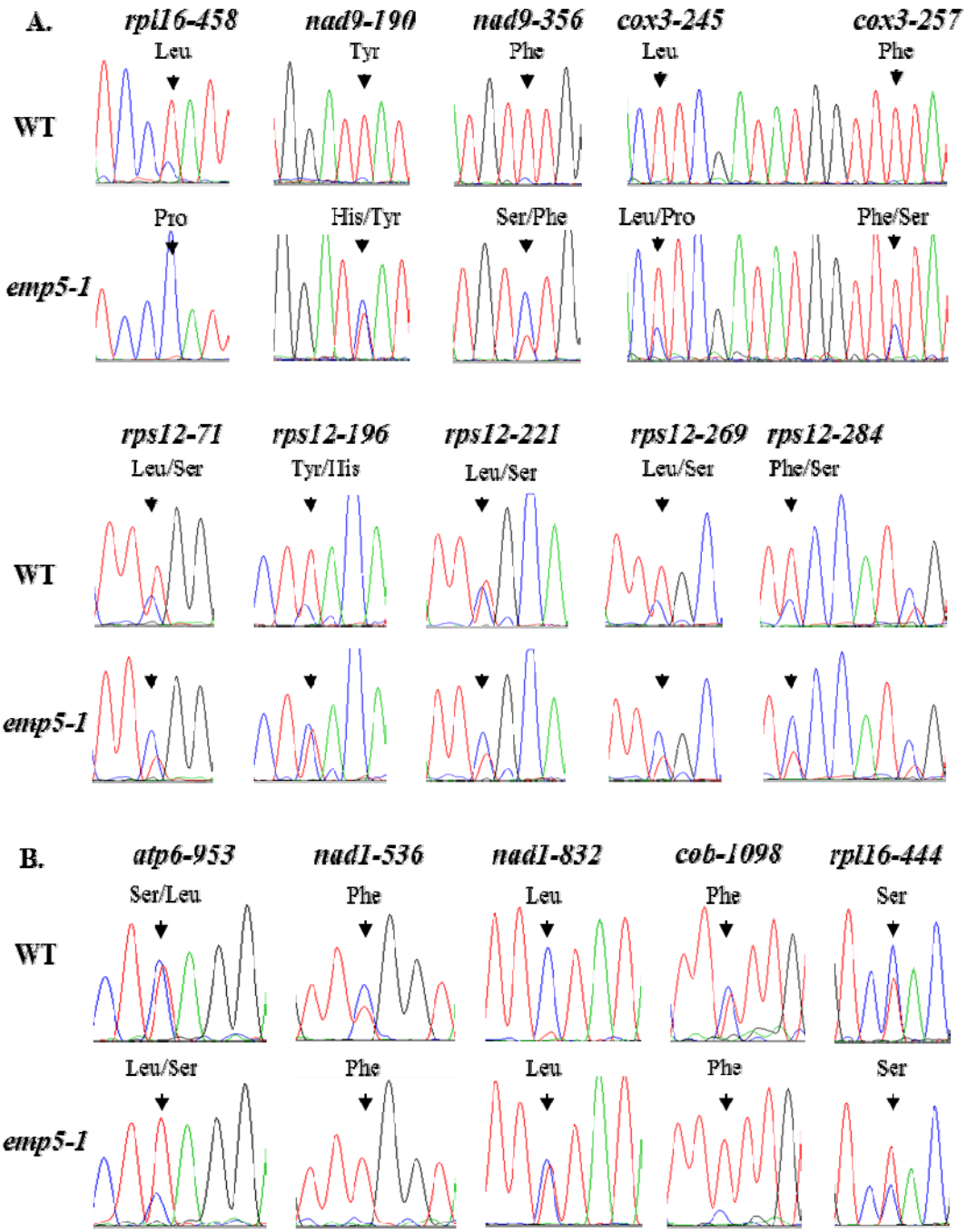

Figure 15. RNA editing defects of mitochondria genes in *emp5-1* mutant.

(A) RNA editing of 10 sites in 4 mitochondrial gene transcripts decreased in *emp5-1*.

(B) RNA editing of 5 sites in 4 mitochondrial gene transcripts increased in *emp5-1*. Sequence chromatogram of PCR amplified wild-type cDNA or *emp5-1* cDNA of the editing sites is shown. The position of RNA editing represents the name of transcripts and the edited C position. The amino acid change is indicated at the top, and most coding amino acid was put in the front in these overlap peak sites. Arrowheads indicate the editing sites.

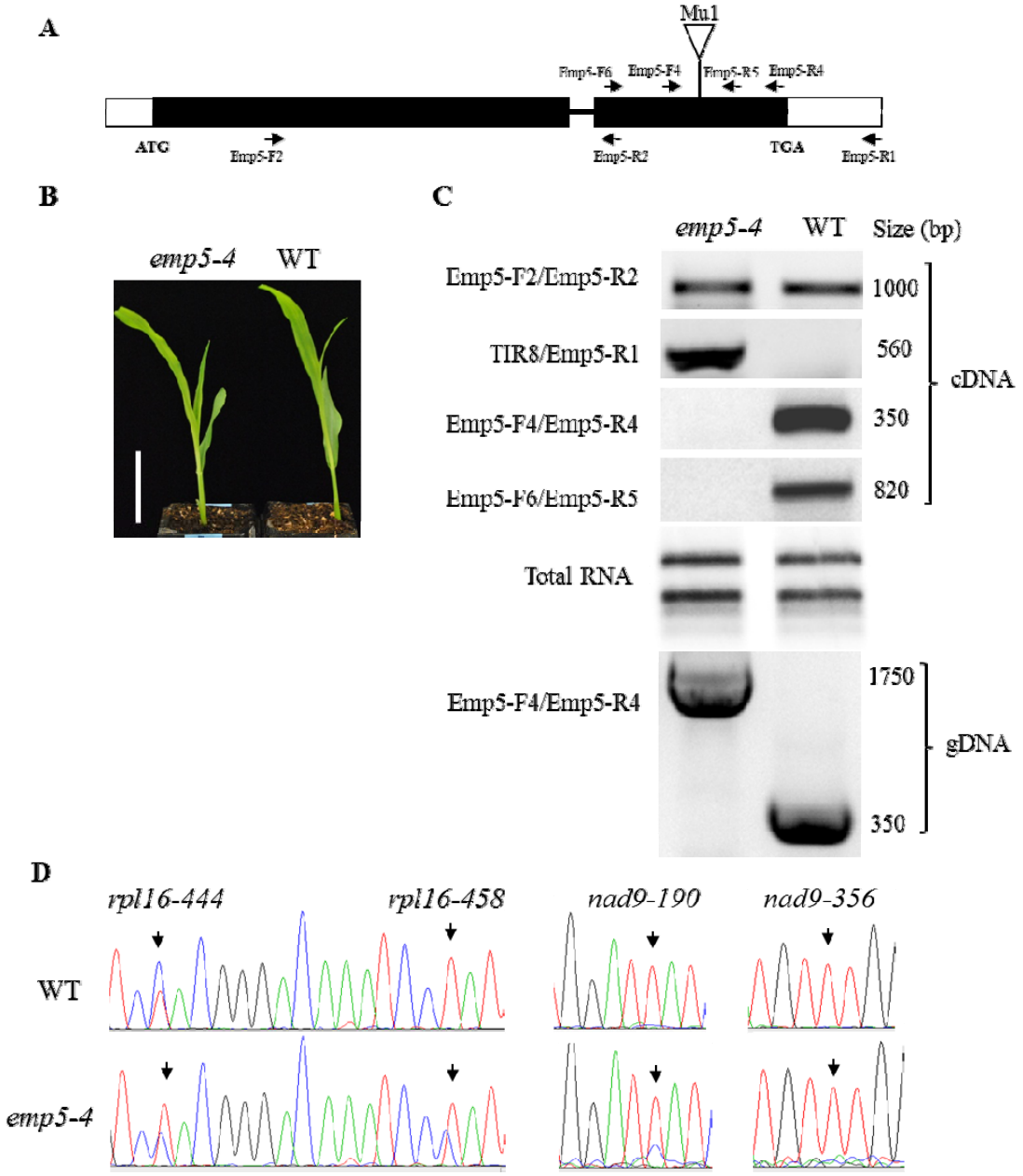

Figure 16. *emp5-4* is a weak allele.

(A) Gene structure of *Emp5* and location of *Mu1* insertion in *emp5-4* was shown. The position of primers used in RT-PCR Emp5-F2 and Emp5-R2, Emp5-F6 and Emp5-R5, Emp5-F4 and Emp5-R4, and Emp5-R1, were shown.

(B) The seedling phenotype of *emp5-4* is indistinguishable from the WT. Bars=5 cm.

(C) Expression pattern of *Emp5* in *emp5-4* and WT seedling leave was shown by RT-PCR. Primer Emp5-F2 and Emp5-R2, TIR8b and Emp5-R1, Emp5-F4 and Emp5-R4, Emp5-F6 and Emp5-R5 were used in RT-PCR respectively. PCR reaction with Emp5-F4 and Emp5-R4 primers on genomic DNA was used as a control.

(D) Sequence chromatogram of PCR amplified *emp5-4* cDNA of the *rpl16* editing sites is shown, compared with WT. The position of RNA editing represents the name of transcripts and the edited C position. Arrowheads indicate the editing sites.

4.7 Functional Analysis of the Rice *OsEmp5* Gene

The rice OsEMP5 protein is a putative ortholog of EMP5 in maize, sharing a high degree of similarity (84%) with EMP5 and containing 11 PPR repeats in the N-terminus, and E/E+/DYW domains in the C-terminus (Figure 12). To address whether OsEMP5 has a conserved function similar to the EMP5 in maize, we characterized the OsEMP5 in rice. First, we determined the subcellular localization of OsEMP5 by expressing $OsEMP5^{N372}$:GFP fusion in tobacco leaf epidermal cells. Confocal laser scanning microscopy analysis revealed that the GFP signal of $OsEMP5^{N372}$:GFP was co-localized with MitoTracker Red in mitochondria (Figure 17), confirming that OsEMP5 is targeted to the mitochondrion. To study the function of OsEMP5, OsEmp5 RNAi transgenic rice was created. It is likely that severe knock-down transgenic lines would be lethal considering the essential role of EMP5 in maize, therefore only weak knock-down lines were likely generated. About twenty lines of transgenic rice were obtained and three representational lines were used for further molecular analysis. Southern analysis of the three *OsEmp5* RNAi transgenic lines (line 19, 23 and 33), using the *hygromycin phosphotransferase (hpt)* gene as a probe, confirmed that they were independent transgenic lines, each carrying one copy of transgene (Figure 18A). RT-PCR analysis revealed the *OsEmp5* expression level was significantly decreased in lines 19 and 23, and slightly decreased in line 33 when comparing to the WT (Figure 18B). The transgenic plants grew much slower than the WT at the seedling stage. But after the seedling stage, the plants gradually recovered and grew to normal adult plants (Figure 18C). The T1 progeny of all these three transgenic lines segregated defective seeds as different ratio (Figure 18C). The ratio of defective seeds in line19 is 24.7% (199:807), line 23 is 22.3% (93:417), and line 33 is 9% (53:584). The severity of seed phenotype was roughly consistent with the suppressed *OsEmp5* expression level. This result indicated that similar to *Emp5* in maize, *OsEmp5* is essential to seed development in rice.

To reveal the effect on mitochondrial RNA editing in these OsEmp5-RNAi transgenic lines, mitochondrial *rpl16* transcripts were amplified and sequenced in the

three *OsEmp5* RNAi transgenic lines and WT plants to compare RNA editing difference. This analysis revealed that less than 50% *rpl16*-458 site in all three transgenic lines was edited, comparing to 100% editing in WT (Figure 19). For *rpl16*-444 site, the editing level was increased in the three transgenic lines, which is similar to the maize *emp5-1* allele (Figure 19). This result confirmed that similar to maize EMP5, OsEMP5 is required for the editing of *rpl16*-458 in rice mitochondria.

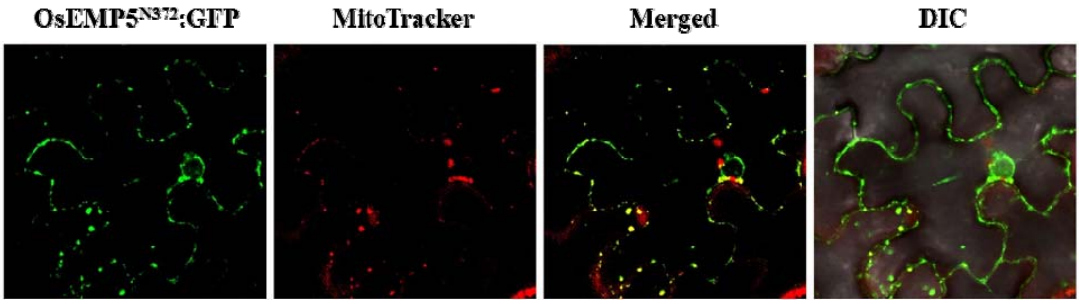

Figure 17. OsEMP5 is a mitochondrion localized PPR-DYW protein. OsEMP5^{N372}:GFP fusion protein was transient expressed in tobacco epidermal cells. Fluorescence signals from OsEMP5^{N372}:GFP (green) and Mito Tracker stained mitochondria (red) were detected by confocal microscopy.

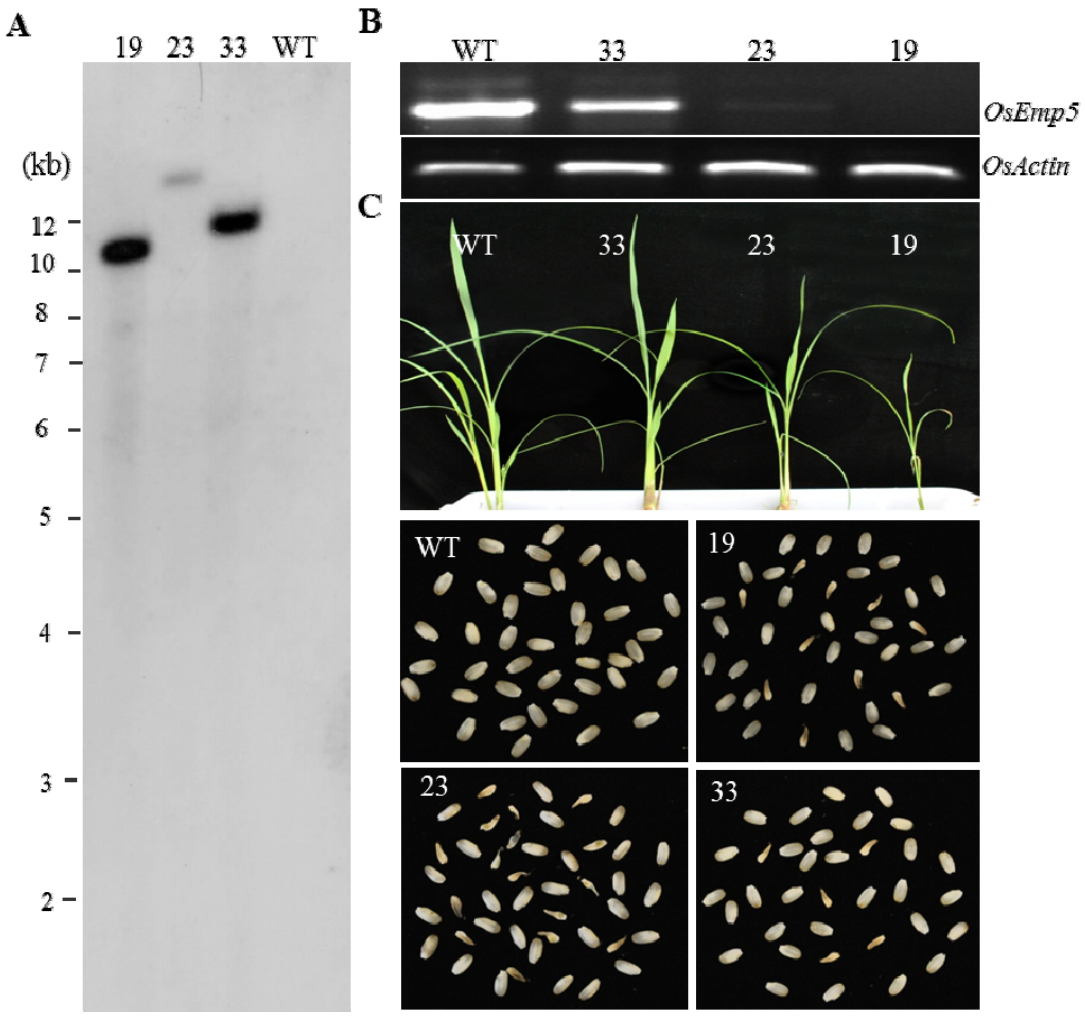

Figure 18. RNAi knock-down of *OsEmp5* expression in transgenic rice.

(A) Southern analysis of independent OsEmp5 RNAi transgenic lines (line 19, 23 and 33) by using the *hpt* gene as a probe.

(B) RT-PCR analysis of endogenous *OsEmp5* expression in 3 transgenic lines. *OsActin* (X15865.1) was used as control. Primer Osactin-F, Osactin-R, OsEmp5-F and OsEmp5-R (Supplementary Table 1) were used in RT-PCR.

(C) Phenotypes of OsEmp5 RNAi transgenic plants at seedling stage and in T1 progeny segregating defective seeds.

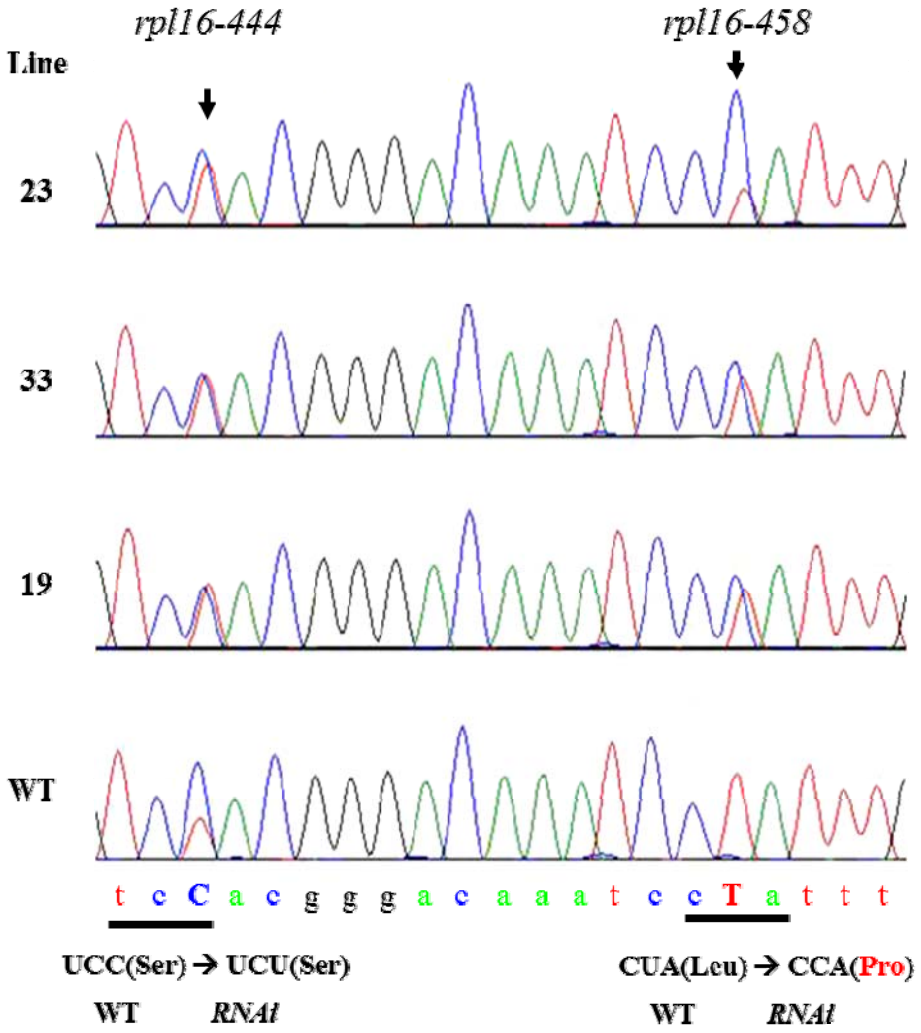

Figure 19. Analysis of mitochondrial *rpl16* editing in OsEmp5-RNAi transgenic lines.

Sequence chromatogram of PCR amplified wild-type cDNA or 3 *OsEmp5* RNAi transgenic lines cDNA of the *rpl16* editing sites is shown. The position of RNA editing represents the name of transcripts and the edited C position. Arrowheads indicate the editing sites. The amino acid change is indicated at the bottom.

Chapter V.

Discussion

5.1 Abortion of *emp5-1* Mutant Seed Development is Caused by Defective Mitochondrial RNA Editing

The *emp* mutants are defined by a dramatic reduction in embryo and endosperm size, yet possess a normal pericarp (Sheridan and Neuffer, 1980). In maize, three mutants with a similar phenotype have been characterized. *Empty pericarp 2* encodes a repressor of a heat shock response in seeds, mutation of which unleashes a heat shock response causing seed development abortion (Fu et al., 2002). *Emp4* encodes a PPR protein that is required for normal level expression of several mitochondrial genes (Gutierrez-Marcos et al., 2007). Recently, MPPR6, a PPR gene which mutation causes an empty pericarp phenotype, is required for maturation and translation initiation of *rps3* mRNA in mitochondria (Manavski et al., 2012). In this study, we show that *Emp5* encodes a DYW subclass PPR protein that functions in the editing of several maize mitochondrial gene transcripts. The cloning of *Emp5* is supported by multiple independent insertions in the *Emp5* gene that condition a typical *emp* phenotype and also by the further functional analysis in editing. And such function in mitochondrial editing is conserved in its ortholog *OsEmp5* in rice. The results indicates that the deficiency of editing in these mRNAs compromises the mitochondrial function, causing the embryo and endosperm development to be arrested at transition stage. Since the pericarp tissue is maternal, growth of the pericarp was not expected to be affected in heterozygous *Emp5/emp5* plants. Thus, the mutant kernels are appeared as *empty pericarp*. Considering the critical functions of mitochondria to both the embryo and the endosperm, the *emp* mutants should be enriched with genes that have key functions in mitochondria. In maize, the *emp* mutants form a distinct class which will be ideal genetic materials for dissecting the mitochondrial PPR functions.

In the severe allele of *emp5-1*, the editing of *rpl16*-458 was completely abolished, and the other 9 editing sites in 3 different transcripts (*nad9, cox3* and *rps12*) are also diminished. *rpl16* and *rps12* encode ribosomal proteins of the mitochondrial translation machinery. *nad9* and *cox3* are required for complex I and Complex IV

function in electron transport chain respectively. However, a significant portion of correctly edited transcripts of *nad9, cox3* and *rps12* still exist. It is not clear of the impact on seed development as a result of reduced editing in these three transcripts. A fraction of the correctly edited mRNA may be sufficient for mitochondrial function to complete embryogenesis. This is indicated in other mitochondrial genes such as *cob*. A complete loss of *cob*-908 editing severely reduces the growth in maize, whereas a residual level of *cob*-908 editing is sufficient to assure normal plant growth in *Arabidopsis* (Sosso et al., 2012a). In addition, our editing analysis in maize mitochondrial transcripts showed that partial editing for a particular site is common in WT mitochondrial transcripts. This argues a possibility that the reduced editing in the 9 sites of three transcripts (*nad9, cox3* and *rps12*) may not be the major cause for the *emp* phenotype. Instead, the complete loss of a single *rpl16*-458 editing site may be attributable to the abortion of *emp5* mutant seed development.

The *rpl16* transcript lacking *rpl16*-458 editing translates to a protein with a proline instead of leucine residue at position 153, which may severely compromise the RPL16 function in mitochondrial protein translation. Proline residue often acts as a structural disruptor in protein secondary structure such as alpha helices and beta sheets. Thus, this change may severely affect the structure and the function of the RPL16 protein. There are several cases where an unedited site coding for proline cause severe impact on plant growth and development. The maize mitochondrial *cob*-908 is normally C-to-U edited by PPR2263 to render a proline to leucine change. The unedited *cob*-908 in *ppr2263* mutants caused defects in seed development and seedling growth (Sosso et al., 2012a). In the *slg1* mutant of *Arabidopsis*, RNA editing in a single site *nad3*-250 in mitochondria is abolished, resulting in a codon for proline instead of the edited for serine. This single amino acid mutation caused slow growth and delayed development (Yuan and Liu, 2012a). The mutations of *OGR1, SLO1* and *MEF11* also lead to absence of RNA editing of mitochondrial transcripts, which all resulted in leucine to proline changes in *nad4*-416, *nad4*-449 and *cox3*-422 respectively (Kim et al., 2009; Sung et al., 2010; Verbitskiy et al., 2010). All three

mutants exhibited significant defects in plant growth and development. Conceivably, genetic screens will enrich the identification of changes critical to protein functions. These cases demonstrate that the editing for transition from proline to other amino acids is critical to the function of multiple mitochondrial proteins.

The *rpl16* gene has been proved to be essential to mitochondrial gene expression as a key component in the translation machinery. Thus far, a complete deletion of the *rpl16* gene in mitochondria has not been isolated, presumably due to its lethality (Newton et al., 1996). In *Arabidopsis maternal distorted leaf* (*mdl*) mutant, rearrangement in two mtDNA fragments associated with the *rps3-rpl16* polycistron resulted in a deletion of part of the intron and exon b of *rps3* that is upstream of *rpl16* but without affecting the *rpl16* coding region. The mutant showed poor growth, distorted rough leaves and aborted flowering organs (Sakamoto et al., 1996). A deletion of the 5' UTR sequences including the promoter for the transcription unit of the *rps3-rpl16* polycistron in maize *nonchromosomal stripe 3 (ncs3)* and *ncs4* mutants causes severe stunted and striped leaves, and male fertility respectively (Hunt and Newton, 1991; Newton et al., 1996). The mRNA abundance corresponding to the *rps3/rpll6* coding region was specifically reduced in these mutants. Mitochondrial protein synthesis was dramatically reduced in severely affected mutant plants. These results reveal that the RPL16 protein is essential for protein synthesis in mitochondria, which consequently will be crucial for plant development. Therefore, the failed development in both embryo and endosperm in the *emp5* mutant is likely due to a loss of *rpl16*-458 editing.

5.2 Increased Editing in the *emp5* Mutant

The mechanism how RNA editing is carried out in plastids and mitochondria is not known. PPR proteins are considered as the specificity determinant that recognizes different transcripts (Shikanai, 2006; Chateigner-Boutin and Small, 2010). A single PPR protein such as EMP5 responsible for multiple editing sites in multiple transcripts has been reported in several PPR-DYW proteins such as MEF1, MEF11

and OGR1 in *Arabidopsis* and rice (Kim et al., 2009; Zehrmann et al., 2009; Verbitskiy et al., 2010). A considerate amount of RNA was found still correctly edited in the *emp5* null mutant, raising a reasonable possibility that overlapping editing on the same site may be mediated by different PPR proteins. In addition, we found that five editing sites on four mitochondrial transcripts were increased in the *emp5-1* mutants, interspersing in *atp6, nad1, cob* and *rpl16* transcripts (Figure 15B). This result was also confirmed in *OsEmp5* RNAi knock down transgenic rice where *rpl16*-444 editing level is increased in all three transgenic lines compared to the WT. Intriguingly, the editing in these sites do not change the amino acid coding except in *atp6*-953 where editing converts Ser to Leu. This phenomenon has also been reported in PPR protein REME1, mutation of which reduces the editing of *nad2*-558 and *orfX*-552, but increases the editing extent in at least two sites, *matR*-1771 and *rpl5*-92 (Bentolila et al., 2010). It seems that PPR proteins may work as a positive and negative regulator of organelle RNA editing at the same time. One possibility is that over-lapping functions by PPR proteins in editing a single site exist in plants such that the deficiency in one PPR leads to a compensational expression of another PPR protein that edits another set of sites, overlapping but not identical. In this case, some sites will be edited more than the WT as a result of the compensational PPR proteins. Another possibility is that although a *Mu2* is inserted into the *Emp5* gene, it can still be transcribed and translated into a truncated protein that recognizes different RNA substrates. For the *emp5-1* allele, the *Mu2* insertion disrupted the 6 of the 11 PPR motifs, allowing five PPR repeats remained in the mutant protein if it can be translated (Figure 11A).

5.3 Substrate Specifying Sequences of EMP5 Are Not Conserved

Two PPR proteins, CRR4 and RF1, were reported for possessing an RNA binding activity in vitro without other factors (Okuda et al., 2006; Kazama et al., 2008), leading to a hypothesis that PPR proteins can specifically recognize the cis-element of

an editing site. Some RNA editing sites were shown to have conserved sequences upstream of the editing sites (Karcher et al., 2008; Kobayashi et al., 2008; Sosso et al., 2012a). However, these conserved sequences were derived by comparing merely two editing sites. In cases where a PPR protein is involved in the editing of multiple sites such as OGR1, MEF1, MEF11 and CRR22, no conserved sequences in the corresponding region can be identified (Kim et al., 2009; Okuda et al., 2009; Zehrmann et al., 2009; Verbitskiy et al., 2010). In CLB19, the sequences surrounding the two editing sites showed little sequence similarity (Chateigner-Boutin et al., 2008). In an attempt to identify conserved sequences for EMP5 edited sites, we aligned the adjacent sequences near the editing sites (region from -40 to +20). We did not find any conserved sequences except most of the -1 base is T (Figure 20). The likely possibility is that PPR proteins recognize a specific secondary structure of the transcripts, not the primary sequence. It is hypothesized that PPR editing factors can only distinguish pyrimidines from purines and, at some positions, must be able to recognize specific bases (Hammani et al., 2009). Recently, Barkan et al used computational methods to infer a PPR-RNA recognition mechanism where a combination of position 6 amino acid residue in a PPR motif and the 1' position of the next PPR motif recognizes one nucleotide in the RNA substrate (Barkan et al., 2012). Another group further showed the involvement of an additional amino acid (residue 1) in RNA recognition (Yagi et al., 2013). As such, the tandem PPR motifs are decoded to a nucleotide sequence that specifies the substrate RNA of the PPR protein. It was validated by recoding a PPR protein to bind novel RNA sequences *in vitro*. However, we failed to make the connection between EMP5 and *rpl16.*

5.4 The E+ and DYW Motif of EMP5 Is not Essential for This Protein Function

PPR proteins are classified into P, PPR-E, and PPR-DYW subclasses based on presence of additional C-terminal motifs (Lurin et al., 2004). The PPR repeats are proposed to recognize the target RNA sequences (Shikanai, 2006). In MEF11, the second PPR repeat was shown to be crucial for the specific editing of *cox3*-422,

nad4-124 and *ccb203*-344 (Verbitskiy et al., 2010). The DYW domain which showed significant similarity to cytidine deaminases was proposed to have catalytic editing activity (Salone et al., 2007). Indeed, its presence is correlated with presence of RNA editing in plant evolution (Fujii and Small, 2011). Although PPR proteins such as CRR4, CLB19, MEF9 and SLO1, all lacking a DYW domain, still possess the function of C-to-U RNA editing (Kotera et al., 2005; Chateigner-Boutin et al., 2008; Sung et al., 2010; Takenaka, 2010), a recent report shows that the DYW domain can be supplied in trans to the CRR4 protein and it is essential for the editing (Boussardon et al., 2012). Therefore, the hypothesis that the DYW domain functions as the deaminase enzyme remains. Up to this work, all the PPR proteins involved in editing contain an E domain, but the E domain lacks any obvious catalytic characteristics, suggesting that the E domain is indispensable for editing with unknown function (Shikanai, 2006; Okuda et al., 2007). It is possible that the E domain mediates protein-protein interaction to recruit another protein with deaminase activity, which may include PPR-DYW proteins.

In the *emp5-4* allele, a *Mu1* element is inserted in the middle of E+ motif, disrupting the E+ motif and DYW motif. RT-PCR analysis using several sets of primers revealed the presence of two transcripts where one contained the 5' region of the *Emp5* gene and ended inside the *Mu1*, and the other contained the 3' region of the *Emp5* gene, but no transcript across the *Mu1* element existed. We analyzed the two transcripts for ORFs that would possibly encode proteins with a scenario like the PPR-E protein with a trans-supplied DYW protein. As shown in Figure 21, the translation of first transcript is predicted to terminate by a stop codon TAA, only adding two amino acids encoded by the *Mu1* TIR sequence. This will produce a truncated EMP5 protein, similar to a PPR-E protein, and if translated, it should be able to target to mitochondria. The second transcript predicts only one ORF in frame with the DYW domain. The translation adds five amino acid residues (MAIIS) from the *Mu1* sequence to the N-terminus of this hypothetical protein. However, the N-terminus sequence did not predict a mitochondrion signal peptide by currently

available algorisms. Therefore, it is highly unlikely that the protein would target to the mitochondrion, even if it can be translated. Alternatively spliced transcripts were not detected although we intentionally set the conditions favoring the amplification of potentially spliced *emp5-4* transcripts. This leads us to conclude that the insertion in the *emp5-4* allele causes a possible deletion of the E+ domain and the entire DYW domain. However, a truncated version of the *EMP5-4* protein may be produced. Genetic analysis indicated that the *emp5-4* allele appeared normal in growth and development. And most of the editing events targeted by EMP5 showed similar editing level comparing with WT, except the editing of *rpl16*-458 which is considered critical showed partial editing. In addition, *rpl16*-444 and *nad9*-190 editing efficiency was also changed slightly. The presence of editing in *emp5-4* that lacks the E+ and DYW motif indicates that E+ and DYW domains are not essential for EMP5 editing function. This conclusion is consistent with the plastid-located CRR22, CRR28, and OTP28, and the mitochondrial factor MEF3 and MEF11 (Okuda et al., 2009; Okuda et al., 2010; Verbitskiy et al., 2010; Zehrmann et al., 2011; Verbitskiy et al., 2012). However, the DYW domain of MEF1 cannot be destroyed without severe effects on its function in editing (Zehrmann et al., 2009). Therefore, it is possible that EMP5 associates with a DYW-container partner, which can almost completely complement E+ and DYW motif truncation and carry out editing, although inefficiently. But for MEF1, no such a DYW-container partner to associate, so the DYW domain cannot be deleted or mutated.

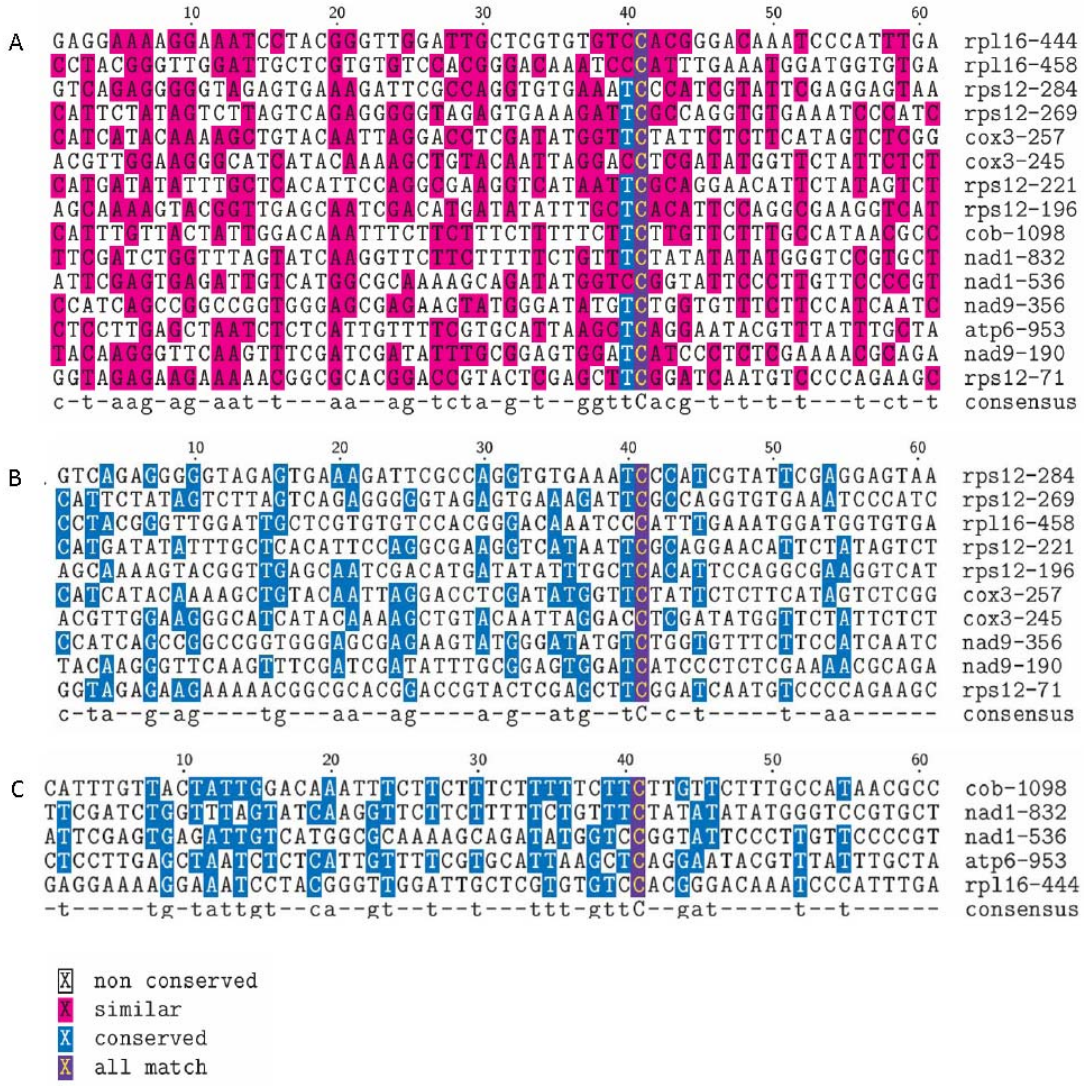

Figure 20. Analysis of consensus sequences in the 5' region of the EMP5 editing sites

A. Alignment of -40 to +20 sequences of all 15 RNA editing sites in 7 mitochondrial gene transcripts changed in *emp5-1*.

B. Alignment of -40 to +20 sequences of 10 RNA editing sites in 4 mitochondrial gene transcripts decreased in *emp5-1*.

C. Alignment of -40 to +20 sequences of 5 RNA editing sites in 4 mitochondrial gene transcripts increased in *emp5-1*.

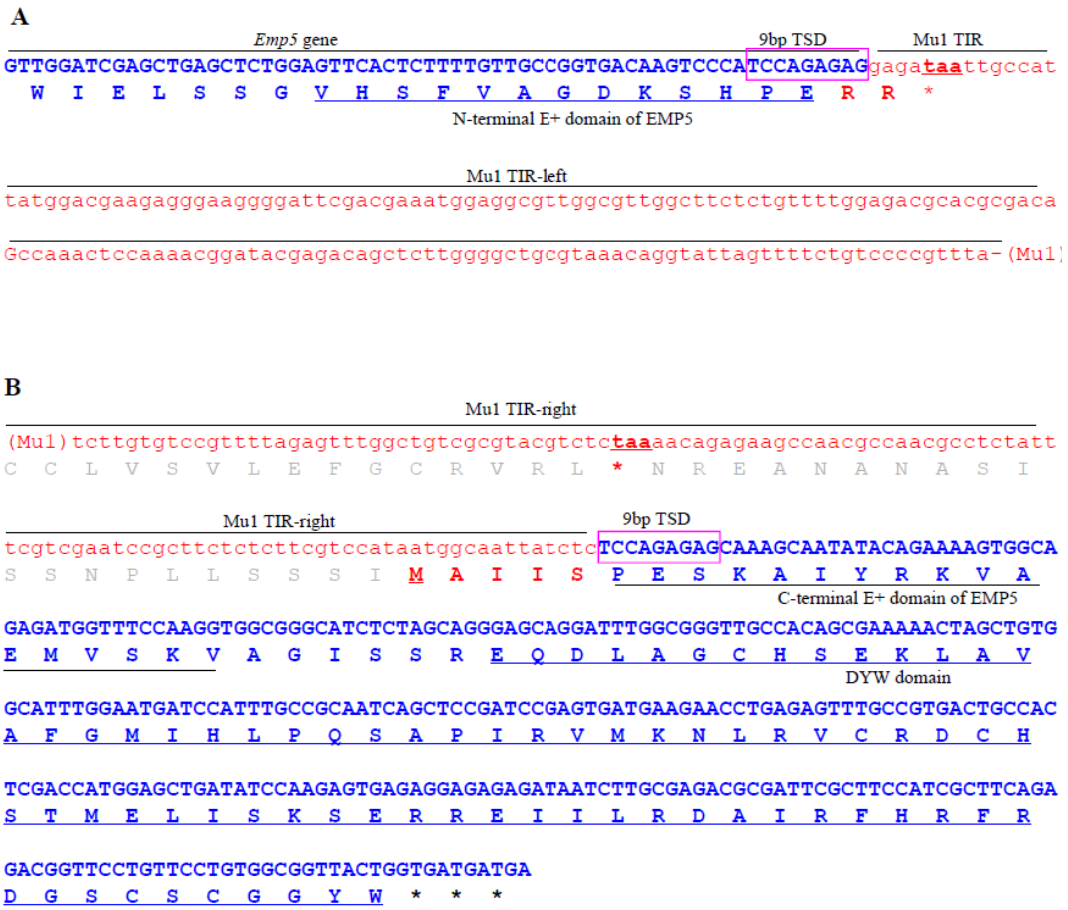

Figure 21. The hypothetical proteins as predicted based on the two transcripts detected in the *emp5-4* mutant seeds.

A. The 5' transcript of *emp5-4* predicts a putative truncated protein with addition of two Arg residues to the C-terminus before a stop codon from the *Mu1* TIR sequence. *Emp5* sequence is highlighted in blue and the *Mu1* sequence is highlighted in red. The 9bp TSD is marked.

B. The 3' transcript of *emp5-4* predicts a putative protein with addition of five amino acid residues (MAIIS) to the N-terminus of a DYW motif containing protein.

Table 2. Primers used in this study.

Primer name	Primer sequence	Use
TIR6	AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC	Inverse PCR
Mu1-62	CCCTTCCCTCTTCGTCCATAAT	Inverse PCR
Emp5-RTF1	GCCAGTAACGTGCAAAGATGAG	Cloning full length *Emp5* cDNA
Emp5-R1	TAGGGTTTGTTCCTGGTCGTTT	Cloning full length *Emp5* cDNA, RT-PCR
Emp5-ENF1	CACCATGGAGGCCTTCTACCTCCA	Construct full length *Emp5:GFP* and $Emp5^{N469}$*:GFP* vector
Emp5-ENR1	CCAGTAACCGCCACAGGAACAG	Construct full length *Emp5:GFP* vector
Emp5-R2	ATCAACATCCCCTGACCTAGCA	Construct $Emp5^{N469}$*:GFP* vector; RT-PCR; genotyping
Emp5-F2	GTTGTTCCTGTCCAACTCGGTC	RT-PCR
TIR8b	CGCCTCCATTTCGTCGAATCCSCTT	Genotyping
Emp5-GTF1	GCTCTGACCCACGATCTTTCTT	Cloning full length *Emp5* gDNA
Emp5-GTR1	CAGTCTTGATTCGGTCGCATAG	Cloning full length *Emp5* gDNA
Emp5-R5	CCTGCTCCCTGCTAGAGATG	RT-PCR
Emp5-F6	GCTAGGTCAGGGGATGTTGA	RT-PCR
Emp5-R4	GTCTCTGAAGCGATGGAAGC	RT-PCR
Emp5-F4	CGTGAAGAAAGAGCCTGGTC	RT-PCR
nad3-F1	CTTTCCTATGTCCTTCCCCC	Gene amplification and sequencing in RNA editing analysis
nad3-R1	GAGGAGAGCGAGAGAACGAA	Gene amplification and sequencing in RNA editing analysis
nad4L-F1	CTGACATTCCATGTTTCCGA	Gene amplification and sequencing in RNA editing analysis
nad4L-R1	GAAGAGAACGAAAGGAGAACAGA	Gene amplification and sequencing in RNA editing analysis
nad6-F1	TGGAAAAACCAAACCCACAT	Gene amplification and sequencing in RNA editing analysis
nad6-R1	CAAGTTCCCTTGGCGTAGTC	Gene amplification and sequencing in RNA editing analysis
nad9-F1	AGCAAGAAGCGGAACAAAAA	Gene amplification and sequencing in RNA editing analysis
nad9-R1	TATTGATTTGTCCCCTCCCC	Gene amplification and sequencing in RNA editing analysis
nad7-F1	GTTTTGGCTCGCAATAAAGC	Gene amplification and sequencing in RNA editing analysis
nad7-R1	CAGGTGGGACAAGCTCTAGG	Gene amplification and sequencing in RNA editing analysis
nad4-F1	CAGTCACCCGGAGAAGATTT	Gene amplification and sequencing in RNA editing analysis
nad4-R1	TAATTTGGCGCCTGATTGAC	Gene amplification and sequencing in RNA editing analysis
nad4-F2	TTCTCCAAACAGGAACCACC	Gene sequencing in RNA editing analysis
nad1-F1	GGCCCGATCATGAGTGAATA	Gene amplification and sequencing in RNA editing analysis
nad1-R1	GCCCCCTTCAGAAGAAACTT	Gene amplification and sequencing in RNA editing analysis
nad2-F1	GACGGAGGAGAGGAAATGAA	Gene amplification and sequencing in RNA editing analysis
nad2-R1	GCAGTCCACCCTTTCTTTGA	Gene amplification and sequencing in RNA editing analysis
nad5-F1	CGCTCGAACATTGTCTGATT	Gene amplification and sequencing in RNA editing analysis
nad5-R1	GTCCTGGCAAGCTCCTACAG	Gene amplification and sequencing in RNA editing analysis
nad5-F2	GGTTTACACGACTTCAGGCG	Gene sequencing in RNA editing analysis
nad5-R2	GAGCAGCAAACTCGGATTCG	Gene sequencing in RNA editing analysis
rps3-F1	GCAGAAAGGGGCAAAAGTAA	Gene amplification and sequencing in RNA editing analysis
rps3-R1	TCGCGACCCCTACTACATCT	Gene amplification and sequencing in RNA editing analysis
rps3-F2	TCAAGCATCCGAAATACGCC	Gene sequencing in RNA editing analysis
rps13-F1	TCATGATGATTAAGGGAAGAGTGA	Gene amplification and sequencing in RNA editing analysis

rps13-R1	TTGAATTGAACAGTGTGATTGAT	Gene amplification and sequencing in RNA editing analysis
rps12-F1	CTAGCTGCTTCCATATCGCC	Gene amplification and sequencing in RNA editing analysis
rps12-R1	CGGATCGGGAGTAACCACTA	Gene amplification and sequencing in RNA editing analysis
rps12-ct-F1	TGTACGGTTCTGTAGAGGGACA	Gene amplification and sequencing in RNA editing analysis
rps12-ct-R1	TCCGTTTTCTTTTTATAAGGGC	Gene amplification and sequencing in RNA editing analysis
rps12-ct-F2	TGCCTTACGTAAAGTTGCCA	Gene sequencing in RNA editing analysis
rps7-ct-F1	TTGAACCTCTTTCACGCTCA	Gene amplification and sequencing in RNA editing analysis
rps7-ct-R1	TTCCGATCGAGATGTATGGA	Gene amplification and sequencing in RNA editing analysis
rps7-F1	TTCGTTGGAAAAACCTACGC	Gene amplification and sequencing in RNA editing analysis
rps7-R1	ATGAGGAAGGCCGATTTTCT	Gene amplification and sequencing in RNA editing analysis
rps4-F1	AGAGTTGGGTTCGATTCCCT	Gene amplification and sequencing in RNA editing analysis
rps4-R1	AGCGACTAGGCCGATCTTTT	Gene amplification and sequencing in RNA editing analysis
rps2B-F1	TCCATGGACCCACGTAAAAT	Gene amplification and sequencing in RNA editing analysis
rps2B-R1	GGCCCCTCTCTGATAAGGAA	Gene amplification and sequencing in RNA editing analysis
rps2A-F1	CAGGAAAGATATTTGCCCCA	Gene amplification and sequencing in RNA editing analysis
rps2A-R1	CCTGTATCTCCGGAAACGAA	Gene amplification and sequencing in RNA editing analysis
rps2A-F2	AGAACAGCAATCAACGGGC	Gene sequencing in RNA editing analysis
rps1-F1	AAGGTGGGCTTCGGATTATT	Gene amplification and sequencing in RNA editing analysis
rps1-R1	TCTTCAGTTTTACGCTTACGCT	Gene amplification and sequencing in RNA editing analysis
rpl16-F1	GGTTTTTCCCCACTAACCAA	Gene amplification and sequencing in RNA editing analysis
rpl16-R1	GGGTGCGGAAATAGCTAGAA	Gene amplification and sequencing in RNA editing analysis
atp1-F1	CGTTGCTGGTGAAGAAGCAT	Gene amplification and sequencing in RNA editing analysis
atp1-R1	AAAAGCGGATTTATCCATCG	Gene amplification and sequencing in RNA editing analysis
atp1-R2	ACCGCAGACTTGTTTCATAGC	Gene sequencing in RNA editing analysis
atp4-F1	AGCCACGTGCTCTAATCCTC	Gene amplification and sequencing in RNA editing analysis
atp4-R1	TCCCTTTCTCTTGGAGCAGA	Gene amplification and sequencing in RNA editing analysis
atp6-F1	CCAAGTCTCTTTTGGGAGCA	Gene amplification and sequencing in RNA editing analysis
atp6-R1	GGCTCCTCGTTTTTATGCAA	Gene amplification and sequencing in RNA editing analysis
atp6-F2	ATCATTATCCCTGGAGGCGG	Gene sequencing in RNA editing analysis
atp6-F3	CCTTGCATCTCGGTCACTTT	Gene sequencing in RNA editing analysis
atp8-F1	GGCAAGGATCCTCAGTCCTA	Gene amplification and sequencing in RNA editing analysis
atp8-R1	GAGGGTTGGTTTGATTGGAA	Gene amplification and sequencing in RNA editing analysis
atp9-F1	AGGGGCCTCGTCATCTCTAT	Gene amplification and sequencing in RNA editing analysis
atp9-R1	TAGTTGCGAAGGAAAAGCGT	Gene amplification and sequencing in RNA editing analysis
ccmB-F1	AGCCGTCGAAGTGAATGAAT	Gene amplification and sequencing in RNA editing analysis
ccmB-R1	AACGGCTTTTCCATGACTTG	Gene amplification and sequencing in RNA editing analysis
ccmB-F2	CTTTCCTCCCGAACCTTTTC	Gene amplification and sequencing in RNA editing analysis
ccmB-R2	GAAAAGGTTCGGGAGGAAAG	Gene sequencing in RNA editing analysis
ccmC-F1	ACTTGCAAGGCAAGGAAAAA	Gene amplification and sequencing in RNA editing analysis
ccmC-R1	CCATGGATGCTTTAGCGAGT	Gene amplification and sequencing in RNA editing analysis
ccmFC-F1	GAGAAGCTCAAATCGAACGG	Gene amplification and sequencing in RNA editing analysis
ccmFC-R1	CGCAGCCACTATTTTGACTC	Gene amplification and sequencing in RNA editing analysis
ccmFN-F1	TGAAGATTGTAAGGCGTTTCC	Gene amplification and sequencing in RNA editing analysis

ccmFN-R1	GGATCATCCTGTGGTTACCG	Gene amplification and sequencing in RNA editing analysis
ccmFN-F2	GAACAGCAGATTGACGGAGC	Gene sequencing in RNA editing analysis
cob-F1	ATCAAGGCAAGGGGGTAAAT	Gene amplification and sequencing in RNA editing analysis
cob-R1	GGTGTGATCAGTCTCATCCG	Gene amplification and sequencing in RNA editing analysis
cob-F2	TGTTTGGTGTCTCGGAGTTGT	Gene sequencing in RNA editing analysis
cob-F3	CTACCGATCCATGCCATTCT	Gene sequencing in RNA editing analysis
cox1-F1	GGCCCCTCTCTGATAAGGTT	Gene amplification and sequencing in RNA editing analysis
cox1-R1	GTTAAGGCAAAGCCCAAACA	Gene amplification and sequencing in RNA editing analysis
cox1-F2	GGTTGTTGCCACCAAGTCTC	Gene sequencing in RNA editing analysis
cox1-F3	TTTTGGTTCTTCGGTCATCC	Gene sequencing in RNA editing analysis
cox1-F4	TTCTATGGGAGCCGTTTTTG	Gene sequencing in RNA editing analysis
cox2-F1	GTCCTACTTCTGGTGCTGCC	Gene amplification and sequencing in RNA editing analysis
cox2-R1	GAGAATTGCATTTCCGCTTC	Gene amplification and sequencing in RNA editing analysis
cox3-F1	TCAATCCACTTATTCGTTCCC	Gene amplification and sequencing in RNA editing analysis
cox3-R1	GTTTACATACAACCGGGGCA	Gene amplification and sequencing in RNA editing analysis
Mat-r-F1	AACGCCTGTTCGCTAAAATC	Gene amplification and sequencing in RNA editing analysis
Mat-r-R1	AGGCTTTGCTCCCCTTTTT	Gene amplification and sequencing in RNA editing analysis
Mat-r-F2	CCAAGACAACAGAGCCCTCA	Gene sequencing in RNA editing analysis
Mat-r-R2	ACTATGTCTCTGTCGCTGACG	Gene sequencing in RNA editing analysis
mttB(orfX)-F	TTGGTTTAGAATTGCTCGGG	Gene amplification and sequencing in RNA editing analysis
mttB(orfX)-R	AGGGGGAACCCTACCGAC	Gene amplification and sequencing in RNA editing analysis
OsEmp5-KpnI	GGGGTACCCACTGTCATGGGGAACTCTTG	Rice pTCK303-OsEMP5 RNAi vector construction
OsEmp5-BamHI	CGGGATCCCATAGGACGCTGACGAAGTG	Rice pTCK303-OsEMP5 RNAi vector construction
OsEmp5-SpeI	GGACTAGTCACTGTCATGGGGAACTCTTG	Rice pTCK303-OsEMP5 RNAi vector construction
OsEmp5-SacI	GGAGCTCCATAGGACGCTGACGAAGTG	Rice pTCK303-OsEMP5 RNAi vector construction
Osactin-F	TCTGGCATCACACCTTCTACA	Rice *OsEmp5* RT-PCR analysis
Osactin-R	GGAAACGCTCAGACCAAT	Rice *OsEmp5* RT-PCR analysis
OsEmp5-F	GTCATGGGGAACTCTTGCG	Rice *OsEmp5* RT-PCR analysis
OsEmp5-R	GAGATTTCAGTATCAGTGGTCGC	Rice *OsEmp5* RT-PCR analysis
OsEmp5-F2	CACCGAATCTAAACAGGAACTAACC	Construct $OsEmp5^{N372}$:*GFP* vector
OsEmp5-R2	ATCTTCCATACAGCCAGAGTTGA	Construct $OsEmp5^{N372}$:*GFP* vector
Hpt-F1	GGTGAGTTCAGGCTTTTTCAT	Amplify probe for Southern analysis of rice transgenic lines
Hpt-R1	AATTAATTCGGGGGATCTGG	Amplify probe for Southern analysis of rice transgenic lines
Hpt-F3	GACCTGATGCAGCTCTCGGAG	Arabidopsis transgenic screening
Hpt-R3	TGCTCCATACAAGCCAACCACG	Arabidopsis transgenic screening

Chapter VI.

Conclusion

The goal of this research is to clone the *Emp5* gene, analyze its function and elucidate the potential mechanism by which EMP5 regulates embryogenesis and endosperm development in maize. By using molecular, genetic and biochemistry approaches, we have successfully cloned this *Emp5* gene, and demonstrated that *Emp5* encodes a PPR-DYW protein that is required for the editing of multiple transcripts in mitochondria and the editing events, particularly the C-to-U editing at the *rpl16-458* site, are critical to the mitochondrial functions and hence to seed development in maize. Based on the results of this study, the following conclusions can be drawn:

1. *Emp5* encodes a mitochondrion-targeted PPR-DYW subclass protein. Motif prediction analysis of EMP5 protein sequence by the algorithm TPRpred (http://tprpred.tuebingen.mpg.de/tprpred) revealed that it contains 11 PPR motifs, and the C-terminal region contain an E, an E+, and a DYW domains. Confocal microscopy analysis of the transgenic Arabidopsis with $EMP5^{N469}$-GFP fusion expression and pea chloroplast import assay revealed that EMP5 is localized in mitochondria.

2. EMP5 is required for mitochondrial RNA editing in maize. Loss of the EMP5 function abolishes the C-to-U editing of *rpl16*-458 (100% edited in the wildtype), decreases the editing at nine sites in *nad9, cox3* and *rps12*, and surprisingly increases the editing at five sites of *atp6, nad1, cob* and *rpl16* transcripts.

3. Targeting sequences of EMP5 are not conserved. Alignment of the adjacent sequences near all editing sites (region from -40 to +20) showed that there is no any conserved sequence except most of -1 base is T. The likely possibility is that PPR proteins recognize a specific secondary structure of the transcripts, or with other recognition mode, not the primary sequence.

4. The E+ and DYW motif of EMP5 is not essential for this protein function. *emp5-4 which* lacking the E+ and DYW domain still retains the substrate specificity and editing function, only at reduced efficiency.

5. Abortion of *emp5-1* mutant seed development is caused by defective

mitochondrial RNA editing. Loss of *rpl16*-458 editing results in a proline instead of leucine residue at position 153 of EMP5, which may severely affect the structure and the function of the rpl16 protein. rpl16 protein is essential for protein synthesis in mitochondria, which consequently will be crucial for plant development. Therefore, the failed development in both embryo and endosperm in the *emp5* mutant is likely due to a loss of *rpl16*-458 editing.

6. Rice ortholog OsEMP5 has a conserved function similar to the EMP5 in maize. OsEMP5 is required for the editing of *rpl16*-458 in rice mitochondria, and it is also essential to seed development in rice.

Note:

A part of this dissertation has been published in *The Plant Cell*:

Liu, Y.J., Xiu, Z.H., Meeley R., and Tan, B.C.* (2013) *Empty Pericarp5* encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *The Plant cell* 25,868-883.

References

Alfonzo, J.D., Thiemann, O., and Simpson, L. (1997). The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. Nucleic acids research **25,** 3751-3759.

Arts, G.J., and Benne, R. (1996). Mechanism and evolution of RNA editing in kinetoplastida. Biochimica et biophysica acta **1307,** 39-54.

Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A. (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. Plant molecular biology **42,** 603-613.

Barkan, A., Walker, M., Nolasco, M., and Johnson, D. (1994). A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. The EMBO journal **13,** 3170-3181.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS genetics **8,** e1002910.

Becraft, P.W., Li, K., Dey, N., and Asuncion-Crabb, Y. (2002). The maize *dek1* gene functions in embryonic pattern formation and cell fate specification. Development **129,** 5217-5225.

Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. (2008). The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Molecular and cellular biology **28,** 5337-5347.

Bellaoui, M., Pelletier, G., and Budar, F. (1997). The steady-state level of mRNA from the Ogura cytoplasmic male sterility locus in Brassica cybrids is determined post-transcriptionally by its 3' region. The EMBO journal **16,** 5057-5068.

Bemer, M., Heijmans, K., Airoldi, C., Davies, B., and Angenent, G.C. (2010). An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. Plant physiology **154,** 287-300.

Benne, R. (1994). RNA editing in trypanosomes. European journal of biochemistry / FEBS **221,** 9-23.

Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell **46,** 819-826.

Bensen, R.J., Johal, G.S., Crane, V.C., Tossberg, J.T., Schnable, P.S., Meeley, R.B., and Briggs, S.P. (1995). Cloning and characterization of the maize *An1* gene. The Plant cell **7,** 75-84.

Bentolila, S., Knight, W., and Hanson, M. (2010). Natural variation in Arabidopsis leads to the identification of REME1, a pentatricopeptide repeat-DYW protein controlling the editing of mitochondrial transcripts. Plant physiology **154,** 1966-1982.

Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., and Hanson, M.R. (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proceedings of the National Academy of Sciences of the United States of America **109,** E1453-1461.

Berger, F. (2003). Endosperm: the crossroad of seed development. Current opinion in plant biology **6,** 42-50.

Blum, B., Bakalara, N., and Simpson, L. (1990). A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell **60,** 189-198.

Bollenbach, T.J., Lange, H., Gutierrez, R., Erhardt, M., Stern, D.B., and Gagliardi, D. (2005). RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. Nucleic acids research **33,** 2751-2763.

Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C. (2012). Two Interacting Proteins Are Necessary for the Editing of the *NdhD*-1 Site in Arabidopsis Plastids. The Plant cell 9, 3684-3694.

Brangeon, J., Sabar, M., Gutierres, S., Combettes, B., Bove, J., Gendy, C., Chetrit, P., Des Francs-Small, C.C., Pla, M., Vedel, F., and De Paepe, R. (2000). Defective splicing of the first *nad4* intron is associated with lack of several complex I subunits in the Nicotiana sylvestris NMS1 nuclear mutant. The Plant journal **21,** 269-280.

Chateigner-Boutin, A.L., and Small, I. (2010). Plant RNA editing. RNA biology **7,** 213-219.

Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S., and Small, I. (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. The Plant journal **65,** 532-542.

Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-Garcia, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., Jimenez, L.F., Lurin, C., Small, I., and Leon, P. (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. The Plant journal **56,** 590-602.

Chaudhuri, S., and Maliga, P. (1996). Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site. The EMBO journal **15,** 5958-5964.

Cho, Y., Qiu, Y.L., Kuhlman, P., and Palmer, J.D. (1998). Explosive invasion of plant mitochondria by a group I intron. Proceedings of the National Academy of Sciences of the United States of America **95,** 14244-14249.

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 physiology **136,** 3486-3503.

Cline, K. (1986). Import of Proteins into Chloroplasts - Membrane Integration of a Thylakoid Precursor Protein Reconstituted in Chloroplast Lysates. Journal of Biological Chemistry **261,** 4804-4810.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. The Plant journal **16,** 735-743.

Covello, P.S., and Gray, M.W. (1989). RNA editing in plant mitochondria. Nature **341,** 662-666.

Cushing, D.A., Forsthoefel, N.R., Gestaut, D.R., and Vernon, D.M. (2005). Arabidopsis *emb175* and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. Planta **221,** 424-436.

de Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H., and Small, I. (2008). The pentatricopeptide repeat gene OTP51 with two

LAGLIDADG motifs is required for the cis-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana.* The Plant journal **56,** 157-168.

Ding, Y.H., Liu, N.Y., Tang, Z.S., Liu, J., and Yang, W.C. (2006). Arabidopsis *GLUTAMINE-RICH PROTEIN23* is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. The Plant cell **18,** 815-830.

Dombrowski, S., Brennicke, A., and Binder, S. (1997). 3'-inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. Embo Journal **16,** 5069-5076.

Fisk, D.G., Walker, M.B., and Barkan, A. (1999). Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. The EMBO journal **18,** 2621-2630.

Fu, S., Meeley, R., and Scanlon, M.J. (2002). *Empty pericarp2* encodes a negative regulator of the heat shock response and is required for maize embryogenesis. The Plant cell **14,** 3119-3132.

Fujii, S., and Small, I. (2011). The evolution of RNA editing and pentatricopeptide repeat genes. The New phytologist **191,** 37-47.

Garcia, D., Fitz Gerald, J.N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. The Plant cell **17,** 52-60.

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jurgens, G., and Berger, F. (2003). Arabidopsis haiku mutants reveal new controls of seed size by endosperm. Plant physiology **131,** 1661-1670.

Golden, T.A., Schauer, S.E., Lang, J.D., Pien, S., Mushegian, A.R., Grossniklaus, U., Meinke, D.W., and Ray, A. (2002). SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY, a Dicer homolog, is a maternal effect gene required for embryo development in Arabidopsis. Plant physiology **130,** 808-822.

Grennan, A.K. (2011). To thy proteins be true: RNA editing in plants. Plant physiology **156,** 453-454.

Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., and Grienenberger, J.M. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature **341,** 660-662.

Gutierrez-Marcos, J.F., Dal Pra, M., Giulini, A., Costa, L.M., Gavazzi, G., Cordelier, S., Sellam, O., Tatout, C., Paul, W., Perez, P., Dickinson, H.G., and Consonni, G. (2007). *empty pericarp4* encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. The Plant cell **19,** 196-210.

Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T., and Small, I. (2009). A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. The Plant cell **21,** 3686-3699.

Hammani, K., des Francs-Small, C.C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A., and Small, I. (2011). The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in Arabidopsis mitochondria. The Journal of biological chemistry **286,** 21361-21371.

Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A. (1989). RNA editing in plant

mitochondria. Science **246,** 1632-1634.

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

Hueros, G., Royo, J., Maitz, M., Salamini, F., and Thompson, R.D. (1999). Evidence for factors regulating transfer cell-specific expression in maize endosperm. Plant molecular biology **41,** 403-414.

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

Jonietz, C., Forner, J., Hildebrandt, T., and Binder, S. (2011). RNA PROCESSINGFACTOR 3 is crucial for the accumulation of mature *ccmC* transcripts in mitochondria of *Arabidopsis thaliana* accession Columbia. Plant physiology **3**, 1430-1439

Kable, M.L., Heidmann, S., and Stuart, K.D. (1997). RNA editing: Getting U into RNA. Trends in biochemical sciences **22,** 162-166.

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*). Current genetics **22,** 463-470.

Karcher, D., Kahlau, S., and Bock, R. (2008). Faithful editing of a tomato-specific mRNA editing site in transgenic tobacco chloroplasts. RNA **14,** 217-224.

Kazama, T., Nakamura, T., Watanabe, M., Sugita, M., and Toriyama, K. (2008). Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. The Plant journal **55,** 619-628.

Khrouchtchova, A., Monde, R.A., and Barkan, A. (2012). A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. RNA **6**, 1197-1209

Kim, S.R., Yang, J.I., Moon, S., Ryu, C.H., An, K., Kim, K.M., Yim, J., and An, G. (2009). Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. The Plant journal **59,** 738-749.

Knoop, V. (2011). When you can't trust the DNA: RNA editing changes transcript sequences. Cellular and molecular life sciences : CMLS **68,** 567-586.

Kobayashi, Y., Matsuo, M., Sakamoto, K., Wakasugi, T., Yamada, K., and Obokata, J. (2008). Two RNA editing sites with cis-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. Nucleic acids research **36,** 311-318.

Kohler, C., and Makarevich, G. (2006). Epigenetic mechanisms governing seed development in plants. EMBO reports **7,** 1223-1227.

Kotera, E., Tasaka, M., and Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature **433,** 326-330.

Kuhn, J., Tengler, U., and Binder, S. (2001). Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. Molecular and cellular biology **21,** 731-742.

Lambowitz, A.M., and Zimmerly, S. (2004). Mobile group II introns. Annual review of genetics **38,** 1-35.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K.,

Harada, J.J., and Goldberg, R.B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America **107,** 8063-8070.

Lid, S.E., Gruis, D., Jung, R., Lorentzen, J.A., Ananiev, E., Chamberlin, M., Niu, X., Meeley, R., Nichols, S., and Olsen, O.A. (2002). The *defective kernel 1* (*dek1*) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. Proceedings of the National Academy of Sciences of the United States of America **99,** 5460-5465.

Lu, Y., Li, C., Wang, H., Chen, H., Berg, H., and Xia, Y. (2011). AtPPR2, an Arabidopsis pentatricopeptide repeat protein, binds to plastid 23S rRNA and plays an important role in the first mitotic division during gametogenesis and in cell proliferation during embryogenesis. The Plant journal **67,** 13-25.

Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., and Chaudhury, A. (2005). *MINISEED3 (MINI3)*, a WRKY family gene, and *HAIKU2 (IKU2)*, a leucine-rich repeat (LRR) *KINASE* gene, are regulators of seed size in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **102,** 17531-17536.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. The Plant cell **16,** 2089-2103.

Ma, Z.R., and Dooner, H.K. (2004). A mutation in the nuclear-encoded plastid ribosomal protein S9 leads to early embryo lethality in maize. The Plant Journal **37,** 92-103.

Magnard, J.L., Heckel, T., Massonneau, A., Wisniewski, J.P., Cordelier, S., Lassagne, H., Perez, P., Dumas, C., and Rogowsky, P.M. (2004). Morphogenesis of maize embryos requires *ZmPRPL35-1* encoding a plastid ribosomal protein. Plant physiology **134,** 649-663.

Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R. (2012). An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of *rps3* mRNA in Maize Mitochondria. The Plant cell **24,** 3087-3105.

Martin, J.R., Harwood, J.H., McCaffery, M., Fernandez, D.E., and Cline, K. (2009). Localization and integration of thylakoid protein translocase subunit cpTatC. The Plant journal **58,** 831-842.

McCarty, D.R., Settles, A.M., Suzuki, M., Tan, B.C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., Messing, J., Koch, K.E., and Hannah, L.C. (2005). Steady-state transposon mutagenesis in inbred maize. The Plant journal **44,** 52-61.

Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003). HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. The Plant cell **15,** 1480-1495.

Mohr, G., and Lambowitz, A.M. (2003). Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. Nucleic acids research **31,** 647-652.

Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G. (2003). RNA-binding properties of

HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. European journal of biochemistry / FEBS **270,** 4070-4081.

Neuffer, M.G., and Sheridan, W.F. (1980). Defective kernel mutants of maize. I. Genetic and lethality studies. Genetics **95,** 929-944.

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. Developmental genetics **19,** 277-286.

Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006). A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. The Journal of biological chemistry **281,** 37661-37667.

Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. Proceedings of the National Academy of Sciences of the United States of America **104,** 8178-8183.

Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. The Plant cell **21,** 146-156.

Okuda, K., Hammani, K., Tanz, S.K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2010). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid ndhB and ndhG transcripts. The Plant journal **61,** 339-349.

Olsen, O.A. (2001). ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. Annual review of plant physiology and plant molecular biology **52,** 233-267.

Ostheimer, G.J., Williams-Carrier, R., Belcher, S., Osborne, E., Gierke, J., and Barkan, A. (2003). Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. The EMBO journal **22,** 3919-3929.

Pate, J.S., and Gunning, B.E.S. (1972). Transfer Cells. Ann Rev Plant Physio **23,** 173-&.

Perrin, R., Lange, H., Grienenberger, J.M., and Gagliardi, D. (2004a). AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in *Arabidopsis thaliana* mitochondria. Nucleic acids research **32,** 5174-5182.

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

Pfalz, J., Bayraktar, O.A., Prikryl, J., and Barkan, A. (2009). Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. The EMBO journal **28,** 2042-2052.

Pien, S., and Grossniklaus, U. (2007). Polycomb group and trithorax group proteins in Arabidopsis. Biochimica et biophysica acta **1769,** 375-382.

Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. Proceedings of the National Academy of Sciences of the United States of America **108,** 415-420.

Saalaoui, E., Litvak, S., and Araya, A. (1990). The Apocytochrome-B from an Alloplasmic Line of Wheat (Triticum-Aestivum, Cytoplasm-T-Timopheevi) Exists in 2 Differently Expressed Forms. Plant Science **66,** 237-246.

Saha, D., Prasad, A.M., and Srinivasan, R. (2007). Pentatricopeptide repeat proteins and their emerging roles in plants. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale **45,** 521-534.

Sakamoto, W., Kondo, H., Murata, M., and Motoyoshi, F. (1996). Altered mitochondrial gene expression in a maternal distorted leaf mutant of Arabidopsis induced by chloroplast mutator. The Plant cell **8,** 1377-1390.

Salone, V., Rudinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. FEBS letters **581,** 4132-4138.

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in plant science **13,** 663-670.

Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2005). RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. The Plant cell **17,** 2791-2804.

Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas, A., and Barkan, A. (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast *rps12* pre-mRNA. The Plant cell **18,** 2650-2663.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science **326,** 1112-1115.

Schulz, P., and Jensen, W.A. (1971). Capsella embryogenesis: the chalazal proliferating tissue. Journal of cell science **8,** 201-227.

Shen, Y., Li, C., McCarty, D.R., Meeley, R., and Tan, B.C. (2013). *Embryo defective12* encodes the plastid initiation factor 3 and is essential to embryogenesis in maize. The Plant journal. DOI: 10.1111/tpj.12161

Sheridan, W.F., and Neuffer, M.G. (1980). Defective Kernel Mutants of Maize II. Morphological and Embryo Culture Studies. Genetics **95,** 945-960.

Sheridan, W.F., and Clark, J.K. (1993). Mutational Analysis of Morphogenesis of the Maize Embryo. The Plant Journal **3,** 347-358.

Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. Cellular and molecular life sciences : CMLS **63,** 698-708.

Sieber, P., Gheyselinck, J., Gross-Hardt, R., Laux, T., Grossniklaus, U., and Schneitz, K. (2004). Pattern formation during early ovule development in *Arabidopsis thaliana*. Developmental biology **273,** 321-334.

Small, I.D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends in biochemical sciences **25,** 46-47.

Sosso, D., Mbelo, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., Dauzat, M., Heurtevin, L., Guyon, V., Takenaka, M., and Rogowsky, P.M. (2012). PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. The Plant cell **24,**

676-691.

Sung, T.Y., Tseng, C.C., and Hsieh, M.H. (2010). The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria. The Plant journal **3**, 499-511

Takenaka, M. (2010). MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the nad7 transcript in mitochondria of Arabidopsis. Plant physiology **152,** 939-947.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Hartel, B., and Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proceedings of the National Academy of Sciences of the United States of America **109,** 5104-5109.

Torres-Ruiz, R.A., Lohner, A., and Jurgens, G. (1996). The *GURKE* gene is required for normal organization of the apical region in the Arabidopsis embryo. The Plant journal **10,** 1005-1016.

Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., Patton, D., and Meinke, D. (2003). The Arabidopsis SeedGenes Project. Nucleic acids research **31,** 90-93.

Tzafrir, I., Pena-Muralla, R., Dickerman, A., Berg, M., Rogers, R., Hutchens, S., Sweeney, T.C., McElver, J., Aux, G., Patton, D., and Meinke, D. (2004). Identification of genes required for embryo development in Arabidopsis. Plant physiology **135,** 1206-1220.

van Herpen, T.W.J.M., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J., and Beekwilder, J. (2010). Nicotiana benthamiana as a Production Platform for Artemisinin Precursors. PloS one **5**, e14222

Verbitskiy, D., Zehrmann, A., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2010). The PPR protein encoded by the *LOVASTATIN INSENSITIVE 1* gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. The Plant journal **61,** 446-455.

Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., Hartel, B., and Takenaka, M. (2012). The E-Class PPR Protein MEF3 of *Arabidopsis thaliana* Can Also Function in Mitochondrial RNA Editing With an Additional DYW Domain. Plant and Cell Physiology **53,** 358-367.

Wahleithner, J.A., MacFarlane, J.L., and Wolstenholme, D.R. (1990). A sequence encoding a maturase-related protein in a group II intron of a plant mitochondrial *nad1* gene. Proceedings of the National Academy of Sciences of the United States of America **87,** 548-552.

Wang, C.X., Barry, J.K., Min, Z., Tordsen, G., Rao, A.G., and Olsen, O.A. (2003). The calpain domain of the maize DEK1 protein contains the conserved catalytic triad and functions as a cysteine proteinase. Journal of Biological Chemistry **278,** 34467-34474.

Williams, P.M., and Barkan, A. (2003). A chloroplast-localized PPR protein required for plastid ribosome accumulation. The Plant journal **36,** 675-686.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PloS one **8,** e57286.

Yuan, H., and Liu, D. (2012). Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in

Arabidopsis. The Plant journal **70,** 432-444.

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

Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., and Takenaka, M. (2011). PPR proteins network as site-specific RNA editing factors in plant organelles. RNA biology **8,** 67-70.

Zhou, W.B., Cheng, Y.X., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I., and Huang, J.R. (2009). The Arabidopsis gene *YS1* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. The Plant journal **58,** 82-96.

Zoschke, R., Kroeger, T., Belcher, S., Schottler, M.A., Barkan, A., and Schmitz-Linneweber, C. (2012). The Pentatricopeptide Repeat-SMR Protein ATP4 promotes translation of the chloroplast atpB/E mRNA. The Plant journal 4,547-558