TRANS-ACTING FACTORS ESSENTIAL FOR PLANT ORGANELLE RNA EDITING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tao Sun

August 2014

TRANS-ACTING FACTORS ESSENTIAL FOR PLANT ORGANELLE RNA EDITING

TAO SUN, Ph.D.

Cornell University 2014

In higher plants, RNA editing is a C-to-U conversion that corrects chloroplast and mitochondrial transcripts that are otherwise defective. Although plant RNA editing has been known for over two decades, the molecular mechanism is poorly understood. Until recently, all the known *trans*-acting factors were members of the Pentatricopeptide Repeat (PPR) protein family, which serve as recognition factors via specific interaction with *cis*-elements upstream of the C targets. An additional editing factor, RIP1, was identified by a proteomics study. RIP1 is a dual-targeted protein that selectively interacts with PPR editing factors and affects 14 editing events in chloroplasts and over 400 editing events in mitochondria. RIP1 belongs to a small protein family, 5 members of which were later shown to be major editing factors. Homology searching with the RIP protein led to the discovery of ORRM1, a hybrid protein which possesses a RIP-like domain at its N terminus and an RNA Recognition Motif (RRM) domain at its C terminus. Loss of ORRM1 results in editing defects in multiple plastid sites. A transient complementation assay indicates that the editing activity of ORRM1 is carried by the RRM, which places it in a different family than RIP proteins. Additional members of the ORRM1 family might be involved in plant RNA editing. A plastid-targeted protein immunoprecipitated with a functional epitope-tagged ORRM1. Loss of this protein leads to editing defects at many plastid sites,

most of which are also controlled by ORRM1. Homology searches with this plastid protein identified several related proteins which are all organelle-targeted. The function of this new family still needs further investigation. So far, four types of *trans*-acting factors have been identified for plant organelle RNA editing, which reveals an unexpected complexity of the editing machinery.

BIOGRAPHICAL SKETCH

Tao Sun was born in Laizhou, a small coast city in northern China in 1986. His father was working in a marine resources company and his mother was a history teacher in a middle school. Like most Chinese kids born in 1980s he is the only child of the family.

Tao spent a lot of time during his childhood with his grandparents who were living in the rural area during which he developed strong interests in nature. As he grew older, science became his favorite subject, especially biology. He was fascinated by the beauty of life and decided to become a biologist.

In 2005, he was admitted to Zhejiang University, one of the top universities in China. He was selected to enroll in Cho Kochen Honors College, an elite program where students were trained in multidisciplines including mathematics, physics, organic chemistry, phycology and geography. In his junior year, he picked biological sciences as his major and started working as an undergraduate researcher in the school of medicine advised by Professor Tianhua Zhou on mammalian cell cycle regulation.

In 2008, he joined an overseas summer internship program and worked on agrobacteria and plants interactions in Stanton Gelvin's lab at Purdue University. In 2009 after he obtained his bachelor's degree from Zhejiang University, he came back to the United States to pursue his Ph.D in Plant Biology at Cornell. Fascinated by plant RNA editing, he decided to join Maureen Hanson's lab to decipher the unknown mechanism that allows plants to precisely edit their transcripts. He enjoyed doing research and the exciting moments of great discoveries. His work identified and

characterized two novel factors, which greatly expanded our knowledge of plant editing mechanism. In 2013 he received Hsien Wu and Daisy Yen Wu Scholarship /The Liu Memorial Award for outstanding student of Chinese descent and the Barbara McClintock Award for outstanding students in plant science.

In his spare time, he likes jogging and swimming as well as kayaking. He is also a big fan of Sci-Fi and comedy. He likes hanging out with his friends and has been grateful for their company all through his Ph.D studies at Cornell.

After graduation Tao is going to Stanford medical school for postdoctoral training to expand his research interest into human RNA editing and related diseases. In the future, he sees himself as a scientist working towards better understanding of human genetics and pushing forward the frontiers of healthcare.

To my parents and grandparents who have been unconditionally

loving and supporting me

ACKNOWLEDGMENTS

I first would like to thank my advisor Professor Maureen Hanson for having me in her lab and giving me such a wonderful project. She has been very accessible, helpful and patient as a student advisor despite her routinely busy schedules. She is a dedicated and insightful scientist who sets a great example for me.

I also would like to acknowledge Stephane Bentolila and Wade Heller, two Hanson lab members who I have been working with. Wade is one of the smartest graduate students I have ever seen. He never hid from problems but rather look for solutions by troubleshooting. His optimistic attitude greatly inspired me. His discovery of RIP1 is the cornerstone of the project that we have been working on ever since. Stephane has been a co-advisor to me. He is very nice and very dedicated. It has been great fun to work with him and share exciting moments of discovery.

My thanks also go to my committee members--Klaas van Wijk and Sandy Lazarowitz who have been very supportive of my research. Our collaboration with the van Wijk lab on proteomics has been fruitful and a great training for me. The protoplasts transfection protocol kindly provided by the Lazarowitz lab turned out to be one of the most powerful tools I routinely used for my research. I want to specially thank Mike Scanlon for serving on my committee for my dissertation.

I want to thank the whole Hanson lab, especially Lin Lin for helping me with cloning and protein purification, Myat Lin for advising on scientific and career questions, Xiaowen Shi for helping me finish up the project, Wenfang Li for handling my mutants. I thank Kenji Nishimura from the van Wijk lab for sharing his

biochemistry tricks, Dr. Jian Hua and Dr. Adriene Roeder for sharing valuable plasmids and plant materials, and Dr Ailong Ke for advising me on enzymatic analysis.

I want to thank the Wu family and Liu family for establishing Hsien Wu and Daisy Yen Wu Scholarship /The Liu Memorial Award to recognize outstanding students of Chinese descent and Dr. Robert Rabson for establishing the Barbara McClintock Award for outstanding plant science major students. I am very honored to receive both awards and they will be my great encouragement for my future career.

Last but not least, I want to thank my parents, grandparents and my dear friends for the support and the joy they brought to me over the years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3FS 3Flag-strepII

EMSA Electrophoresis Mobility Shift Assay

MORF Multiple Organellar RNA editing Factor

ORRM1 Organelle RNA Recognition Motif protein 1

PPE Poisoned Primer Extension

PPR Pentotricopeptide repeat

RanBP2 Ran Binding Protein 2

RIP RNA editing factor Interacting Protein

RRM RNA recognition motif

VAR3 Variegated 3

Y2H Yeast 2 Hybrid

NDH NADH Dehydrogenase

Chapter 1

Introduction

First discovered in *Trypanosome brucei* (1), RNA editing is known to be a common RNA processing step in various species from viruses to plants and animals (2–6). Two major forms of RNA editing have been observed: insertion/deletion editing and nucleotide conversion editing. Trypanosome mitochondria undergo insertion/deletion editing, in which non-genomically encoded uridines are either inserted into or deleted from the transcripts (1). RNA editing is essential for trypanosome mitochondrial genome expression since massive editing creates functional coding sequences that otherwise would have been defective. Guide RNAs are required to specify editing patterns and the reaction is catalyzed by the RNA editing core complex (7). Nucleotide conversion editing, on the other hand, does not involve alteration of the RNA backbones, but rather nucleotide modifications that are usually catalyzed by a deamination activity (6, 8). Two well-studied examples are C-to-U editing of *Apolipoprotein-B* (*apoB*) and A-to-I editing in humans. Unedited *apoB* is translated into protein APOB100 in liver while editing of *apoB* creates a premature UAA stop codon, giving rise to a smaller protein isoform APOB48 in small intestine. Human C-to-U editing requires the deaminase Apolipoprotein-B-mRNA-editing-enzyme-1 (APOBEC1) and a cofactor APOBEC1-Complementation-Factor (ACF) which binds to the 3' mooring sequence of the C target (9, 10). Recently, another factor called RNA-Binding-Motif-protein-47 (RBM47) was shown to be required for this process and it can substitute for ACF in the editing complex (11). A-to-I editing is

catalyzed by Adenosine Deaminases Acting on RNA (ADARs). Inosines created by editing are recognized as Guanine by the translation machinery. A-to-I editing in the nervous system affects many mRNAs that encode neurologically important membrane channels and receptors (12–14). An explosion in the number of editing sites has been reported recently, implicating a more general role of A-to-I editing in regulating gene expression (15).

RNA editing in higher plants is a C-to-U type conversion which only occurs in plastids and mitochondria (16). Typical land plant plastids have around 30 edited Cs while mitochondria have over 500. Due to the sensitive next generation sequencing technique, 37 plastid and 619 mitochondrial editing sites have been identified in the model plant *Arabidopsis thaliana* (17). Although the editing enzyme remains elusive, plant RNA editing is believed to be deamination or transamination (18).

With a majority of editing events resulting in non-synonymous substitutions, RNA editing restores a codon for a conserved amino acid or creates a start codon or a stop codon. However, unlike C-to-U editing in humans, no evidence supports the hypothesis that plant RNA editing contributes to diversification of functional organelle proteins. So it is believed that the purpose of plant RNA editing is to correct defective transcripts at the RNA level.

Loss of editing can be deleterious for plants (19, 20), while off-target editing could theoretically introduce undesirable mutations. The proper C target is selected by a sophisticated combination of *cis*-elements and *trans*-factors. Editing *cis*-elements are short RNA sequences adjacent to the C target. Both *in vivo* and *in vitro* studies have mapped editing *cis*-elements usually to between 20 nucleotides upstream and a few nucleotides downstream of the target C (21–24). All known *trans*-factors are nuclear-

encoded proteins (25, 26). The first identified editing factor, Chlororespiratory Reduction 4 (CRR4), was found through a genetic screen of mutagenized Arabidopsis based on an NADH dehydrogenase phenotype. CRR4 encodes a Pentatricopeptide Repeat (PPR) protein, loss of which disrupts editing at one particular site of the chloroplast *ndhD* transcript (27). This breakthrough was followed by numerous reports of other PPR proteins playing a role in RNA editing. So far, approximately 20 plastid PPR proteins and 10 mitochondrial PPR proteins have been established as editing factors.

Pentatricopeptide Repeat proteins

PPR proteins carry PPR motifs, which are degenerate ~35 amino acid sequence present as repeats. The Arabidopsis genome encodes 458 PPR proteins, making it one of the largest protein families in plants (28, 29). PPR proteins act as site recognition factors for editing via the interaction between the PPR tract and the editing *cis*-element (30, 31). Depending on the similarity across *cis*-elements, one PPR protein can recognize one to several *cis*-elements, thus specifying 1 to several editing events. Loss of a PPR editing factor can result in loss of editing for these particular sites. Recently, both a code for PPR-RNA recognition and the crystal structure of a PPR protein, PPR10, have been reported (32–35). They agreed with each other in that the 6th amino acid in one PPR motif and the 1st amino acid in the next motif recognize one specific nucleotide on the RNA *cis*-element. In addition, the binding region is always four nucleotides upstream of the target C, implicating the presence of a molecular ruler for editing. The identification of the code is expected to bring about a revolutionary change to the field. With this combinatorial code, one can predict *in silico* the RNA

binding partner of a given PPR protein instead of performing laborious genetic screening.

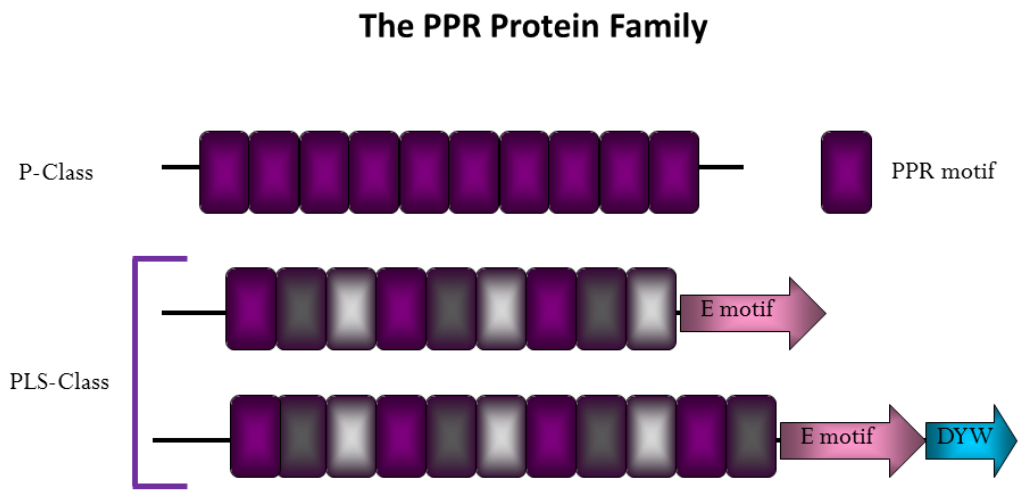


Figure 1.1. Schematic diagram of PPR protein family. Blocks represent PPR motifs. Color variation indicates various lengths of PPR motif in P-L-S class. E and DYW domains are indicated by arrows (28).

Based on the length of the PPR motif and the presence of C-terminal domains, PPR proteins are divided into two subclasses-P class and P-L-S class (Figure 1.1). While P class proteins have canonical 35aa repeats, P-L-S proteins have variable length motifs and always contain an additional C terminal domain. All identified editing PPR proteins belong to the P-L-S class, which possess the E domain and sometimes a DYW domain.

The function of C-terminal domains, which exist in all editing PPR proteins, is largely unknown. The E domain has been shown to be essential to editing, disruption of which affects editing (36). Recently, the essential region for this domain has been further mapped to a 15-amino acid region designated the PG region (37). Although both organelles employ PPR proteins as editing factors, the E domains across organelles are not well conserved and in some cases are not exchangable (38).

The DYW domain shares sequence similarity with canonical cytidine deaminases, thus is considered as the best candidate for the editing enzyme. Indeed, expansion of the DYW domain in the PPR family is tightly correlated with emergence of new RNA editing sites (39). Given that the PPR tract binds to RNA, a simple "one editing event—one PPR protein" model was postulated, analogous to the human C-to-U editosome, which contains an RNA binding factor and a catalytic factor. However this model was challenged by several observations. The DYW domain has been shown to have endoribonuclease activity *in vitro* instead of deaminase activity (40). Not all editing PPR proteins have a DYW domain. More surprisingly, deletion of the DYW domain from CRR22 and CRR28 does not affect their function in RNA editing (41). Apparently, a more sophisticated machinery is involved in plant RNA editing. One possibility is that the DYW domain can be supplied *in trans* by another PPR protein. Discovery of a truncated PPR protein named DYW1 strengthened this theory (42). Both DYW1 and a PPR recognition factor CRR4 is required for *ndhD-2* editing. While CRR4 contains an E domain, DYW1 lacks PPR motifs and possesses only a partial E domain and a DYW domain. CRR4 interacts with DYW1 *in planta* and a CRR4-DYW1 hybrid protein complements a *crr4/dyw1* double mutant. This discovery demonstrates that DYW can be provided *in trans* if it is missing from the *cis*-element binding PPR protein. It also raises the question whether this is a common scenario for organelle editing. AtECB2 and RARE1 have both been reported to affect *accD-794* editing (43, 44). However, according to the PPR-RNA recognition code, only RARE1 is the *bona fide* recognition factor for this particular site (45). It will be of interest to examine whetherAtECB2 is actually a DYW donor instead. However, the donor is expected to be less specific, given that very few PPR proteins have been shown to

control the same editing events. In fact, with the first PPR protein providing specificity, the second PPR protein could potentially be any DYW- containing partner. Another piece of evidence supporting the DYW motif as the deaminase comes from a biochemical discovery. Recombinant DYW1 and ELI1 proteins bind two zinc ions (37), as is expected for cytidine deaminases (46). One of the zinc ions is sitting in the active center that is shared by most cytidine deaminases, which is coordinated in a tetrahedral configuration by a histidine or cysteine residue ((H/C)XE) and two cysteine residues (CXXC). The binding site of the other zinc ion is still unknown. Incorporation of zinc ions into the DYW domain implies that the DYW motif may be the catalytic factor, although more direct evidence is needed.

RIP/MORF protein family

The complexity of the plant editing mechanism was further demonstrated by several biochemical experiments. RARE1 (Required for *AccD* RNA Editing 1) is a PPR protein that specifies editing of chloroplast *accD-794* (43). While RARE1 from which the transit sequence has been removed is around 72kD, a size exclusion chromatography assay showed that it is in a complex 200kD to 400kD in size (47). Similarly, when a maize chloroplast extract was fractionated by size exclusion chromatography, fractions corresponding to 200kD~400kD had editing activity for *rpoB-467* substrate in an *in vitro* editing assay (Charles Bullerwell, unpublished). These observations provide evidence that plant organelle RNA editing is carried out by a protein complex- "editosome" rather than a single PPR protein.

The first discovered non-PPR components of the editosome are RIP/MORF proteins (47, 48). RNA editing factor Interacting Protein 1 (RIP1) was found by a

proteomics study of a RARE1 co-immunoprecipitate, while Multiple Organeller RNA editing Factor1 (MORF1) was found through an EMS mutant screening for mitochondrial editing defects. RIP1 and MORF1 belong to the same family, which has 10 family members in Arabidopsis. All family members are predicted to be organelle targeted, except for RIP10 which is likely to be a pseudogene.

The first member of the *RIP* family was identified as *DAG* (Differentiation and Greening) from a transposon mutant in *Antirrhinum majus* (49). The unstable mutagenized plants have white leaves with revertant green sectors. Chloroplasts in the white sectors are defective. *DAG* encodes a plastid targeted protein which is required for the chloroplast *rpoB* expression. Similarly, *RIP2* was characterized as *DAL* (*DAG*-like) in other reports as well (50, 51). Mutation of *RIP2* results in an albino phenotype and a role in rRNA processing was implicated (48).

Strikingly, loss of one single RIP/MORF protein leads to defective editing at a massive scale, different from any PPR editing factors, which only control one to several editing events. Analysis of a *RIP1* knock-down mutant by next generation sequencing of RT-PCR products showed editing defects at over 400 mitochondria sites and 11 chloroplast sites, among which over 200 mitochondrial sites have a major loss of editing. In chloroplasts, almost every editing event is affected by mutation of *RIP2* or *RIP9*, indicating both proteins are very important components of the plastid editosome. Another remarkable difference between PPR factors and RIP/MORF factors is that editing at particular sites is generally completely lost in PPR mutants, while editing defects caused by RIP/MORF mutation vary from complete disruption to mild reduction. To examine the editing of hundreds of sites, a Strand and Transcript Specific PCR seq (STS-Seq) method that takes advantage of next generation

sequencing technology was invented and shown to be powerful for analysis of large numbers of editing sites (17). In summary, RIP1 is the major mitochondrial editing factor while RIP3 and RIP8 are moderately important factors for mitochondria. In plastids, RIP2 and RIP9 are the major players while RIP1 plays a minor role. Other members of the RIP/MORF family only mildly affect a very small portion of editing sites, suggesting they might not have a direct role in RNA editing.

Table 1.1. RIP/MORF family editing factors. Naming system of RIP and MORF is listed for comparison. Subcellular localization is predicted by Target P. Protein interaction data is from CCDB(http://interactome.dfci.harvard.edu)

RIP/MORF Nomenclature	Accession number	Predicted Localization	Mutant phenotype	Role in editing	Note
RIP1/MORF8	At3g15000	plastids+ mitochondria	Knock-down dwarf; Null mutant probably lethal	Major mitochondria editing factor; minor plastid editing factor	Cobalt binding protein
RIP2/MORF2	At2g33430	plastids	Yellow seedling	Major plastid editing factor	First known as DAL, in rRNA processing
RIP3/MORF3	At3g06790	mitochondria	Slightly delayed development	Moderate mitochondria editing factor	
RIP4/MORF4	At5g44780	mitochondria	No macroscopic defect	Minor effect on mitochondrial editing, may be indirect	
RIP5/MORF5	At1g32580	mitochondria	No macroscopic defect	Minor effect on mitochondrial editing, may be indirect	
RIP6/MORF6	At2g35240	mitochondria	No macroscopic defect	Minor effect on mitochondrial editing, may be indirect	
RIP7/MORF7	At1g72530	mitochondria	No macroscopic defect	Minor effect on mitochondrial editing, may be indirect	
RIP8/MORF1	At4g20020	mitochondria	Null mutant lethal	Moderate mitochondria editing factor	
RIP9/MORF9	At1g11430	plastids	Green cotyledon white true leaves Veriagation	Major plastid editing factor	Interacts with RIP6
RIP10/MORF10	At1g53260		N/A	no	Possibly pesudogene

RIP/MORF proteins can promiscuously interact with PPR editing factors, which may reflect how these proteins interact in editosomes. The PPR motifs of

RARE1, but not the E domain or DYW domain, mediates the interaction with RIP1. In fact, only a few PPR motifs from the N terminus are sufficient for this interaction (Chapter 2). However, it is not known yet whether this holds true for all PPR-RIP interactions. Another intriguing feature of RIP proteins is that they can form homodimers or heterodimers, which might explain some genetic interactions between *RIP* genes. For instance, both RIP2 and RIP9 can form homodimers in yeast and interact with each other as well. Almost every plastid editing site is affected when RIP2 or RIP9 is mutated. At *ndhD-2*, mutation of either RIP2 or RIP9 leads to a total disruption of editing. It is possible a RIP2/RIP9 heterodimer is required for this particular editing event. On the other hand, *petL-5* editing is lost in the *rip9* mutant while in the *rip2* mutant, 65% editing of *petL-5* remains. One explanation is that RIP9 is an essential component for the *petL-5* editosome and both RIP2/RIP9 heterodimer and RIP9/RIP9 homodimer are functionally competent, but the heterodimer is preferred. However, why and how RIP proteins are distributed across different editosomes is not yet understood.

Although no known domain is found within RIP/MORF proteins, a motif scanning prediction showed that they contain motifs of unknown function. In addition to the conserved N termini, some RIP proteins have extended C termini. RIP1 has a 150 amino acid proline-rich region at the C terminus, while RIP8 has a glycine- and proline-rich region. The unique C terminus is not required for RIP1 interaction with RARE1 (47). So far it is not known whether the unique C termini in RIP proteins are necessary for editing.

ORRM family

Organelle RNA Recognition Motif protein 1 (ORRM1), an outlier of the RIP family was found through homology search. Compared to other RIP domain-containing proteins, ORRM1 is strikingly different. It possesses two duplicated but truncated RIP domains on the N terminus, and an RNA Recognition Motif (RRM) near its C terminus. Targeted to plastids, ORRM1 is an essential factor for most plastid editing events (52). An Arabidopsis null mutant of *ORRM1* has major loss of editing at 12 sites and various defects of editing at 9 other sites. Surprisingly, this particular mutant does not exhibit a macroscopic phenotype under greenhouse growth conditions. This is largely explained by the fact that most affected editing sites reside within the transcripts encoding subunits of the NADH Dehydrogenase (NDH) complex. A plant without the NDH complex has normal appearance under regular conditions. In an orthologous mutant in maize, loss of *ORRM1* leads to a more severe phenotype--pale green leaves and seedling lethality (52). Different Cs are selected for editing in maize chloroplasts than in Arabidopsis chloroplasts. Although the function of *ORRM1* is highly conserved in two species, maize ORRM1 affects editing of some sites which are only editable in maize. The severe defect in maize *orrm1* is likely caused by defective editing of maize C targets that are important for the function of the affected maize proteins.

Although first identified through the RIP-RIP domain, ORRM1 was shown to carry its editing activity at its RRM motif. In a transient protoplast complementation assay, the RRM motif alone, rather than the RIP-RIP portion, was able to rescue the mutant's editing defects (52). This unexpected result placed ORRM1 in a different category of editing factor family from RIP proteins. In fact, involvement of RRMs in

RNA editing is not unprecedented. In humans, the APOBEC1 Complementation Factor (ACF) contains three RNA Recognition motifs that bind the *apoB* mooring sequence during the C-to-U editing (9, 10). In plants, RRM-containing protein CP31 has also been implicated in plastid RNA editing. Immunodepletion of CP31 inhibited *psbL* and *ndhB in vitro* editing (53), and null mutants also showed plastid editing defects at multiple sites (54). However, a direct role of CP31 in editing is questioned. First, editing events are site-specific, but defects in *cp31* are transcript-dependent. Second, none of the sites completely loses editing in *cp31* (54). Third, plastid transcript abundance is greatly reduced in *cp31*, and CP31 was later shown to be an RNA stability factor (55). Thus the editing defect seen in *cp31* is likely to be a secondary effect due to transcript instability. Conversely, no obvious change of transcript abundance was observed in the *orrm1* mutant. In addition, the editing sites ORRM1 controls are not transcript-specific. For example, editing of *ndhB-467* is disrupted in *orrm1* but *ndhB-1481* on the same transcript is not affected. On the contrary, editing sites recognized by the same PPR factor are similarly affected in *orrm1*. For instance, two sites recognized by CRR28, *ndhB-467* and *ndhD-878* are both affected by *orrm1*, while none of the three sites recognized by OTP84 (*ndhF-290*, *ndhB-1481*, *psbZ-50*) is affected in *orrm1*. Thus ORRM1 is believed to play a direct role in plastid editing (52).

Through the RIP-RIP domain, ORRM1 selectively interacts with PPR editing factors. The RRM increases binding affinity with some PPR proteins but RRM alone is not sufficient for the interaction (52). It is not yet known how the RRM motif without the RIP-RIP region is fully competent for complementation of editing in an *orrm1* mutant. The RRM of ORRM1 may be interacting with some other critical

components of the editosome in order to function.

RRMs are known to be able to bind to RNAs. ORRM1 binds near some, if not all, *cis*-elements *in vitro* (52). Whether RRM binding is relevant *in vivo* to RNA editing is not known. ORRM1 affects editing at 24 plastid sites. It is unlikely that ORRM1 has specific affinity for 24 *cis*-elements. It is possible that the binding partner of ORRM1 is protein rather than RNAs.

The Arabidopsis genome encodes 196 RRM containing proteins, while ORRM1 belongs to a distinct clade in which many are glycine-rich proteins and small RNA binding proteins. ORRM1 is the only one with a RIP-RIP domain (52). It is unknown yet if ORRM1 acquired this domain through recombination or other ORRM-like proteins lost their RIP parts during selection. Nevertheless, the importance of the RRM domain in editing raises the question whether other RRM proteins are also important components of the editosome. Screening of ORRM family mutants should provide more information on their possible involvement in RNA editing.

Accessory components-PPO1 and OCP3

Recently, a protoporphyrinogen IX oxidase 1 (PPO1) in the tetrapyrrole biosynthesis pathway was shown to have a surprising role in plastid RNA editing (56). In the *ppo1* null mutants, defective editing was seen at 18 plastid sites, most of which encode subunits of the NDH complex. PPO1 interacts with plastid RIP/MORF proteins as well as two PPR editing factors, CRR28 and OTP82. What is interesting is that the function of PPO1 in editing is independent of its function as an oxidase, since PPO1 without the enzyme region is able to complement editing defects in the mutant. This finding implicates PPO1 as playing a direct role in plastid editing. Notably,

except for *ndhD-2* site, none of the affected sites completely loses editing. This indicates that PPO1 is not an essential editing factor, but rather an accessory factor that facilitates editing of *ndh* transcripts, perhaps thus regulating NDH complex function.

Another regulatory factor for NDH activity, OCP3 (Overexpressor of Cationic Peroxidase 3), might also be involved in plastid RNA editing (57). In the *ocp3* mutant and silenced plants, multiple editing sites on the *ndhB* transcript showed reduced extent of editing. Cyclic electron flow is impaired in the mutant, indicating compromised NDH activity. This might be caused by the less efficient editing of *ndhB*, although it is only mildly affected in the mutant. An intriguing observation is that in *ocp3* as well as in some PPR editing factor mutants— *crr2* and *crr21*, impaired NDH activity accompanied enhanced resistance to fungal infection (57). Although how OCP3 associates with the editing apparatus is still unknown, these observations in the mutant lines provide new insights into the unexpected complexity of plant RNA editosome.

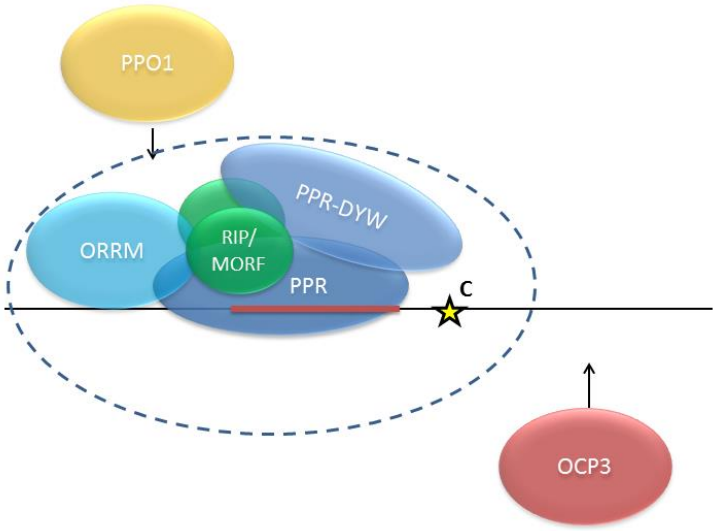


Figure 1.2. Schematic diagram of current model of a plant RNA editosome. Red line represents *cis*-element. Star represents editable cytidine. One PPR protein recognizes the cis-element and a second PPR protein provides DYW domain. Two RIP/MORF proteins exist in the complex and interact with PPR protein. Accessory factors PPO1 and OCP3 are not included in the core complex, which is circled in dashed line.

Strategy to identify unknown trans-acting factors

At least 3 types of editing factors are required for plastid RNA editing—PPR proteins, RIP/MORF proteins and ORRMs, which presumably constitute the core complex of editing. However, whether they are sufficient for editing is still an open question. In Chapter 4, I will describe another novel editing factor—VAR3, a zinc finger protein that was immunoprecipitated with ORRM1 and found to be required for plastid editing. Apparently, unknown factors still exist in the editing complex.

In order to identify unknown factors, researchers have employed various strategies. Forward genetics is powerful in identifying many mitochondrial editing factors (48, 58). However, it is laborious and has limits, especially for lethal mutations. Comparative genomics has been another way to identify PPR editing factors, which is

based on the variation of C targets across species (37, 43). Now, with the recently released RNA-PPR recognition code, finding PPR editing factors is not as great a challenge as it was previously (32). What is more difficult is identification of the other components of the editing complex.

ORRM1 was found through homology search against RIP family, although it turned out to be a different type of editing factor (52). A similar strategy can still be applied. For instance, a homology search can be performed against the unique RRM domain of ORRM1, and a mutant screening for this set of candidates might identify more similar factors.

With more and more editing factors available, co-expression analysis will be more powerful and accurate. CREF3, CREF7 as well as OCP3 were all found based on this strategy (45, 57). However, one drawback is that if the query protein is multi-functional, it will be difficult to filter out the noise from the real editing candidates.

Epitope-tagging—immunoprecipitation—proteomic analysis is another powerful strategy, especially for multi-component complexes. RIP1 and VAR3 were identified through this approach [(47) and Chapter 4]. Important considerations are how to produce a protein that is functional despite an epitope tag and how to distinguish top candidates from proteins that have bound nonspecifically.

Although many proteins might be involved in plant RNA editing, only a limited number of them are expected to be essential components of the core complex, given the 200-400kD molecular size of editosome. A full set of essential components should be able to reconstitute editing activity *in vitro.* Plant RNA editing is an important processing step for organelle *transcripts.* Elucidation of the plant editing machinery will undoubtedly help us better understand plant organelle gene expression.

Both the RNA recognition factor and the unknown enzymatic factor in the editing complex are potentially reprogrammable. Thus the editing complex can be engineered to powerful tools for gene expression control for both agricultural and pharmaceutical purposes.

References

1. Benne R et al. (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.

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

3. Hiesel R, Wissinger B, Schuster W, Brennicke A (1989) RNA editing in plant mitochondria. *Science* 246:1632–1634.

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

5. Thomas SM, Lamb RA, Paterson RG (1988) Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* 54:891–902.

6. Chen SH et al. (1987) Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* 238:363–366.

7. Aphasizhev R, Aphasizheva I (2014) Mitochondrial RNA editing in trypanosomes: Small RNAs in control. *Biochimie* 100:125–131.

8. Sommer B, Köhler M, Sprengel R, Seeburg PH (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* 67:11–19.

9. Mehta A, Driscoll DM (1998) A sequence-specific RNA-binding protein complements Apobec-1 to edit *Apolipoprotein B* mRNA. *Mol Cell Biol* 18:4426–4432.

10. Mehta A, Driscoll DM (2002) Identification of domains in apobec-1 complementation factor required for RNA binding and *apolipoprotein-B* mRNA editing. *RNA* 8:69–82.

11. Fossat N et al. (2014) C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO . doi: 10.15252/embr.201438450*

12. Dracheva S et al. (2008) Increased serotonin 2C receptor mRNA editing: a possible risk factor for suicide. *Mol Psychiatry* 13:1001–1010.

13. Jepson JEC, Reenan RA (2008) RNA editing in regulating gene expression in the brain. *Biochim Biophys Acta* 1779:459–470.

14. Tan BZ, Huang H, Lam R, Soong TW (2009) Dynamic regulation of RNA editing of ion channels and receptors in the mammalian nervous system. *Mol Brain* 2:13.

15. Bazak L et al. (2014) A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res* 24:365–376.

16. Gray MW, Covello PS (1993) RNA editing in plant mitochondria and chloroplasts. *FASEB J* 7:64–71.

17. Bentolila S, Oh J, Hanson MR, Bukowski R (2013) Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet* 9:e1003584.

18. Rajasekhar VK, Mulligan RM (1993) RNA editing in plant mitochondria: [alpha]-phosphate is retained during C-to-U conversion in mRNAs. *Plant Cell* 5:1843–1852.

19. Chateigner-Boutin A-L et al. (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56:590–602.

20. Zhou W et al. (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. *Plant J* 58:82–96.

21. Chateigner-Boutin A-L, Hanson MR (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol Cell Biol* 22:8448–8456.

22. Chaudhuri S, 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. *EMBO J* 15:5958–5964.

23. Hayes ML, Hanson MR (2007) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA* 13:281–288.

24. Reed ML, Peeters NM, Hanson MR (2001) A single alteration 20 nt 5′ to an editing target inhibits chloroplast RNA editing in vivo. *Nucleic Acids Res* 29:1507–1513.

25. Halter CP, Peeters NM, Hanson MR (2004) RNA editing in ribosome-less plastids of *iojap* maize. *Curr Genet* 45:331–337.

26. Zeltz P, Hess WR, Neckermann K, Borner T, Kossel H (1993) Editing of the chloroplast *rpoB* transcript is independent of chloroplast translation and shows different patterns in barley and maize. *EMBO J* 12:4291–4296.

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

28. Lurin C et al. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell Online* 16:2089–2103.

29. Colcombet J et al. (2013) Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. *RNA Biol* 10:1557–1575.

30. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281:37661–37667.

31. Tasaki E, Hattori M, Sugita M (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J* 62:560–570.

32. Barkan A et al. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8:e1002910.

33. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T (2013) Elucidation of the RNA recognition code for Pentatricopeptide Repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* 8:e57286.

34. Takenaka M, Zehrmann A, Brennicke A, Graichen K (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* 8:e65343.

35. Yin P et al. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504:168–171.

36. Okuda K, Myouga F, Motohashi R, Shinozaki K, Shikanai T (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci* 104:8178–8183.

37. Hayes ML, Giang K, Berhane B, Mulligan RM (2013) Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem* 288:36519–36529.

38. Chateigner-Boutin A-L et al. (2013) The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J* 74:935–945.

39. Salone V et al. (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581:4132–4138.

40. Nakamura T, Sugita M (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* 582:4163–4168.

41. Okuda K et al. (2009) Pentatricopeptide Repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell Online* 21:146–156.

42. Boussardon C et al. (2012) Two interacting proteins are necessary for the editing of the *ndhD*-1 site in Arabidopsis plastids. *Plant Cell Online* 24:3684–3694.

43. Robbins JC, Heller WP, Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the Arabidopsis chloroplast *accD* transcript. *RNA* 15:1142–1153.

44. Yu Q-B, Jiang Y, Chong K, Yang Z-N (2009) AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in Arabidopsis thaliana. *Plant J* 59:1011–1023.

45. Yagi Y et al. (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol* 10:1419–1425.

46. Hegeman CE, Hayes ML, Hanson MR (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in Arabidopsis in vitro. *Plant J* 42:124–132.

47. Bentolila S et al. (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci* 109:E1453–E1461.

48. Takenaka M et al. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci* 109:5104–5109.

49. Chatterjee M et al. (1996) *DAG*, a gene required for chloroplast differentiation and palisade development in Antirrhinum majus. *EMBO J* 15:4194–4207.

50. Babiychuk E, Fuangthong M, Montagu MV, Inzé D, Kushnir S (1997) Efficient gene tagging in Arabidopsis thaliana using a gene trap approach. *Proc Natl Acad Sci* 94:12722–12727.

51. Bisanz C et al. (2003) The Arabidopsis nuclear *DAL* gene encodes a chloroplast protein which is required for the maturation of the plastid ribosomal RNAs and is essential for chloroplast differentiation. *Plant Mol Biol* 51:651–663.

52. Sun T et al. (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc Natl Acad Sci* 110:E1169–E1178.

53. Hirose T, Sugiura M (2001) Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *EMBO J* 20:1144–1152.

54. Tillich M et al. (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci* 106:6002–6007.

55. Kupsch C et al. (2012) Arabidopsis chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell Online* 24:4266–4280.

56. Zhang F et al. (2014) Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc Natl Acad Sci*:201316183.

57. Garc á-Andrade J, Ram írez V, L ópez A, Vera P (2013) Mediated plastid RNA editing in plant immunity. *PLoS Pathog* 9:e1003713.

58. Zehrmann A, Verbitskiy D, H ärtel B, Brennicke A, Takenaka M (2010) RNA editing competence of trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain. *FEBS Lett* 584:4181–4186.

Chapter 2

RIP1, a member of an *Arabidopsis* protein family, interacts with the protein RARE1 and broadly affects RNA editing

Stephane Bentolila, Wade P.Heller, Tao Sun, Arianne M. Babina, Giulia Friso, Klaas J. van Wijk, and Maureen R. Hanson

This work was originally published in. PNAS. 2012; 109: E1453–E1461.
TS contributed to experimental design, performing experiments and data analysis. TS provided RNA editing analysis results of the chloroplast sites, yeast two-hybrid results, complementation results and subcellular localization results. SB provided editing analysis result of the mitochondrial sites, real-time PCR results. WPH designed the proteomics study and identified the protein. GF and KJW conducted proteomics analysis. SB and MRH wrote the manuscript.

Abstract

Transcripts of plant organelle genes are modified by C-to-U RNA editing, often changing the encoded amino acid predicted from the DNA sequence. Members of the PLS subclass of the pentatricopeptide repeat (PPR) motif-containing family are site-specific recognition factors for either chloroplast or mitochondrial C targets of editing. However, other than PPR proteins and the cis-elements on the organelle transcripts, no other components of the editing machinery in either organelle have previously been identified. The Arabidopsis chloroplast PPR protein RARE1 specifies editing of a C in the *accD* transcript. RARE1 was detected in a complex of >200 kD.

We immunoprecipitated epitope-tagged RARE1 and tandem MS/MS analysis identified a protein of unknown function lacking PPR motifs; we named it RNA-editing factor Interacting Protein 1 (RIP1). Yeast two-hybrid analysis confirmed RIP1 interaction with RARE1, and RIP1-GFP fusions were found in both chloroplasts and mitochondria. Editing assays for all 34 known Arabidopsis chloroplast targets in a *rip1* mutant revealed altered efficiency of 14 editing events. In mitochondria, 266 editing events were found to have reduced efficiency, with major loss of editing at 108 C targets. Virus-induced gene silencing of *RIP1* confirmed the altered editing efficiency. Transient introduction of a WT *RIP1* allele into *rip1* improved the defective RNA editing. The presence of RIP1 in a protein complex along with chloroplast editing factor RARE1 indicates that RIP1 is an important component of the RNA editing apparatus that acts on many chloroplast and mitochondrial C targets.

Introduction

Posttranscriptional C-to-U RNA editing occurs in plastid and plant mitochondrial transcripts. In a typical vascular plant, approximately 30 C targets in chloroplasts and over 500 C targets in mitochondria are targeted for editing (1, 2). The majority of the editing events results in encoding of a different amino acid than the one predicted from the genomic sequence. The editing-encoded amino acid is usually more conserved relative to residues present in homologous proteins in other organisms than the genomically encoded amino acid. Because there is presently no known case in which useful genetic variation results from partial editing of a transcript population,

the current concept is that editing is a correction mechanism for T-to-C mutations that have arisen in plant organelle genomes (1, 3, 4).

Little is known about the molecular apparatus that is responsible for recognizing the correct C target for editing and converting it to U, although plant mitochondrial RNA editing was discovered over 20 years ago (5-7). Cis-elements for recognition of editing sites have been identified proximal and 5' to the nucleotide to be modified (8-10). As few as 22 nt of sequence surrounding the C target is sufficient to specify RNA editing (9). In 2005, a pentatricopeptide repeat (PPR) motif-containing protein termed CRR4 was discovered to be required for editing of the chloroplast *ndhD* start codon (11) and it binds to cis-elements on *ndhD* transcripts *in vitro* (12). Since that time, members of the PPR protein family have been identified as site-specific recognition factors for a number of C targets in either chloroplasts or mitochondria. PPR proteins consist of a tandem array of degenerate 35-aa repeats and can be divided into two major subfamilies based on the nature of their PPR motifs, the P and PLS subfamilies (13). The P subfamily contains a 35-aa motif, whereas the PLS subfamily exhibits longer (L) or shorter (S) variant PPR motifs within the tandem arrays. The PLS subfamily, which is specific to the plant kingdom, can be further separated into smaller subclasses based on two C-terminal motifs, the E and DYW motifs (14). All of the well-characterized organelle editing factors that are required for editing at specific sites are members of the PLS subfamily of PPR proteins (11, 15-29).

Other than the cis-elements and site-specific PPR proteins, the components of the editing machine are unknown. The enzymatic activity that converts C to U remains unidentified, although the DYW domain found in about half of the Arabidopsis PPR

editing factors does contain a sequence similar to the conserved cytidine/deoxycytidylate deaminase motif (30). To identify additional components of the chloroplast editing apparatus in Arabidopsis, we immunoprecipitated an epitope-tagged PPR-DYW protein named RARE1, which is responsible for recognition of a C target in the chloroplast *accD* transcript (21). MS/MS analysis of the co-immunoprecipitated proteins resulted in the identification of a protein of unknown function lacking PPR motifs. Yeast two-hybrid analysis confirmed the interaction of RARE1 and the novel protein, which is named **R**NA-editing factor **I**nteracting **P**rotein 1 (RIP1). Although RIP1 was identified by its interaction with a chloroplast PPR protein, GFP localization experiments revealed its presence in both plastids and mitochondria. Virus-induced gene silencing of *RIP1* resulted in defective editing of both chloroplast and mitochondrial C targets. A homozygous *rip1* mutant line exhibited altered editing of 14 Cs in chloroplast transcripts and impaired editing of 266 of 368 mitochondrial editing sites that were assayed, with major loss of editing of 108 mitochondrial Cs. Transient introduction of a wild-type *RIP1* allele into the mutant resulted in improvement in the defective RNA editing. Our findings indicate that *RIP1*, which belongs to a 10-member gene family, is required for efficient editing at most Arabidopsis mitochondrial editing sites and plays an important role in chloroplast editing as well. Identification of RIP1 is a significant step that will aid additional efforts to understand the mechanism of plant organelle RNA editing.

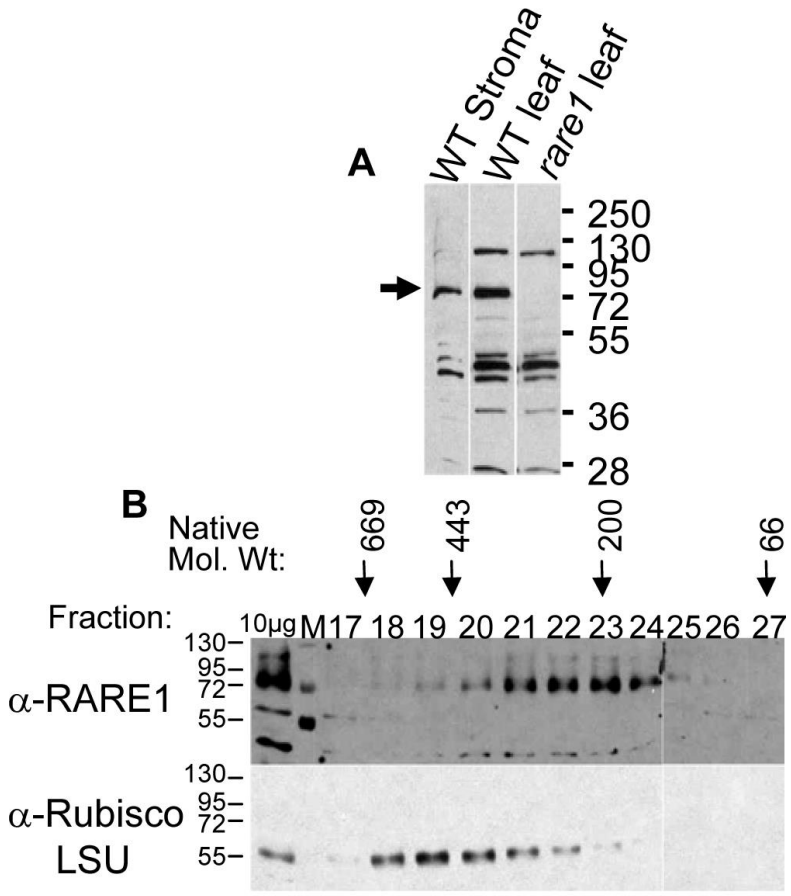


Figure 2.1. RARE1 is part of a protein complex. (A) Immunoblot of wild-type and *rare1* protein extracts using α -RARE1 antibody. α-RARE1 antibody reacts with a 75 kDa protein in wild-type stroma and leaf, which is absent in *rare1* leaf. Arrow indicates RARE1 protein. Loading for all plant protein samples is 20 µg/lane. (B) Size exclusion chromatography fractions of wild-type stroma probed with α-RARE1 antibody or α-Rubisco LSU antibody. An equal volume of each fraction was loaded. An arrow indicates the fraction containing the greatest amount of each size standard.

Results

Identification of RIP1 as a RARE1-interacting Protein

Our previous work reported the identification of RARE1, a plastid editing factor that controls the editing of *accD*-794 (21). We determined that RARE1 is present in a protein complex by performing size exclusion column chromatography on

chloroplast stroma (Figure 2.1). To identify members of this complex, we produced transgenic plants that express RARE1 protein carrying a 3x FLAG tag (RARE1-3xF) (31) (Figure 2.2). Leaf protein extract from transgenic plants was incubated with α-FLAG agarose to isolate the RARE1 complex (Figure 2.3).

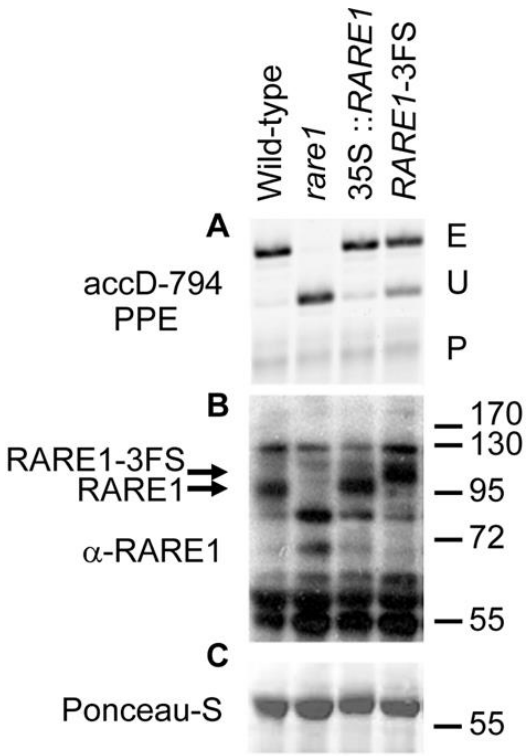


Figure 2.2. A tagged version of RARE1 partially restores the *accD*-794 editing defect in the *rare1* mutant. (A) Acrylamide gel separating the poisoned primer extension (PPE) products obtained from the wild-type, the *rare1* mutant, and two transgenic *rare1* lines transformed with different versions of *RARE1*, 35S :: *RARE1*: wild-type allele under the control of the 35S promoter, *RARE1*-3xF: tagged *RARE1* with 3xFLAG under the control of the native promoter. The PPE products E (edited), U (unedited), and P (Primer) are 34, 30 and 22 nt, respectively. The two constructs, 35S :: *RARE1* and *RARE1*-3xF, restore *accD*-794 editing extent with a decreasing efficiency, as shown by the increasing intensity of the unedited band in the respective lanes. (B) Immunoblot in which 20 μg total leaf protein from each sample was probed with α-RARE1 antibody to determine the relative abundance of RARE1 protein in the individual lines. (C) Ponceau-S stain of Rubisco large subunit demonstrates approximately equal loading.

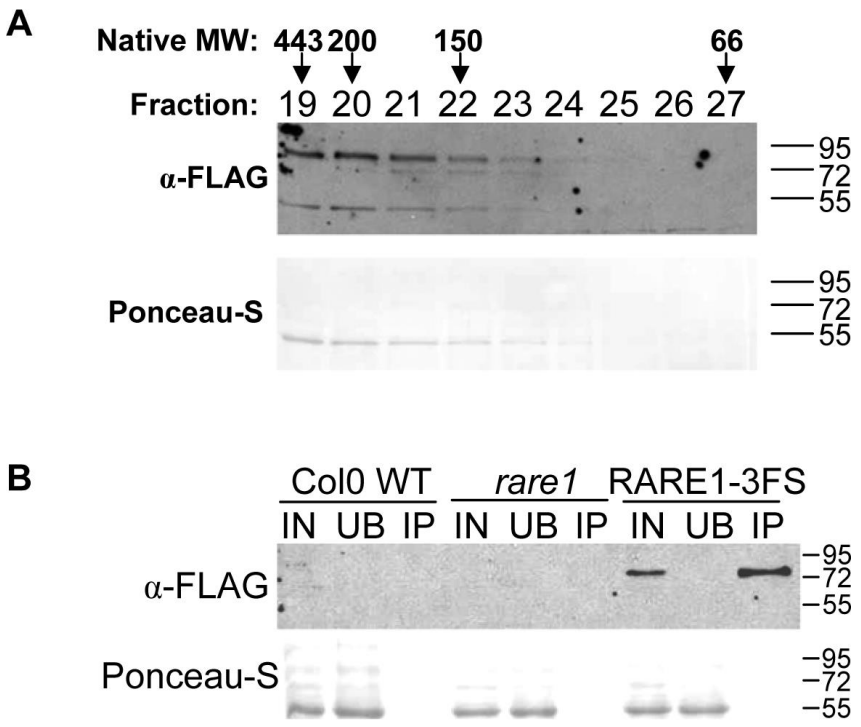


Figure 2.3. Separation and Immunoprecipitation of the RARE1-3xF complex. (A) Extracts of chloroplast stroma in RIPA buffer contain a RARE1-3xF complex of similar size as the previously observed RARE1 complex extracted in KEX Buffer (Figure 1). Size exclusion chromatography fractions of wild-type stroma were probed with α-FLAG antibody, with the peak fraction indicated where the size standards eluted. Due to the different buffer used for the RARE1-3xF extracts, the particular fraction(s) in which size standards and RARE1 complexes eluted are not identical to the chromatography with the native complex. (B) Leaf extracts were treated with α-FLAG antibody, immunoprecipitates were separated by SDS-PAGE, and the immunoblot was probed with α-FLAG antibody. As expected, neither the wild-type nor the mutant react with the α-FLAG antibody. The RARE1-3XF protein is present in the input (IN) and immunoprecipitate (IP) fractions from the transgenic line and depleted in the unbound (UB) fraction. Ponceau-S stain of Rubisco shows equal loading of control and transgenic samples.

The MS data indicated that the protein encoded by At3g15000 was the top candidate RARE1-interacting protein present in the immunoprepitate, because it had the largest number of matches of MS/MS spectra other than RARE1 (Table 2.1).

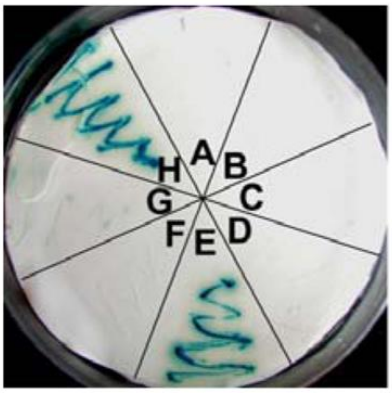
Table 2.1.MS/MS based identification of RIP1 (At3g15000.1) in the co-immunoprecipitate from FLAG-tagged RARE1 (At5g13270.1)

Peptide (a)	Modification (b)	SearchType (c)	# matched MS/MS spectra
RARE1 - At5g13270.1			
ACASLEELNLGK		Full_Tryptic	7
AGLCSNTSIETGIVNMYVK	Oxidation (M)	Full_Tryptic	5
AGVSVSSYSYQCLFEACR		Full_Tryptic	4
AVGLFSGMLASGDKPPSSMYTTLLK	2 Oxidation (M)	Full_Tryptic	2
ELSCSWIQEK		Full_Tryptic	4
FIVGDKHHPQTQEIYEK		Full_Tryptic	1
HVSLVTGHEIVIR		Full_Tryptic	6
KLNEAFEFLQEMDK	Oxidation (M)	Full_Tryptic	1
KPVACTGLMVGYTQAGR	Oxidation (M)	Full_Tryptic	7
LAIAFGLISVHGNAPAPIK		Full_Tryptic	3
LFDEMSELNAVSR	Oxidation (M)	Full_Tryptic	4
LKEFDGFMEGDMFQCNMTER	3 Oxidation (M)	Full_Tryptic	2
LNEAFEFLQEMDK	Oxidation (M)	Full_Tryptic	4
NLELGEIAGEELR		Full_Tryptic	7
SGLLDEALK		Full_Tryptic	3
SLIGSQYGESALITMYSK	Oxidation (M)	Full_Tryptic	4
TTMISAYAEQGILDK	Oxidation (M)	Full_Tryptic	5
RIP1 - At3g15000.1			
TLAQIVGSEDEAR	none	Full_Tryptic	10

The mass spectral data were searched using MASCOT ($p<0.01$; error <6 ppm for precursor ions) against the Arabidopsis database (v.8) downloaded from TAIR. Neither proteins were identified in control samples

(a) Matched peptide sequence from MS/MS spectra, within 6 ppm mass accuracy
(b) Variable peptide methione oxidation
(c) Only full tryptic peptide are allowed

The gene encodes a member of the Differentiation and Greening (DAG) family; mutants in members of this gene family exhibit chloroplast biogenesis defects (32, 33). Yeast two hybrid analysis confirmed the interaction between RARE1 and the protein encoded by At3g15000, which was therefore named RIP1 (Figure 2.4). Serial deletions of both RARE1 and RIP1 established the portions responsible for the interaction on the N termini of the proteins (Figure 2.5).



B

	pDEST32	pDEST22	Interaction
A	empty	RIP1FL	0
B	empty	RIP1ΔcTP	0
C	RARE1ΔcTP	empty	0
D	RARE1ΔcTP	RIP1FL	0
E	RARE1ΔcTP	RIP1ΔcTP	+++
F	Krev1	RalGDS-m2	0
G	Krev1	RalGDS-m1	+
H	Krev1	RalGDS-wt	+++

Figure 2.4. RIP1 interacts with RARE1 *in vivo.* (A) X-gal reporter assay of lacZ transcriptional activation as proof of interaction in a yeast two-hybrid experiment. (B) Table describing the constructs tested for interaction in the yeast two-hybrid analysis and relative degree of lacZ expression. F-H contain control plasmids included with ProQuest kit for a negative, weak and strong protein-protein interaction in F, G and H, respectively. Unless otherwise indicated, pDEST22 and pDEST32 are empty vectors used to demonstrate that there is no autoactivation of lacZ expression when only RARE1- or RIP1-fusion proteins are expressed. RIP1FL denotes full-length RIP1 without cTP removal and RIP1ΔcTP indicates removal of a TargetP-predicted 56 aa cTP.

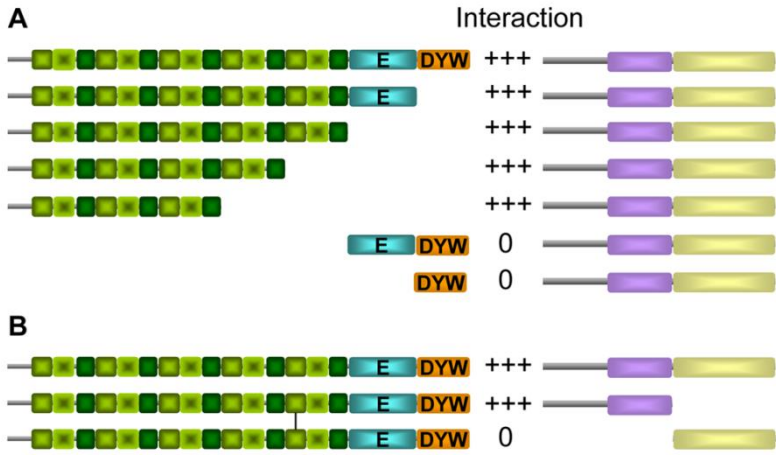


Figure 2.5. Specific regions of RARE1 and RIP1 are responsible for their interaction *in vivo*. (A) Diagram of the serial deletions of RARE1 (left) that were tested in by yeast two-hybrid analysis with RIP1 (right), which is divided into N-terminal (purple) and C-terminal (yellow) regions. Pentatricopeptide motifs: P L S. All the proteins were expressed without predicted transit peptides. Relative degree of *lacZ* expression indicated by + signs.

T-DNA Insertional *rip1* Mutant Exhibits a Dwarf Phenotype and Altered Chloroplast RNA Editing

Two mutant lines with insertions in the *RIP1* locus (Figure 2.6) were obtained from the INRA FLAGdb T-DNA collection (34). Homozygous mutants could not be recovered from the FLAG_607H09 line; possibly the T-DNA insertion in FLAG_607H09 might be lethal because of the complete loss of expression. Homozygous FLAG_150D11 mutants, which have a T-DNA inserted 140 bp upstream of the *RIP1* coding region, exhibit a dwarf phenotype (Figure 2.6D). We measured the level of *RIP1* transcript in the homozygous FLAG_150D11 mutant line and homozygous wild-type siblings by quantitative RT-PCR. The expression of the *RIP1* ORF was found to be increased 4 to 6 fold in the T-DNA mutant compared to the

wild-type (Figure 2.6E). Nevertheless, the proximity of the T-DNA insertion to the open reading frame may result in impaired production of RIP1 protein; abnormal phenotypes have previously been reported in T-DNA insertional mutants that exhibited increased rather than reduced target gene transcript abundance (35).

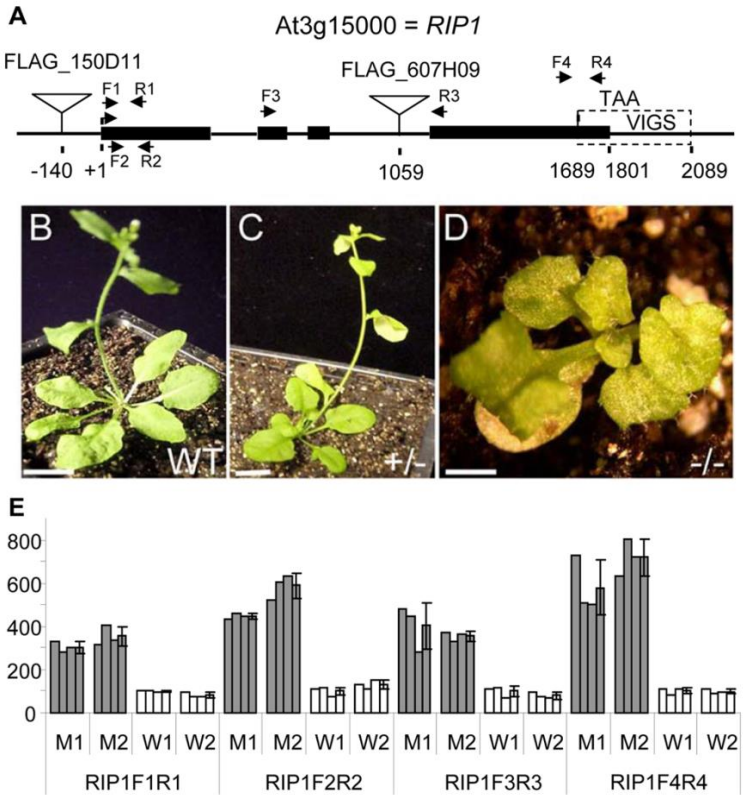


Figure 2.6. A *rip1* mutant exhibits dwarf phenotype and increases in *RIP1* transcript. (A) Map of At3g15000 (*RIP1*) with exons shown as black rectangles, T-DNA insertions shown as triangles, the region used for VIGS indicated, and the location of primers used for quantitative RT-PCR shown as facing arrows. (B-D) WT, heterozygous, and homozygous progeny of a heterozygous plant carrying the FLAG_150D11 insertion. Plants are 32 days old. (Scale bars: B and C, 10 mm; D, 1 mm.) (E) The expression of *RIP1* is increased four- to sixfold in the T-DNA mutant compared with WT. Quantitative RT-PCR measured the level of *RIP1* transcript in two homozygous mutants (M1 and M2) and two homozygous WT siblings (W1and W2). Quantitative RT-PCR assays were replicated three times for each plant. The expression level was arbitrarily set at 100 for W1. SDs are indicated (n=3).

Because RIP1 co-immunoprecipitates and interacts *in vivo* with RARE1, a chloroplast editing factor, we surveyed the editing extent of all known Arabidopsis chloroplast editing sites in segregating progeny for the FLAG_150D11 T-DNA insertion. A poisoned primer extension (PPE) assay is shown in Figure 2.7 for *accD*-794 and the 3 sites showing the most pronounced editing extent variation in the mutant relative to WT. *PetL*-5 and *ndhD*-2 exhibit a significant reduction of editing extent in the mutant (60% and 55%, respectively), whereas *rps12*-(i1)58, a site in the first intron of *rps12*, shows a significant increase of editing extent in the mutant (Figure. 2.7). PPE data for *accD*-794, the site under the control of RARE1, indicate that editing in the homozygous mutant is reduced relative to wild-type as observed for *petL-5* and *ndhD-2*, but to a lesser extent (83% in mutant compared with 98% in WT or a 15% reduction). The mutation is clearly recessive, because the editing extent of the heterozygous plant for these sites is similar to the homozygous wild plants (Figure. 2.7).

Of the 34 known chloroplast C-targets of editing present in Arabidopsis, 14 C targets exhibited significant changes in RNA editing extent between the homozygous WT and the homozygous mutant plants (Table 2.2). 11 of the 14 sites exhibit a decrease in editing extent in the mutant, whereas an increase of editing extent in the mutant is observed for only 3 sites (Table 2.2). The editing extent of the heterozygote was not significantly different from the homozygous WT at any of the chloroplast sites.

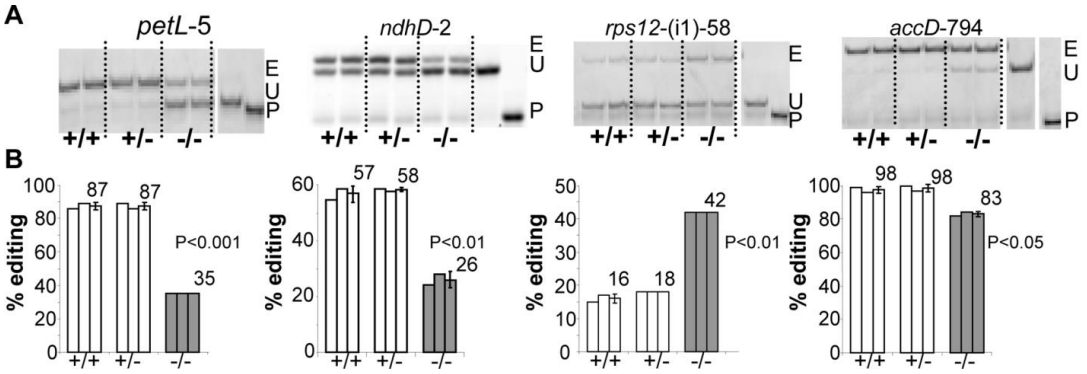


Figure 2.7. Mutation in *RIP1* affects the editing extent of plastid sites. (A) Acrylamide gels separate the PPE products obtained from sibling plants, 2 homozygous WT (+/+), 2 heterozygous (-/+) mutants and two homozygous mutants (-/-).E: edited; P: primer; U: unedited. The name of the site assayed is given above each gel. (B) The quantification of editing extent, derived from the measure of the band's intensity, is represented by a bar below each lane of the acrylamide gels. The average is given for each genotypic class with SD. The sites *petL*-5, *ndhD*-2, and *accD*-794 show a significant decrease of the editing extent in the mutant, representative of the majority of the plastid sites showing an effect of the *RIP1* mutation. The site *rps12*-(i1)-58 shows a significant increase of editing extent in the mutant compared with the WT and heterozygous plants as observed only in two other plastid editing sites.

Table 2.2. Effect of FLAG_150D11 insertion on RNA editing of chloroplast C-targets, ranked by degree of change in editing and grouped by known trans-factors

Trans-factor (if known)	Chloroplast C-target	Genotype +/+	Genotype -/+	Genotype -/-	P value	Δ Editing -/-::+/+
	rps12-(i1)58	16 ± 1	18	42	0.001 * *	**162.5**
	petL-5	87 ± 1	87 ± 1	35	0.0008 * * *	**-60.0**
CRR4	*ndhD*-2	57 ± 2	58	26 ± 2	0.004* *	**-55.4**
CRR28	*ndhB*-467	85 ± 1	85 ± 3	68 ± 2	0.006 **	**-20.6**
CRR28	*ndhD*-878	91 ± 1	91 ± 1	70 ± 9	0.08 ns	-23.0
	rpoC1-488	62 ± 2	57 ± 1	74 ± 4	0.002 * *	**18.4**
RARE1	*accD*-794	98 ±1	98 ±1	83 ±1	0.015 *	**-14.9**
	ndhB-586	94	93 1	84	0.0004***	**-10.8**
	rpoB-2432	83 ±2	85 ± 3	91 ± 1	0.03 *	**9.6**
OTP84	*ndhB*-1481	94	96	89 ± 1	0.04*	**-5.4**
OTP84	*ndhF*-290	98	98 ± 1	95 ± 1	0.029*	**-3.6**
OTP84	*psbZ*-50	94 ± 1	94 ± 2	90 ± 3	0.27 ns	-3.5
CRR21	*ndhD*-383	98	98 ± 1	94 ± 1	0.049*	**-3.6**
OTP82	*ndhB*-836	95	95 ± 1	92	0.03*	**-2.8**
OTP82	*ndhG*-50	77 ± 3	82 ± 1	72 ± 1	0.18 ns	-5.9
	ndhB-830	98 ± 1	98	95	0.03*	**-2.7**
CRR22	*ndhB*-746	98	97 ± 1	96	0.02*	**-1.2**
CRR22	*ndhD*-887	88 ± 2	88 ± 2	73 ± 9	0.16 ns	-16.7
CRR22	*rpoB*-551	50 ± 9	50 ± 3	50 ± 13	0.99 ns	1.0
OTP80	*rpl23*-89	69 ± 2	70 ± 8	60 ± 3	0.07 ns	-13.3
OTP85	*ndhD*-674	92 ± 1	92 ± 1	82 ± 6	0.15 ns	-10.8
CLB19	*rpoA*-200	72 ± 4	67 ± 8	80 ± 3	0.16 ns	10.2
CLB19	*clpP*-559	92 ± 1	92 ± 2	93	0.33 ns	1.2
	ndhB-149	98 ± 4	97 ± 1	90 ± 4	0.1 ns	-8.0
	rps14-80	79 ± 1	77 ± 1	73 ± 1	0.05 ns	-7.6
	matK-640	69 ± 6	64 ± 6	73 ± 7	0.58 ns	5.9
LPA66	*psbF*-77	83	83 ± 3	86 ± 1	0.078 ns	3.6
	ndhB-1255	91 ± 5	93 ± 1	88 ±2	0.5 ns	-3.2
	ndhB-872	87 ± 3	86 ± 1	85	0.39 ns	-2.7
	psbE-214	98	98	96 ± 2	0.21 ns	-2.3
	atpF-92	93 ± 3	97	95 ± 1	0.44 ns	2.3
	accD-1568	60 ± 2	58 ± 7	61 ± 1	0.57 ns	2.2
YS1	*rpoB*-338	97 ± 1	97	95 ± 1	0.22 ns	-1.5
OTP86	*rps14*-149	80	80 ± 1	80	0.25 ns	-0.7

The variation in editing is = 100* (editing extent in-/-- editing extent in +/+)/editing extent in +/+. Minus sign indicates that the editing extent is decreased in the mutant. Significant editing extent variation is given in bold.

RIP1 is Dual-targeted to Both Chloroplasts and Mitochondria.

RIP1 has been previously reported to be located in mitochondria, according to characterizations of the Arabidopsis mitochondrial proteome (36, 37). In addition, the dwarf phenotype of the FLAG_150D11 T-DNA insertional mutant could be indicative of mitochondrial dysfunction. We therefore determined the location of RIP1 by transiently expressing a construct encoding the full length *RIP1* attached to GFP under the control of a 35S promoter into Arabidopsis protoplasts (38). Our observations indicate that RIP1 is dually targeted to chloroplasts and mitochondria (Figure 2.8). Most of the Arabidopsis protoplasts showed RIP1 to be localized in mitochondria (Figure 2.8C). Occasionally we observed RIP1 both in mitochondria and chloroplasts (Figure 2.8G). This observation is reminiscent of a recent report on PPR2263, a maize PPR-DYW that is dually targeted to mitochondria and chloroplasts, with a preference for mitochondria (39).

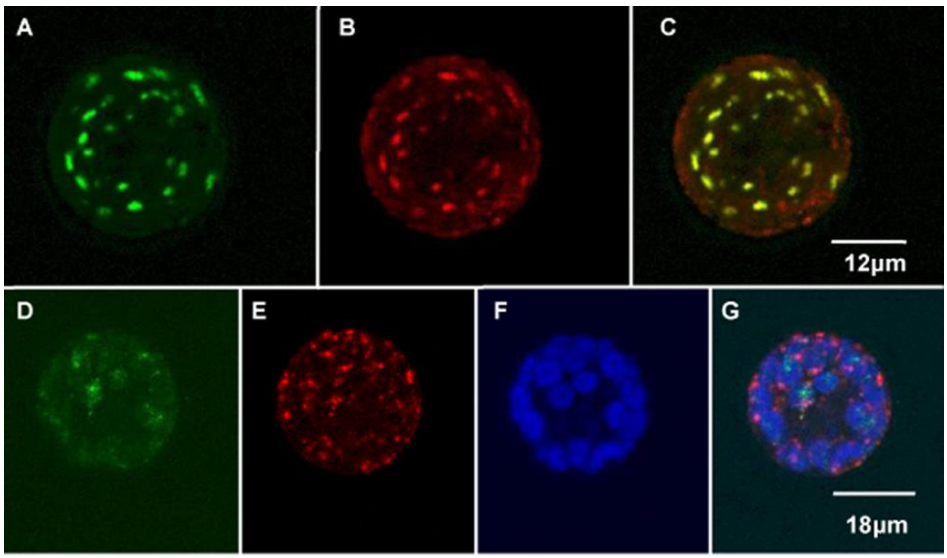


Figure 2.8. RIP1 is dual-targeted to Arabidopsis mitochondria and chloroplasts. Protoplasts prepared from leaves of Arabidopsis accession Col-0 were transfected with a construct encoding an RIP1-GFP fusion protein under the control of a 35S promoter. Protoplasts were examined for fluorescence 16 h after incubation with the construct. (A and D) GFP signal is green (B and E) Mitochondria (red) were labeled with Mitotracker Orange. (C) Merge of GFP and mitochondrial signal is yellow. (F) Chlorophyll autofluorescence is shown in blue. (G) Merge of D-F gives turquoise signals where GFP and chlorophyll overlap and yellow images where GFP and Mitotracker overlap.

To confirm the dual localization of RIP1 to both organelles, we repeated the previous experiment by transfecting *N. benthamiana* protoplasts. In contrast to Arabidopsis protoplasts, all of the transfected *N. benthamiana* protoplasts showed a dual localization of RIP1 to both mitochondria and chloroplasts (Figure 2.9). DAPI staining of the *N. benthamiana* protoplasts showed that some of the small punctuate structures targeted by RIP1-GFP co-localize with nucleoids (Figure 2.9).

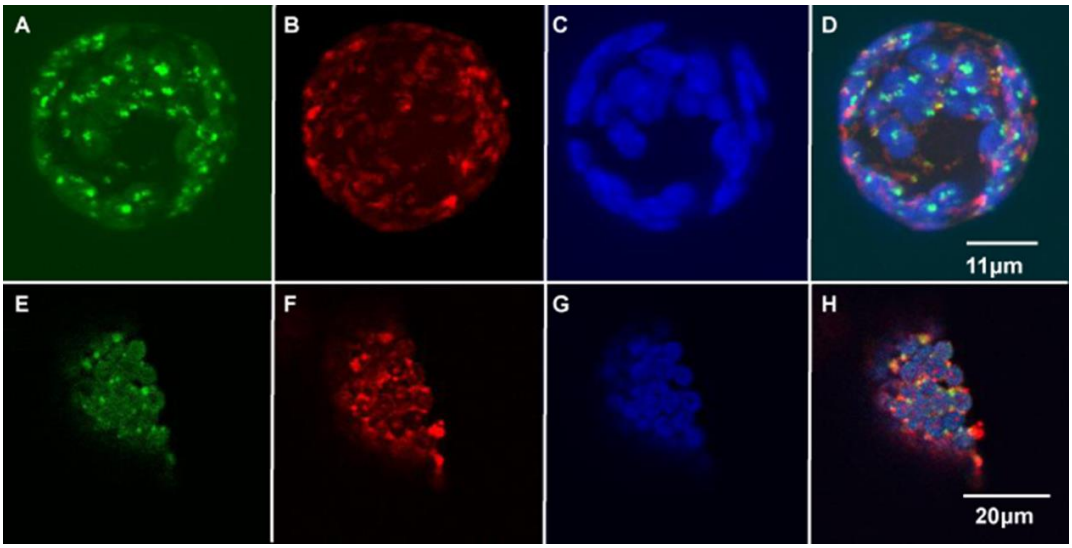


Figure 2.9. RIP1 is dual-targeted to *N. benthamiana* mitochondria and chloroplasts (A-D) and co-localizes with plastid nucleoids (E-H). Protoplasts prepared from leaves of *N. benthamiana* were transfected with a construct encoding a RIP1-GFP fusion protein under the control of a 35S promoter. (A, E) Protoplasts were examined for GFP fluorescence 16 h after incubation with the construct. (B) Mitochondria were detected with Mitotracker Orange. (C, G) Chlorophyll autofluorescence is shown as blue. (D) Merged image shows GFP co-localization within mitochondria (yellow) spots or in chloroplasts (turquoise). (F) DAPI staining of DNA in chloroplast nucleoids (red). (H) Merged images of DAPI and GFP signals (yellow) shows RIP1 to co-localize with nucleoids

rip1 Mutant Exhibits Altered Mitochondrial Editing.

We conducted a bulk sequencing screen of the 33 mitochondrial protein-coding genes known to harbor editing sites by comparing the sequencing electrophoretograms of the RT-PCR products obtained from the homozygous T-DNA mutant with the homozygous WT line. A typical result is shown in Figure 2.10A, where editing extent in the *nad2* transcript is not uniformly affected by the *RIP1* mutation along the transcript.

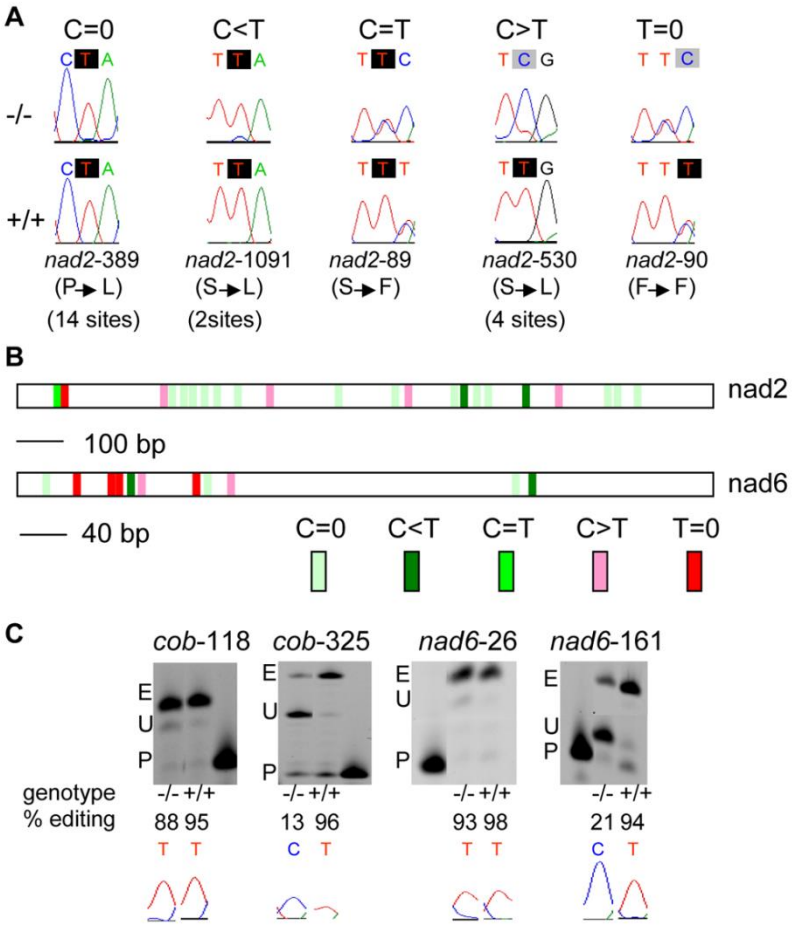


Figure 2.10. Editing extent is not uniformly affected along mitochondrial transcripts in *rip1* mutants. (A) Portions of electrophoretograms from RT-PCR bulk sequencing of *nad2* are shown for the homozygous T-DNA mutant (-/-) and the homozygous wild-type (+/+). Below the electrophoretograms are given the position of the editing site in the *nad2* transcript with the aa change upon editing in between parenthesis, and the number of sites in *nad2* sharing the same molecular phenotype. The editing phenotype of the mutant was classified in one of five categories, above the electrophoretograms, from C=0 (no effect of the mutation on the editing extent) to T=0 (total loss of editing in the mutant). The C target of editing is highlighted by a black shade for T and a grey shade for C, and shown according to its position in the codon. (B) Distribution of the effect of the *RIP1* mutation on the editing extent of mitochondrial sites on *nad2* and *nad6* transcripts. Each site is represented by a block whose background color indicates for the strength of the *rip1* mutation's effect on the editing extent as detected by bulk sequencing. (C) PPE assays confirm the reduction of editing extent of mitochondrial sites in *cob* and *nad6* transcripts previously detected by bulk sequencing. On top are shown the PPE products run on acrylamide gels, with the name and position of the site being assayed above the gel; P: primer, U: unedited, E: edited. Below the gels are shown the electrophoretograms of the editing site.

Table 2.3. Effect of FLAG_150D11 insertion on RNA editing of mitochondrial C-targets

Gene	Effect				
	C=0	C<T	C=T	C>T	T=0
nad1	14	4		1	
nad2	14	2	1	4	1
nad3				3	4
nad4	21	4	3	2	
nad4L		2	2	2	4
nad5	16	4		4	1
nad6	3	2		2	4
nad7	14	2	2	3	1
nad9		4	2	1	
complex I	82	24	10	22	15
cob -complex III	3	3			2
cox2	5	2	1	1	3
cox3		5	1	1	
complex IV	5	7	2	2	3
atp1	1	1	1		
atp4 (orf25)	1	3		3	1
atp6-1	1				
atp9	2	2			
complex V	5	6	1	3	1
ccmB (ccb206)			1	4	27
ccmFn-2 (ccb203)			1	5	4
ccmC (ccb256)	1		1	17	7
ccmFn-1 (ccb382)			1	5	7
ccmFc (ccb452)	4	2	4	1	
cytochrome c biogenesis	5	2	8	32	45
rpl2					1
rpl5		2	1	3	4
rpl16	2	2	1		1
rps3		2	2	1	2
rps4		3	1	1	5
rps7				1	
rps12				5	3
rps14		1			
ribosomal protein	2	10	5	11	16
matR			2	2	5
mttb (OrfX)				6	21
TOTAL	102	52	28	78	108

The five categories of *RIP1* mutation effect on mitochondrial editing, from no effect (C=0) to a total loss of editing (T=0) extent have been presented in text and in Figure 2.10.

The majority of the *nad2* sites, 14 of 22 sites, do not show any reduction in editing extent in the mutant compared with the WT (Table 2.3). However, editing of *nad2*-90 is not detectable in the mutant, because only a C peak is observed at that position (Figure 2.10A). Between these two extremes are detected sites in which editing is reduced to less than one-half of WT, about one-half of WT, or more than one-half of WT as observed in *nad2*-1091, *nad2*-89, and *nad2*-530, respectively (Figure 2.10A).

Table 1 summarizes the results of the bulk sequencing screen by presenting the number of sites for each mitochondrial gene that falls into one of five categories described for *nad2* transcript, from no effect of the *RIP1* mutation to an apparent absence of editing. Of the 33 mitochondrial genes surveyed, only *atp6-1*, which contains one reported editing site at position 475, does not show any dependence on a functional RIP1 for efficient editing. Overall, mutation in *RIP1* affects the editing extent of a very high number of mitochondrial sites; 108 of 368 sites surveyed show a major loss of editing in the mutant (Table 2.3). A very similar number of sites (102 sites) do not show any variation in editing extent in the mutant. A complete list of all the affected mitochondrial C targets of editing among the 368 sites assayed is shown in Dataset S1.

Plant mitochondrial sites in the *rip1* mutant analyzed can be divided into two categories, totally RIP1-dependent (108 of 368 sites or 29%) and totally RIP1-independent (102 of 368 sites or 28%). Although these categories are approximately equal in size in the entire population of genes analyzed, RIP1 seems to play a larger role in editing of transcripts for proteins of certain mitochondrial complexes than

others. For example, transcripts of complex 1 genes exhibit 10% (15/153) C targets affected by the *RIP1* mutation, and 45% (82/153) unaffected. In contrast, the cytochrome c biogenesis complex exhibits 49% C targets (45/92) affected and only 5% (5/92) sites with editing extent that is unaffected (Table 2.3). The effect of *RIP1* mutation on mitochondrial extent does not seem to be related to the location of the C target on the transcript, because there is no apparent pattern in the distribution of the RIP1-dependent and -independent sites along the transcript (Figure 2.10B).

Editing events can be divided into two classes: non-silent (when editing changes the encoded amino acid) or silent (when the amino acid is unchanged). Non-silent sites are predominant in the population of sites surveyed (335 non-silent sites or 91%, Table 2.4). There are somewhat fewer non-silent sites in the group of sites that are strongly affected in the *rip1* mutant than there are in the entire population of surveyed sites [83 % (90 non-silent sites to 108 sites) vs. 91% respectively] (Table 2.4).

We also examined a small selection of editing sites by the PPE assay, which is more precise and sensitive than the RT-PCR/bulk sequencing method that we used to survey the 368 sites in the *rip1* mutant (40). We chose some mitochondrial editing sites that exhibited either no or complete dependence on functional RIP1 (Figure 2.10A). Although no editing of the C targets in *cob*-325 and *nad6*-161 was detected by the less sensitive bulk sequencing method, we found that both exhibit a residual editing extent detectable by PPE (13% and 21%, respectively) (Figure 2.10C). The negative effect of the *rip1* mutation on *cob*-325 and *nad6*-161 is greater than its effect on any chloroplast C targets (Table 2.3). When the editing extent of these two sites was assayed by PPE in homozygotes, heterozygotes, and WT, we found no difference

between heterozygotes and WT, indicating the mutation is completely recessive with respect to editing efficiency at these two C editing targets as well as at other mitochondrial sites (Figure 2.11).

Table 2.4. Effect of FLAG_150D11 insertion on RNA editing extent of mitochondrial C-targets, evaluated by the belonging of each site to one of five categories from C=0 no effect, to T=0 total loss of editing in the mutant as detected by bulk sequencing

gene	C=0				T>C				C=T				C>T				T=0				Total		
	#	position	S	NS	#	position	S	NS	#	position	S	NS	#	position	S	NS	#	position	S	NS		S	NS
atp1	1	1292		Y	1	1178		Y	1	1415		Y									3		3
atp6-1	1	484		Y																	1		1
atp9	2	83		Y	2	53		Y													4		4
		224		Y		167		Y															
ccb203									1	391		Y	5	176		Y	4	65		Y	10		10
														259		Y		208		Y			
														277		Y		226		Y			
														320		Y		356		Y			
														344		Y							
ccb206									1	137		Y	4	149		Y	27	28		Y	32	2	30
														154		Y		71		Y			
														338		Y		80		Y			
														428		Y		128		Y			
																		148	Y				
																		159	Y				
																		160		Y			
																		164		Y			
																		172		Y			
																		179		Y			
																		181		Y			
																		193		Y			
																		194		Y			
																		286		Y			
																		304		Y			
																		367		Y			
																		379		Y			
																		380		Y			
																		424		Y			
																		467		Y			
																		475		Y			
																		476		Y			
																		485		Y			
																		512		Y			
																		514		Y			
																		551		Y			
																		554		Y			
ccb256	1	463		Y					1	133		Y	17	103		Y	7	467		Y	26	1	25
														179		Y		473		Y			
														184		Y		575		Y			
														331		Y		608		Y			
														395		Y		614		Y			
														400		Y		618	Y				
														421		Y		650		Y			
														436		Y							
														446		Y							
														458		Y							
														497		Y							
														521		Y							
														548		Y							
														568		Y							
														619		Y							
														656		Y							
														673		Y							

ccb382							1	262	Y	5	143	Y	7	104	Y	13	1	12
											269	Y		157	Y			
											791	Y		289	Y			
											806	Y		378	Y			
											955	Y		709	Y			
														710	Y			
														779	Y			
ccb452	4	103	Y	2	160	Y	4	122	Y	1	1215	Y				11	1	10
		146	Y		334	Y		155	Y									
		1172	Y					406	Y									
		1280	Y					415	Y									
cob	3	118	Y	3	286	Y							2	325	Y	8	0	8
		908	Y		853	Y								568	Y			
		1084	Y		982	Y												
cox2	5	71	Y	2	581	Y	1	476	Y	1	557	Y	3	27	Y	12	2	10
		253	Y		742	Y								138	Y			
		379	Y											278	Y			
		698	Y															
		721	Y															
cox3				5	112	Y	1	257	Y	1	314	Y				7	0	7
					245	Y												
					311	Y												
					413	Y												
					422	Y												
matR							2	374	Y	2	1593	Y	5	1731	Y	9	2	7
								461	Y		1730	Y		1751	Y			
														1771	Y			
														1807	Y			
														1895	Y			
nad1	14	167	Y	4	265	Y				1	500	Y				19	2	17
		307	Y		571	Y												
		308	Y		635	Y												
		376	Y		755	Y												
		490	Y															
		492	Y															
		493	Y															
		536	Y															
		580	Y															
		674	Y															
		725	Y															
		743	Y															
		823	Y															
		898	Y															
nad2	14	344	Y	2	961	Y	1	89	Y	4	341	Y	1	90	Y	22	2	20
		389	Y		1091	Y					530	Y						
		394	Y								842	Y						
		400	Y								1160	Y						
		427	Y															
		461	Y															
		695	Y															
		821	Y															
		953	Y															
		991	Y															
		995	Y															
		1279	Y															
		1280	Y															
		1309	Y															

nad3											3	83	Y	4	64	Y		7	1	6
												149	Y		211		Y			
												250	Y		212		Y			
															254		Y			
nad4	21	107	Y	4	403		Y	3	659	Y	2	836	Y					30	2	28
		124	Y		977		Y		784	Y		896	Y							
		158	Y		1010		Y		1148	Y										
		164	Y		1101	Y														
		166	Y																	
		197	Y																	
		317	Y																	
		362	Y																	
		376	Y																	
		436	Y																	
		437	Y																	
		449	Y																	
		608	Y																	
		767	Y																	
		1033	Y																	
		1129	Y																	
		1172	Y																	
		1355	Y																	
		1373	Y																	
		1405	Y																	
		1417	Y																	
nad4L				2	188		Y	2	100	Y	2	110	Y	4	41		Y	10	0	10
					197		Y		158	Y		131	Y		55		Y			
															86		Y			
															95		Y			
nad5	16	155	Y	4	272		Y				4	242	Y	1	608		Y	25	2	23
		358	Y		398		Y					374	Y							
		494	Y		725		Y					548	Y							
		553	Y		1610		Y					609	Y							
		598	Y																	
		629	Y																	
		676	Y																	
		713	Y																	
		764	Y																	
		835	Y																	
		863	Y																	
		875	Y																	
		1275	Y																	
		1490	Y																	
		1550	Y																	
		1580	Y																	
nad6	3	26	Y	2	95		Y				2	103	Y	4	53		Y	11	0	11
		169	Y		463		Y					191	Y		88		Y			
		446	Y												89		Y			
															161		Y			
nad7	14	77	Y	2	739		Y	2	698	Y	3	213	Y	1	963	Y		22	3	19
		137	Y		1057		Y		795	Y		316	Y							
		200	Y									724	Y							
		209	Y																	
		244	Y																	
		251	Y																	
		335	Y																	
		344	Y																	
		578	Y																	
		679	Y																	
		734	Y																	
		769	Y																	
		926	Y																	
		1079	Y																	

nad9				4	167	Y	2	298	Y	1	92	Y					7	0	7
					190	Y		328	Y													
					398	Y																
					439	Y																
orf25	1	89	Y	3	248	Y				3	138	Y		1	395	Y	8	2	6
					251	Y					215		Y									
					416	Y					250	Y										
orf114																						
orf240																						
orfX										6	161		Y	21	97		Y	27	5	22
											361		Y		144	Y						
											379		Y		145		Y					
											409		Y		164		Y					
											505		Y		173		Y					
											649		Y		364		Y					
															406		Y					
															407		Y					
															412		Y					
															440		Y					
															474	Y						
															530		Y					
															538		Y					
															552	Y						
															581		Y					
															587		Y					
															643		Y					
															665		Y					
															693	Y						
															700		Y					
															705	Y						
rpl2														1	212		Y	1	0	1		
rpl5				2	317	Y	1	35	Y	3	64		Y	4	47		Y	10	1	9		
					329	Y					169		Y		58	Y						
											512		Y		59		Y					
															92		Y					
rpl16	2	34	Y	2	61	Y	1	440	Y					1	512		Y	6	0	6		
		506	Y		209	Y																
rps3				2	64	Y	2	603	Y	1	887		Y	2	1470	Y		7	2	5		
					1598	Y		1571	Y						1534		Y					
rps4				3	226	Y	1	992	Y	1	377		Y	5	175		Y	10	0	10		
					299	Y									235		Y					
					524	Y									308		Y					
															332		Y					
															967		Y					
rps7										1	332		Y					1	0	1		
rps12										5	104		Y	3	84	Y		8	2	6		
											146		Y		221		Y					
											196		Y		269		Y					
											284		Y									
											285	Y										
rps14				1	194	Y												1	0	1		
Total	102		5 97	52		1 51	28		2 26	78		7 71	108		18 90	368	33 335				
Percentage			5 95			2 98			7 93			9 91			17 83		9 91				

indicate the number of sites found in each of the categories representing the effect of the *rip1* mutation for each mitochondrial gene, and position indicate the position of the site relative to the start codon.
S and NS refer to the class to which each site belongs, silent or non-silent respectively.

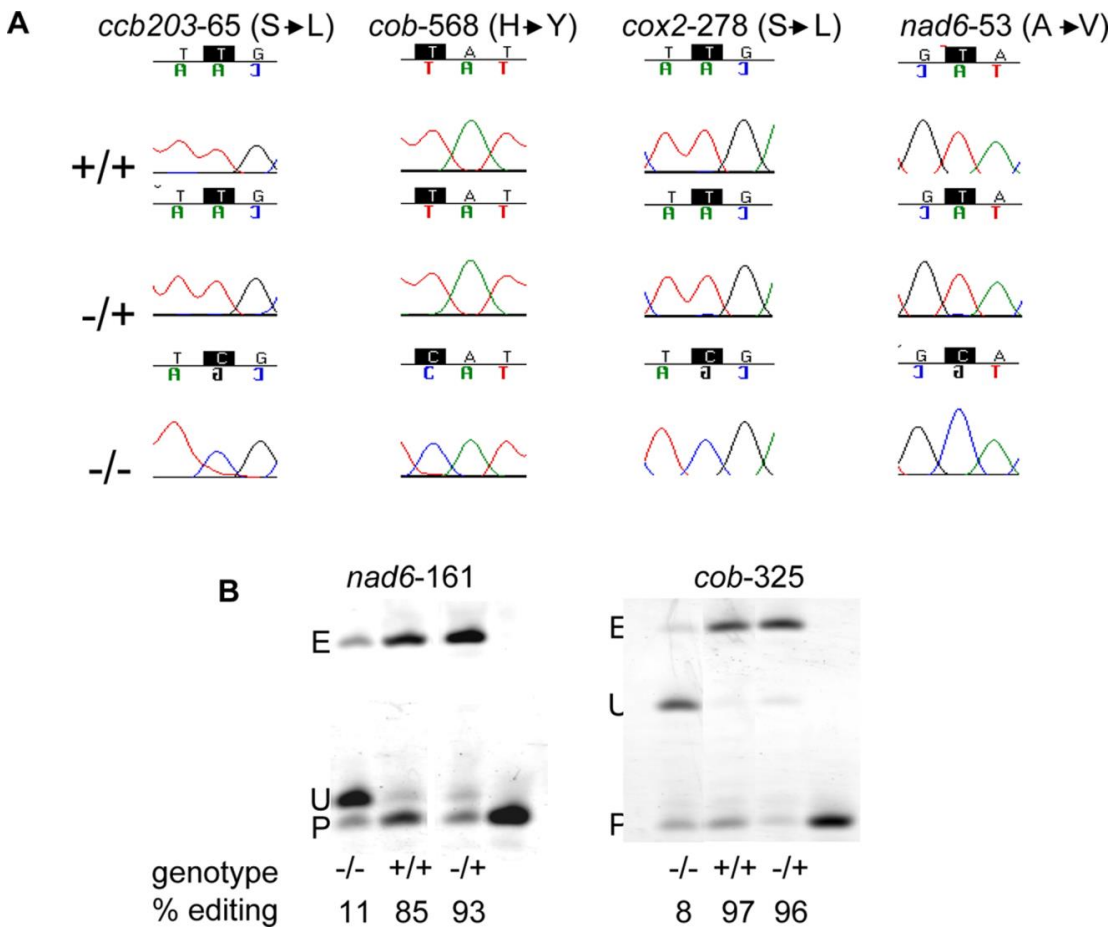


Figure 2.11. *rip1* mutation is recessive in its effect on mitochondrial editing extent. (A) Bulk sequencing electrophoretograms following RT-PCR are shown for 4 sites in different mitochondrial transcripts and for the 3 genotypic classes, top row: homozygous wild-type (+/+), middle row: heterozygous (-/+), bottom row: homozygous mutant (-/-). Above the electrophoretograms is given the name of the editing site (the position of the site is given after the name of the transcript to which it belongs) followed by the aa change upon editing. The edited position is highlighted by a black shade. No difference can be detected between the electrophoretograms of the RT-PCR products derived from wild-type homozygous (+/+) and the heterozygous (-/+) plants. (B) PPE assay confirms *rip1* mutation to be recessive in its effect on mitochondrial site editing extent. No significant difference is found between the editing extent of heterozygous and homozygous wild siblings for sites *nad6*-161 and *cob*-325, two sites that show a very strong reduction of editing extent in the homozygous mutant. P: primer, U: unedited, E: edited

***rip1* Mutation Affects Transcript Abundance.**

We examined the level of a selection of mitochondrial transcripts in the mutant to investigate the possibility of a link between the steady state level of mitochondrial transcripts and the editing extent of their targeted C sites. Among the 10 mitochondrial transcripts assayed by quantitative RT-PCR, 5 transcripts showed a significant increase (approximately four-to- sixfold) in their abundance compared to the WT; 3 transcripts showed a moderate increase (1.3- to 1.5-fold) in the mutant; and 2 transcripts were in similar amount in both the mutant and the WT (Figure2.12). Although *ccmB*, the transcript harboring the highest number of sites whose editing extents are severely affected by *rip1* mutation (Table 2.1), shows the highest increase of transcript abundance in the mutant (Figure 2.12), there is no obvious correlation between the steady state level of transcript and the incidence of *rip1* mutation on the editing extent. For example, *nad9*, which exhibits a similar increase of its transcript abundance in the mutant as *ccmB* (Figure 2.12), does not harbor any site with editing extent that is greatly impaired in the mutant (Table 2.1). Conversely, *ccmFn-2*, whose 4 of 10 targeted sites show an apparent total loss of editing in *rip1* (Table 2.1) experiences only a slight increase of its transcript abundance in the mutant relative to the WT (Figure 2.12C). Eight of the sites on *nad7* transcript show a reduced editing extent in *rip1*, whereas *nad7* abundance is similar to the WT (Figure 2.12D).
A model can be proposed in which a dosage effect is transcript-specific, and therefore. a slight increase of *ccmB* transcript abundance is sufficient to have an effect on the editing extent of some of its sites. In this model, some of the recognition trans-factors directing the specific editing site of targeted C sites, are in limiting amounts, and

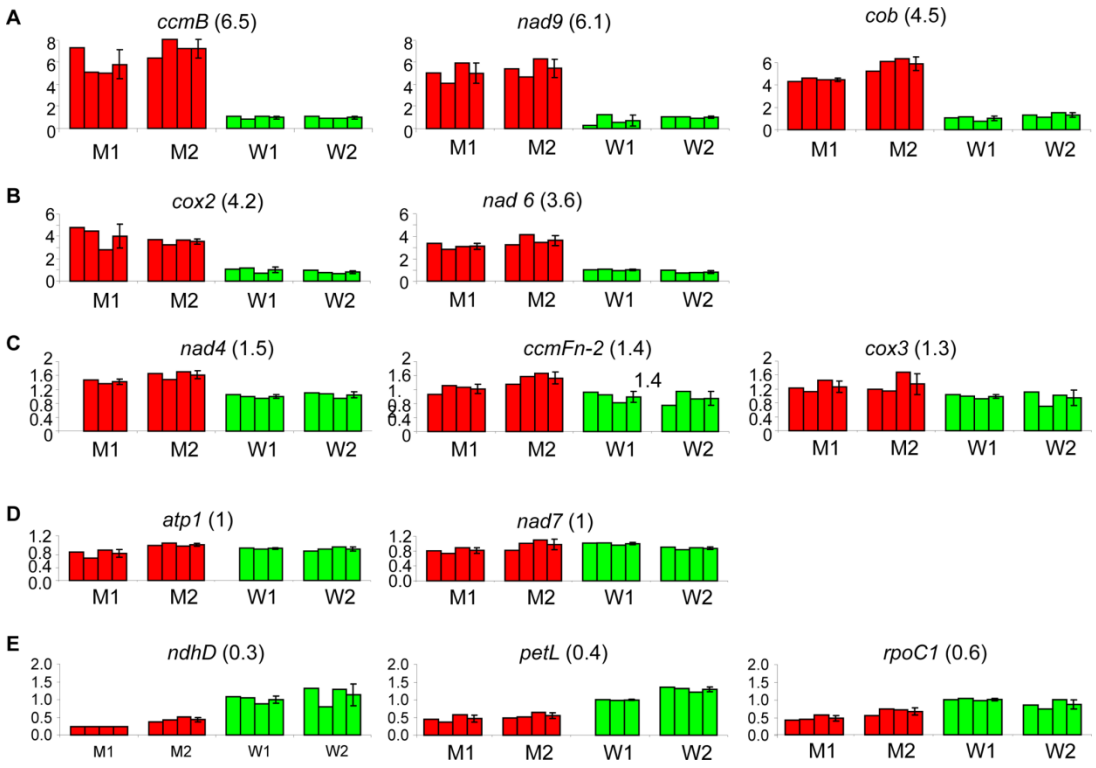


Figure 2.12. Transcript abundance is differentially affected in *rip1* organelles. (A-D) qRT-PCR shows a variable steady state level of mitochondrial transcripts in *rip1* (M1, M2*)* relative to wild-type (W1, W2). The level of *rip1* mitochondrial transcripts is markedly increased (A-B), slightly increased (C) or unchanged (D) when compared to wild-type. (E) By contrast to mitochondrial transcripts, qRT-PCR indicates a significantly decreased amount of transcript in *rip1* for plastid genes *ndhD*, *petL* and *rpoC1*. The values are means of three replicates normalized to W1, with error bars representing S.D. The number in parenthesis close to the gene refers to the ratio of transcript expression (M/W).

therefore, even a slight increase of the transcript abundance might deplete these recognition factors resulting in an apparent reduction in editing efficiency. However this possibility is refuted by a close examination of the *nad4*, *ccmFn-2*, and *cox3* transcript abundances, which show a very similar and slight increase in the *rip1* mutant (Figure 2.12C). These three transcripts each possess a site, *cox3*-422, *nad4*-124 and *ccb203* (*ccmFn-2*)--344, recognized by the same recognition factor, MEF11 (27). In the *rip1*mutant, the editing extent of *ccmFn-2*-344 is severely reduced, whereas the

editing extent of *cox3*-422 is only slightly reduced; *nad4*-124 does not show any detectable difference in editing extent between *rip1* and the WT (Table 2.4). In conclusion, our mutant analysis data clearly indicate independence of the editing extent of the sites carried by a mitochondrial transcript and its abundance.

Unlike the mitochondrial transcripts, the three plastid transcripts assayed by quantitative RT-PCR all show a reduction of steady state level in *rip1* compared with WT (Figure 2.12E). Similar to editing of mitochondrial transcripts, there is no clear connection between the editing extent of plastid sites and the amount of transcript that carries these sites. The *ndhD* and *petL* sites show a decrease of editing extent in *rip1*, whereas *rpoC*1-488 editing extent is significantly increased in the mutant (Table 2.2).

Editing Defects in the *rip1* Mutant Differ from the Minor Defects Seen in Other Types of Mutants

We investigated organelle editing in two mutants that mimic the growth phenotype of the *rip1* mutant and are compromised in some aspects of organelle RNA metabolism or organelle biogenesis. Mutant tissue was available in the chloroplast polynucleotide phosphorylase, which has a major role in maturing mRNA and rRNA 3′-ends, but also participates in RNA degradation through exonucleolytic digestion and polyadenylation (41). We obtained a second mutant that was affected in the gene encoding a chloroplast envelope membrane protein containing a putative LrgB domain, which has been reported recently to play an important role in *Arabidopsis thaliana* chloroplast development (42). Examination of the editing extent in two null mutants of the genes encoding these plastid proteins shows no difference from the WT in the editing extent of *nad6*-161 and *cob*-325, two mitochondrial sites that show a

drastic reduction of editing extent in *rip1* (Figure2.13). Among the five plastid sites whose editing extent showed the largest variation in the *rip1* mutant, we observed only an increase of editing extent in the null *pnp* and *lrgB* mutants, for certain sites. *NdhD*-2, with an editing extent in the *rip1* mutant that is about one-half the amount observed in WT (Table 2.2), shows an increase of editing extent in both mutants (110% and 40% in *pnp* and *lrgB* mutants, respectively) (Figure 2.13). *AccD*-794 and *petL*-5 exhibit an increase of editing extent only in the *pnp* mutant; whereas *rpoC1*-488 editing extent is markedly increased only in the *lrgB* mutant. *Rps12*-158 editing extent is invariant in both the *pnp* and *lrgB* plants.

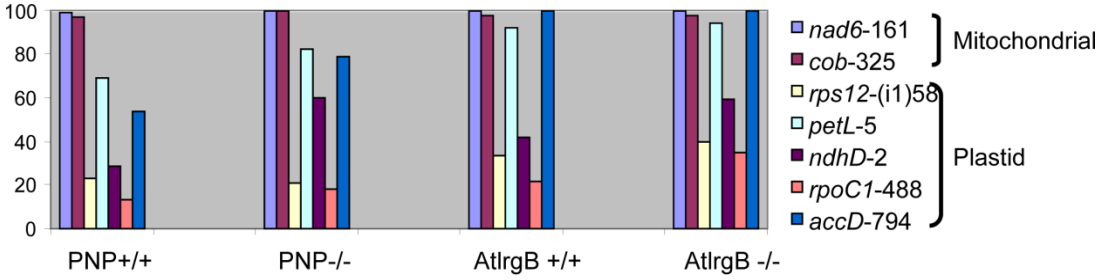


Figure 2.13. Editing extent of some plastid sites show an increase in *pnp* and *lrgB* null mutants that are impaired in plastid RNA metabolism and plastid biogenesis, respectively. PPE assay reveals an invariant level of editing extent at two mitochondrial sites in *pnp* and *lrgB* mutants, while plastid *ndhD-2* shows a marked increase of editing extent in both mutants. Editing extents of *petL*-5 and *accD*-794 are increased only in the *pnp* mutant, whereas *rpoC1*-488 editing extent increases only in the *lrgB* mutant.

Virus-induced Gene Silencing (VIGS) of *RIP1* Affects Chloroplast and Mitochondrial Editing Efficiency.

Because additional mutant lines with a second independent T-DNA insertion in *RIP1* were not available, we silenced *RIP1* by virus-induced gene silencing (VIGS) to confirm that the effect on RNA editing was specifically attributable to a defective

RIP1 gene. Two types of control plants were used in this experiment, uninoculated plants and plants inoculated with a silencing vector containing only GFP. Quantitative RT-PCR showed that the level of *RIP1* transcript in *RIP1*-silenced plants was reduced to 38% of the level detected in uninoculated plants (Figure 2.14). Unexpectedly, the level of *RIP1* transcript in GFP-silenced plants was increased to about two times the level in uninoculated plants. Both *RIP1*- and GFP-silenced plants show a significant reduction of GFP transcript compared with the uninoculated plants (87% and 95%, respectively) (Figure 2.14).

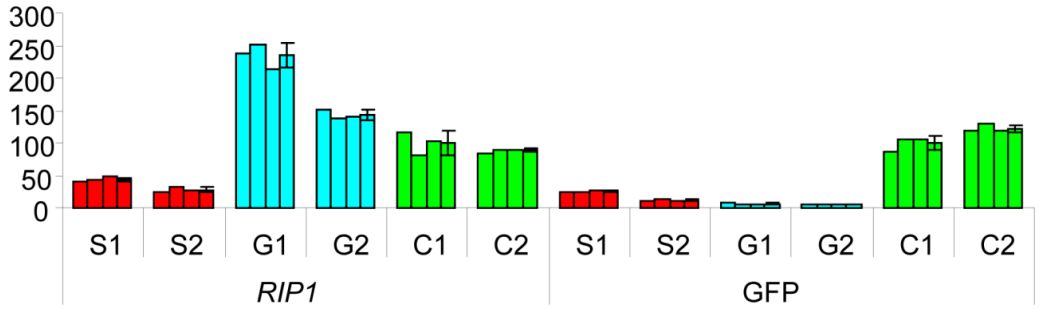


Figure 2.14. *RIP1* silencing results in a significant decrease of *RIP1* transcript level. Virus induced gene silencing (VIGS) of *RIP1* results in a significant decrease of *RIP1* transcript level in silenced plants (S) relative to control plants (G, C). Quantitative RT-PCR measured the level of *RIP1* and GFP expression in two *RIP1*-silenced plants (S1, S2), two GFP-silenced plants (G1, G2) and two uninoculated plants (C1, C2). The level of RIP1 and GFP cDNAs was arbitrarily fixed at 100 for C1. GFP is used as a marker for silencing in VIGS experiment and as such is significantly decreased in both *RIP1* silenced and GFP silenced plants. *RIP1* transcript level is significantly decreased only in *RIP1* silenced plants.

PPE assays on transcripts from uninoculated and silenced plants showed that silencing of *RIP1* results in an average 18 % decrease in *petL*-5 editing extent and a 24% increase in *rps12*-(i1)-58 editing extent compared with the uninoculated plants ($P<0.01$ and $P<0.05$, respectively) (Figure 2.15). A decrease in *petL*-5 editing and an

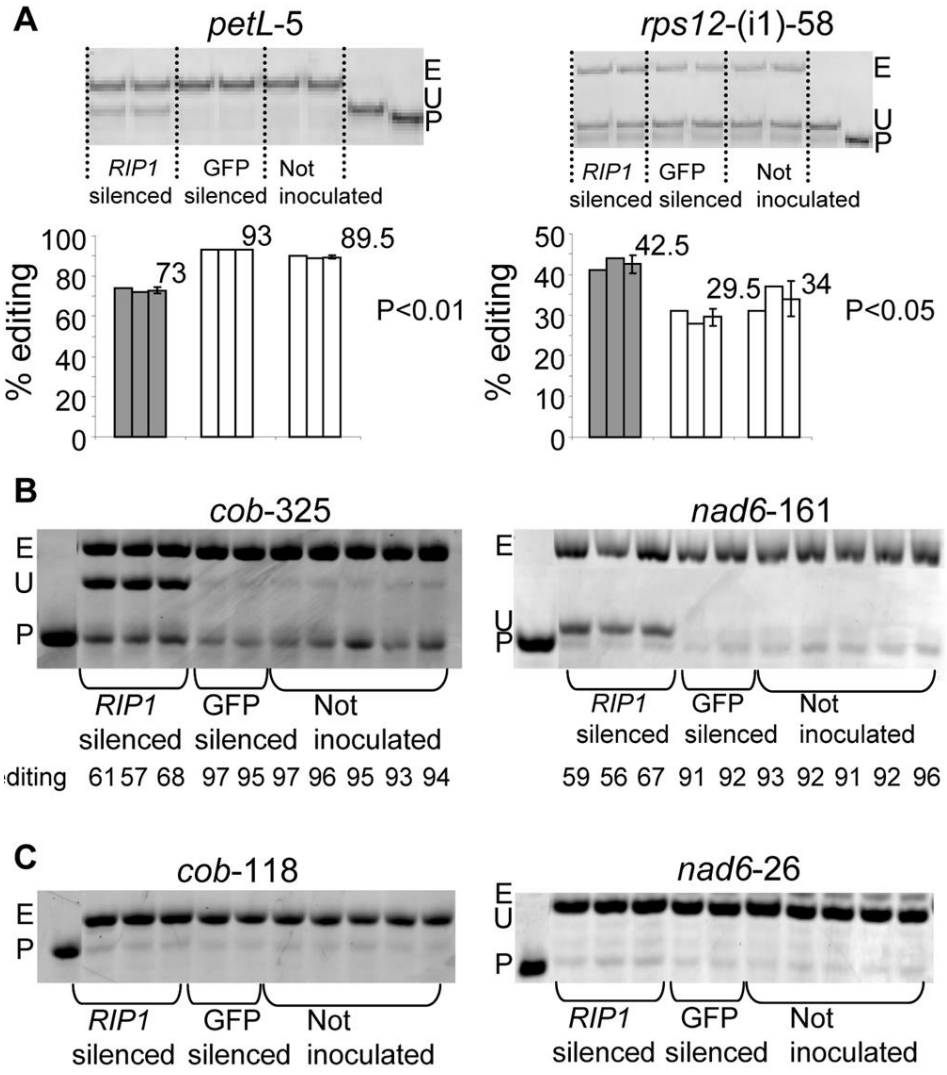


Figure 2.15. *RIP1* silencing recapitulates the effect of *rip1* mutation on editing extent of organelle sites. (A) PPE assays on plastid sites with quantification of the editing extent represented by a bar below each lane. The average is given with standard deviations on the right of each group of plants: *RIP1*-silenced, and two sets of controls, GFP-silenced and uninoculated plants. *petL-5* and *rps12*-intron C targets were chosen for assay because they exhibit reduction and increase, respectively, in the *rip1* T-DNA mutants. (B) PPE assays on mitochondrial sites *cob*-325 and *nad6*-161 in *RIP1* silenced plants compared to the two sets of controls. *Cob*-325 and *nad6*-161 are C targets that also show a very strong reduction of editing extent in *rip1* mutants. (C) *RIP1* silencing does not induce any change in the editing extent of *cob*-118 and *nad6*-26, two sites whose editing extent was also not affected in the *RIP1* mutant. P: primer, U: unedited, E: edited. (The unedited band is not detectable on the *cob*-118 PPE gel).

increase in *rps12*-(i1)-58 editing was likewise observed in the T-DNA mutant (Table 2.2), although, as expected, the effect of *rip1* knockdown in silenced plants is less than in the mutant. No significant difference in editing extent was detected between the GFP-silenced and uninoculated plants (Figure 2.15).The editing extent for *rps12*-(i1)-58 in the uninoculated WT siblings was 16% (Figure 2.7) whereas the uninoculated GFP plants exhibited a 34% editing efficiency (Figure 2.15). This discrepancy results from the fact that the WT siblings of the mutant are in Wassilewskija background whereas our Arabidopsis GFP line used for VIGS is in Columbia background. PPE assay on mitochondrial transcripts from uninoculated and silenced plants also confirms the variation in mitochondrial editing extent observed in the T-DNA mutant. *Cob*-325 and *nad6*-161 exhibit a very significant reduction of editing extent in *RIP1*-silenced plants compared with uninoculated control plants [35% ($P<10^{-4}$) and 34% ($P<10^{-4}$), respectively] (Figure 2.15B). As expected, *cob*-118 and *nad6*-26, two C targets with editing efficiency that are not affected in the *rip1* mutant, do not show any decrease of editing extent in *RIP1*-silenced plants (Figure 2.15C). Not all C targets with editing that was greatly impaired editing in the *rip1* mutant exhibited detectable reduction in the silenced plants. For example, no effect on *nad6*-88 and *nad6*-89 editing extent was detected in knockdown plants although they were strongly affected in the mutant (Figure 2.16). There may be sufficient RIP1 present in the silenced plants to allow editing of some sites to proceed normally.

Transient Expression of *RIP1* in the *rip1* Mutant Seedling Partially Complements the Defect in Mitochondrial and Chloroplast RNA Editing.

We used a transient transformation system in which an Agrobacterium strain is vacuum infiltrated directly into young Arabidopsis seedlings (43). *rip1* homozygous mutant seedlings were infiltrated with an Agrobacterium strain carrying a binary vector containing a 35S-RIP1 construct. Three days post-infiltration, PPE analysis showed that editing extent in the mitochondrial *nad6*-161 and *cob*-325 and plastid *accD*-794 sites was significantly increased in transfected vs. not transfected seedlings [80%, 86%(P<0.01), and 17% (P< 0.02), respectively] (Figure 2.17). This experiment provides additional evidence for the necessity of functional *RIP1* for efficient mitochondrial and plastid RNA editing.

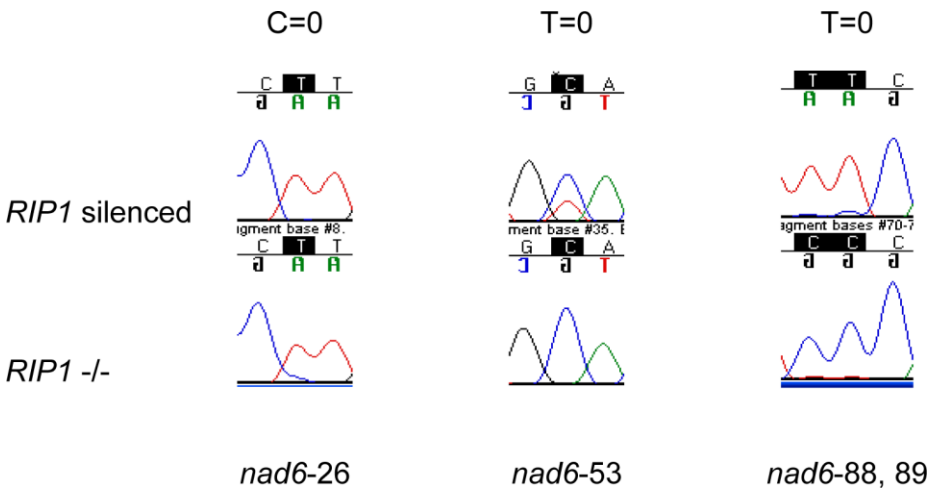


Figure 2.16. *RIP1* silencing affects only sites exhibiting a strong *RIP1* dependence in the mutant. *RIP1* silencing affects only the editing extent of sites showing a strong *RIP1* dependence in the mutant (T=0), but the reverse is not true; some sites showing strong *RIP1* dependence are not affected by *RIP1* silencing. On the left is shown an electrophoretogram of a *RIP1*-independent site whose editing extent is not affected by *RIP1* mutation nor by *RIP1* silencing. In the middle is an electrophoretogram of a *RIP1* dependent site whose editing extent is both affected by *RIP1* mutation and silencing. On the right the electrophoretogram of two *RIP1* dependent sites, *nad6*-88 and 89, show no detectable reduction of editing extent in a *RIP1* silenced plant but still exhibit a total loss of editing in the mutant.

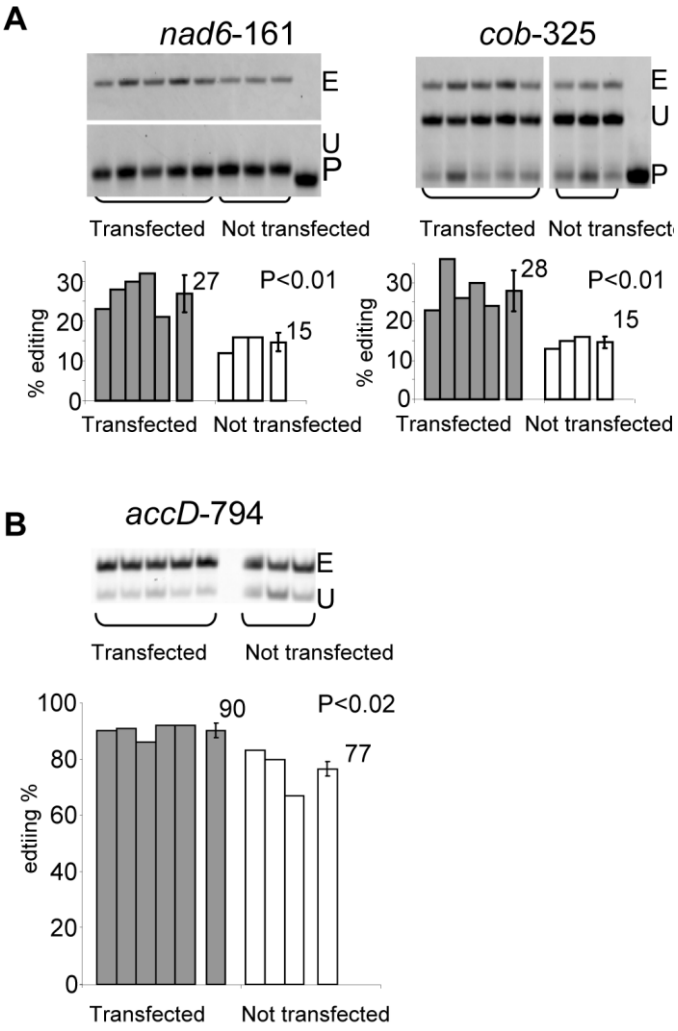


Figure 2.17. Transfection of *rip1* mutant with a wild-type version of *RIP1* partially complements the editing defect in both organelles. (A) Transfection increases editing extent of mitochondrial *nad6*-161 and *cob*-325. (B) Transfection increases editing extent of plastidl *accD*-794. On top of each panel are shown the PPE products obtained from plants either transfected or not with a construct containing a functional copy of *RIP1* under the control of a 35S promoter. Below the gels are graphs depicting the quantification of editing extent for each lane; on the right of each group the average is shown with s.d. P: primer, U: unedited, E: edited

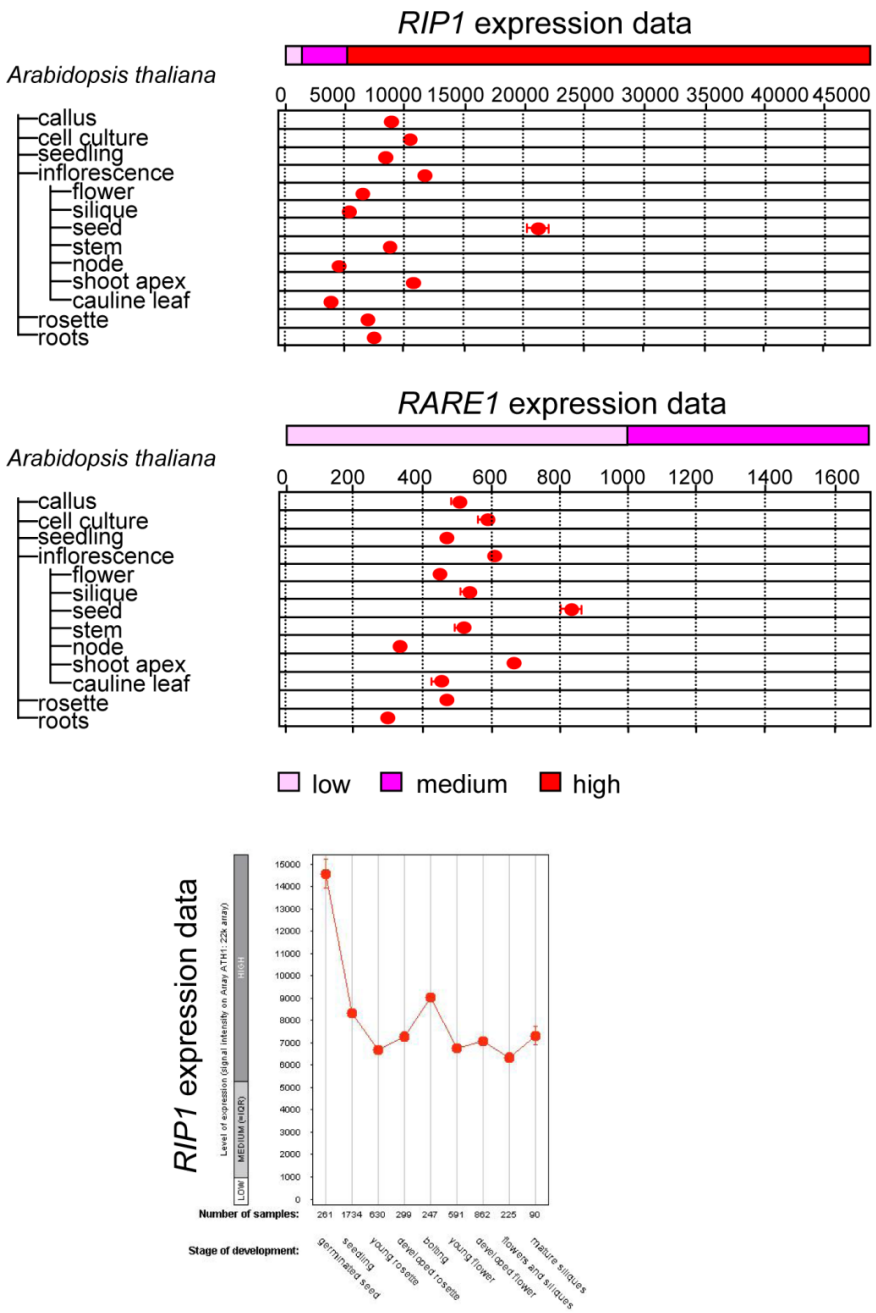


Figure 2.18. *RIP1* is highly expressed throughout plant development. Expression data were obtained from Genevestigator, a database and web-browser data mining interface for Affymetrix GeneChip data (https://www.genevestigator.ethz.ch). Although their gene products interact, *RIP1* is expressed at much higher level than *RARE1*.

Discussion

Unlike other proteins known to affect editing efficiency at specific C targets of organelle transcripts, RIP1 is not a member of the PPR family. RIP1 affects editing efficiency of hundreds of C targets, whereas known PPR proteins specify editing at one to seven C targets on either chloroplast or mitochondrial transcripts. The high level of *RIP1* transcript expression in both green and non-green tissues and across many developmental stages is consistent with an essential function for RIP1 in a variety of plastid types as well as mitochondria (Figure 2.18). The presence of homologs to the RIP1 family members in monocot, such as rice and maize, shows its evolution before the split between eudicot and monocots and emphasizes the importance of this family in plant biology. The disparity in level of transcript expression of RIP1 vs. a typical PPR protein may be explained if RIP1 is required in many editing complexes, whereas PPR proteins are needed only for recognition of a small number of target transcripts.

RIP1 has been observed to affect both chloroplast and mitochondrial editing, whereas PPR protein editing factors have been shown to affect editing in either chloroplasts or mitochondria, but not both organelles. Recently, OTP87, a PPR-E editing factor, was shown to be dual-targeted to mitochondria and chloroplasts; however, although mutation of *OTP87* was shown to affect mitochondrial editing, no effect on chloroplast editing extent in the mutant was reported (29). RIP1 is among about 50 dual- targeted proteins in plants reported to date; 16 of these proteins are amino acyl-tRNA syntetases (reviewed in (44)). One theory holds that dual targeting may comprise a means of inter-organelle communication: sending the same protein to

both organelles may ensure that the protein's activity occurs in a coordinated manner (44).

RIP1 belongs to a small family of Arabidopsis proteins that contains 10 members, all of which are predicted to be targeted to chloroplast or mitochondria. RIP1 has been annotated as "similar to DAG protein," which refers to the *Antirhinnum majus* protein, DAG; it was shown to affect expression of the *rpoB*, encoding a subunit of the chloroplast RNA polymerase, affect accumulation of plastid-targeted nuclear gene products targeted to plastids, and control chloroplast development at a very early stage (32). The molecular function of DAG was not investigated in the report of its discovery. The only member of the Arabidopsis protein family characterized to date is DAG-LIKE 1, AT2G33430.1, mutants of which have a yellow phenotype and have been shown to have defects in chloroplast *rrn* operon processing (33). The possible involvement of other members of the RIP1 family in RNA editing obviously merits experimental investigation.

RIP1 is a Positive and Negative Regulator of Plastid RNA editing.

RIP1 controls RNA editing of 14 of 34 Arabidopsis editing sites in the chloroplast sites (Table 2.2). Editing extent of these 14 sites was significantly altered in the *rip1* T-DNA insertional mutant (Figure 2.7, Table 2.2), but editing of most chloroplast sites was unaffected. Because there are fewer plastid sites than mitochondrial sites, the amount of RIP1 in the mutant may still be sufficient to accomplish much of its editing function in the plastid, especially if RIP1 import into chloroplasts is more efficient than entry into mitochondria.

Three plastid sites exhibited an increase of editing extent in the *rip1* T-DNA insertional mutant (Table 2.2). This unexpected effect of the *rip1* mutation was also confirmed for *rps12*-(i1)-58 in *RIP1*-silenced plants (Figure 2.15A). The increased rate of editing of *rps12*-(i1)-58 in *RIP1*-silenced plants was less pronounced than in *rip1* mutant plants. The expression of *RIP1* is expected to be less affected in a silenced plant than a homozygous mutant. In addition to *RIP1*, mutations in two PPR protein-encoding genes have been shown to increase editing extent. Mutation of *REME1*, which encodes a PPR-DYW protein, negatively affects editing of both *nad2*-558 and *orfX*-552, but also increases editing extent in at least two sites, *matR*-1771 and *rpl5*-92 (23). A null mutant of *PPR596*, which encodes a PPR-P protein, showed an increase of editing extent in several sites in the *rps4* transcript (45). Site-specific inhibition of editing by protein factors has previously been reported in the apolipoprotein B (*apoB)* RNA editing system in mammals. Antisense inhibition of expression of GRY-RBP or CUGBP2 in McA cells led to a 2 to 3-fold increase in endogenous *apo*B RNA editing, suggesting that both these factors may participate in the *apoB* editing complex as negative regulators *in vivo* (46, 47). In contrast to these *apoB* factors, RIP1, is able to promote editing at many sites while inhibiting editing at a few other sites.

Examination of editing extent in two mutants impaired in either plastid RNA metabolism or plastid biogenesis suggests that the increase in editing observed for some plastid sites in *rip1* might be an indirect effect of the mutation (Figure 2.13 and Table 2.2). Another indication that plastid editing activity might be indirectly compromised by *RIP1* mutation is the observation that plastid-encoded polymerase (PEP) transcripts (e.g. *petL*, *ndhB*) show decreased editing whereas most nucleus-encoded polymerase (NEP) transcripts (e.g. *rpoA*, *rpoB*, *rpoC*) show increased editing

(Table 2.2). However, If the plastid editing effect observed in the *rip1* mutant was indirectly caused by PEP dysfunction, we would expect to observe increase in transcripts abundance generated by NEP which is generally observed for mutants impaired in PEP activity (e.g.(48)). Our data clearly disprove this model, because the *rpoC1* (NEP-generated) transcript level is reduced in *rip1*, as well as levels of PEP transcripts *ndhD* and *petL* (Figure 2.12). In addition, the *accD* (NEP-generated) transcript exhibits a reduction of editing extent at site 794 in *rip1*, which is unlike other NEP transcripts (Table 2.2). More importantly, the current view that genes of photosystems I and II are completely dependent on PEP transcription, and a few housekeeping genes, including the *rpoB* operon, are transcribed exclusively from NEP promoters, has been recently challenged in a study of the barley chloroplast transcriptome (49). In this study, which included a PEP-lacking plastid mutant, Zhelyazkova et al.(49) observed that most genes, including genes coding for photosynthesis proteins, have both NEP and PEP promoters.

It remains possible that many of the minor alterations in editing extent of plastid and mitochondrial sites could possibly be caused by indirect effects on transcript or trans-factor abundance. For example, altered organelle metabolism could potentially affect the abundance of particular PPR protein editing factors and thus result in minor differences in editing extent at specific sites. However, we have also shown, in the case of the editing trans-factor MEF11, that the sites on which it operates are differentially affected in the *rip1* mutant.

RIP1 and RARE1 Function in an Editing Complex

The current model for RNA editing holds that PPR proteins recognize cis-elements near C targets of editing, serving as molecular adaptors that recruit an editing activity to specific transcripts (50). CRR4, a PPR-E protein necessary for the editing of the plastid *ndhD* transcript, directly interacts with the transcript in the area surrounding the targeted C for editing (12). Two other PPR editing *trans*-factors, the *Physcomitrella patens* PpPPR_71 and *A. thaliana* OTP87 have been shown to bind the RNA sequence surrounding their target editing sites (51, 29). However, the identity of proteins that act in conjunction with PPR proteins to convert C targets to U targets is unknown.

Deamination is the favored process to explain the base modification because the sugar phosphate backbone and the nucleotide base are retained during C to U conversion (52). The DYW domain found in about half of the Arabidopsis PPR editing factors has been suggested to be catalyzing the editing activity based on the similarity of one of its motifs to the conserved cytidine deaminase motif (30). In addition, the phylogenetic distribution of the DYW domain in plant taxa is strictly correlated with RNA editing (30). However, about one-half of the Arabidopsis editing factors lack the DYW domain; moreover mutant complementation experiments with a truncated DYW protein has proved that the DYW domain is dispensable for editing (53). More importantly, *in vitro* experiments with recombinant DYW proteins failed to detect any RNA editing activity (53, 54). Examination of the RIP1 sequence for motifs with a web-based tool (http://www.genome.jp/tools/motif/) did not detect any known motif; in particular, the conserved cytidine deaminase motif is absent from RIP1. However, it remains possible that a complex of RIP1 and one or more PPR proteins could constitute an editing activity.

The co-immunoprecipitation and yeast two-hybrid interaction of RIP1 and RARE1 indicates that they are present in the same protein complex in chloroplasts. The C-terminal PPR repeats are not needed for interaction with RIP1, and the C-terminal portion of RIP1 is dispensable for interaction with RARE1 in yeast two-hybrid analysis. Perhaps the C-terminal PPR repeats are involved in interaction with RNA cis-elements, whereas the N-terminal repeats mediate interaction with RIP1. Supporting this model is the recent finding that two PPR motifs are sufficient to bind to an RNA target *in vitro* (55).

RIP1 Controls the Editing Extent of Many More Mitochondrial C Targets than any identified PPR Protein Editing Factor.

The editing extent of 33 mitochondrial genes in the *RIP1* T-DNA homozygous mutant exhibited a variable decrease, from undetectable editing for 108 sites by bulk sequencing to a mild decrease in 52 sites (Table 2.1). The intermediate level of decrease in editing in the mutant ranges from severe in 78 sites to moderate in 28 sites (Table 2.1). Thus the number of mitochondrial sites with editing that is affected by RIP1 equals 266, and represents roughly 70% of the sites assayed. A residual editing extent was detected by the sensitive PPE method at two sites, *cob*-325 and *nad6*-161, although no editing was detected by bulk sequencing (Figure 2.10C). To study the function of *RIP1,* we have used a hypomorphic allele with an upstream insertion that likely allows accumulation of some active RIP1 protein. A low level of editing at most affected C targets may explain why the *rip1* homozygous mutant is viable.

Examination of transcript level by quantitative RT-PCR clearly shows that the changes in mitochondrial editing observed in *rip1* cannot be trivially explained by

changes in RNA abundance. Mutation in *RIP1* has a generally positive effect on mitochondrial transcript levels as previously observed in respiratory mutants (24, 39).

All of the currently identified Arabidopsis mitochondrial PPR protein editing factors belong to the 152-member PLS subfamily, of which 65 members contain only the C-terminal E domain and 87 members exhibit both the E and DYW domains (14). Approximately 2/3, or about 100 of these proteins are predicted to be targeted to mitochondria. Among the 13 Arabidopsis members of the PLS subfamily reported to be mitochondrial editing factors (22-29), only five members have been observed to control the editing extent at more than one site. MEF1 and MEF11 control the editing of three sites (22, 27), whereas REME1, SLO1 and OTP87 control two sites each (23, 24, 29). Whether 100 PPR proteins are sufficient to recognize the over 500 C targets of editing in Arabidopsis mitochondrial is presently unknown. Some PPR proteins are known to have roles in other aspects of RNA metabolism instead of editing (56-60).

Although mutation of *RIP1* negatively affects the editing extent of a large number of mitochondrial C targets, editing of some C targets was not affected at all. It is possible that RIP1 interacts with only a subset of PPR proteins that interact with target RNAs, whereas other members of the RIP1 family interact with a different subset of PPR proteins to stimulate editing at other C targets. The discovery of the important role of RIP1 in mitochondrial editing will open new inquiry into the functions of its 10-member gene family.

Table 2.5. Oligonucleotides used in this study

Name	Sequence (5' to 3')	Purpose
Rare1-159F	CCCGAATTCCCACTATTGATCATTATGATTGT	159 aa F
Rare1-159/194R	CGAGTCGAGGTCAATCAAGAAGCTGTTCTCTTCT	159/194 aa R
Rare1-194F	5'CCCGAATTCCACTATTGATCATTATGATTGT	194 aa F
Rare1_+100F	GGATCCATGTCGAGCACTTCTTCTCCGTCT	Δ 33 aa cTP
3XFLAG-StrepIIF1	TAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGTCGGCGCCGGTT	PCR1 F
3XFLAG-StrepIIF2	CCCGGGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATT	PCR2 F
3XFLAG-StrepIIR	TGACAAGGTCGGCGCCGGTTGGTCTCATCCTCAATTTGAAAAATAAGAGCTC	PCR1/2 R
Rare1F	TCCATCAACTATGACGATTCTCACTGT	Full-length F
Rare1_+2259R	TCACCAGTAATCGTTGCAAGAACA	Full-length R
L5-Rare1_+2256R	ACCTCCACCAGATCCCCAGTAATCGTTGCAAGAAC	L5-3FS fusion
Rare1_-311F	GCCGCCATTTGAGAGGAGG	Native promoter
Rare1_+1933F	CACCATCCACAAACTCAGGAG	genotyping
At3g15000_-442F	GTCACACATTTTCACCAAATTGACC	genotyping
At3g15000_+99R	GGCGAGAGGAGCAGATGAAG	genotyping
FLAG_LB4	CGTGTGCCAGGTGCCCACGGAATAGT	genotyping
At3g15000_+856F	GGTAGTTGCTTTGCTCGTCC	genotyping
At3g15000_+1334R	GGCCTCCTGCCATGTTCT	genotyping
FLAG_Tag3	CTGATACCAGACGTTGCCCGCATAA	genotyping
At3g15000_VF	ACCCCCACAGAACAACAA	Rip1-VIGS F
At3g15000_VR	AATCCCGTTTAATGCAGAA	Rip1-VIGS R
At3g15000_+169F	ATGGGCGGCCTTGTGTCTGTC	Δ 56 aa cTP
At3g15000_+1F	ATGGCGACGCATACCATTTCTCG	Full-length F
At3g15000_+1188R	TTAACCCTGGTAGGGGTTGCC	Full-length R
At3g15000_+1185R	ACCACCACCAGAACCCTGGTAGGGGTTGCCACT	RIP1-stop
cob-118	GCAATCTTAGTTATTGGTGGGGGTTCGG	PPE
cob-325	TTGTGGTTTACCTTCATATTTTTCGTGGTC	PPE
nad6-26	CAACCATCAAACCAGAGACCAAAGC	PPE
nad6-161	GGTCTCGACTTCTTCGCTATGATCTTCC	PPE
RARE1_+1956R	CTAGATCTCCTGAGTTTGTGGATGGTG	Y2H
RARE1_+1933F	CACCATCCACAAACTCAGGAGATC	Y2H
RARE1_+1674R	CTAGCTCATTGCATCAGGTTAAAAGG	Y2H
RARE1_+1651F	CCTTTTGAACCTGATGCAATGAGC	Y2H
RARE1_+1050R	CTACCATGAAAATATCGGGGTTGTC	Y2H
RARE1_+1368R	CTAGTTTGGCTCACGGATTTCTTGAAATGC	Y2H
At3g15000_+702R	CTATCTTCTCCTCTCAAAGTTTCTGCT	Y2H
At3g15000_+703F	TCTTCTCCTCTCAAAGTTTCTGCT	Y2H
RIP1-F1	ATGGCGACGCATACCATTTCTCG	qRT-PCR
RIP1-R1	ACGCCGGAGATTTGGCGAGAG	qRT-PCR
RIP1-F2	ACCGGCGAAATCTCTTTCGTTTCT	qRT-PCR
RIP1-R2	ACAAGGCCGCCACGGAAAAC	qRT-PCR
RIP1-F3	GCTTTTGGGGCACTTGTGTCAGAA	qRT-PCR
RIP1-R3	CAGCCTTCCCATCGATGAAAGGTT	qRT-PCR
RIP1-F4	GGAGCACCCCCACAGAACAACAA	qRT-PCR
RIP1-R4	GTAGGGGTTGCCACTGCCATCC	qRT-PCR
At2g28390-F	AACTCTATGCAGCATTTGATCCACT	qRT-PCR
At2g28390-R	TGATTGCATATCTTTATCGCCATC	qRT-PCR
GFP-F	ATGGCCCTGTCCTTTTACCAGACA	qRT-PCR
GFP-R	TAATCCCAGCAGCTGTTACAAACTCAAG	qRT-PCR
atp1-F1	TCCCGCGGGAAAGGCTATGCT	qRT-PCR
atp1-R1	TCCCAGGGGCTTTCACTTCGACA	qRT-PCR
ccb203-F1	TCCGGATTGCTAGCTCCCGTTCAT	qRT-PCR
ccb203-R1	CTTCGCGCCACAACCATCTCTTTT	qRT-PCR
ccb206-F1	GATTCGGATCCCTCCGTTGTTTC	qRT-PCR

ccb206-R1	GAATAACCCGGTGACCCACCAA	qRT-PCR
cob-F1	TGGGGGTTCGGTCCGTTAGCT	qRT-PCR
cob-R1	GCAACCAGCCCCCTTCAACATC	qRT-PCR
cox2-F1	TACCCCGTCCCCATGGGCAATAGT	qRT-PCR
cox2-R1	AGTGGCGCCTAGCCGTTGAGAGC	qRT-PCR
cox3-F1	GTGGCGCGATGTTCTACGTGAAT	qRT-PCR
cox3-R1	TCTACCGCAGGTGCCAAAGAAGA	qRT-PCR
nad4-F1	TTTCGCCGTCAAAGTGCCTATG	qRT-PCR
nad4-R1	CGCTTCGGGAAACATGGGTATT	qRT-PCR
nad6-F1	TCGCGACACTTCAGGTTTACTTC	qRT-PCR
nad6-R1	TCTTCGTGAATCTCCGCTATTTG	qRT-PCR
nad7-F1	CCGGCAACCGTATCTGGAAACA	qRT-PCR
nad7-R1	TTCGCGAATCCCAGCATACCC	qRT-PCR
nad9-F1	TGCGGAGTTGATCATCCCTCTCGA	qRT-PCR
nad9-R1	CCGGCCGGCTGATGGAAATAGA	qRT-PCR
ndhD-F1	CAACATCTCCCGGTCAACGTAATT	qRT-PCR
ndhD-R1	CAGCGCCAATAAATCCATGAGAA	qRT-PCR
petL-F1	AAAAAAACATATTTTATTGAGTCCCTTCATG	qRT-PCR
petL-R1	GACCAATAAACAGAACTGAGGTTATAG	qRT-PCR
rpoC1-F1	GGGCGGGTGCTATCCGAGAAC	qRT-PCR
rpoC1-R1	TCCCCGTAGGCCCTTCTTCTCC	qRT-PCR

MATERIALS AND METHODS

Genotyping All genotyping was done by PCR with BioMix Red (Bioline, Taunton, MA). For amplification of RARE1 in transgenic plants, primer Rare1_+1933F and the 3xFLAG-StrepIIR primer were used. For genotyping of FLAG_150D11 line, the WT allele and T-DNA alleles were amplified with primer pairs At3g15000_-442F with At3g15000_+99R, or At3g15000_-442F with FLAG_LB4, respectively (Table 2.5). Likewise, for the FLAG_607H09 line, the primer pairs were At3g15000_+856F with At3g15000_+1334R and FLAG_Tag3 with At3g15000_+1334R. Both lines were obtained from the INRA FLAGdb T-DNA collection (34).

VIGS VIGS of At3g15000 using a GFP co-silencing marker as in refs. 21, and 61 was performed with CATMA primers (62) At3g15000_VF and At3g15000_VR (Table 2.5). Tissue was collected 18 days post inoculation.

Analysis of RNA Editing by PPE

All 34 known Arabidopsis chloroplast RNA editing C-targets (63, 64) were assayed as in ref. 21. Mitochondrial RNA editing sites were assayed by RT-PCR bulk sequencing using primers described in refs. 65 and 66. PPE analysis on mitochondrial sites *cob*-118, *cob*-325, *nad6*-26 and *nad6*-161 was conducted as in ref. 23 with primers cob-118, cob325, nad6-26 and nad6-161 (Table 2.5).

Transient Transformation of *rip1* Seedlings

Production of the binary vector: *RIP1* ORF was transferred from the gateway entry clone G67651 (ABRC, Ohio State University) into the binary vector pH7RWG2.0 (67) by recombination using LR Clonase II (Invitrogen). After sequence verification, the plasmid was transformed into *Agrobacterium tumefaciens* GV3101.

Transformation of rip1 seedlings: sterile seeds from *RIP1*- FLAG_150D11-T-DNA heterozygous plant were germinated on MS plates. After 2 weeks, the homozygous mutant plants, which were distinguishable from the other progeny because of their dwarf phenotype, were collected onto new MS plates and subjected to Agrobacterium infiltration according to the protocol in ref. 43. DNA and RNA were collected three days post-infiltration. DNA genotyping confirmed the visual assignment of the mutant seedlings based on the dwarf phenotype.

Subcellular Localization of RIP1

The *RIP1* ORF minus the stop codon was amplified from the clone G67651 with primers At3g15000_+1F and At3g15000_+1185R (Table 2.5), and cloned into pCR8/GW/TOPO (Invitrogen). After sequence verification, the insert was transferred

into the pEARLEY GATE103 vector (68) by recombination using LR Clonase II (Invitrogen), creating a RIP1-GFP fusion driven by the 35S promoter. Protoplasts from Arabidopsis Col-0 accession or *N. benthamiana* were transfected with the plasmid using the protocol in ref. 40. Protoplasts were checked for fluorescence under the confocal microscope 16 h after incubation with the plasmid. Protoplasts were incubated with MitoTracker Orange CM-H_2TMRos (Invitrogen) at a final concentration of 500nM for 30 min, centrifuged and resuspended in dye-free medium. Images were acquired using a Leica SP2 confocal microscope. For chloroplast nucleoid staining, *N. benthamiana* protoplasts were incubated with 3μg/ml DAPI (Sigma) for 5 min before visualized using a Zeiss 710 confocal microscope.

Real-time Quantitative RT-PCR Conditions and Analysis

Quantitative RT-PCR was performed as described in ref.69. All of the primers used for quantitative RT-PCR are given in Table 2.5. The results of the quantitative RT-PCR analysis were normalized using the gene At2g28390, which has been shown to be a superior reference gene for transcript normalization in Arabidopsis (70).

Generation of α-RARE1 Antibody

A 159 aa polypeptide spanning the last short PPR motif, the E domain and the beginning of the DYW domain of RARE1 (71), was expressed in *E. coli* strain Rosetta (DE3) (EMD Novagen, Madison WI) by cloning into vector pGEX-6p3. Primers Rare1-159F and Rare1-159/194R (Table 2.5) were used to amplify the fragment by PCR, which was cloned into vector pCR2.1/TOPO (Invitrogen, Carlsbad, CA), before subcloning the EcoRI-SalI fragment into pGEX6p3. Following sonic disruption of the

cells, the GST-RARE1 fusion protein was purified on Glutathione-Agarose (Sigma-Aldrich, St. Louis, MO) according to the manufacturer's recommended protocol, except after binding, RARE1 was proteolytically cleaved from GST using PreScission Protease (GE Healthcare, Piscataway, NJ). The eluted protein was used as an antigen for production of rabbit polyclonal antisera (PRFAL, Canadensis, PA). A 194 aa recombinant polypeptide, including the 159 aa region above, but with an additional PPR repeat on the N-terminus, was produced in a similar fashion, using instead as a forward primer Rare1-194F (Table 2.5). Immuno-affinity chromatography using the SulfoLink kit (Thermo Fisher Pierce, Rockford, IL) was used to purify α-RARE1 according to the manufacturer's recommended protocol.

Generation of Transgenic Plants Expressing Affinity-tagged RARE1

Transformation vector pBI121 (72) was modified to contain an affinity tag C-terminally fused to a Gateway cassette in place of the GusA gene. The affinity tag we used contains a sequence encoding the the 3xFLAG epitope (Sigma) 5' to a sequence encoding the StrepII epitope (IBA, St. Louis, MO) with a 4 aa V-G-A-G linker (73). Two rounds of PCR with overlapping primers were used to generate the fusion tag: first 3xFLAG-StrepIIF1 and 3xFLAG-StrepIIR and secondly with 3xFLAG-StrepIIF2 and 3xFLAG-StrepII R (Table 2.5). The resulting 117 nt fragment was cloned into pCR2.1/TOPO, and a SmaI-SacI fragment was used to replace the GusA of pBI121 cut with the same two enzymes. For overexpression (35S promoter) constructs, the GWb cassette (Invitrogen) was inserted at the SmaI site. For native promoter constructs, the CaMV 35S promoter was first removed using HindIII and XbaI before inserting the GWb cassette.

Full-length *RARE1* for overexpression was cloned by PCR using primers Rare1F and Rare1_+2259R for untagged constructs or L5- Rare1_+2256R for making fusion proteins with a 5 aa linker (L5) encoding G-S-G-G-G, which had been successfully used in (74). For native promoter constructs, 311 bp 5' of the start codon was amplified using Rare1_-311F in combination with the primers above (Table 2.5). All *RARE1* PCR products were cloned to pCR8/GW/TOPO (Invitrogen) and fragments were recombined into the modified pBI121 vectors above using LR Clonase II (Invitrogen). After sequence verification, the plasmids were transformed into *Agrobacterium tumefaciens* GV3101 and floral dip transformation of *rare1* homozygous mutants (WiscDsLox330H10) or (GABI_167A04) was performed as in (75). Transgenic plants were selected on MS agar plus 50 μg/ml kanamycin and 100 μg/ml carbenicillin. Sequencing of the transgene in the plants containing RARE1-3xFS under the control of the native promoter revealed that after the 3xFLAG sequences, a frameshift had occurred that affected the StrepII sequence. Thus only the FLAG sequences were used as an epitope tag.

Immunoblotting

10 or 12% Tris-Glycine (Protogel, National Diagnostics, Atlanta, GA), or 4-12% Bis-Tris NuPAGE (Invitrogen) polyacrylamide gels were used for SDS-PAGE (76). Proteins were electroblotted to nitrocellulose using a Mini-Protean II cell (BioRad, Hercules, CA), blocked with 5% powdered milk. When probed with α-RARE1 or α-Rubisco LSU (77) goat α-Rabbit IgG-HRP (GE Healthcare) secondary antibody was used for detection; otherwise, α-FLAG M2-HRP (Sigma-Aldrich) was used according to manufacturer's protocol.

Size Exclusion Chromatography

Stromal protein (0.5mg) was prepared as in (78), dialyzed against KEX buffer (30 mM HEPES-KOH, pH 8.0, 200 mM KOAc, 10 mM $Mg(OAc)_2$, and 5 mM DTT) (79), clarified by micro-centrifugation and 0.4 μm filtered before fractionation over Superdex-200 resin (GE Healthcare) with KEX buffer. Flow was maintained by use of a peristaltic pump and fractions of approximately 0.3 ml were collected. As KEX buffer was found to precipitate in 2X Laemmli sample buffer (76), protein from individual fractions was purified using the SDS-Page Sample Prep Kit (Thermo Fisher Pierce), and 50% of the indicated fractions were subjected to SDS-Page. Calibration of the Superdex column was performed with standards from Sigma MWGF1000 Kit, including carbonic anhydrase, bovine serum albumin, alcohol dehydrogenase, β-Amylase, apoferritin, thyroglobulin and Blue Dextran corresponding to 29, 66, 150, 200, 443, 669 and 2,000 kDa, respectively. Standards were run one at a time over the column, and protein concentration was measured by measuring absorbance at 260 nm. For size exclusion chromatography of 3xFLAG-tagged RARE1, the buffer used was RIPA (formulation in immunoprecipitation section), and 1 mg total leaf protein prepared in this buffer was fractionated.

Immunoprecipitation

For immunoprecipitation with the α-RARE1 antibody, the Dynabeads Protein-A Kit (Invitrogen) was used according to manufacturer's protocol. Antibody was crosslinked to the beads using 5 mM Bis(sulfosuccinimidyl)suberate (Thermo Fisher, Waltham, MA) prior to addition of 2 mg leaf extract per immunoprecipitation. Total leaf protein

extracts were prepared by powdering with a mortar and pestle in liquid nitrogen prior to extraction in RIPA lysis/binding buffer (50mM Tris-HCl, pH 7.4, 150mM NaCl, 1mM EDTA, 1% TritonX-100, 25 mM 2-mercaptoethanol and 1X Complete Protease Inhibitor Cocktail [Roche, Indianapolis, IN]) and subsequent pelleting of insoluble material by centrifugation. After washing with supplied Wash Buffer, the immunoprecipitate was eluted in NuPAGE LDS Sample Buffer plus Reducing Reagent (Invitrogen).

3xFLAG immunoprecipitation was performed as in (80), except α-FLAG M2 Magnetic Agarose (Sigma-Aldrich) was mixed with10 mg total leaf extract prepared as above (without 2-mercaptoethanol) and elution was done with 2 M $MgCl_2$, 50 mM Tris pH 8, 150 mM NaCl, and 0.5 % CHAPS (addition of CHAPS as in (81)). $MgCl_2$ concentration was reduced 3-fold by the addition of TBS, and proteins were precipitated by adding 3 volumes of acetone. Proteins were resuspended in 2X Laemmli sample buffer and were resolved by SDS-PAGE as above. Staining was performed with SilverSNAP (Thermo Fisher) or SyproRuby protein gel stain (Invitrogen).

Proteome Analysis by NanoLC-LTQ-Orbitrap Mass Spectrometry

Each gel lane was cut in seven slices. Proteins were digested with trypsin and the extracted peptides were analyzed by nanoLC-LTQ-Orbitrap mass spectrometry using data dependent acquisition and dynamic exclusion, as described in (82). Peak lists (mgf format) were generated using DTA supercharge (v1.19) software (http://msquant.sourceforge.net/) and searched with Mascot v2.2 (Matrix Science) against the Arabidopsis genome (ATH v8) supplemented with the plastid-encoded

proteins and mitochondrial-encoded proteins. Details for calibration and control of false positive rate can be found in (12). Mass spectrometry-based information of all identified proteins was extracted from the Mascot search pages and filtered for significance (*e.g.* minimum ion scores, etc), ambiguities and shared spectra as described in (82).

Protein-protein Interaction Verification *in vivo*

Yeast two-hybrid analysis was performed with the ProQuest Two-Hybrid System (Invitrogen), using Gateway-ready cDNA clone G67651 for At3g15000 obtained from the Arabidopsis Biological Resource Center (ABRC, The Ohio State University). Additionally At3g15000 was cloned without a putative transit peptide of 56 aa, using primers At3g15000_+169F and At3g15000_+1188R (Table 2.5). These clones were used for LR Clonase II recombination reactions with pDEST22, generating GAL4 transcriptional activation domain fusions with each. RARE1 without a putative transit peptide of 33 aa was cloned using RARE1_+100F and RARE1_+2259R primers (Table 2.5) and TOPO cloned in pCR8/GW/TOPO before recombination into pDEST32, thereby fusing it to the GAL4 DNA-binding domain. *Saccharomyces cerevisiae* strain Mav203 was transformed using the recommended protocol and transformants were selected on SD dropout media lacking leucine and tryptophan (Sunrise Science Products, San Diego, CA). The X-Gal reporter assay was done according to the suggested protocol.

ACKNOWLEDGMENTS

We thank Carol Bayles for instruction in confocal microscopy and Lin Lin for excellent technical assistance. We also thank Benoit Castandet for providing us with RNA from *pnp* and its WT sibling. This work was supported by National Science Foundation grants from Molecular and Cellular Bioscience, Genetics Mechanisms Program Grants 1020636 (to S.B.), 0716888 (to M.R.H.) and 0929423 (to M.R.H.) and Integrative Organismal Systems Grant 0922560 (to K.J.v.W.).

References

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

2. Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA Metabolism. *Annu Rev Plant Biol* 61:

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

4. Chateigner-Boutin AL, Small I (2010) Plant RNA editing. *RNA Biol* 7: 213-219.

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

6. Gualberto JM, Lamattina L, Bonnard G, Weil JH, Grienenberger JM (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341: 660-662.

7. Hiesel R, Wissinger B, Schuster W, Brennicke A (1989) RNA editing in plant mitochondria. *Science* 246: 1632-1634.

8. Bock R, Hermann M, Kossel H (1996) *In vivo* dissection of *cis*-acting determinants for plastid RNA editing. *EMBO J* 15: 5052-5059.

9. Chaudhuri S, 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. *EMBO J* 15: 5958-5964.

10. Hayes ML, Reed ML, Hegeman CE, Hanson MR (2006) Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript *in vivo* and *in vitro*. *Nucleic Acids Res* 34: 3742-3754.

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

12. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281: 37661-37667.

13. Small ID, Peeters N (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25: 46-47.

14. Lurin C, et al. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089-2103.

15. Okuda K, Myouga F, Motohashi R, Shinozaki K, Shikanai T (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci USA* 104: 8178-8183.

16. Chateigner-Boutin AL, et al. (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56: 590-602.

17. Zhou W, et al. (2008) The arabidopsis gene YS1 encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J* 58:82-96.

18. Cai W, et al. (2009) LPA66 is required for editing *psbF* chloroplast transcripts in arabidopsis. *Plant Physiol* 150: 1260-1271.

19. Hammani K, et al. (2009) A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21:3686-3699.

20. Okuda K, et al. (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. *Plant J*61:339-349.

21. Robbins JC, Heller WP, Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the arabidopsis chloroplast *accD* transcript. *RNA* 15: 1142-1153.

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

23. Bentolila S, Knight WE, Hanson MR (2010) Natural variation in arabidopsis leads to the identification of REME1, a pentatricopeptide repeat-DYW protein controlling the editing of mitochondrial transcripts. *Plant Physiol* 154:1966-1982.

24. Sung TY, Tseng CC, Hsieh MH (2010) The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria. *Plant J* 63:499-511.

25. 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 Physiol* 152: 939-947.

26. Takenaka M, Verbitskiy D, Zehrmann A, Brennicke A (2010) Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana. J Biol Chem* 285: 27122-27129.

27. Verbitskiy D, Zehrmann A, van der Merwe JA, Brennicke A, 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. Plant J* 61: 446-455.

28. Verbitskiy D, Hartel B, Zehrmann A, Brennicke A, Takenaka M (2011) The DYW-E-PPR protein MEF14 is required for RNA editing at site *matR*-1895 in mitochondria of *Arabidopsis thaliana. FEBS Lett* 585: 700-704.

29. Hammani K, et al. (2011) The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in Arabidopsis mitochondria. *J Biol Chem* 286: 21361-21371.

30. Salone V, et al. (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581: 4132-4138.

31. Gillman JD, Bentolila S, Hanson MR (2007) The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. *Plant J* 49: 217-227.

32. Chatterjee M, et al. (1996) *DAG*, a gene required for chloroplast differentiation and palisade development in *Antirrhinum majus. EMBO J* 15: 4194-4207.

33. Bisanz C, et al. (2003) The arabidopsis nuclear DAL gene encodes a chloroplast protein which is required for the maturation of the plastid ribosomal RNAs and is essential for chloroplast differentiation. *Plant Mol Biol* 51: 651-663.

34. Samson F, et al. (2002) FLAGdb/FST: A database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Res* 30: 94-97.

35. Wang YH (2008) How effective is T-DNA insertional mutagenesis in *Arabidopsis* ?. *J Biochem Tech* 1: 11-20.

36. Kruft V, Eubel H, Jansch L, Werhahn W, Braun HP (2001) Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol* 127: 1694-1710.

37. Heazlewood JL, et al. (2004) Experimental analysis of the arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16: 241-56.

38. Yoo SD, Cho YH, Sheen J (2007) Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat Protoc* 2: 1565-1572.

39. Sosso D, et al. (2012) PPR2263, a DYW-subgroup pentatricopeptide repeat protein, is required for mitochondrial *nad5* and *cob* transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* Feb 7. [Epub ahead of print].

40. Peeters NM, Hanson MR (2002) Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA* 8: 497-511.

41. Germain A, et al. (2011) Mutational analysis of Arabidopsis chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation. *Plant J* 67: 381-394.

42. Yang Y, et al. (2012) A chloroplast envelope membrane protein containing a putative LrgB domain related to the control of bacterial death and lysis is required for chloroplast development in *Arabidopsis thaliana*. *New Phytol* 193: 81-95.

43. Marion J, et al. (2008) Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of Arabidopsis seedlings. *Plant J* 56: 169-179.

44. Carrie C, Giraud E, Whelan J (2009) Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *FEBS J* 276: 1187-1195.

45. Doniwa Y, et al. (2010) The involvement of a PPR protein of the P subfamily in partial RNA editing of an Arabidopsis mitochondrial transcript. *Gene* 454: 39-46.

46. Anant S, et al. (2001) Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing. CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1 complementation factor. *J Biol Chem* 276: 47338-51.

47. Blanc V, et al. (2001) Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing. *J Biol Chem* 276: 10272-83.

48. Kindgren P, et al. (2011) The plastid redox insensitive 2 mutant of Arabidopsis is impaired in PEP activity and high light-dependent plastid redox signalling to the nucleus. *Plant J* Dec 28. [Epub ahead of print].

49. Zhelyazkova P, et al. (2012) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24: 123-136.

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

51. Tasaki E, Hattori M, Sugita M (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J* 62: 560-570.

52. Rajasekhar VK, Mulligan RM (1993) RNA editing in plant mitochondria: [alpha]-phosphate is retained during C-to-U conversion in mRNAs. *Plant Cell* 5: 1843-1852.

53. Okuda K, et al. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* 21: 146-156.

54. Nakamura T, Sugita M (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* 582: 4163-4168.

55. Kobayashi K, et al. (2011) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res* doi: 10.1093/nar/gkr1084

56. Schmitz-Linneweber C, et al. (2006) A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast *rps12* pre-mRNA. *Plant Cell* 18: 2650-2663.

57. de Longevialle AF, et al. (2007) The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial *nad1* intron 1 in *Arabidopsis thaliana*. *Plant Cell* 19: 3256-3265.

58. Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol Cell Biol* 28: 5337-5347.

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

60. Chateigner-Boutin AL, et al. (2011) OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J* 65: 532-542.

61. Burch-Smith TM, Schiff M, Liu Y, Dinesh-Kumar SP (2006) Efficient virus-induced gene silencing in arabidopsis. *Plant Physiol* 142: 21-27.

62. Crowe ML, et al. (2003) CATMA: A complete Arabidopsis GST database. *Nucleic Acids Res* 31: 156-158.

63. Tillich M, et al. (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J* 43: 708-715.

64. Chateigner-Boutin AL, Small I (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* 35: e114.

65. Bentolila S, Chateigner-Boutin AL, Hanson MR (2005) Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in Arabidopsis. *Plant Physiol* 139: 2006-2016.

66. Bentolila S, Elliott LE, Hanson MR (2008) Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* 178: 1693-1708.

67. Karimi M, Inze D & Depicker A (2002) GATEWAY vectors for agrobacterium-mediated plant transformation. *Trends Plant Sci* 7: 193-195.

68. Earley KW, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* 45: 616-629.

69. Bentolila S, Stefanov S (2010) A re-evaluation of rice mitochondrial evolution based on the complete sequence of male fertile and male sterile mitochondrial genomes. *Plant Physiol* DOI:10.1104/pp.111.190231

70. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol* 139: 5-17.

71. Robbins JC, Heller WP, Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the arabidopsis chloroplast *accD* transcript. *RNA* 15: 1142-1153.

72. Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* 6: 3901-3907.

73. Witte CP, Noel LD, Gielbert J, Parker JE, Romeis T (2004) Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. *Plant Mol Biol* 55: 135-147.

74. Kwok EY, Hanson MR (2004) GFP-labelled rubisco and aspartate aminotransferase are present in plastid stromules and traffic between plastids. *J Exp Bot* 55: 595-604.

75. Zhang X, Henriques R, Lin SS, Niu QW, Chua NH (2006) Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc* 1: 641-646.

76. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.

77. Makino A, Mae T, Ohira K (1983) Photosynthesis and ribulose 1,5-bisphosphate carboxylase in rice leaves: changes in photosynthesis and enzymes involved in carbon assimilation from leaf development through senescence. *Plant Physiol* 73: 1002-1007.

78. Hegeman CE, Hayes ML, Hanson MR (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in Arabidopsis *in vitro*. *Plant J* 42: 124-32.

79. Jenkins BD, Barkan A (2001) Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. *EMBO J* 20: 872-879.

80. Gillman JD, Bentolila S, Hanson MR (2007) The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. *Plant J* 49: 217-227.

81. Williams-Carrier R, Kroeger T, Barkan A (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 14: 1930-1941.

82. Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ (2010) Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: Large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol* 152: 1219-1250.

Chapter 3

An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize

Tao Sun, Arnaud Germain, Ludovic Giloteaux, Kamel Hammani, Alice Barkan, Maureen R Hanson, and Stéphane Bentolila

This works was originally published in **PNAS.** 2013. 110:E1169-78.

TS contributed to experimental design, performing experiments and data analysis. TS provided mutant RNA editing analysis results, yeast two-hybrid results, complementation results. SB provided protein motif scanning results. AG provided NGS data analysis. LG provided phylogenetic analysis. KH and AB contributed EMSA results. SB and MRH wrote the manuscript.

Abstract

Plant RNA editing modifies cytidines (C) to uridines (U) at specific sites in the transcripts of both mitochondria and plastids. Specific targeting of particular Cs is achieved by pentatricopeptide (PPR) proteins that recognize *cis* elements upstream of the C that is edited. Members of the RNA-editing factor interacting protein (RIP) family in Arabidopsis have recently been shown to be essential components of the

plant editosome. We have identified a gene that contains a pair of truncated RIP domains (RIP-RIP). Unlike any previously described RIP family member, the encoded protein carries an RNA recognition motif (RRM) at its C terminus and has therefore been named Organelle RRM protein 1 (ORRM1). ORRM1 is an essential plastid editing factor; in Arabidopsis and maize mutants, RNA editing is impaired at particular sites, with an almost complete loss of editing for 12 sites in Arabidopsis and 9 sites in maize. Transfection of Arabidopsis *orrm1* mutant protoplasts with constructs encoding a region encompassing the RIP-RIP domain or a region spanning the RRM domain of ORRM1 demonstrated that the RRM domain is sufficient for the editing function of ORRM1 *in vitro*. According to a yeast two-hybrid assay, ORRM1 interacts selectively with PPR trans-factors via its RIP-RIP domain. Phylogenetic analysis reveals that the RRM in ORRM1 clusters with a clade of RRM proteins that are targeted to organelles. Taken together, these results suggest that other members of the ORRM family may likewise function in RNA editing.

Introduction

The nucleotide sequences of RNAs are altered co- or post-transcriptionally through RNA editing, a form of RNA processing that differs from capping, splicing or 3' end formation. First discovered in the mitochondrial RNAs of kinetoplastid protozoa, this phenomenon has been observed in a wide range of organisms and can affect the mRNAs, tRNAs and rRNAs present in all cellular compartments (reviewed

in (1)). Nucleotides can be inserted, deleted, or modified through RNA editing. In flowering plants, RNA editing is restricted to organelle transcripts and modifies specific cytidines (C) to uridine (U). The reverse editing reaction, U to C, is also found in a few plant lineages. In Arabidopsis, 34 plastid Cs and over 500 mitochondrial Cs have been reported to be edited (2-4). The current consensus view is that RNA editing corrects at the post-transcriptional level mutations that have occurred in plant organelle genomes (5). The absence of editing in some mutants leads to the production of improper proteins that can result in seedling lethality (6).

Despite the discovery of plant RNA editing more than twenty years ago (7-9), only some of the components of the plant editosome are known. *Cis*-elements needed for recognition of C targets are usually found within 30 nt of the C to be edited (10-13). Recognition of the *cis*-elements is performed by members of the PLS sub-class of the large pentatricopeptide repeat (PPR)-containing family of proteins (14). However, the enzyme catalyzing the editing reaction, presumably by deamination, is still unknown, though suspicion has fallen on the DYW domain present in some PPR proteins because it contains residues similar to the conserved cytidine deaminase motif (15). The elusiveness of the enzyme responsible for plant RNA editing (16-18) suggests that some important components of the editing machinery are still to be identified.

Recently, members of the RNA-editing interacting protein (RIP) family in Arabidopsis have been discovered to be *trans*-factors essential for editing. We identified Arabidopsis dual-targeted protein RIP1, an essential plant editing factor that is required for the editing of numerous Cs both in plastids and mitochondria (19). A

rip1 mutant plant exhibited reduced editing efficiency at 266 mitochondrial C targets, with a major loss of editing for 108. RIP1 is a member of a small protein family that contains 10 members. Other members of the RIP family have also been shown to be required for organelle editing (20). RIP proteins are able to interact selectively with PPR *trans*-actors and also with each other (19, 20); however, their function in the plant editosome remains unclear.

Here we report the identification of a unique protein that is both a member of the RIP family and the RRM-containing family. This protein carries an RNA recognition motif (RRM) at its C terminus, unlike any other RIP domain-containing proteins. The RRM is the most widespread motif involved in RNA binding and is found in all kingdoms (21). However, the RRM domain of this unique protein is most similar to the domain present in an identifiable clade of RRM proteins, most of which are either known to be localized or are predicted to be targeted to plant organelles. We therefore refer to this RRM subfamily as the organellar RRM (ORRM) family. As the founding member of the family, At3g20930 has been named ORRM1. Identification of ORRM1 as an editing factor implicates a new class of RRM-containing proteins as potentially involved in RNA editing as well as other aspects of organelle RNA metabolism.

Results

Identification of a protein carrying truncated RIP domains.

A blastp search for homologs to the RIP1 protein returned the 10 previously reported members of the RIP family (19, 20) as well as a new RIP family member, encoded by the gene At3g20930 (Figure 3.1A). This protein was not previously described as either a RIP or MORF protein (19, 20). We used the MEME software (22) to identify 4 highly significant motifs in the RIP family (Figure 1B). The RIP block can be defined by the following series: motif 1-gap-motif 2-motif 3-motif 4 (Figure 3.1A). The distal motif 4 found in RIP7 is below the threshold of e-10 (p-value = 9.6 e-6). Most RIP proteins possess a complete RIP block (Figure 3.1C). The new member of the RIP family encoded by At3g20930 exhibits a duplication of truncated RIP blocks; the first block, from amino acid 89 to amino acid 147, contains a degenerate motif 1 plus motifs 2 and 3, while the second RIP block from amino acid 163 to amino acid 250 contains motifs 1-3. Both RIP blocks in At3g20930 lack motif 4 (Figure3.1C).

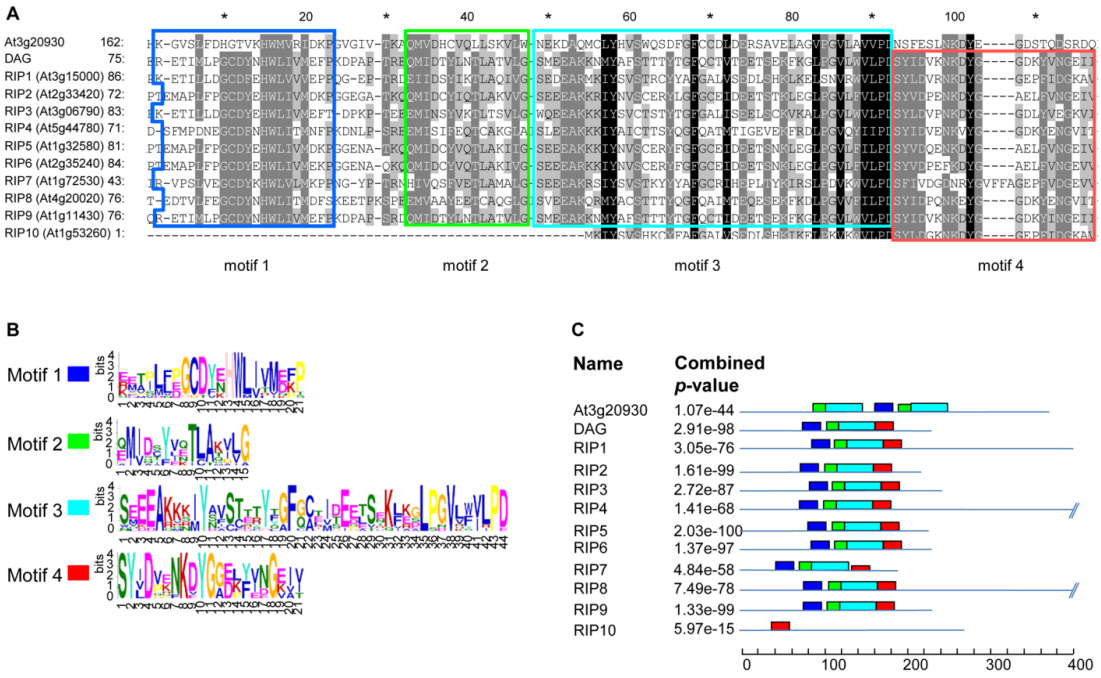


Figure 3.1. The protein encoded by At3g20930 belongs to the RIP family and contains a pair of truncated RIP domains. (A) Alignment of the conserved regions in the Arabidopsis RIP proteins (RIP1 to RIP10), the DAG protein from *Antirrhinum majus* (Genbank CAA65064) and the protein encoded by At3g20930 was performed using T-Coffee Version_9.03, and displayed using GeneDoc with the conserved residue shading mode and similarity groups enabled. Overlaid on the aligned sequences are the 4 motifs detected by the MEME software Version_4.9.0 . (B) The RIP domain contains 4 motifs uncovered by MEME. The settings were 6aa<width<100aa, maximum number of motifs to find=4. All 4 motifs are highly significant (E-values: 7.8e-208, 5.5e-111, 3.7e-86, 4.2e-52). For each motif is given its sequence logo showing the likelihood of residue at each position. (C) The combined motif diagrams are shown for each of the RIP protein, the DAG protein, and the protein encoded by At3g20930. The height of a motif is truncated when its p-value is > 1e-10, for example, motif 3 for RIP7.

At3g20930 protein contains an RNA recognition motif at its C terminus and belongs to a clade of RRM proteins.

A motif search with Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) identified the presence of a RNA recognition motif (RRM) at the C terminus of the protein. The RRM domain is approximately 80 amino acids long and contains two short consensus sequences, RNP1 (octamer) and RNP2 (hexamer) that are characteristic of RRMs (Figure 3.2A). In Arabidopsis, 196 RRM-containing proteins were previously identified through an *in silico* search for the RRM motif (23). A blast search using the RRM domain of the protein encoded by At3g20930 revealed that this domain was more closely related to the RRM found in two distinct families described by Lorkovic and Barta (2002), the Glycine-rich RNA-binding proteins (GR-RBP) and the small RNA-binding proteins (S-RBP). A common feature of these two protein families is their similar domain organization with one N-terminal RRM and a C-terminal extension. GR-RBPs are represented by eight members, while 15 proteins were annotated as S-RBP (23). At the time of the Lorkovic and Barta's (2002) report, Vermel et al. (2002) (24) identified by biochemical means a family of mitochondrial-specific RRM-containing proteins that they named mitochondrial RNA-binding proteins (mRBP). The eleven mRBPs belong to either the GR-RBP or the S-RBP family. Figure 3.2A illustrates the strong similarity between the RRM domains found in the At3g20930 product, the GR-RBPs, and the mRBPs. To verify the similarity between the RRM domains found in the At3g20930 encoded product, the GR-RBPs, and the mRBPs, we aligned them with the RRM of a protein encoded by

At5g46840,which does not belong to any of these sub-families (Figure 3.2A).

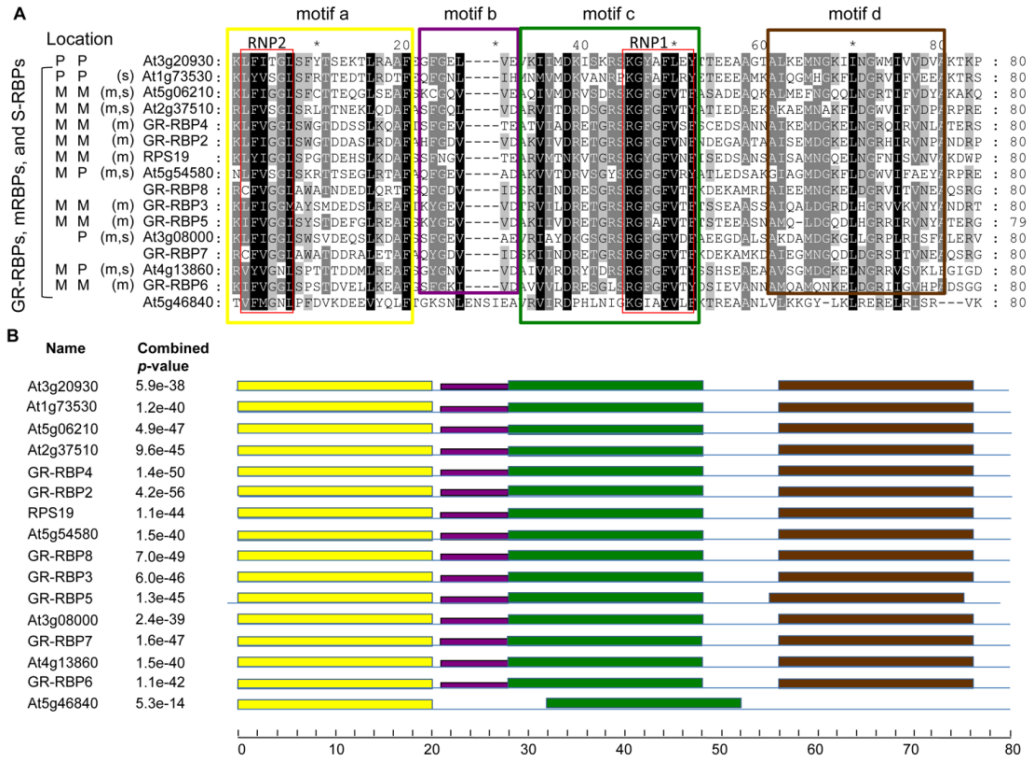


Figure 3.2. The RRM domain found at the C-terminus of ORRM 1shows most similarity to the RRMs from glycine rich (GR), mitochondrial (m), and small (S), RNA-binding proteins (RBP). (A) RRMs from the protein encoded by At3g20930, the GR-RBPs, the mRBPs, and the S-RBPs were aligned by T-Coffee Version_9.03, and displayed using GeneDoc with the conserved residue shading mode and similarity groups enabled. Depending on the database of protein domains searched, prosite or pfam, the RRM motif was located at position 282-360 with a E-value of 1.3e-11, or at position 284-350 with a E-value of 2.2e-24, respectively. Overlaid on the aligned sequences are the 4 motifs detected by the MEME software Version_4.9.0. Location on the left of each protein refers to the subcellular location predicted by Predotar or TargetP, P: plastid, M: mitochondrion. In parenthesis preceding the name of the protein is given the annotation (m): mitochondrial (24), (s): small (23). (B) Combined p-values and block motifs computed by the MEME software indicated that the RRMs fromORRM1, the GR-RBPs, the mRBPs, and the S-RBPs belong to the same family defined by motifs a, b, c, and d. The product encoded by At5g46480, a RRM-containing protein (motifs a and c only) does not belong to this family.

When we used the MEME software with the number of motifs set at 4, and width greater than 5 but less than 20 amino acids, 4 motifs were identified in this set of RRMs that can define the new sub-family related to the RRM present in the protein containing two RIP motifs. Two of the four motifs, motif a and motif c, are found in the RRM domains of the new sub-family but also in the RRM domain of the protein encoded by At5g46840 (Figure 3.2B). In contrast, motif b and motif d are specific to the RRM domains found in the product encoded by the RIP-family protein At3g20930, as well as in the GR-RBPs and the mRBPs (Figure 3.2B). All of the proteins shown in Figure 3.2A are predicted to be located in plastids or mitochondria. We have therefore named the group of organelle-targeted proteins containing the 4 motifs described in Figure 3.2 the Organelle RRM- containing (ORRM) family. The protein encoded by At3g20930 is hereafter designated as ORRM1.

A Pfam domain search using At3g20930 (ORRM1) as a query identified 642 RRM containing regions in the *A. thaliana* genome. As examples, putative RNA-binding proteins, poly(A)-binding ribonucleoproteins, splicing-factors, U2 small nuclear ribonucleoproteins, *Arabidopsis-mei2-like* proteins and chloroplast ribonucleoproteins were retrieved. Several clearly identifiable clusters in the phylogenetic tree can be distinguished, suggesting that RRMs can be classified according to groups of proteins of the same function, such as U2 small nuclear ribonucleoproteins (snRP), poly(A)-binding proteins (PABP), splicing factors (RSP), and chloroplast RNA-binding proteins (CP) (Figure 3.3).

ORRM1 appears to form a monophyletic group with members of the Glycine-Rich RNA binding proteins (Figure 3.3). The topology of the tree constructed by the

UPGMA, MP, ML and ME methods was not different from that of the Neighbor-Joining tree presented here, supporting a consistent grouping of the proteins.

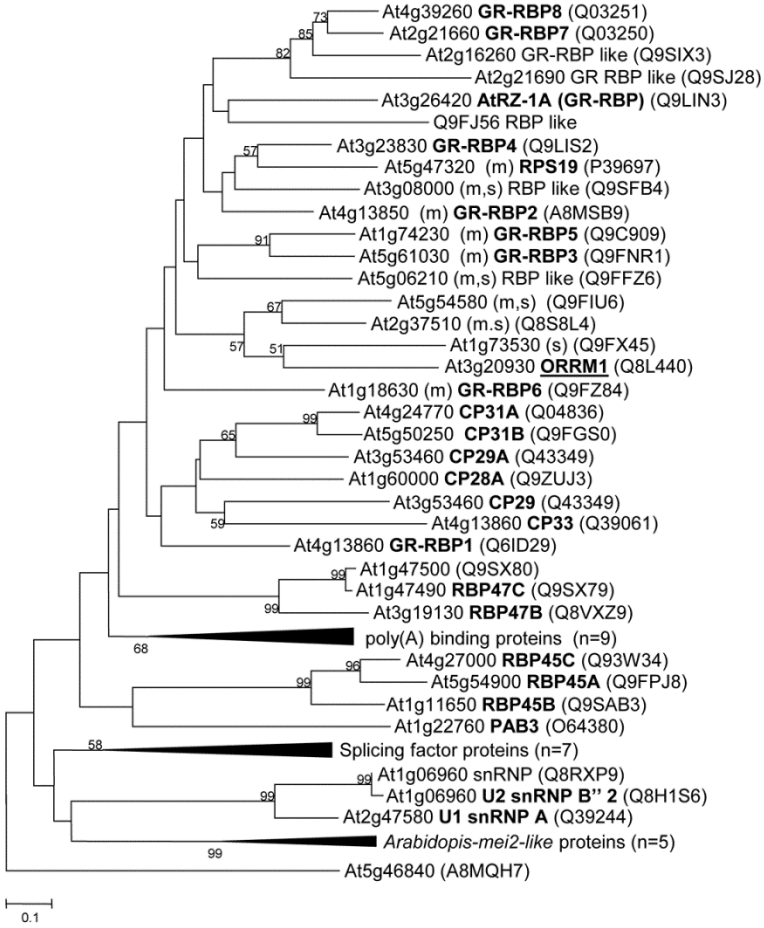


Figure 3.3. Phylogenetic tree based on the amino-acid sequences of the RRM motifs in RRM-containing proteins (84 amino-acids considered). The tree was inferred using the Neighbor-Joining method, and evolutionary distances were computed using the Poisson correction method. The scale bar corresponds to 0.1 substitutions per site. GR= Glycine-Rich; RBP = RNA binding protein; CP= chloroplast ribonucleoprotein; PABP = poly(A) binding protein; snRP = small nuclear ribonucleoprotein. In the ORRM1 clade, figure in parenthesis the annotation (s) for small given in Lorkovic and Barta (23), or (m) for mitochondrial given in Vermel et al. (24).

A T-DNA insertional mutant in *ORRM1* exhibits severe defects in plastid editing

We obtained an Arabidopsis mutant from the ABRC stock collection and verified that it was homozygous for a T-DNA insertion in the first exon of *ORRM1* (Figure 3.4A; SALK_072648, designated here as *orrm1*). The homozygous mutant did

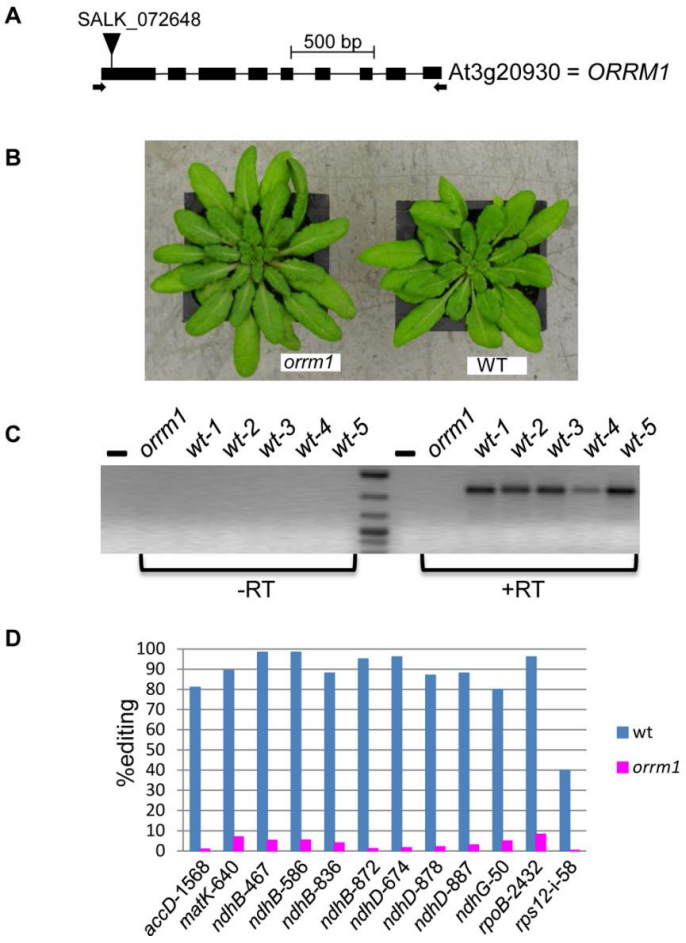


Figure 3.4. A T-DNA insertional mutant in *ORRM1* is severely impaired in plastid editing. (A) Schematic representation of the model gene for *ORRM1* with exons represented as squares and introns as lines, the T-DNA insertion is shown as a triangle in the first exon. The primers used for the RT-PCR are indicated by arrows. (B) The homozygous mutant plant (left) does not show any phenotypic defect compared to a Columbia wild-type plant (right) when grown under growth-room conditions. (C) No expression of ORRM1 is detectable by RT-PCR after 45 cycles in the *orrm1* mutant although expression is readily observed in wild-type (D) 13 plastid sites show a severe reduction of editing extent (Δ*ORRM1*) > 90% in the *ORRM1* T-DNA insertional mutant. Because the *orrm1* homozygous mutant line was in a Columbia background, five wild-type siblings of other *rip* family insertional mutants that were in a Columbia background served as positive controls.

not show any phenotypic defect when grown under growth room conditions (Figure3.4B). No expression of *ormm1* was detected by RT-PCR (Figure 3.4C). We examined the organelle transcriptome of the mutant for editing defects because other proteins carrying RIP domains have been shown to be editing factors (19, 20). We analyzed plastid RNA editing extent with a new methodology based on RNA-seq. Briefly, total RNA is isolated from leaves and RT-PCR products corresponding to known organelle genes are obtained by using gene-specific primers. The products are mixed in equimolar ratio, sheared and used as templates to produce an Illumina TruSeq library. This RNA-seq analysis demonstrated that ORRM1 is a plastid editing factor; 12 among 34 plastid sites exhibit a severe reduction of editing extent in the mutant relative to the wild-type (Figure3.4D).

In addition to the 12 sites where editing is reduced by 90% or more, nine plastid sites exhibit a reduction of editing extent between 10 and 90% in the mutant. Thus 62% (21/34) of the plastid editing sites are under the control of ORRM1 (Table 3.1). We confirmed the gene-specific RT-PCR results on the *orrm1* mutant by performing RNA-seq on total plastid RNA. For this purpose, total RNA was extracted from chloroplasts purified from mutant and the wild-type, and reverse transcribed using random hexamers. The number of reads per chloroplast gene ranged from ~20 to ~5000 (Table 3.1). The numbers of reads are much higher in the gene-specific RT-PCR- generated Illumina library, with averages of ~7000 and ~11000 for the wild-type and the mutant, respectively (Table 3.1). Despite the difference in depth coverage between the gene-specific and total plastid RNA Illumina libraries, the reductions of

editing extent in the mutant are highly consistent between the two assays (Table 3.1, Figure 3.5).

Table 3.1. Effect of the T-DNA insertional mutation in ORRM1 on the editing extent of plastid sites

Gene	Position	Gene specific							Total plastid RNA						
		Editing extent			Number of reads*				Editing extent			Number of reads			
					Wt		*orrm1*					Wt		*orrm1*	
		Wt	*orrm1*	Δ*orrm1*	C	T	C	T	WT	*orrm1*	Δ*orrm1*	C	T	C	T
accD	794	1.00	1.00	0.00	70	17812	24	5148	0.95	0.97	-0.02	5	102	3	113
accD	1568	0.81	0.01	0.99	2937	12252	20671	117	0.67	0.04	0.95	19	38	54	2
atpF	92	0.97	0.97	0.00	1920	62144	1825	54857	0.85	0.88	-0.04	182	1057	156	1194
clpP	559	0.90	0.24	0.73	1144	9814	19229	6043	0.85	0.40	0.53	21	121	98	66
matK	640	0.89	0.07	0.92	2633	21095	573	41	0.85	0.01	0.99	18	102	93	1
ndhB	149	0.99	0.99	0.00	315	36606	107	17738	0.93	0.97	-0.05	32	429	24	908
ndhB	467	0.98	0.05	0.95	625	31831	29209	1473	0.91	0.04	0.95	54	515	835	36
ndhB	586	0.98	0.05	0.95	612	33777	27689	1460	0.92	0.06	0.93	40	447	583	40
ndhB	746	0.98	0.46	0.53	375	20070	13551	11499	0.94	0.49	0.48	26	436	336	329
ndhB	830	0.94	0.28	0.71	461	7731	19234	7351	0.84	0.22	0.73	31	159	371	107
ndhB	836	0.88	0.04	0.95	965	6806	25408	1122	0.73	0.08	0.90	30	83	407	33
ndhB	872	0.95	0.01	0.99	685	11974	27525	191	0.85	0.01	0.99	23	129	360	3
ndhB	1255	1.00	0.40	0.60	204	40709	12844	8474	0.95	0.42	0.55	20	376	233	172
ndhB	1481	0.98	0.99	0.00	268	15846	320	23271	0.88	0.90	-0.03	43	310	53	478
ndhD	2	0.46	0.21	0.54	71538	60341	11478	3086	0.54	0.31	0.43	99	118	230	103
ndhD	383	0.99	0.99	0.00	392	35163	124	11839	0.98	0.97	0.01	12	674	21	640
ndhD	674	0.96	0.02	0.98	1346	28753	12917	199	0.88	0.02	0.97	63	446	721	17
ndhD	878	0.87	0.02	0.97	6529	45066	15494	388	0.86	0.03	0.96	112	679	968	32
ndhD	887	0.88	0.03	0.96	6077	45838	15359	504	0.85	0.03	0.97	104	573	723	21
ndhF	290	0.99	0.99	0.00	844	66974	204	20210	1.00	0.99	0.01	0	55	1	67
ndhG	50	0.80	0.05	0.94	5111	20447	10297	553	0.77	0.07	0.90	41	141	248	20
petL	5	0.94	0.94	0.00	296	4371	480	7642	0.63	0.65	-0.04	10	17	8	15
psbE	214	1.00	1.00	0.00	112	28930	382	87546	0.99	0.99	-0.01	33	2232	26	3253
psbF	77	0.99	0.98	0.01	12	1764	213	12773	0.96	0.98	-0.02	73	1957	54	3026
psbZ	50	0.94	0.96	-0.02	5129	79997	2383	53887	0.94	0.95	-0.01	121	1906	76	1479
rpl23	89	0.85	0.81	0.04	1010	5632	1746	7562	0.66	0.75	-0.13	25	49	26	76
rpoA	200	0.81	0.22	0.73	10415	44159	7017	1927	0.40	0.22	0.44	111	75	107	31
rpoB	338	0.93	0.80	0.14	1141	16137	2668	10601	0.83	0.77	0.08	5	25	6	20
rpoB	551	0.96	0.25	0.74	1086	23061	6419	2086	0.83	0.22	0.74	4	20	25	7
rpoB	2432	0.96	0.08	0.92	2608	68572	4126	343	0.81	0.20	0.75	8	35	24	6
rpoC1	488	0.21	0.22	-0.05	36084	9359	71377	19698	0.19	0.25	-0.32	198	46	133	44
rps12	i-58	0.40	0.00	0.99	17174	11391	5454	18	0.33	0.00	1.00	14	7	25	0
rps14	80	0.94	0.97	-0.03	1662	26479	105	3384	0.96	0.98	-0.02	75	1702	99	5057
rps14	149	0.92	0.60	0.35	1645	18939	992	1494	0.90	0.75	0.16	200	1757	1081	3264

Editing extent: EE = T/(C+T)

Δ*orrm1*: variation of editing extent in *orrm1* mutant = (EE (wt)-EE (*orrm1*))/(EE(wt)

* C (WT) = C (wt-1) + C (wt-2)+ C (wt-3)+ C (wt-4)+ C (wt-5), T (WT) = T (wt-1) + T (wt-2)+ T (wt-3)+ T (wt-4)+ T (wt-5)

orrm1 was sequenced twice, the number of reads are the sums of the two sequencing experiments

In order to verify the suitability of the RNA-seq method for assaying RNA editing, we performed poisoned primer extension (PPE) on a selection of transcripts

and compared the results to the gene-specific RNA-seq and total plastid RNA-seq data (Figure 3.6).

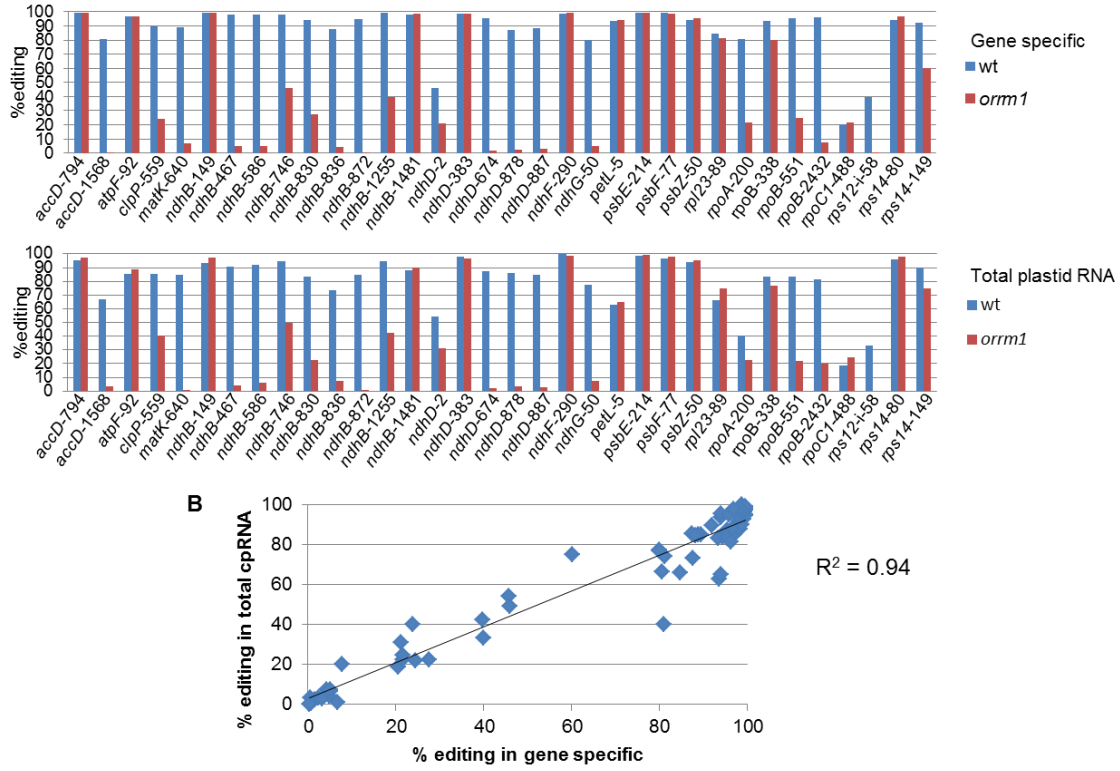


Figure 3.5. The reduction of plastid editing extent in the At3g20930 T-DNA insertional mutant is consistently detected in different RNA-seq experiments. (A) Comparison of the editing extent of the 34 plastid sites in a wild-type (wt) and *orrm1* mutant plants. The values of editing extent were obtained from gene-specific (upper panel) or total plastid RNA (lower panel). (B) A correlation >0.9 exists between level of editing extent evaluated on cDNAs corresponding to gene transcripts (gene specific) and cDNAs corresponding to the whole plastid transcriptome (total plastid RNA).

In the PPE assay, cDNA serves as a template for an extension reaction in the presence of a dideoxy G (ddG). When ddG is incorporated, the extension stops at the first unedited C that is encountered by the enzyme. The products of edited vs. unedited

transcripts will differ in size; the amount of each product can be accurately monitored on gels.

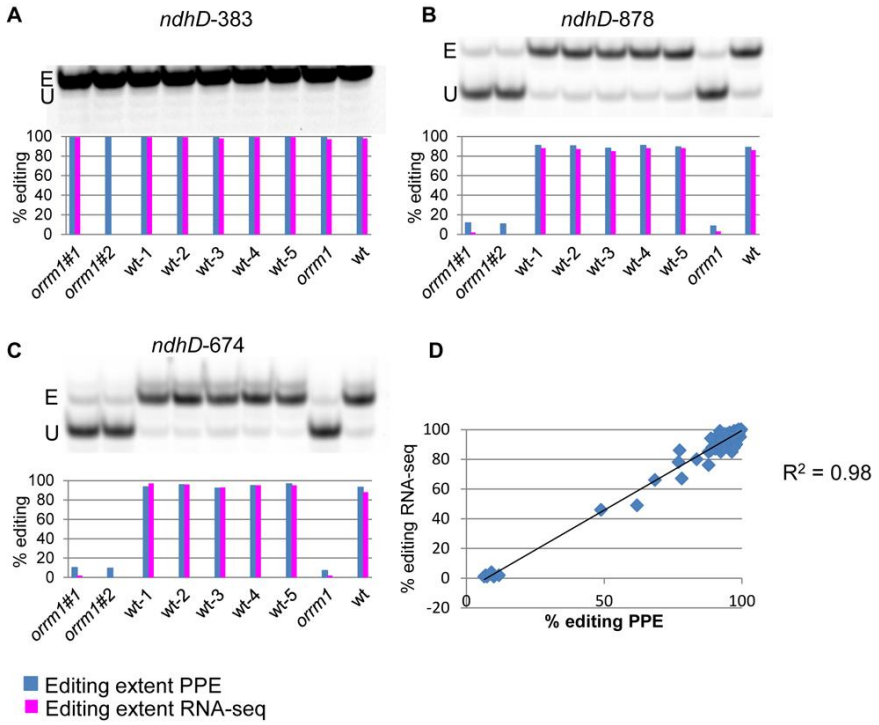


Figure 3.6. PPE assay validates the editing extents derived from RNA-seq. (A) (B) (C) Acrylamide gels separate the PPE products obtained from samples used in this study. E, edited; U, unedited. The name of the site assayed is given above each gel. The quantification of editing extent derived from the measure of the band's intensity is represented by a bar below each lane of the acrylamide gels (blue diagonal background). By way of comparison, the editing extent derived from RNA-seq is represented by a magenta bar. RNA-seq was performed on gene specific cDNAs for *orrm1#1*, wt-1, wt-2, wt-3, wt-4, and wt-5, and on total plastid RNA for *orrm1* and wt. *orrm1#2* was not analyzed by RNA-seq. (D) The correlation between the editing extent values derived from PPE assay and RNA-seq verifies RNA-seq is a sound method to determine editing extent. The correlation was calculated by plotting 72 points (8 samples x 9 PPE gels).

As an example, we show the PPE data for three C targets of editing in the *ndhD* transcript, one whose editing extent is unaffected by the *orrm1* mutation

(Figure3.6A), while the other two exhibit almost complete loss of editing (Figures. 3.6B and 3.6C).

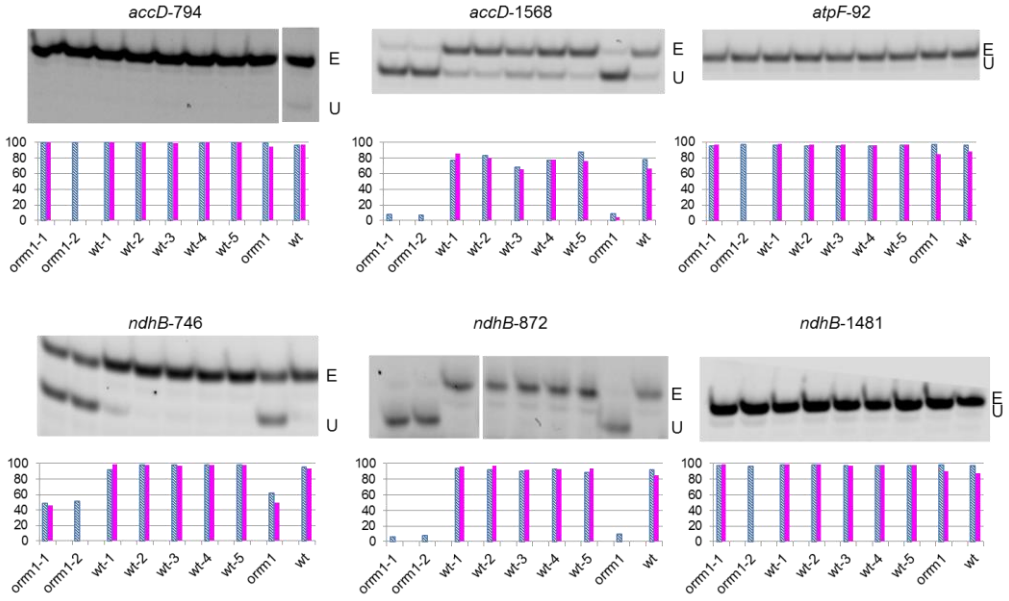


Figure 3.7. The PPE assay confirms the plastid editing defects detected in the *orrm1* mutant by RNA-seq. Acrylamide gels separate the PPE products obtained from samples used in this study. E, edited; U, unedited. The name of the site assayed is given above each gel. The quantification of editing extent derived from the measure of the band's intensity is represented by a bar below each lane of the acrylamide gels (blue diagonal background). As a way of comparison, the editing extent derived from RNA-seq is represented by a magenta bar. RNA-seq was performed on gene specific cDNAs for *orrm1-1*, wt-1, wt-2, wt-3, wt-4, and wt-5, and on total plastid RNA for *orrm1* and wt. *Orm1-1* and *orrm1-2* are two homozygous mutant plants.

We performed PPE assays on 6 additional transcripts (Figure 3.7) and demonstrate the consistency of the two assays by graphing the RNA-seq editing extent data against the data from the PPE assay (Figure 3.6D).

We also surveyed the mitochondrial transcriptome of the *orrm1* mutant with gene-specific primers; none of the 574 mitochondrial sites assayed showed a

significant difference in editing extent between the mutant and the wild-type. Thus, ORRM1 is an editing factor that is specific to plastids.

RNA editing defects are detected in *orrm1* chloroplast transcripts that do not differ in abundance from wild-type transcripts

Although our gene-specific RNA-seq method does not provide any information on transcript abundance, our RNA-seq experiments using total chloroplast RNA do allow us to quantify relative abundance of transcripts from different genes. The number of total plastid RNA reads corresponding to each plastid transcript in the *orrm1* and wild-type total plastid RNA data exhibit little variation (Table 3.1), indicating that changes in RNA abundance does not explain the effect of the mutation on editing at specific C targets. To verify that abundance of transcripts carrying Cs affected in editing extent does not vary greatly between the mutant and wild-type, we performed RNA blots with total chloroplast RNA from *orrm1* and wild-type plants (Figure 3.8). We used three probes corresponding to the *matK*, *ndhB*, and *ndhD* genes, whose transcripts carry C targets with reduced editing in the *orrm1* mutant. These blots demonstrate that there is no difference in the complexity of the RNA profile or abundance of particular transcript species between wild-type and *orrm1* (Figure 3.8). In addition, different Cs located on the same transcript sometimes vary greatly in their editing extent in the *orrm1* mutant.

Both *ndhB* and *ndhD* carry a C target that is unaffected in the *orrm1* mutant,

while the other Cs in the *ndhB* and *ndhD* exhibit major reduction in editing efficiency in the mutant (Figure 3.5, Fig 3.7).

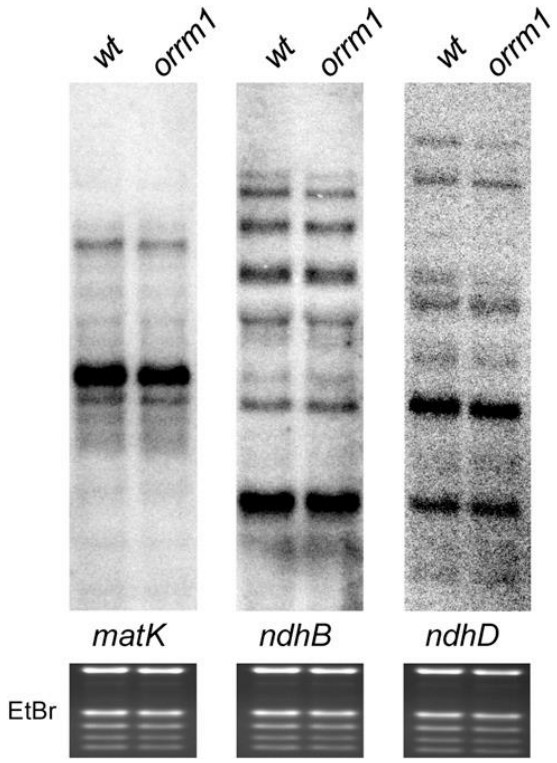


Figure 3.8.RNA blots demonstrate the absence of change of transcript abundance in the *orrm1* mutant. Below each blot is given the name of the transcript corresponding to the probe used. Below each blot is shown the EtBr gel as a control for equal loading of the wild-type and orrm1 RNAs.

The RRM domain can rescue the editing defect in *orrm1* protoplasts.

We determined whether the T-DNA insertional mutation could be complemented by transient expression of *ORRM1* under the control of a 35S promoter in mutant protoplasts. PPE assay demonstrated that mutant protoplasts transfected with the construct carrying the full-length *ORRM1* exhibited significant increase in editing extent (Figure 3.9A, lane F). The extent of editing for *ndhB*-872 and *rps12*-i58 in

transfected mutant protoplasts, 43% and 27% respectively, is sufficient to observe a very distinct edited product band on the PPE gel when compared to untransfected protoplasts (Figure 3.9A, lane NT). As expected in a transient expression assay, the

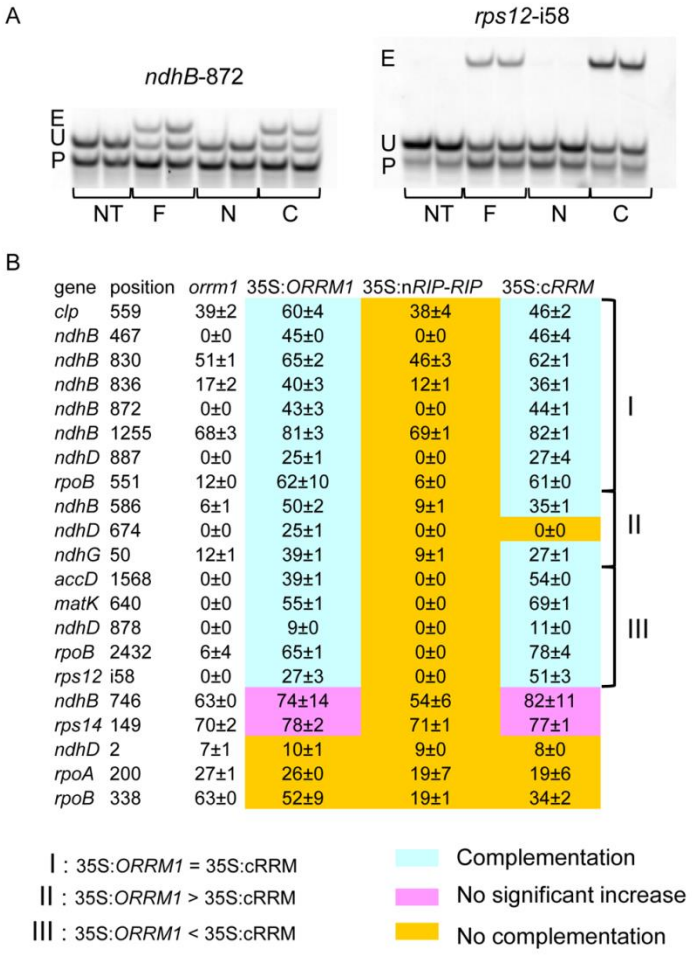


Figure 3.9. ORRM1 is able to complement *orrm1* protoplasts with its RRM domain, not its RIP domain. (A) PPE products from not transfected protoplasts (NT), protoplasts transfected with a construct encoding a full length (F), the N terminal portion (N) that contains the RIP-RIP domain, or the C terminal portion (C) that contains the RRM domain. Above each gel is given the name of the editing site (gene-position), E: edited, U: unedited, P: primer. The presence of the edited bands is only observed in protoplasts transfected with the full-length construct or with a construct encoding the RRM. (B) 23 plastid sites showing a decrease in the *orrm1* mutant were assayed for complementation in the transfected mutant protoplasts. Among the 16 sites complemented by the full length construct, 15 also exhibit a complementation by the construct encoding the RRM domain of ORRM1.

level of editing extent of *ndhB*-872 and *rps12*-i58 in the transfected protoplasts does not reach the level observed in the wild-type plant, which is 95% and 40%, respectively (Table 3.1). 76% (16/21) of the plastid sites assayed that showed a reduction or a lack of editing extent in the mutant exhibited a significant increase of their editing extent in the transfected protoplasts with the full-length *ORRM1* (Figure 3.9B).

Similar transfection experiments were performed with constructs encoding either the N-terminal portion of ORRM1 that contains the duplicated RIP-RIP region (Figure 3.9A, lane N) or the C-terminal portion, which carries the RRM domain (Figure 3.9A, lane C). Of the two truncated constructs tested, only the construct encoding the RRM domain was able to complement the editing defect of the mutant (Figure 3.9A, lane C). Among the 16 sites partially complemented by the full-length *ORRM1*, 15 showed a significant increase of editing extent in the mutant protoplasts upon transfection with the construct encoding the RRM domain (Figure 3.9B). At three sites, the full-length construct was able to complement the editing defect more efficiently than did the RRM construct. Among these sites, *ndhD*-674 is the only one for which no effect of the RRM construct was observed (Figure 3.9B). The RRM construct more efficiently complemented 5 sites than did the full-length construct; among these sites, *rps12*-i58 shows almost twice the amount of edited transcripts in protoplasts transfected with the RRM construct than with the full-length construct (Figure 3.9A). The increase of editing extent in transfected protoplasts was below the significance threshold for two sites, *ndhB*-746 and *rps14*-149. An absence of effect

from the transfection with either the full-length or the RRM construct was observed in only 3 of the assayed sites (Figure 3.9B).

The maize ORRM1 ortholog is required for the editing of both orthologous and maize-specific sites.

Maize mutants with *Mu* transposon insertions in the ortholog of ORRM1 (Zm-*orrm1*) were recovered during the identification of causal mutations in a large collection of non-photosynthetic mutants (http://pml.uoregon.edu/photosyntheticml.html). The Zm-*orrm1* mutants originally came to our attention due to their unusual spectrum of protein deficiencies (see below), which could not easily be explained by defects in known chloroplast biogenesis genes. Therefore, the mutants were selected for gene identification with a high-throughput method for sequencing *Mu* insertion sites (25) (see Methods). Two alleles were identified, both with an insertion in the first exon (Figure 3.10A). Complementation crosses yielded heteroalleic mutant progeny (Zm-*orrm1-1*/Zm-*orrm1-2*) displaying a pale green phenotype (Figure 3.10B). The mutant progeny of complementation crosses were used for the molecular analyses summarized below, as phenotypes expressed in this material must result from disruption of the Zm-*orrm1* gene. These heteroallelic mutants will be referred to hereafter as Zm-*orrm1* mutants.

Defects in the major photosynthetic enzyme complexes were profiled in Zm-*orrm1* mutants by quantifying one core subunit of each complex: PetD of the cytochrome b_6f complex, PsaD of photosystem I, PsbA of photosystem II, RbcL of

Rubisco, and AtpB of the plastid ATP synthase (Figure 3.10C). The accumulation of these proteins is known to parallel that of other closely associated subunits in the same

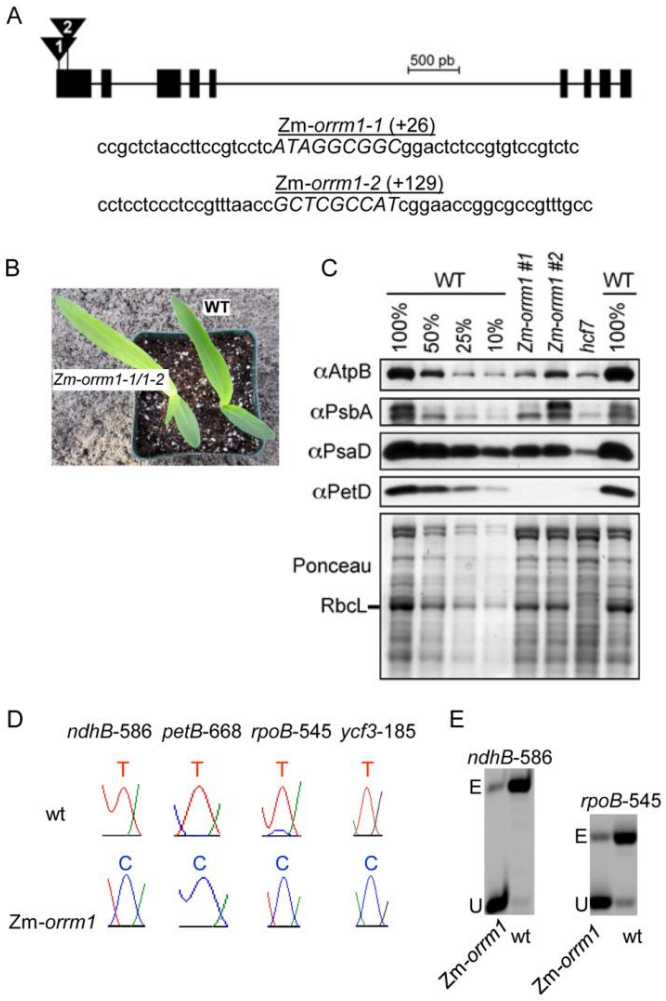


Figure 3.10. The maize orthologous gene to *ORRM1* encodes a plastid editing factor. (A) gene model of Zm-*ORRM1* with exons (squares) and introns (lines). The 2 independent *Mu* insertions are shown as triangles in the first exon, with the 9 bp target site duplication shown in capital letters . The insertions in exon 1 are at the indicated position with respect to the start codon (B) Zm-*orrm1-1/1-2* complementation cross progeny mutant plant is a photosynthetic mutant exhibiting a pale green phenotype (left) compared to the wild-type (right). (C) Immunoblot analysis of photosynthetic enzyme accumulation in *Zm-orrm1-1/Zm-orrm1-2* mutants. An immunoblot of total leaf extracts (5 μg or the indicated dilutions) was probed with antibodies to the indicated proteins. The same blot stained with Ponceau S is shown below; the band corresponding to the large subunit of Rubisco (RbcL) is marked. The Zm-*orrm1* mutants are siblings derived from the same complementation cross (genotype Zm-

orrm1-1/Zm-*orrm1-2*). *hcf7* is a previously-described mutant with a global decrease in chloroplast translation (65). (D) Bulk-sequencing electrophoretograms of RT-PCR products from *Zmorrm1* and a wild-type sibling at 4 plastid sites show a total loss of editing in Zm-*orrm1* as no edited peak (T) is detectable. (E) PPE assay reveals a residual editing extent in Zm-*orrm1* mutant plant for *ndhB*-586 and *rpoB*-545, 7% and 14% respectively.

complex. Zm-*orrm1* mutants have a severe deficiency for PetD (<<10% of normal levels), a moderate deficiency for PsaD (roughly 25% of normal levels), and mild, somewhat variable reductions in RbcL, AtpB, and PsbA (40-80% of normal levels). In light of the editing defects observed in Arabidopsis *orrm1* mutants, a reasonable hypothesis was that these protein deficiencies result from defects in chloroplast RNA editing. Indeed, when all 27 known edited nucleotides in maize chloroplasts were assayed by bulk sequencing of RT-PCR products, multiple RNA editing defects were detected in the Zm-*orrm1* mutant (Figure 3.11).

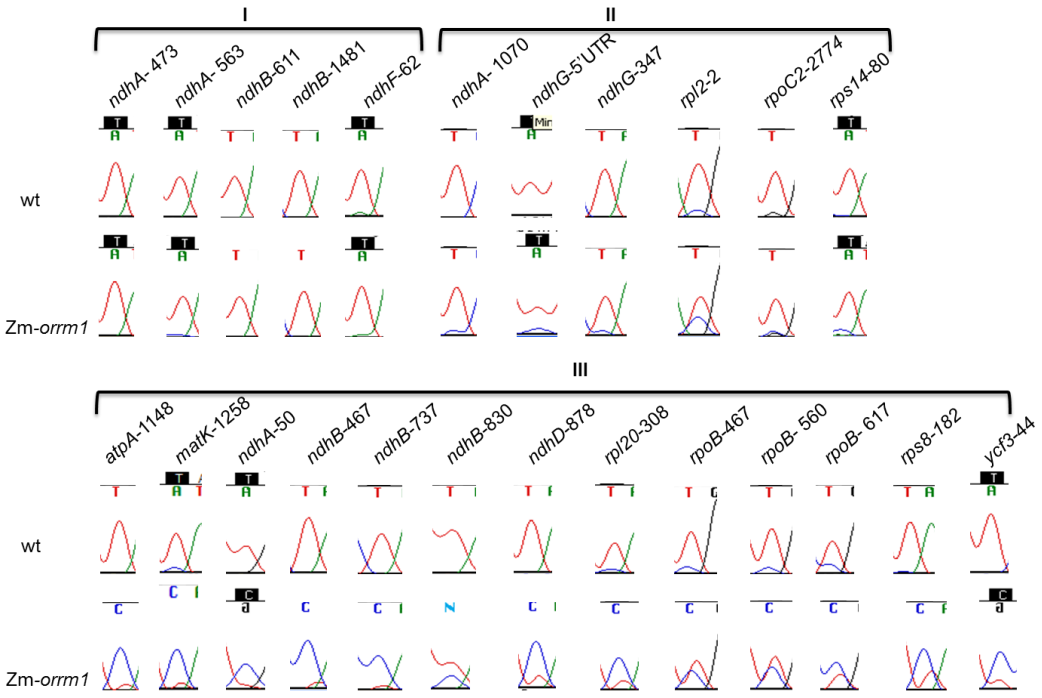


Figure 3.11. Plastid editing sites in the Zm-*orrm1* mutant either do not show a reduction of editing extent (I), or show a slight (II) or pronounced (III) reduction of editing extent when compared to the wild-type plant. Bulk-sequencing electrophoretograms of RT-PCR products obtained from wt (top), and Zm-*orrm1* (bottom) plants. Above each electrophoretogram is given the editing site. Notice the difference in the height of the T (edited, red) and C (unedited, blue) peaks between wt plant and the Zm-*orrm1* mutant particularly for editing sites belonging to the III category.

Four sites, *ndhB*-586, *petB*-668, *rpoB*-545, and *ycf3*-185 exhibited a complete loss of editing (Figure 3.10D). Only five plastid sites did not show any reduction of editing extent in the Zm-*orrm1* mutant (Figure 3.11, Table S2). Maize and Arabidopsis plastid transcripts share 7 common editing sites so we more precisely assayed the editing extent of these sites in the Zm-*orrm1* mutant by a PPE assay using fluorescent primers designed for Arabidopsis (Figure 3.10E; Fig 3.12). The PPE assay is more sensitive than the bulk-sequencing assay, and indicates residual editing of the *ndhB*-586 and *rpoB*-545 sites in the maize mutant (Figure 3.10E).

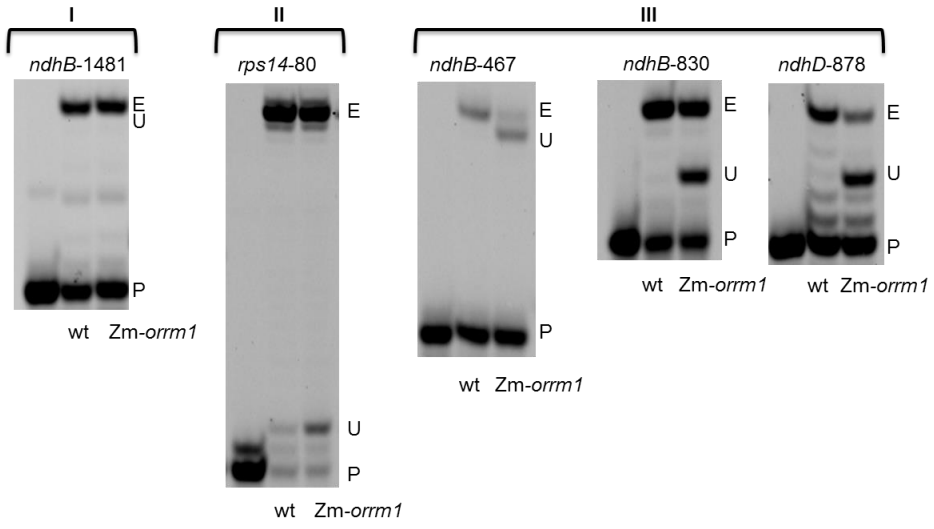


Figure 3.12. PPE assay confirms the plastid editing sites in the Zm-*orrm1* mutant to either show an absence of reduction of editing extent (I), or to show a slight (II) or

pronounced (III) reduction of editing extent when compared to the wild-type plant. Acrylamide gels separate the PPE products obtained from the wild-type (wt) and the the Zm-*orrm1* mutant plant. E, edited; U, unedited; P, primer. The name of the site assayed is given above each gel.

Table 3.2. Editing extent of the plastid sites in the maize Zm-*orrm1* mutant

Gene	Position in maize cds	Position in Arabidopsis cds	Editing extent								
			Bulk			PPE			Arabidopsis[1]		
			Zm-*orrm1*	wt	ΔZm-*orrm1*	Zm-*orrm1*	wt	ΔZm-*orrm1*	At-*orrm1*	Col	ΔAt-*orrm1*
atpA	1148	T in Arabidopsis	0.1	1	0.90						
matk	1258	T in Arabidopsis	0.1	0.9	0.89						
ndhA	50	T in Arabidopsis	0.1	1	0.90						
ndhA	473	T in Arabidopsis	1	1	0.00						
ndhA	563	T in Arabidopsis	1	1	0.00						
ndhA	1070	T in Arabidopsis	0.9	1	0.10						
ndhB	467	467	0.1	1	0.90	0.31	0.97	0.68	0.05	0.98	0.95
ndhB	586	586	0	1	1.00	0.07	0.98	0.93	0.05	0.98	0.95
ndhB	611	no editing in Arabidopsis	1	1	0.00						
ndhB	737	T in Arabidopsis	0.1	1	0.90						
ndhB	830	830	0.6	1	0.40	0.65	0.99	0.34	0.28	0.94	0.71
ndhB	1481	1481	1	1	0.00	0.99	0.99	0.00	0.99	0.98	0
ndhD	878	878	0.2	1	0.80	0.29	0.96	0.70	0.02	0.87	0.97
ndhF	62	T in Arabidopsis	1	1	0.00						
ndhG	347	T in Arabidopsis	0.9	1	0.10						
ndhG 5'UTR			0.9	1	0.10						
petB	668	T in Arabidopsis	0	1	1.00						
rpl2	2	T in Arabidopsis	0.7	0.9	0.22						
rpl20	308	T in Arabidopsis	0.2	0.9	0.78						
rpoB	467	T in Arabidopsis	0.6	0.9	0.33						
rpoB	545	551	0	0.9	1.00	0.14	0.89	0.84	0.25	0.96	0.74
rpoB	560	T in Arabidopsis	0.6	0.9	0.33						
rpoB	617	T in Arabidopsis	0.4	0.8	0.50						
rpoC2	2774	T in Arabidopsis	0.9	1	0.10						
rps14	80	80	0.9	1	0.10	0.9	0.97	0.07	0.97	0.94	-0.03
rps8	182	T in Arabidopsis	0.3	1	0.70						
ycf3	44	T in Arabidopsis	0.2	1	0.80						
ycf3	185	T in Arabidopsis	0	1	1.00						

The editing extent values in bulk sequencing are approximate measurements based on the surface of the peaks of unedited (C peak) and unedited (T peak), editing extent: EE= T/(C+T). Editing extent measurements by PPE come from the measure of the intensity of the extension products bands
Δ*orrm1*: variation of editing extent in orrm1 mutant = (EE (wt)-EE (orrm1))/(EE(wt)

[1]Editing extent in Arabidopsis corresponds to the gene specific RNA library.

These RNA editing defects correlate well with the protein deficiencies in Zm-*orrm1* mutants. The failure to edit *petB*-668 could account for the severe PetD deficiency, as this editing event is essential for the accumulation of the cytochrome b_6f complex (26). The reduction in *ycf3*-185 and *ycf3*-44 editing is likely to account for the loss of PsaD, as Ycf3 is required for the assembly and accumulation of photosystem I (27, 28). The minor loss of other proteins likely results from compromised chloroplast transcription and translation, possibly due to defects in the

editing of *rpoB*, *rpl20*, and *rps8* (encoding subunits of the plastid RNA polymerase, large ribosomal subunit, and small ribosomal subunit, respectively).

The maize and Arabidopsis orthologous *ORRM1* genes evidently play similar roles in editing of the C targets that are shared between the two species. Sites such as *ndhB*-586 exhibit a pronounced reduction of editing extent in both Arabidopsis and maize mutants (Tables 3.1, 3.2). In contrast, the editing extents of *ndhB*-1481 and *rps14*-80 exhibit little or no change in either the Arabidopsis or maize mutants (Tables 3.1, 3.2).

The similarity in editing function between the maize and Arabidopsis orthologs was further analyzed by transfecting Arabidopsis mutant protoplasts with the Zm-*ORRM1* under the control of either the 35S or the cytomegalovirus immediate-early promoter (CMV) promoter. The CMV promoter is used in mammalian expression systems; however, apparently it is able to drive the expression of Zm-*ORRM1* in Arabidopsis protoplasts (Figure 3.13A). Unexpectedly the editing extent of six plastid sites was significantly higher in mutant protoplasts transfected with the CMV construct than with the 35S promoter-containing construct (Figure 3.13A). In fact, the editing of *rpoA*-200 was partially complemented by the CMV maize construct but not with the 35S construct carrying the maize gene (Figure 3.13A). Expression of *Zm-ORRM1* from the CMV promoter may be higher than the 35S promoter in these experiments. The majority of the plastid sites experienced a significant increase of editing extent in transfected mutant protoplasts independently of the promoter used in the construct (Figure 3.13A).

Taken together, the similarity in the editing defects in the maize and Arabidopsis mutants and the transfection experiments with the Zm-*ORRM1* strongly support a similar function in editing for the maize and Arabidopsis genes. The maize and Arabidopsis RRM domains are highly conserved (Figure 3.13B), which likely explains the cross-species complementation of function.

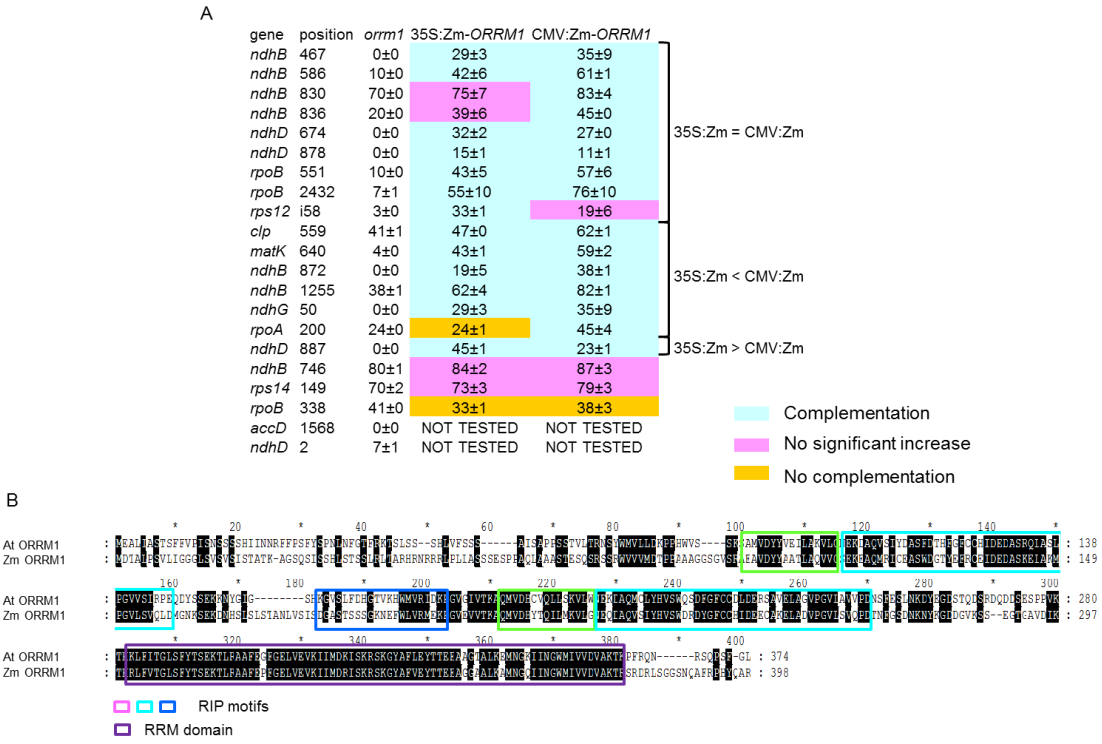
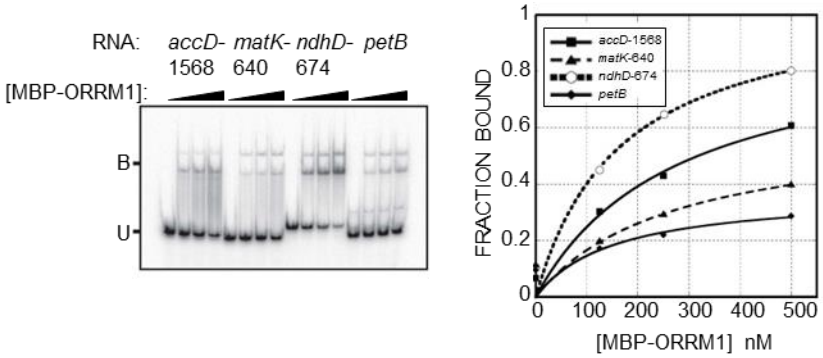


Figure 3.13. Zm-ORRM1 is functionally and structurally similar to At-ORRM1. (A) Zm-ORRM1 is able to complement Arabidopsis orrm1 mutant protoplasts when expressed either under a 35S or a CMV promoter. (B) Alignment of At-ORRM1 and Zm-ORRM1 shows highly conserved RRM domains between the two proteins

Recombinant At-ORRM1 binds near several ORRM1-dependent editing sites *in vitro.*

ORRM1 includes an RRM domain and so was anticipated to be an RNA-binding protein. However, ORRM1 could potentially act in one of two ways: (i) it could bind RNA non-specifically, relying on recruitment to specific editing sites by interaction with a PPR specificity factor; or (ii) it could contribute to editing site choice by binding with specificity near its targets. To address these alternatives, we performed RNA binding assays with purified recombinant ORRM1 fused to maltose binding protein (MBP-ORRM1) (see Methods). Gel mobility shift assays were used to monitor binding to synthetic RNAs mapping between -40 and +19 with respect to the ORRM1-dependent editing sites *ndhD*-674, *accD*-1568, and *matK*-640. Two sequences lacking editing sites were used as negative controls: a 60-mer spanning the Arabidopsis *petB* 5'UTR and a 50-mer from the *psbH* 5'UTR. MBP-ORRM1 consistently bound with much higher affinity to the *ndhD* RNA than to the negative controls (Figure 3.14 and Figure 3.15). MBP-ORRM1 also bound



RNA:

ndhD-674 CUAUUUUAUAUUGGAUUUCUUAUUGCUUUUGCCGUCAAAUCACCGAUUAUACCUUUACAU
accD-1568 UUCAAAUCAUUUUUUUUUUAUCGAUAUUCUUGUUUACUACUCAGUAAACCUCUAUCAACAA
matK-640 AAUUUUGAAUCCAAGAUUUUUCUUGUUCUUAUAUAAUUCUCAUGUAUGUGAAUACGAAUC
petB UGGUAGUUCGACCGCGAAAUUUUUUUCUGCAUUGUAUAUUUCCGGAAUAUGAGUGUGUGA

Figure 3.14. RNA binding activity of recombinant ORRM1. Gel mobility shift assays were performed with MBP-ORRM1 at the indicated concentrations, together with the radiolabeled RNAs shown below (edited site underlined). The *petB* sequence is not edited and serves as a negative control. MBP has been shown to lack RNA binding activity under the conditions used here (66). B, bound RNA; U, unbound RNA.

preferentially to the *accD* RNA in comparison to the *petB* negative control, albeit with lower affinity than it bound to the *ndhD* RNA. However, MBP-ORRM1 exhibited only minimal preference for the *matK* RNA under the binding conditions explored.

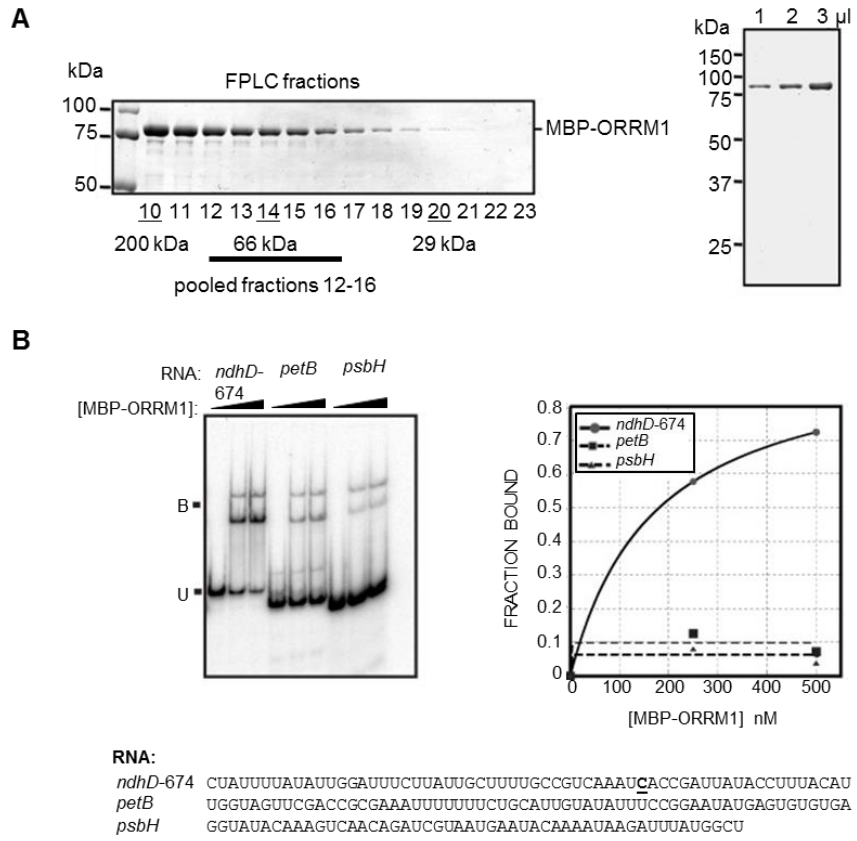


Figure 3.15. Additional evidence for binding of ORRM1 to specific RNA substrates. (A) Purification of recombinant MBP-ORRM1. MBP-ORRM1 was purified by amylose affinity chromatography followed by size fractionation in a Superdex 200 gel filtration column. Superdex 200 column fractions were analyzed by SDS-PAGE and staining with Coomassie Blue (left panel). Fractions 12-16 were pooled, and dialyzed

against storage buffer. The purity of the final preparation is shown in the gel to the right, which was also stained with Coomassie Blue. (B) Additional gel mobility shift assays with MBP-ORRM1. Assays were performed as in Figure 8 except that NaCl was present at 200 mM.

The boundaries of the *cis*-element required to specify *matK*-640 editing is not known, so it remains possible that ORRM1 interacts specifically with RNA outside the assayed region. Taken together, these results support the view that ORRM1 has intrinsic specificity for sequences near at least some of its RNA targets.

ORRM1 interacts with an editing recognition *trans*-factor through its duplicated RIP moiety.

We have previously shown that RIP1 interacts via its RIP moiety with RARE1, a PPR-DYW *trans*-factor that controls the editing of the plastid editing site *accD*-794 (19). A yeast two-hybrid (Y2H) test was performed between ORRM1 and a series of PPR-PLS *trans*-factors known to control the editing extent of plastid sites. We chose the PPR-PLS motif-containing proteins to be tested based on the effect of the *ORRM1* mutation on the editing extent of the sites they control. CRR28 is required for the editing of *ndhB*-467 and *ndhD*-878 (16) and OTP82 is needed for editing of *ndhG*-50 and *ndhB*-836 (17). The editing extent of these sites is severely reduced to more than 90% in the *orrm1* mutant (Table 3.1). In the Y2H analysis we also included RARE1 and OTP84, which are required for the editing of *accD*-794, *ndhF*-290, *ndhB*-1481and *psbZ*-50, respectively (29, 30). The editing extent of *accD*-794, *ndhF*-290, *ndhB*-1481 and *psbZ*-50 in the *orrm1* mutant is identical to the wild-type plant (Table 3.1).

CRR28 and OTP82 interacted with ORRM1 in the Y2H assay, whereas RARE1 and OTP84 did not (Figure 3.16A). The lack of interaction between ORRM1 and the latter two PPR-PLS editing *trans*- factors is expected given the absence in the *orrm1* mutant of an effect on the editing extent of the sites they control (Table 3.1, Figure 3.16A). By testing constructs with either the RIP or the RRM region of ORRM1, we determined that the RIP region is sufficient to interact with CRR28 and OTP82, whereas the RRM domain is not (Figure 3.16B).

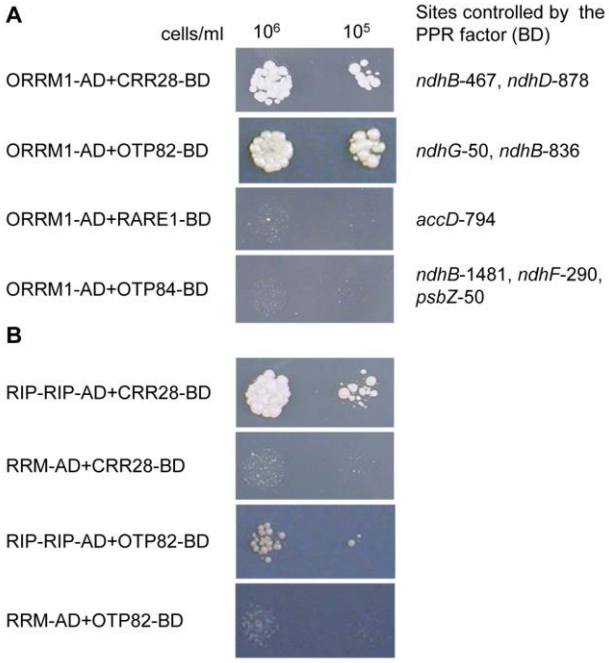


Figure 3.16. ORRM1 interacts selectively with PPR-PLS recognition *trans*-factors via its RIP domain in a yeast two-hybrid assay. (A) Yeast colonies were able to grow in selective media (-histidine) only when ORRM1 fused to the activating domain (AD) and CRR28 or OTP82 fused to the binding domain (BD) were co-expressed into transformed yeast. (B) The interaction between ORRM1 and CRR28 or OTP82 is mediated by the RIP domain of ORRM1 as yeast colonies were able to grow in selective media (-histidine) only when the RIP region of ORRM1 fused to the activating domain (AD) and CRR28 or OTP82 fused to the binding domain (BD) were co-expressed into transformed yeast. No yeast growth was observed when the RIP-AD fusion protein was substituted by the RRM-AD fusion protein. The interaction between the RIP domain of ORRM1 and OTP82 (B, third panel form the top) is

weaker than the interaction of the full length ORRM1 (A, second panel from the top). None of the constructs used in these experiments showed autoactivation for HIS3 reporter.

Discussion

We report here the identification and characterization of ORRM1, a new editing factor that controls the editing extent of 62% of the plastid Cs targeted for editing in Arabidopsis and 81% of C targets in maize. We have demonstrated that ORRM1 is a true editing factor because the effect cannot be explained as an indirect effect caused by changes in RNA transcript abundance. Reduced transcript abundance can indirectly affect the level of editing extent, if RNA is degraded before it is edited (31), or if transcript levels increase to levels high enough to saturate the editing machinery (32). We have verified that there is no significant difference in transcript abundance between the wild-type and the Arabidopsis *orrm1* mutant by performing Northern blots that are consistent with the cpRNA-seq data (Table 3.1). Therefore RNA editing defects do not correlate with transcript abundance in the mutant. In addition, RNA editing defects in the mutant are site-specific, as demonstrated for transcripts with multiple editing sites such as *accD, ndhD* and *ndhB,* which exhibit reduced editing extent of some sites but not others on the same transcript (Figs. 3.5 and 3.7).

The current model for the specificity of the C to be edited in plant organelles postulates two elements, a *cis* sequence primarily upstream of the targeted C and a

trans-factor that recognizes the *cis*-element. The *cis*-acting elements have been delineated to be about 30 nt surrounding the editing site for both organelles (33, 34). Plant site-specific editing factors belong to the PLS subfamily of the pentatricopeptide repeat (PPR) protein family (14). Binding of the *cis*-element and the PPR-PLS *trans*-factor has been demonstrated in several instances (35-37). Recent reports have demonstrated that RIPs, another small class of proteins, are also plant organelle editing trans-factors (19, 20). Although the molecular function of the RIPs remains unknown, they are essential components of the plant RNA editing machinery; *rip* mutants exhibit severe defects in organelle editing. We have shown that RIP1 functions in an editing complex with RARE1, a plastid PPR-PLS protein, and that it binds RARE1 via its RIP-containing moiety (19). We have demonstrated in this study that the portion of ORRM1 that is similar to the RIP family is able to bind to PPR-PLS motifs that are found on a *trans*-factor that controls the editing of sites for which ORRM1 is required. The specificity of ORRM1 in the editing of particular sites might thus sometimes be achieved through binding to particular PPR-PLS recognition factors. We have demonstrated that CRR28 can interact directly with ORRM1, but interactions with other PPR proteins might be indirect, mediated through binding to other RIPs, as some RIPs have been shown to interact with each other (20).

ORRM1, unlike true members of the RIP family, possesses a duplicated set of truncated RIP motifs (Figure 3.1). This unconventional structure, coupled with the presence of the RRM domain not present in other members of the RIP family, suggests that the gene encoding this protein might have originated through recombination during evolution. There are numerous examples of associations of the RRM with other

domains; 21% of the RRMs in eukaryotic proteins are found in association with other domains (21). For example, in Arabidopsis, the mitochondrial ribosomal protein RPS19 is nuclear-encoded and carries an N-terminal RRM. The RPS19 protein is thought to have originated from a fusion from a genomic RRM-encoding gene and a mitochondrial *rps19* gene that was transferred to the nucleus (38). A MAST search for sequences in the non-redundant protein database with the Arabidopsis RIP motif defined by the MEME software (Figure 3.1) returned many proteins; however, the ones carrying the twin truncated RIP domains, both in dicots and monocots, always contain a downstream RRM domain (Figure 3.17). The results here show that the fusion between the twin RIP domains and the RRM predates the monocot/dicot split and strongly suggest that the ancestral gene was involved in RNA editing.

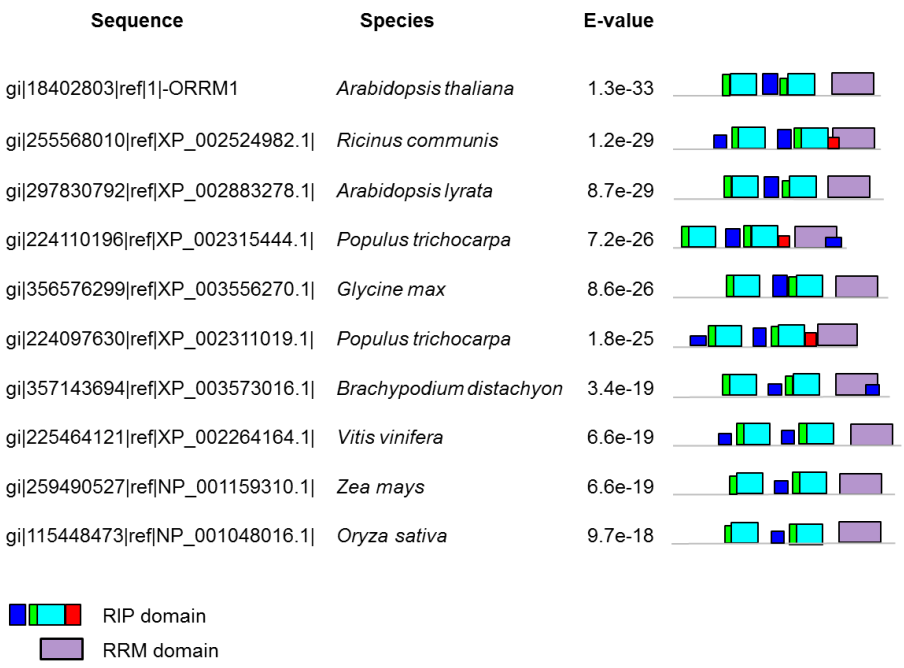


Figure 3.17. Known proteins that contain a twin truncated RIP-RIP always carry a RRM domain. The RIP domain defined by MEME was used to interrogate a non-redundant protein database with the MAST search program. All the significant hits

carrying a set of truncated RIP-RIP, like ORRM1, also contain a downstream RRM domain.

Among the 11 RIP motif-containing proteins found in Arabidopsis, ORRM1 is the only member that has a known domain in addition to the RIP motif. The RRM domain present at the C terminus of ORRM1 is one of the most common protein domains in eukaryotes, and its involvement has been demonstrated in many post-transcriptional events, such as pre-mRNA processing, splicing, mRNA stability, and RNA editing (21). The mammalian apobec-1 complementation factor (ACF1) contains three RRM domains; with apobec-1, which carries the cytidine deaminase activity, ACF1 constitutes the minimal editosome needed for editing of *apo-B* mRNA *in vitro* (39). ACF binds to the *apo-B* mRNA *in vitro* and *in vivo* and is thought to attach to the mooring sequence of *apo-B* mRNA and to dock apobec-1 to deaminate its target cytidine (39). Though an RRM-containing protein is involved in mammalian editing, the function of ACF1 is most analogous to the PPR proteins' C target recognition role in plant editing.

Complementation of the editing defect in the *orrm1* mutant by the sole RRM domain of ORRM1 was an unexpected observation. We speculate that the rescue of the editing defect by the RRM domain at a number of sites when protoplasts are transfected may be due to the high level of expression often achieved during transient expression. The RNA binding studies (Figs. 3.14 and 3.15) indicate that the ORRM1 RRM exhibits at least some specificity for particular RNA sequences. In wild-type organelles, perhaps interaction of the RIP domains with PPR proteins plays a role in bringing the RRM domain in close proximity to the relevant RNA sequence.

In addition to ORRM1, another RRM-containing protein named CP31 has been implicated in plastid editing. CP31 belongs to a small family of ten chloroplast ribonucleoproteins (cpRNPs), all of which contain a twin RRM and an acidic amino-terminal domain (40, 41), but are in a different clade than ORRM1 (Figure 3.3). CP31 was reported by Hirose and Sugiura (42) to be a common factor for editing of *psbL* and *ndhB* mRNAs *in vitro*. Immunodepletion of CP31 from the editing extract resulted in the inhibition of editing of *psbL* and *ndhB* mRNAs. More recently, a null mutant of *CP31A*, one of the two paralogues found in Arabidopsis, was shown to exhibit multiple specific editing defects in chloroplast transcripts (43). However, Tillich et al (43) also observed that CP31 was responsible for the stability of specific chloroplast mRNAs, because almost no *ndhF* mRNA could be detected, and other chloroplast mRNAs were also depleted in *cp31a* mutants. In contrast no transcripts were reduced in amount in *orrm1* (Table 3.1). The editing defect in *cp31a* mutant and the *cp31a/cp31b* double mutant is much less severe than ones that we observed in the orrm1 mutant; an edited peak was detectable in the electrophoretograms of RT-PCR bulk sequences surrounding the editing sites most affected in *cp31a/cp31b* mutant [Figure 3.7 in (43)]. Bulk sequencing is a much less sensitive editing assay than either RNA-seq or PPE. If we had chosen bulk sequencing to assay the editing extent in the *orrm1* mutant plant, there would not have seen any detectable edited peak for the 12 sites whose editing extent in the *orrm1* mutant is < 0.1 (Table 3.1). Recently, Kupsch et al. (44) found that CP31A associates with large transcript pools and confers cold stress tolerance by influencing multiple chloroplast RNA processing events (44). The authors indicate that relative to its effect on RNA stability, the effect of CP31A on editing extent is minor.

The RRM domain of ORRM1 is most related to RRMs found in glycine-rich RNA-binding proteins (GR-RBPs) and mitochondrial RNA-binding proteins (mRBPs) as well as RRMs found in a group referred to as small RBPs (S-RBPs) (Figure 2),(23). GR-RBPs have been shown to be involved in the plant's response to environmental stresses, particularly cold (45, 46). However, little is known about the molecular function of either GR-RBPs or mRBPs. Recently a rice GR-RBP protein named GRP162, which is likely orthologous to either Arabidopsis GR-RBP7 or GR-RBP8, was shown to be part of a restoration of fertility complex (47). GRP162 interacts *in vivo* with RF5, a PPR protein encoded by the fertility restorer gene *Rf5* to the Hong-Lian cytoplasmic male sterility (CMS). GRP162 was also shown to bind *in vitro* and *in vivo* to *atp6-orfH79*, the CMS-associated transcript (47), which is cleaved in the fertility-restored line. Like GRP162, ORRM1 interacts with a PPR protein and can bind to a transcript targeted for editing.

Table 3.3. Annotation and subcellular localization of the RRM-containing proteins related to ORRM1

Gene		Protein annotation[1]	E-value [2]	Location[3]			
				GFP	MS/MS	Predotar	TargetP
ORRM2	At1g73530	RRM similar to S-RBP	5.00E-23		P	P	P
ORRM3	At5g06210	mRBP, S-RBP11	2.00E-20			M	M
ORRM4	At4g20030		4.00E-19		N	P	P
ORRM5	At2g37510	mRBP, S-RBP9	5.00E-19			M	M
ORRM6	At3g46020	S-RBP4	5.00E-19			M	M
ORRM7	At3g23830	mRBP1b, GR-RBP4	8.00E-19	M		M	M
ORRM8	At2g27330		2.00E-18			ER	M
ORRM9	At4g13850	mRBP1a, GR-RBP2	2.00E-18		M	M	M
ORRM10	At3g26420	AtRZ-1A (GR-RBP)	5.00E-17	N	C, Px, Pl M		
	At5g47320	mRBP, RPS19	6.00E-17		M	M	M
ORRM11	At5g54580	mRBP, S-RBP12	1.00E-16	M		M	P
ORRM12	At4g39260	GR-RBP8	3.00E-15	N	N, Px, Pl M, P		
ORRM13	At5g61030	mRBP2b, GR-RBP3	2.00E-14		M	M	M
ORRM14	At5g59860	S-RBP13	3.00E-14				
ORRM15	At1g74230	mRBP2a, GR-RBP5	2.00E-13		M	M	M

[1] annotation from Lorkovic and Barta (2002) and Vermel et al. (2002)

[2] e-value obtained when the RRM motif from *ORRM1* is used as a query

[3] subcellular location from the SUBA database (Heazlewood at al. 2007). C: cytosol, ER: endoplasmic reticulum, M: mitochondrion, N: nucleus, P: plastid, Pl M: plasma membrane, Px: peroxysome

ORRM1 is the only well-characterized member of the ORRM clade of Arabidopsis proteins (Fig 3.3). Among the 15 proteins whose RRM domains are most similar to the one found in ORRM1 there are seven GR-RBPs, eight mRBPs, and six S-RBPs (Table 3.3). There is overlap of the annotated mRBPs with both GR-RBPs and S-RBPs. Ten of these proteins are predicted to be targeted to either the plastid or the mitochondrion by both Predotar and TargetP, two prediction programs for subcellular localization of proteins (Table 3.3) (48, 49). Five of these proteins that have a strong *in silico* prediction for organelle targeting were found in the respective organelle by proteome MS/MS analysis (Table 3.3). None of these proteins have known functions except for the ribosomal protein RPS19. The ORRM sub-family of RRM proteins are obvious targets for further analysis in order to determine whether other family members are involved in plastid or mitochondrial editing.

Materials and Methods

Plant material. The Arabidopsis T-DNA insertion line SALK_072648 was obtained from the ABRC stock center. The wild-type plants come from several segregating T-DNA populations, all in the Col-0 background which is similar to SALK_072648. WiscDsLox419C10 provided the wt plant for the total plastid RNA-seq while SAIL156A04, SAIL731D08, SALK016801, SALK114438, and GK-109E12 provided the wt plants for the gene specific RNA-seq, wt-1, wt-2, wt-3, wt-4, and wt-5 respectively. Plants were grown in 14 h of light/10 h of dark under full-spectrum fluorescent lights in a growth room at 26 ℃. Genotyping was done by PCR with

Qiagen Taq PCR master mix and primers listed in Table 4.4. Bulk-sequencing of the PCR product specific for the T-DNA insertion was done at Cornell University Life Sciences Core Laboratories Center.

The Zm-*orrm1-1* mutant was originally detected during the profiling of pigment and protein defects in maize mutants in the Photosynthetic Mutant Library (50) (http://pml.uoregon.edu/photosyntheticml.html): homozygous mutants were pale green and seedling lethal, with strongly reduced levels of photosystem I and cytochrome b6f proteins, and modest losses of Rubisco, ATP Synthase and photosystem II proteins. An Illumina-based method (25) was used to identify Mu insertions that cosegregate with the mutant phenotype. An insertion in gene GRMZM5G899787 emerged from this analysis as the best candidate for the causal mutation because of the exonic location of the insertion and the fact that the gene encodes a predicted chloroplast protein related to proteins known to be involved in chloroplast gene expression. A second allele was identified during the large-scale sequencing of *Mu* insertions in each mutant in the PML collection. Complementation crosses between plants heterozygous for the two alleles yielded chlorophyll deficient progeny whose protein deficiencies were similar to those in the parental alleles (see Figure 3.9), confirming that these defects result from disruption of GRMZM5G899787. GRMZM5G899787 is predicted to be the ortholog of *ORRM1* (At3g20930) by two independent ortholog prediction algorithms: OrthoMCL employed at the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/), and the Ensembl pipeline employed at Gramene (http://www.gramene.org). Protein

extraction, RNA extraction, and immunoblotting were performed as described previously (51).

Table 4.4. Oligonucleotides used in this study

Primers	Sequences	Purpose
SALK_072648-LP	TGAACGATTTTATGATTGACGG	genotyping
SALK_072648-RP	AACCCGAAATGGGTATCAAAG	genotyping
ORRM1_1F	ATG GAA GCT CTT ATT GCT TCC ACT TC	complementation
ORRM1_1F_CACC	caccATG GAA GCT CTT ATT GCT TCC ACT TCC	complementation
ORRM1_822R	cta TGA ATC ATC TTG ATC TCT TGA ATC TTG CGT	truncation/Y2H/complementation
RecA_1F_CACC	caccATGGATTCACAGCTAGTCTTGTCTCTG	complementation
RecA_ORRM1-C	CTT TGT CTT TAC GGG AGG AGA CTC GTC GCG ATC GAA TTC AGA ACT GAT TTT GTG GGA G	complementation
ORRM1_823F	GAGTCTCCTCCCGTAAAGACAAAG	truncation/Y2H/complementation
ORRM1_R	CTA GAG CCC GAA ACT TGG TTG	Y2H/complementation
OTP84_133F	GCCTCCGCCGTTTCTGGCGCA	Y2H
OTP84_R	TCA CCA ATA GTC TCC ACA GGA GCA	Y2H
CRR28_121F	GCCTCCACCGCCGGTAACCAT	Y2H
CRR28_R	CTA CCA GTA GTC TAA ACA AGA GCA GGA	Y2H
ORRM1_151F	CTCGTCTTCTCATCTTCTGCAATT	Y2H
OTP81_127F	CTCCGACAACTAAAGCAAACCCAT	Y2H
OTP81_R	TCA CCA GAA ATC GTT ACA GGA ACA	Y2H
OTP82-292F	AACCTGTTGATTTGGAACACGATGTTT	Y2H
OTP82-R	CTACCAGTAGTCATTGCAGGAACAAACA	Y2H
RARE1_100F	GGATCCATGTCGAGCACTTCTTCTCCGTCT	Y2H
RARE1_R	TCACCAAGTAATCGTTGCAAGAACA	Y2H
ORRM1_163F_BamHI	tatataggatccTCTTCTGCAATTTCCGCACCGCCT	protein expression
ORRM1_R_SalI	tatatagtcgacCTA GAG CCC GAA ACT TGG TTG ACT	protein expression
ndhB-F	TTTTATGTGGTGCTAACGATTTAA	probe for RNA blot
ndhB-R	AATCGCAATAATCGGGTTCATT	probe for RNA blot
ndhD-F	CATGTGGGGTGGAAAGAAAC	probe for RNA blot
ndhD-R	AGCGCCAATAAATCCATGAG	probe for RNA blot
atpA-maize-F1	GAGCCGCTAAATTAAATTCTCTTT	RT-PCR bulk sequencing
atpA-maize-R1	ATCCTCTCGTTCCGGTATAAATAG	RT-PCR bulk sequencing
ndhA-maize-F1	TAGGGTAGAGGTAGAAACTATCAA	RT-PCR bulk sequencing
ndhA-maize-R1	ATTCTGCCAAAGAAGAAATTAGAA	RT-PCR bulk sequencing
ndhA-maize-F2	TTTGGCAGAATGTGAAAGATTACC	RT-PCR bulk sequencing
ndhA-maize-R2	GAACCCAGTTAGCATAGGGAACAT	RT-PCR bulk sequencing
ndhB-maize-F1	CTATCCGTAGAGTACATTGAATGT	RT-PCR bulk sequencing
ndhB-maize-R1	CTAAAAGAGGGTATCCTGAGC	RT-PCR bulk sequencing
ndhD-maize-F1	TGAACCCGGATTAGATTTAGAAAG	RT-PCR bulk sequencing
ndhD-maize-R1	CCGTTCCCGCCAAGAAA	RT-PCR bulk sequencing
ndhF-maize-F1	ATCAATATGCCTGGGTAATTCCTC	RT-PCR bulk sequencing

ndhF-maize-R1	ATAGGAACACATTCCCACAAGTTC	RT-PCR bulk sequencing
ndhG-maize-F1	GTCAGTTCATGAAAAATTTTATAC	RT-PCR bulk sequencing
ndhG-maize-R1	GAGCCATAGTAATTGCACCTATTA	RT-PCR bulk sequencing
petB-maize-F1	GAGGCCAACTTTGGTTGGTTAATC	RT-PCR bulk sequencing
petB-maize-R1	GGACCCGAAATACCTTGCTTACG	RT-PCR bulk sequencing
rpl2-maize-F1	CAACCGGGTTATTCTATTCCACTT	RT-PCR bulk sequencing
rpl2-maize-R1	TAGCCCCTCTGGGATGTAAAATAT	RT-PCR bulk sequencing
rpl20-maize-F1	GGGGGCTCATTTAAGACTTAA	RT-PCR bulk sequencing
rpl20-maize-R1	TTGGAAATCGTGTAAAGATTATTT	RT-PCR bulk sequencing
rpoB-maize-F1	TATCCGCGAGATTAATTTTTGGTT	RT-PCR bulk sequencing
rpoB-maize-R1	TGTAATTCCTCGCATAAGGACTCC	RT-PCR bulk sequencing
rpoC2-maize-F1	CATATCTTGCCGAGATCCTCATCC	RT-PCR bulk sequencing
rpoC2-maize-R1	CGCGAATTAGATCATTTGTTTTTA	RT-PCR bulk sequencing
rps8-maize-F1	GGCAAGGACACTATTGCTGATTTA	RT-PCR bulk sequencing
rps8-maize-R1	CTCCCCCAATTCTGTTTAGTCG	RT-PCR bulk sequencing
rps14-maize-F1	GGAGAAGAAGCGGCAGAAATT	RT-PCR bulk sequencing
rps14-maize-R1	GTCCGGATAGCCCAAAGTCTC	RT-PCR bulk sequencing
ycf3-maize-F1	ATGCCTAGATCCCGTATAAATGG	RT-PCR bulk sequencing
ycf3-maize-R1	TTCCGAATCGCCCTGTAGAA	RT-PCR bulk sequencing

Phylogenetic analysis

Protein alignments were achieved by using ClustalX 2.1 (52) and adjusted manually. The construction of phylogenetic trees was performed with MEGA5 (53). The presented tree was inferred using the Neighbor-Joining method (54) and evolutionary distances were computed using the Poisson correction method. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. Trees were also constructed using the UPGMA, maximum-likelihood (ML), maximum-parsimony (MP) and minimum-evolution (ME) methods available on the MEGA5 software. 1000 bootstrap replications were performed to determine the confidence level of the phylogenetic tree topology. Only representative of RRMs were

used to construct the tree to avoid over-representation of certain groups, which would change the tree artificially.

Measure of editing extent

RNA extraction and RT-PCR methods were as described in (4) and chloroplast isolation as described in Hayes and Hanson (55). Primers to amplify the mitochondrial and plastid transcripts have been described (4, 29, 56). Analysis of RNA editing by PPE was done as in (29). The editing extent in maize plastid genes was measured primarily by bulk-sequencing of RT-PCR products amplified with primers listed in Table 4.4.

Measure of editing extent by RNA-seq was done by sequencing two kinds of templates, either cDNAs corresponding to organelle gene transcripts and amplified with organelle gene-specific primers, or cDNAs corresponding to the whole plastid transcriptome and reverse transcribed with random hexamers. Gene-specific organelle cDNAs were quantified, mixed in equimolar ratio, and sheared by sonication; the sheared cDNA entered the workflow of low-throughput protocol for TruSeq™ RNA Sample Preparation Guide at the step of performing end repair. cDNAs corresponding to the whole plastid transcriptome were obtained by using total RNA prepared from plastid purified fraction on a percoll gradient (55); the RNA entered the workflow of low-throughput protocol for TruSeq™ RNA Sample Preparation Guide at the step of "elute, fragment, prime" RNA.

In the analysis, we used post-filter (PF) Illumina reads. After trimming the low-quality bases from both ends using the default settings of the seqtk trimfq program (https://github.com/lh3/seqtk), the resulting reads were aligned to the NCBI *Arabidopsis thaliana* chloroplast genomic template (NC_000932) using the tophat program (57) with the default settings of 2 mismatches allowed per read. The C to T editing sites were determined using a combination of the programs samtools (58) and bedtools (59) and excel spreadsheets. The criteria were as followed: i) the reference allele was C ii) the two major alleles were C and T, iii) the sum of all alleles' depth (if any) was at most 20% of the depth of the second major allele, iv) total C+T read depth was at least 20 and v) the T fraction (T fraction = T/(C+T)) was ≥ 5%.

RNA blots

RNA gel blot analysis was performed as described in Germain et al. (60). Primers used to make the probes are shown in Table S4.

Constructs used in this study (all the primers are listed in Table S4)

Complementation constructs. ORRM1_1F_CACC and ORRM1_822R were used to amplify the N terminal *ORRM1*, followed by TOPO cloning into pENTR/SD/D vector (Invitrogen). RecA _1F_CACC, RecA_ORRM1-C, ORRM1_823F, ORRM1_R were used in an overlapping PCR to amplify *ORRM1* C terminus fused with a RecA transit peptide sequence, followed by TOPO cloning into pENTR/SD/D vector. ORRM1_1F

and ORRM1_R were used to amplify the full length *ORRM1* coding sequence with the stop codon. These vectors were used in LR Clonase II recombination reactions with pEXSG-EYFP (61) to generate the full-length, N terminal, and C terminal *ORRM1* constructs driven by a 35S promoter.

Maize ORRM1 complementation constructs. Maize cDNA clone Zm_BFb0091M02 was obtained from the Maize Full Length cDNA Project. BP reaction was performed using pDONR201 (Invitrogen) and the cDNA clone, followed by LR reaction with pEXSG-EYFP (61) to clone the cDNA under the CaMV 35S promoter.

Yeast Two Hybrid assay constructs. Coding sequences of mature PPR proteins and full length, N-terminus, or C-terminus encoding portions of *ORRM1* were amplified with respectively, and cloned into the PCR8/TOPO/GW vector. These sequences were then shuttled into the pGADT7GW and pGBKT7GW vectors (62) to create GAL4 activation domain (AD) fusion and DNA binding domain (BD) fusion respectively.

Protein expression constructs. The mature *ORRM1* coding sequence was cloned by PCR using ORRM1_163F_BamHI and ORRM1_R_SalI. The PCR product was cloned into the pMal-TEV vector (63) using restriction digestion and ligation.

Yeast Two-hybrid assay

Two different mating types, α and a, of yeast stain PJ69-4 were used for transformation. The transformation procedures were performed following the original paper (64). SD-leu-trp-his amino acid dropout media (Sunrise Science, CA) were used

to test the interaction. Yeast harboring both bait and prey constructs grown in liquid culture were diluted with sterile water to cell density $1x10^6$, $1x10^5$ cells/ml before spotting onto the plates. The picture of the growth was taken 3 days later. Each Yeast Two Hybrid construct was paired with either AD or BD empty vector to test autoactivation in yeast using the same method with interaction assay. No autoactivation of HIS3 reporter was observed for the constructs used in this paper.

Protoplast transfection

These assays were performed as in (19).

Expression and purification of recombinant ORRM1

MBP-ORRM1 was expressed in *E. coli* from the pMal-TEV vector, enriched by amylose affinity chromatography and further purified by gel filtration chromatography using the method described previously for MBP-APO1 (51). The purity of the final preparation is illustrated in Figure 3.13.

RNA Binding Assays

Synthetic RNAs (Integrated DNA Technologies) were 5′-end–labeled with [γ-32P]–ATP and T4 polynucleotide kinase, and purified on denaturing polyacrylamide gels.

RNA binding reactions contained 100 mM NaCl, 40 mM Tris-HCl pH 7.5, 4 mM DTT, 0.1 mg/ml BSA, 0.5 mg/ml heparin, 10% glycerol, 10 units RNAsin (Promega), RNAs at 15 pM and recombinant protein at the following concentrations: 0, 125, 250, and 500 nM. Reactions were incubated for 20 min at 25 ℃ and resolved on 5% native polyacrylamide gels. Results were visualized on a Storm phosphorimager. Data quantification was performed with ImageQuant (Molecular Dynamics).

ACKNOWLEDGMENTS

We thank Lin Lin for excellent technical assistance, especially with protein expression in *E. coli*. We are grateful to Margarita Rojas (University of Oregon) for performing the RNA binding experiments, and to Rosalind Williams-Carrier and Susan Belcher (University of Oregon) for identifying the Zm-*orrm1* mutants. This work was supported by the National Science Foundation from the Molecular and Cellular Biosciences, Gene and Genome Systems, grants MCB-1020636 (to S.B.), MCB-0929423 (to M.R.H.), National Science Foundation Grant IOS-0922560 (to A.B), and European Molecular Biology Organization Long-Term Fellowship (to K.H.)

References

1. Gott JM, Emeson RB (2000) Functions and mechanisms of RNA editing. *Annu Rev Genet* 34: 499-531.

2. Chateigner-Boutin AL, Small I (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* 35: e114.

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

4. Bentolila S, Elliott LE, Hanson MR (2008) Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* 178: 1693-1708.

5. Gray MW, Covello PS (1993) RNA editing in plant mitochondria and chloroplasts. *Faseb J* 7: 64-71.

6. Chateigner-Boutin AL, et al. (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56: 590-602.

7. Covello PS, Gray MW (1989) RNA editing in plant mitochondria. *Nature* 341: 662-6.

8. Gualberto JM, Lamattina L, Bonnard G, Weil JH, Grienenberger JM (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341: 660-662.

9. Hiesel R, Wissinger B, Schuster W, Brennicke A (1989) RNA editing in plant mitochondria. *Science* 246: 1632-1634.

10. Bock R, Hermann M, Kossel H (1996) *In vivo* dissection of *cis*-acting determinants for plastid RNA editing. *EMBO J* 15: 5052-9.

11. Choury D, Farre JC, Jordana X, Araya A (2004) Different patterns in the recognition of editing sites in plant mitochondria. *Nucleic Acids Res* 32: 6397-6406.

12. Miyamoto T, Obokata J, Sugiura M (2002) Recognition of RNA editing sites is directed by unique proteins in chloroplasts: Biochemical identification of *cis*-acting elements and *trans*-acting factors involved in RNA editing in tobacco and pea chloroplasts. *Mol Cell Biol* 22: 6726-6734.

13. Hayes ML, Hanson MR (2007) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA* 13: 281-288.

14. Lurin C, et al. (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089-2103.

15. Salone V, et al. (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581: 4132-4138.

16. Okuda K, et al. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* 21: 146-156.

17. Okuda K, et al. (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. *Plant J* 61: 339-349.

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

19. Bentolila S, et al. (2012) RIP1, a member of an *Arabidopsis* protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci USA* 109: E1453-61.

20. Takenaka M, et al. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci USA* 109: 5104-5109.

21. Maris C, Dominguez C, Allain FH (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272: 2118-2131.

22. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37: W202-8.

23. Lorkovic ZJ, Barta A (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Res* 30: 623-635.

24. Vermel M, et al. (2002) A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc Natl Acad Sci USA* 99: 5866-5871.

25. Williams-Carrier R, et al. (2010) Use of illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy mutator lines of maize. *Plant J* 63: 167-177.

26. Zito F, Kuras R, Choquet Y, Kossel H, Wollman FA (1997) Mutations of cytochrome b6 in *Chlamydomonas reinhardtii* disclose the functional significance for

a proline to leucine conversion by petB editing in maize and tobacco. *Plant Mol Biol* 33: 79-86.

27. Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix JD (1997) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J* 16: 6095-6104.

28. Ruf S, Kossel H, Bock R (1997) Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I-related gene. *J Cell Biol* 139: 95-102.

29. Robbins JC, Heller WP, Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast *accD* transcript. *RNA* 15: 1142-1153.

30. Hammani K, *et al* (2009) A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21: 3686-3699.

31. Lu B, Hanson MR (1992) A single nuclear gene specifies the abundance and extent of RNA editing of a plant mitochondrial transcript. *Nucleic Acids Res* 20: 5699-703.

32. Chateigner-Boutin AL, Hanson MR (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared *cis* elements. *Mol Cell Biol* 22: 8448-56.

33. Hayes ML, Reed ML, Hegeman CE, Hanson MR (2006) Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript *in vivo* and *in vitro*. *Nucleic Acids Res* 34: 3742-3754.

34. Neuwirt J, Takenaka M, van der Merwe JA, Brennicke A (2005) An *in vitro* RNA editing system from cauliflower mitochondria: Editing site recognition parameters can vary in different plant species. *RNA* 11: 1563-70.

35. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281: 37661-37667.

36. Hammani K, et al. (2011) The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in *Arabidopsis* mitochondria. *J Biol Chem* 286: 21361-21371.

37. Tasaki E, Hattori M, Sugita M (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J* 62: 560-570.

38. Alba MM, Pages M (1998) Plant proteins containing the RNA recognition motif. *Trends in Plant Science* 3: 15-21.

39. Mehta A, Kinter MT, Sherman NE, Driscoll DM (2000) Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol Cell Biol* 20: 1846-54.

40. Ye LH, et al. (1991) Diversity of a ribonucleoprotein family in tobacco chloroplasts: Two new chloroplast ribonucleoproteins and a phylogenetic tree of ten chloroplast RNA-binding domains. *Nucleic Acids Res* 19: 6485-6490.

41. Ruwe H, Kupsch C, Teubner M, Schmitz-Linneweber C (2011) The RNA-recognition motif in chloroplasts. *J Plant Physiol* 168: 1361-1371.

42. Hirose T, Sugiura M (2001) Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: Development of a chloroplast *in vitro* RNA editing system. *Embo J* 20: 1144-52.

43. Tillich M, *et al* (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci USA* 106: 6002-6007.

44. Kupsch C, et al. (2012) Arabidopsis chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell* 24: 4266-4280.

45. Kim YO, Kim JS, Kang H (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. *Plant J* 42: 890-900.

46. Kim JS, et al. (2007) Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*. *Nucleic Acids Res* 35: 506-516.

47. Hu J, et al. (2012) The rice pentatricopeptide repeat protein RF5 restores fertility in hong-lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. *Plant Cell* 24: 109-122.

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

49. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005-1016.

50. Stern DB, Hanson MR, Barkan A (2004) Genetics and genomics of chloroplast biogenesis: Maize as a model system. *Trends Plant Sci* 9: 293-301.

51. Watkins KP, et al. (2011) APO1 promotes the splicing of chloroplast group II introns and harbors a plant-specific zinc-dependent RNA binding domain. *Plant Cell* 23: 1082-1092.

52. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.

53. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.

54. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.

55. Hayes ML, Hanson MR (2007) Assay of editing of exogenous RNAs in chloroplast extracts of *Arabidopsis*, maize, pea, and tobacco. *Methods Enzymol* 424: 459-482.

56. Bentolila S, Knight WE, Hanson MR (2010) Natural variation in *Arabidopsis* leads to the identification of REME1, a pentatricopeptide repeat-DYW protein controlling the editing of mitochondrial transcripts. *Plant Physiol* 154: 1966-1982.

57. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics* 25: 1105-1111.

58. Li H, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079.

59. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842.

60. Germain A, et al. (2011) Mutational analysis of *Arabidopsis* chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation. *Plant J* 67: 381-394.

61. Jakoby MJ, et al. (2006) Analysis of the subcellular localization, function, and proteolytic control of the *Arabidopsis* cyclin-dependent kinase inhibitor ICK1/KRP1. *Plant Physiol* 141: 1293-1305.

62. Horak J, et al. (2008) The *Arabidopsis thaliana* response regulator ARR22 is a putative AHP phospho-histidine phosphatase expressed in the chalaza of developing seeds. *BMC Plant Biol* 8: 77.

63. Williams-Carrier R, Kroeger T, Barkan A (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 14: 1930-1941.

64. Gietz RD, Schiestl RH, Willems AR, Woods RA (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* 11: 355-360.

65. Barkan A (1993) Nuclear mutants of maize with defects in chloroplast polysome assembly have altered chloroplast RNA metabolism. *Plant Cell* 5: 389-402.

66. Kroeger TS, Watkins KP, Friso G, van Wijk KJ, Barkan A (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. Proc Natl Acad Sci USA 106: 4537-4542.

Chapter 4

An ORRM1-binding zinc finger protein, VAR3 is a novel plastid editing factor

Introduction

In higher plants RNA editing converts specific cytidines to uridines in the organeller transcripts (1–3). A typical land plant has around 30 C targets in chloroplasts and over 500 C targets in mitochondria. Most of the editing events occur in the coding region which is believed to be a correction mechanism to restore functional mRNAs in chloroplasts and mitochondria (4).

Although plant RNA editing has been known for over twenty years, the components of the editing machinery are still largely unknown. Specificity of editing is achieved through the recognition between a PPR *trans-acting* factor and *cis*-element 5' adjacent to the editable cytidine (5, 6). Recently both the PPR-RNA recognition code and the crystal structure of a PPR protein were released. One would be able to predict a given PPR protein's binding sequence without laborious genetic screening (7, 8). The DYW domains present in the C termini of some editing PPR proteins share sequence similarity with known cytidine deaminases and therefore the DYW domain is believed to the best candidate for editing enzyme (9). However, some researchers found that the DYW domain is dispensable for editing and attempts to show its deaminase activity have failed (10, 11). A new possibility is suggested by the

discovery of DYW1, which supplies a DYW domain to a non-DYW PPR protein CRR4 *in trans* (12). Whether this is a common scenario for all editing events requires further study.

Discovery of additional editing factors—RIP/MORF protein—showed that the plant RNA editing machinery is much more complicated than anticipated (13, 14). Members of the RIP/MORF protein family participate in up to hundreds of editing events, making them more general factors compared to PPR proteins (14). RIP/MORF proteins can selectively bind to PPR proteins and also form homo-/hetero-dimers. The N-terminal regions of these proteins are highly conserved but no characterized domain can be found, so their exact role in editing is still unknown.

Recently we reported a new plastid editing factor, ORRM1, containing both a duplicated RIP-RIP domain and an RRM (RNA Recognition Motif) domain (15). Its editing activity is carried on the RRM domain, which places it in a different family from the RIP/MORF proteins. ORRM1 belongs to a distinct clade of RRM containing proteins, so it is likely that other members of the clade are also editing factors. In order to identify remaining components of the editosome (200-400kD) (14), I performed a co-immunoprecipitation assay for an epitope tagged ORRM1. Proteomics analysis found one ORRM1 interacting protein—VAR3 (variegated 3). Here I report that VAR3 is a novel plastid editing factor.

Results

A C-terminal tag disrupts the editing activity of ORRM1 but an N-terminal tag does not

In order to study ORRM1 interacting proteins, I generated constructs to epitope-tagged ORRM1 *in planta*. The tag that was used is a three tandem Flag tag fused with a strepII tag, which will be addressed as 3FS in the rest of this dissertation. A construct expressing ORRM1 in fusion with 3FS at its C-terminus was made (Figure 4.1A), and tested in an *orrm1* protoplast transfection assay to examine whether addition of the tag maintains ORRM1's function. As is shown in Figure 4.1B, no significant increase of editing at *matK-640* was observed when an ORRM1-3FS protein was expressed in *orrm1* protoplasts. On the contrary RecA-RRM, a positive control consisting an RRM motif from ORRM1 fused to RecA transit peptide was able to complement the mutant editing defect (From 11% to 71%). Apparently, the C-terminal tag disrupts the editing activity of ORRM1. As an alternative strategy, an N-terminal tagged ORRM1 construct was made. A transit peptide sequence from RecA was used to replace the predicted endogenous chloroplast transit peptide of ORRM1 (1-54aa). The 3FS tag sequence was cloned between the transit peptide and the mature form of ORRM1 (55-374aa). This construct is designated as RecA-3FS-mORRM1. The protoplast complementation assay showed that this N-terminal tagged ORRM1 can significantly increase the editing extent. As is shown in Figure 4.1C, editing of *matK-640* increased from 11% to 20%. The lower complementation level can be largely attributed to the size of the vector used in this assay. While plasmid harboring RecA-RRM is around 6kb, the N-terminal tagged ORRM1 is on a binary vector around 14kb. The lower complementation level is probably due to lower transfection

efficiency which is typical for large (over 10kb) plasmids (16). This construct was then used in a root transformation of *orrm1* mutant plants. Transgenic plants were obtained from Basta resistance selection and confirmed by PCR. No phenotypic difference was seen between the transgenic plants and the mutant or the wild type *Columbia* (15).

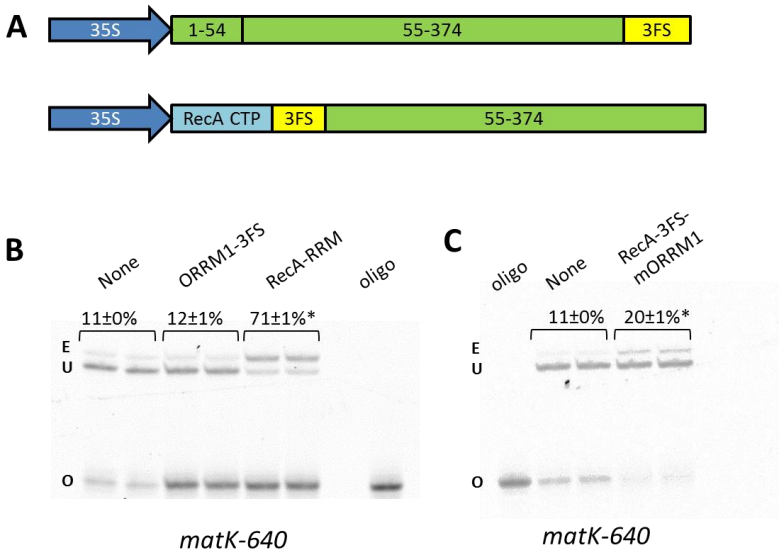


Figure 4.1. Epitope tagged ORRM1 is tested in a transient complementation assay. A, Schematic diagram of the epitope tagged ORRM1 constructs. The predicted transit sequence of ORRM1 is 54 amino acid long located in the N-terminus. Tag was fused to ORRM1 either at the C terminus or between CTP and the 55th amino acid. CTP, Chloroplast Transit Peptide. B, C-terminal tagged ORRM1 could not complement the editing defect at *matK-640* in *orrm1* protoplasts. Editing extent was examined using the poisoned primer extension assay. ORRM1 fused with 3xFlag-strepII at its C-terminus was transiently expressed in *orrm1*protoplasts. A chimeric protein- RRM domain alone from ORRM1 with a transit peptide from RecA was used in this experiment as a positive control. Editing extents were quantified using Image Quant. Asterisk indicates significant difference ($P<0.01$) with the non-transfected control. None, untransfected *orrm1* protoplasts; ORRM1-3FS, *orrm1* protoplasts transfected with construct expressing ORRM1 fused to 3FS at N-terminus; RecA-RRM, *orrm1* protoplasts transfected with construct expressing RecA transit peptide fused to RRM domain of ORRM1. E, edited product; U, unedited product; O, oligo. C, N-terminal tagged ORRM1 can enhance editing of *matK-640* in *orrm1* protoplasts. 3FS was added to the N terminus of mature form of ORRM1. RecA transit peptide sequence

was used to target this chimeric protein to chloroplasts. Star indicates significant difference (P<0.01) with the non-transfected control. None, not transfected *orrm1* protoplasts; RecA-3FS-mORRM1, *orrm1* protoplasts transfected with this construct. E, edited product; U, unedited product; O, oligo.

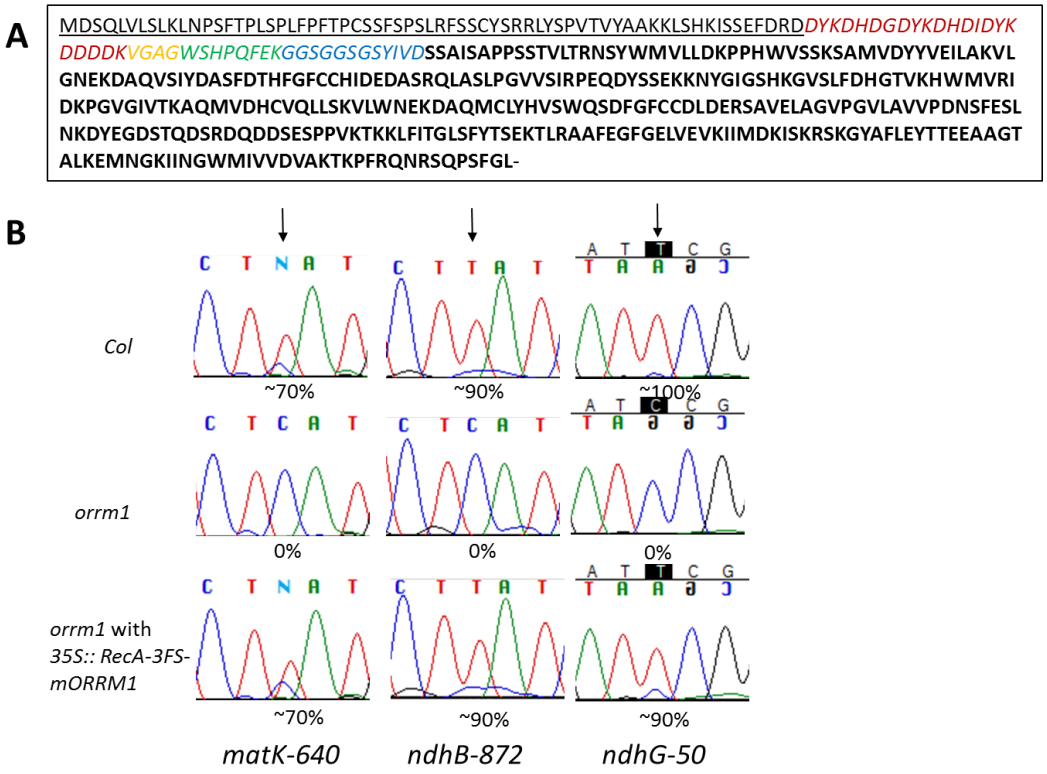


Figure 4.2. Stable integration of 35S::RecA-3FS-mORRM1 into the *orrm1* mutant restores normal editing level in plastids. A, protein sequence of RecA-3FS-mORRM1. Transit peptide sequence from RecA is underlined. Sequence of epitope tags is italic. 3xFlag, spacer, StrepII and Glycine-Serine linker are labeled with red, yellow, green and blue respectively. Sequence of mature ORRM1 without the 54 amino acid transit peptide is bolded. B, stable transformation of *orrm1* with RecA-3FS-mORRM1 under control of a 35S promoter complements the editing defects for plastid sites. Portion of electrophoretograms from RT-PCR bulk sequencing of *matk-640, ndhB-872* and *ndhG-50* is shown for the *Columbia* wild type, *orrm1*, and stable transformant expressing RecA-3xFlag-strepII-mORRM1 under a 35S promoter. The editing sites are indicated with arrows. Complementary strand of the sequenced *ndhG* is shown because sequencing was done from a reverse direction.

RNA was then extracted and used in RT-PCR to amplify the fragments harboring plastid editing sites as in reference (14). Editing extent of *matK-640, ndhB-872* and *ndhG-50,* which lost editing in *orrm1,* was examined by bulk sequencing (Figure 4.2B, reference (15)). Editing of all three sites was restored to wild type level in the RecA-3FS-mORRM1 transgenic plants.

Co-immunoprecipitation of ORRM1 followed by mass spectrometry identified one interacting protein—VAR3

Total leaf proteins were used to perform ORRM1 co-immunoprecipitation. Wild type *Columbia* Arabidopsis was included as a negative control for comparison in order to eliminate non-specific binding proteins. Affinity of strepII tag to the strep-tactin resin was poor probably due to the internal position, so only the Flag tag was used in this work. As is shown in Figure 4.3, the anti-Flag antibody only recognizes one single band from the transgenic plant samples but not from the wild type plant sample. The unique band is slightly bigger than the predicted size of the tagged ORRM1, which is around 42kD. Anti-Flag resins retained almost all tagged ORRM1 protein from the extract (Figure 4.3A). The elutions from both ORRM1 and negative control were separated by a SDS-PAGE gel and silver-stained. The bait 3FS-mORRM1 was clearly seen in the transgenic plant IP but missing in the *Col* negative control. The immunoprecipitates were subject to MS/MS mass spectrometry in order to identify ORRM1-binding proteins. One promising candidate VAR3 (at5g17990)

was selected for further investigation because it had second largest number of matches in MS/MS spectra, only after ORRM1.

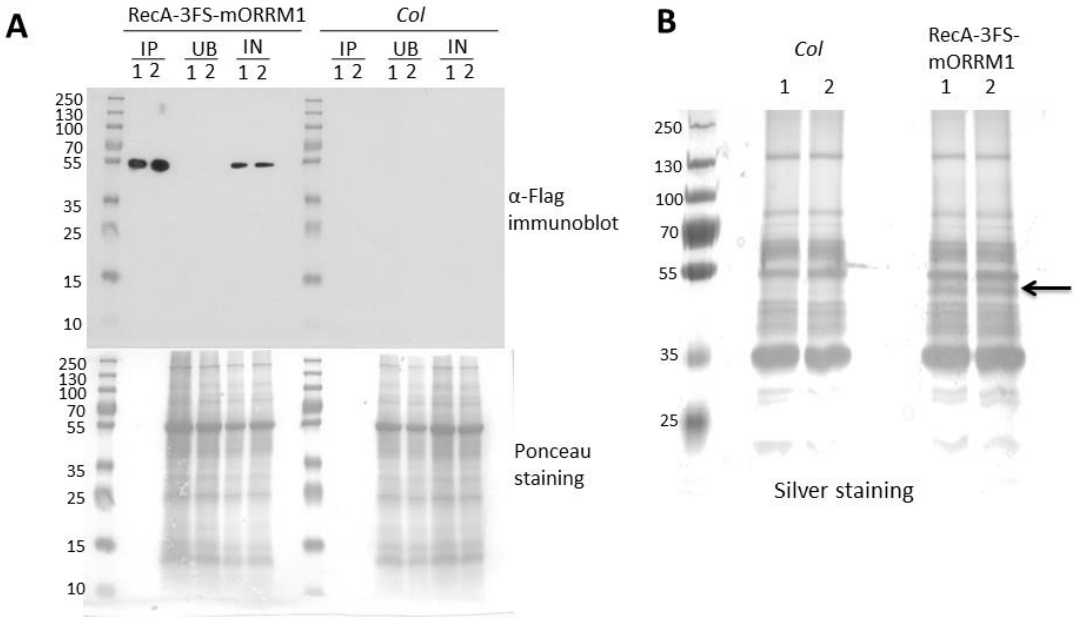


Figure 4.3. Immunoprecipitation of 3FS-ORRM1 using anti-Flag resins. A, immunoblot of immunoprecipitate (IP), unbound (UB) and input (IN) for both *Col* and transgenic RecA-3FS-mORRM1. Around 10ug total protein loaded for IN and UB. 1% of the IP was loaded. Anti-Flag-HRP was used to detect Flag-tagged protein. B, 10% of the IP was separated by 10% Tris-Glycine SDS-PAGE and silver-stained. Arrow indicates tagged ORRM1.

VAR3 protein interacts with ORRM1

Yeast two-hybrid assay was employed to confirm the interaction between VAR3 and ORRM1. Both VAR3 and ORRM1 are plastid targeted proteins, so the transit peptides were removed from each before cloning them into AD/BD fusion constructs.

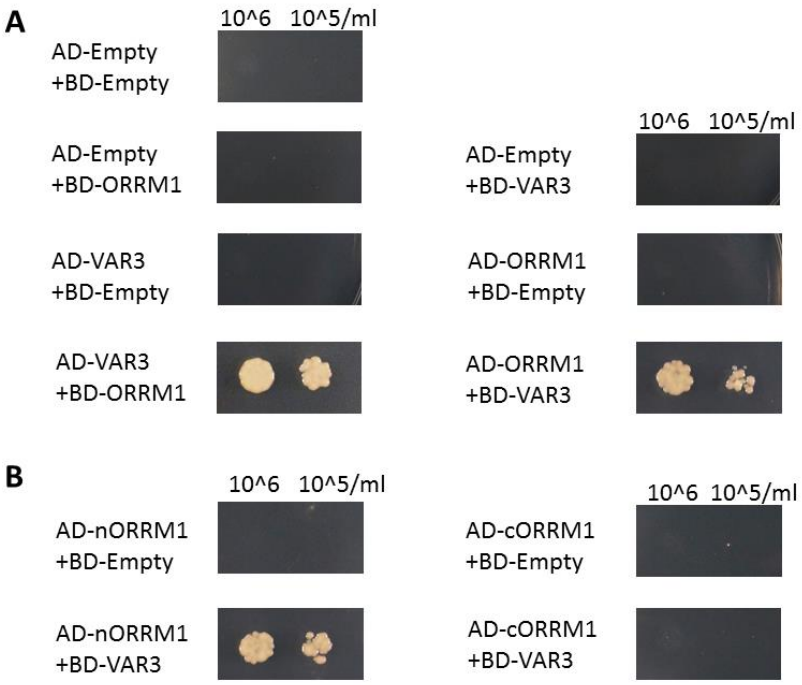


Figure 4.4. VAR3 interacts with ORRM1 in a yeast two-hybrid assay. AD-Empty, pGADT7 empty vector. BD-Empty, pGBKT7 empty vector. Yeast single transformants were mated to make double transformants in order to test interactions. Yeast were grown in -Leucine -Tryptophan double dropout media overnight before they were harvested and diluted into cell density 10^6/ml and 10^5/ml. 10µl of each dilution was spotted onto the -leucine -tryptophan –histidine –adenine quadruple dropout plates. Pictures were taken 3 days after spotting. nORRM1, 55-274th aa of ORRM1 which contains the RIP-RIP domain. cORRM1, 275-374th aa of ORRM1 which contains the RRM domain.

As shown in Figure 4.4A, VAR3 interacts with ORRM1 in yeast. The interaction is not affected by the position of the fusion protein since both AD-VAR3/BD-ORRM1 and its reciprocal pair AD-ORRM1/BD-VAR3 showed interaction, implicating a genuine interaction between these two proteins. ORRM1 was further divided into nORRM1 and cORRM1, encompassing the RIP-RIP and the

RRM domains respectively. nORRM1 but not cORRM1 interacts with VAR3, indicating the RIP-RIP domain actually mediates the interaction with VAR3.

Mutant characterization of VAR3

VAR3 (Variegated 3) was first identified in a *Ds* insertion line in Arabidopsis which showed a yellow green phenotype on the first two leaves and variegation of the other leaves (17). The mutant plants also had defective chloroplast development. *VAR3* encodes a zinc finger protein which contains two tandem zinc finger C2X10C2 domains. Both Target P and Proteomics results indicate that VAR3 is a plastid protein (http://ppdb.tc.cornell.edu). We obtained one T-DNA insertion line from ABRC—SAIL_358_H03 (Figure 4.5A). The insertion of T-DNA in the first exon of *VAR3* gene was verified by sequencing. Strikingly, the homozygous mutant *var3* showed a complete yellow phenotype instead of variegation, as shown in Figure 4.5B. It is possible that in the previously reported mutant, the *Ds* insertion is lost in some tissues due to the instability of the transposon, resulting in phenotypic heterogeneity. Notably, the new leaves of the yellow mutant gradually turned light green as it continued growing to maturity on sucrose media (Figure 4.5C). The first few leaves of the mutant stayed yellow for the mature plant. The partially recovered chlorophyll is sufficient to support autotrophic growth of *var3* in soil (Figure 4.5D).

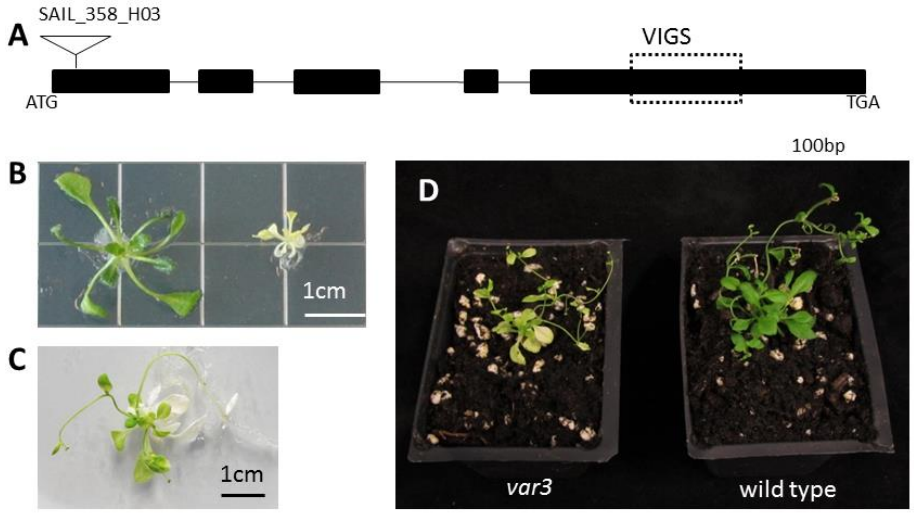


Figure 4.5. *VAR3* (At5g17790) gene structure and mutant phenotype. A, gene structure of *VAR3*. Triangle indicates the location of the T-DNA. Dashed box indicates the gene specific region selected for VIGS (Virus Induced Gene Silencing). B-D, *var3* phenotype. B, plants grown on MS media for 4 weeks. Left, wild type sibling, right, homozygous *var3* mutant. C, new leaves turned light green on a six weeks old *var3* mutant. D, eight weeks old *var3* (left) grows in soil compared to wild type (right).

VAR3 mutation leads to editing defects at most plastid sites

RNA from 4 week-old *var3* homozygous mutants and the siblings was extracted and the editing extent was examined by bulk sequencing as shown in Figure 4.6A.

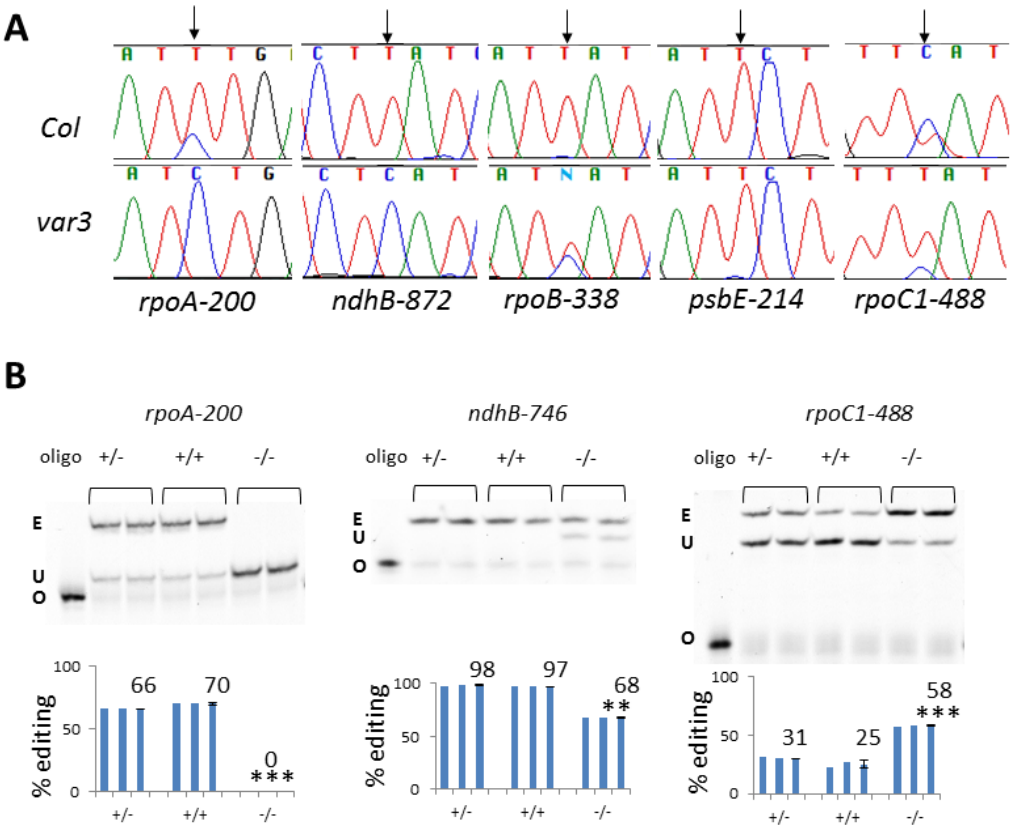


Figure 4.6. RNA editing at multiple plastid sites is affected in *var3*. A, editing of plastid sites is disrupted or enhanced in *var3* demonstrated by bulk sequencing. Portion of electrophoretograms from RT-PCR bulk sequencing is shown. Wild type *Columbia*, upper lane. *var3*, lower lane. Arrows indicate the position of the editable C. B, editing extent is examined by Poisoned Primer Extension assay. Oligos were loaded to the first lane for each gel figure. +/-, heterozygote; +/+, wild type *Col*; -/-, homozygous *var3*. E, edited band; U, unedited band; O, oligo. Significance * (P<0.05), ** (P<0.01), *** (P<0.001)

The *var3* mutation causes various editing defects for some plastid sites. For example, editing of *rpoA-200* and *ndhB-872* is completely lost in *var3* while editing of *rpoB-338* is partially disrupted. No obvious effect was observed for *psbE-214* editing in *var3* mutant. On the contrary, for *rpoC1-488* site, the mutant editing level is up-regulated compared to the wild type. Poisoned primer extension is a more sensitive

method to measure editing extent than bulk sequencing(14, 18). All plastid editing sites were assayed using PPE for both *var3* and its siblings (Figure 4.6B and Table 4.1). The *var3* mutation is clearly recessive since no significant editing difference is seen between heterozygotes and wild type plants. Editing of *rpoA-200* is 0% in *var3* by PPE which confirmed the result from bulk sequencing. *ndhB-746* editing dropped from 97% in wild type to 68% in the mutant. Editing of *rpoC1-488* increases from 25% in wild type to 58% in the mutant which agrees with the previous bulk sequencing result. The complete set of plastid sites data is summarized in the Table 4.1. 14 sites have major loss of editing (>90% decrease in editing) in *var3* and 16 other sites have significantly decreased editing (>5%, $P<0.05$). Although editing defects are massive, editing events on the same transcript are not all affected in the same pattern by *var3*, hence the editing defects are unlikely to be a secondary effect caused by a change in the transcript itself. *ndhD-2* and *ndhD-878* lost over 90% editing in *var3* but *ndhD-383* is not affected at all. On the contrary, for the sites recognized by the same PPR protein, they are mostly affected in the same way by *var3* mutation. *ndhD-878* and *ndhB-467* share the same PPR factor CRR28, and both of them lose over 90% editing in *var3*. It also holds true for *ndhB-836* and *ndhG-50* (OTP82) and *rpoA-200, clpP-559* (CLB19). *ndhF-290 ndhB-1481* and *psbZ-50* are all recognized by OTP84, and in *var3* they are all mildly affected in editing (5%-20%). Given that ORRM1 and VAR3 interact with each other, we also compared the *var3*-dependent sites and *orrm1*-dependent sites to examine if these two factors participate in the same editing events. Indeed, editing of the 14 *VAR3*-dependent sites is all severely affected in *orrm1* mutant. Many other sites mildly affected by *var3* are also *orrm1*-dependent (Table 4.1).

Editing site	Editing PPR protein	Editing extent in var3	Editing extent wild type level	Δediting	Δediting
				var3	orrm1
rpoC1-488	DOT4	72±2	24±2	200	0
rps12-(i1)-58	OTP81/QED1	0±0	28±2	100	99
ndhD-2	CRR4	0±0	56±4	100	57
ndhB-836	OTP82	0±0	95±0	100	95
ndhG-50	OTP82	0±0	84±3	100	94
rpoA-200	CLB19	0±0	71±3	100	73
clpP-559	CLB19	0±0	61±1	100	73
ndhB-1255	CREF7	0±0	99±0	100	60
ndhB-872	QED1	0±0	90±6	100	99
ndhB-467	CRR28	0±0	84±3	100	95
accD-1568	QED1	2±3	77±0	97	99
ndhB-586		5±1	92±1	95	95
ndhD-878	CRR28	5±0	85±2	94	97
ndhB-830	ELI1	7±1	97±2	93	71
ndhB-726		2±0	22±3	91	73
matK-640	QED1	12±2	85±1	86	97
ndhD-674	OTP85	16±1	91±4	82	99
accD-794	RARE1	27±6	95±4	72	0
ndhD-887	CRR22	33±0	83±4	60	96
rpoB-2432	QED1	44±2	82±1	46	92
petL-5		45±1	78±1	42	1
rpoB-551	CRR22	60±0	92±0	35	75
ndhB-746	CRR22	68±1	98±0	31	52
rpl23-89	OTP80	67±4	83±1	22	0
ndhF-290	OTP84	80±0	97±0	18	0
rpoB-338	YS1	76±3	91±0	16	14
ndhB-149		80±0	95±0	11	0
psbF-77	LPA66	88±1	97±1	9	0
ndhB-1481	OTP84	85±1	90±2	6	0
psbZ-50	OTP84	87±0	92±1	5	0
atpF-92		90±0	93±2	3	0
rps14-149	OTP86	93±2	90±1	3	37
ndhD-383	CRR21	96±1	98±1	2	0
rps14-80		93±1	92±1	1	0
psbE-214	CREF3	98±0	98±2	0	0
ndhB-708		0	0	0	99
ndhB-153		n/a	n/a	0	10

Table 4.1. Plastid editing extent in *var3* and *orrm1*. Editing extent was measured using poisoned primer extension. Known PPR factor is listed for corresponding editing sites. Δediting = (editing extent of wild type-editing extent of mutant)/ editing extent of wild type. n/a, not assayed. Editing extent that is significantly (P<0.05) different from wild type level was colored. Green indicates decreased editing level, red indicates increased editing level.

8 sites are only controlled by VAR3 but not ORRM1. Taken together, VAR3 is a genuine plastid editing factor for most plastid sites.

Transient silencing of *VAR3* leads to plastid editing defects

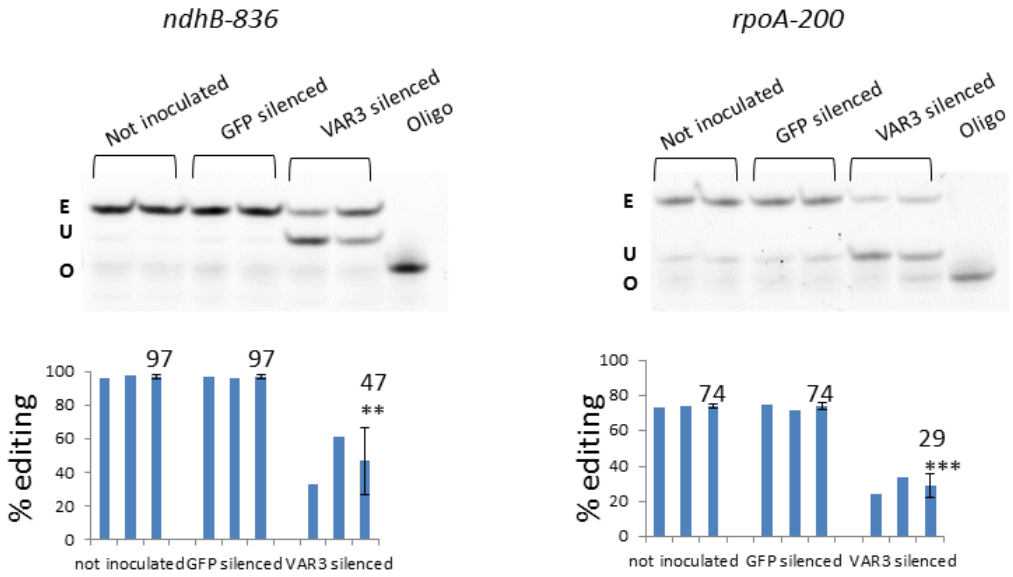


Figure 4.7. Transient silencing of *VAR3* in Arabidopsis results in chloroplast editing defects. Two replicates for each treatment were assayed by poisoned primer extension. Not inoculated, plants that were not inoculated with Agrobacteria. GFP silenced, plants that were inoculated with Agrobacteria harboring a GFP silencing construct. VAR3 silenced, plants that were inoculated with Agrobacteria harboring a GFP and VAR3 co-silencing construct. PPE bands were quantified by image quant software and illustrated in graphs. Average for each group is displayed in a third bar. E, edited band; U, unedited band; O, oligo. Significance ** (P<0.01), ***(P<0.001)

Since a second T-DNA mutant for *VAR3* was not available, we performed Virus Induced Gene Silencing (VIGS) to transiently silence *VAR3* expression in Arabidopsis young seedlings. To monitor the silencing efficiency, a GFP co-silencing marker which is harbored in the VIGS construct was used(19). Agrobacteria carrying either VAR3/GFP co-silencing construct or GFP silencing construct alone were inoculated into 2 weeks old 35S::GFP expressing Arabidopsis seedlings. The plants were grown in long day for another 5 weeks before we analyzed the editing extent of RNA from the successfully silenced plants and not inoculated plants by PPE. GFP silenced plants and untreated plants did not show any difference in editing level (Figure 4.7). *ndhB-836* editing extent decreased from 97% in untreated control to 47% in *VAR3* silenced plants ($P<0.01$). *rpoA-200* editing extent dropped from 74% in untreated control to 29% in *VAR3* silenced plants ($P<0.001$). These results agree with results from mutant, in which editing of both of the sites is abolished. The remaining editing in the silenced plants is probably caused by residual expression of *VAR3*.

Transient expression of *VAR3* in *var3* protoplasts complements the editing defects

Although the young *var3* mutant has severely defective phenotype, the plant continues to grow to maturity and gradually recovers some of the chlorophyll. In a 7 weeks old *var3* plant, the old leaves remain pale yellow while the new leaves are light green (Figure 4.5C). To investigate whether editing defects are rescued in the light green leaves, editing of RNA extracted from both the pale yellow leaves and the light green leaves were analyzed by bulk sequencing.

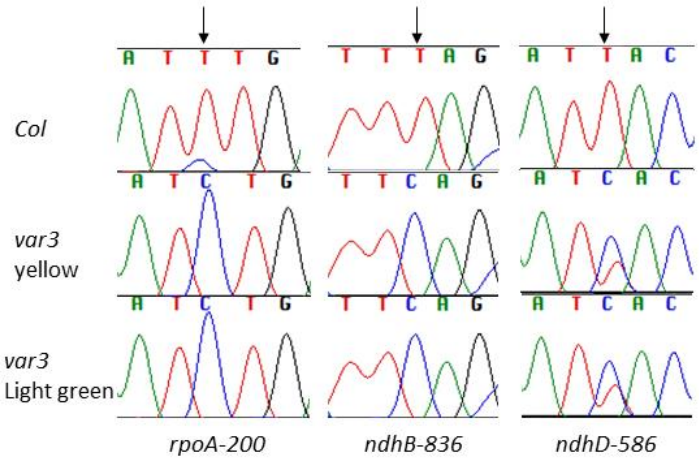


Figure 4.8. Plastid editing is not recovered in light green leaves of *var3*. RNA from yellow leaves and light green leaves of 8 weeks old *var3* plants were used in RT-PCR and then bulk sequenced. Portions of electrophoretograms are shown. Arrow indicates the position of editable C target.

Although pigmentation has been partially recovered in light green leaves, plastid editing is still defective in those leaves compared to wild type *Col* (Figure 4.8). No obvious difference in editing between yellow leaf and green leaf was observed. These green leaves were then used to prepare protoplasts. *VAR3* was cloned into a pSAT4a vector to create a plant transient expression vector driven by 35S promoter. 35S::VAR3 was then used to transfect *var3* protoplasts for complementation. A chloroplast target YFP construct was included as a negative control. Transfection efficiency is over 50% for the YFP control group. RNA was extracted from protoplasts 2 days after the transfection and analyzed in PPE to examine the editing efficiency (Figure 4.9). No significant difference in editing was seen between the untransfected control and the 35S::CP-YFP transfected control. Introduction of the 35S::VAR3 significantly increase the editing level for all the sites we tested. *rpoA-200*

increased from 3% to 21%, *ndhB-836* from 19% to 31% and *rps12-(i1)-58* from 3% to 15%. This confirms that the editing defects in *var3* mutant are truly caused by loss of *VAR3*.

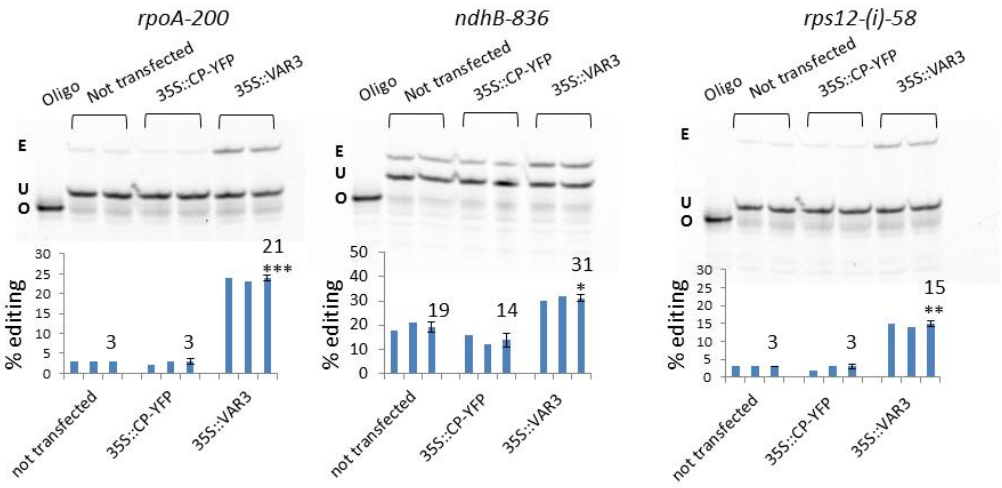


Figure 4.9. Transient expression of *VAR3* under a 35S promoter in *var3* mutant protoplasts complements the editing defects. Two repeats of each treatment were assayed by Poisoned Primer Extension (PPE). PPE bands were quantified by image quant software and illustrated in graphs. Average for each group is displayed in a third bar. Not transfected, *var3* mutant protoplasts that were not transfected; 35S::CP-YFP, *var3* protoplasts transfected with a construct expressing plastid targeted YFP driven by 35S promoter; 35S::VAR3, *var3* protoplasts transfected with a construct expressing *VAR3* driven by 35S promoter. E, edited band; U, unedited band; O, oligo. Significance * (P<0.05), ** (P<0.01), *** (P<0.001)

VAR3 interacts with other components of chloroplast editosome

Up to now, three major types of factors have been identified for chloroplast editing apparatus: PPR proteins, RIP/MORF proteins and ORRM1. Both RIP and

ORRM1 have been reported to interact with PPR proteins. In addition, RIP proteins can form both homodimer and heterodimers. In order to test if VAR3 binds to other components of editing complex and if VAR3 dimerizes, we performed a series of Yeast two-hybrid assays. First, VAR3 fused to AD and BD were used to test the dimerization (Figure 4.10A). VAR3 with either fusion does not show any auto-activation for HIS and ADE reporters while yeast with AD-VAR3/BD-VAR3 is

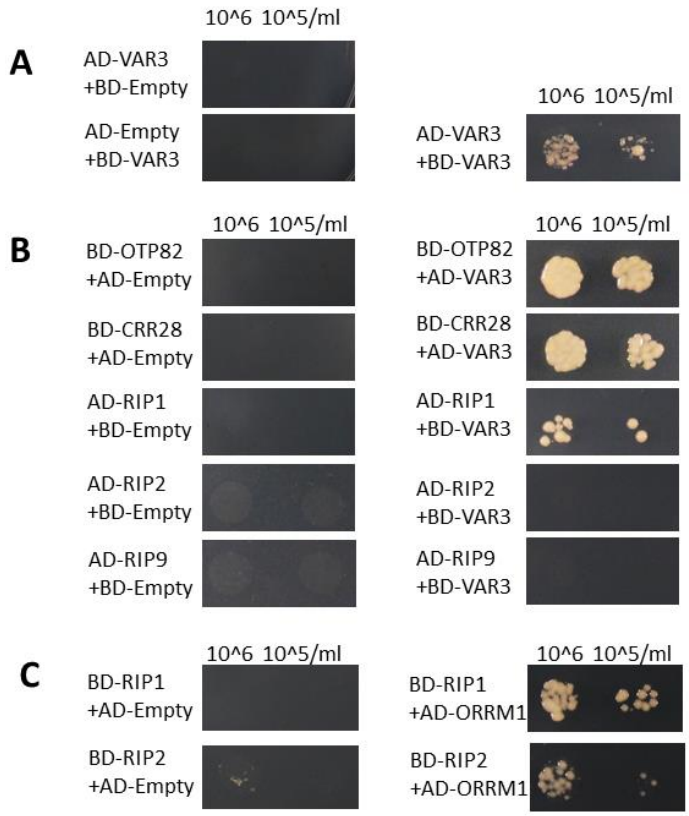
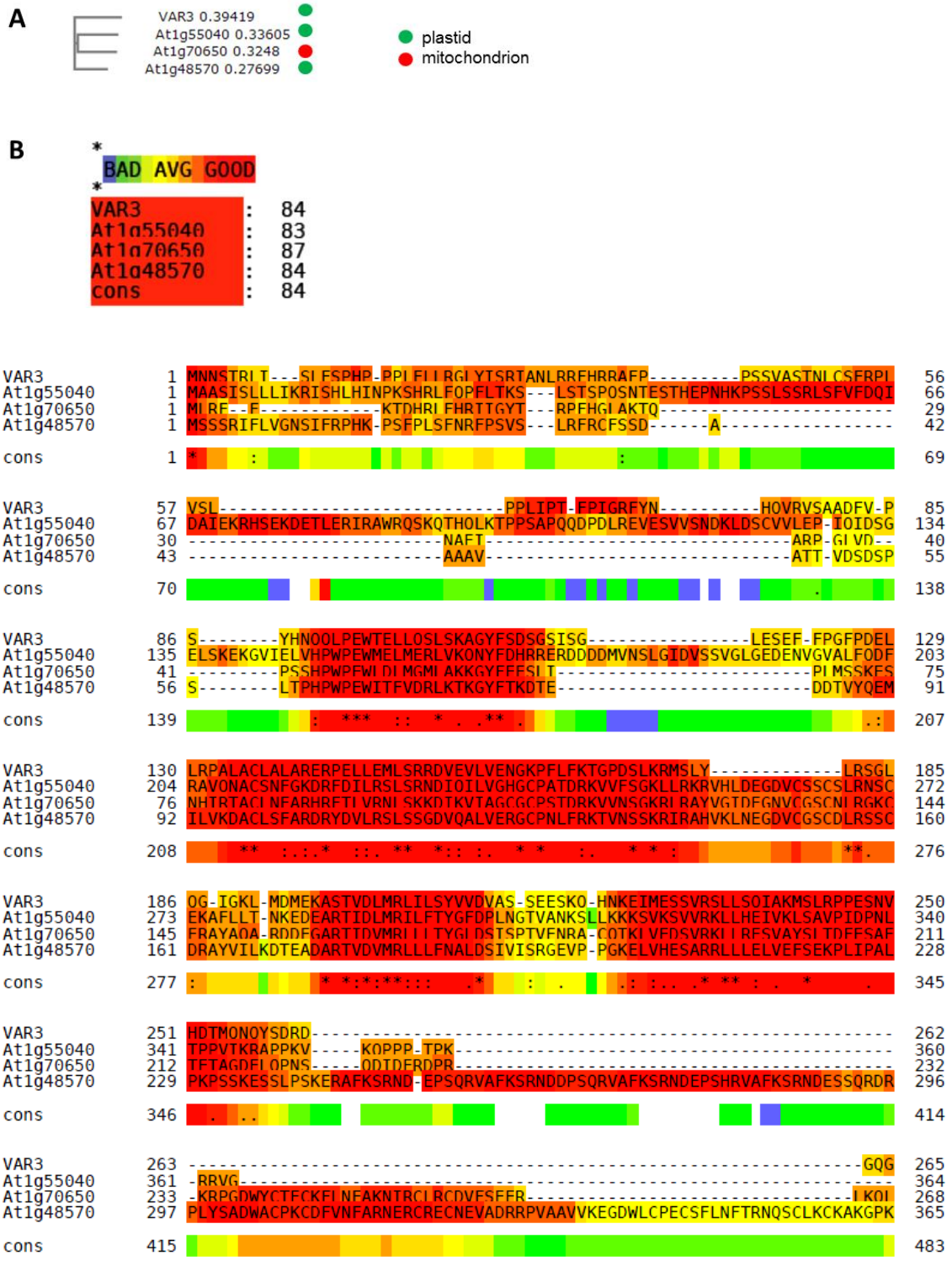


Figure 4.10. VAR3 interacts with other components of the editing complex. A, VAR3 dimerizes. B, VAR3 interacts with OTP82 CRR28 and RIP1 but not RIP2 or RIP9. C, ORRM1 interacts with RIP1 and RIP2. All interactions were tested on –Leu-Trp-His-Ade dropout media. Picture was taken 3 days after spotting.

able to grow on histidine and adenine deficient media indicating interaction between two proteins. Subsequently, interactions between VAR3 and PPR proteins and RIP1 were tested. VAR3 interacts with OTP82 and CRR28 as shown in Figure 4.10B. It also has a weaker interaction with RIP1 protein as fewer colonies are seen in the RIP1/VAR3 testing yeast (Figure 4.10B). However, no interaction was observed between VAR3 and RIP2 or RIP9 (Figure 4.10B). RIP2 and RIP9 are major components of the plastid editing complex. It is possible that VAR3 associates with RIP2 and RIP9 via ORRM1. To test this hypothesis, we performed yeast 2-hybrid assay for ORRM1 and RIP proteins (Figure 4.10C). Both RIP1 and RIP2 can interact with ORRM1. RIP9 fused with GAL4 binding domain strongly autoactivates *HIS* and *ADE* reporters so it was not tested in this experiment. ORMM1 with GAL4 activation domain has been tested in Figure 4.4A and showed no autoactivation. Thus, ORRM1 mediates interaction between VAR3 and RIP2 inside the plastid editosome.

VAR3 belong to a small family in Arabidopsis

VAR3 contains two tandem zinc finger motifs called RanBP2 type zinc finger (X2GDWICX2CX3NFARRX2CXRCX2-PRPEX2; pFAM00641) which was characterized in Ran Binding Protein 2 (RanBP2) (17). Ran is a small GTPase and RanBP2 is a nucleoporin that binds Ran via the zinc finger motifs (20). Three highly similar RanBP2 zinc finger proteins were found in the Arabidopsis protein database in a Blast search with VAR3.


A
VAR3 0.39419
At1g55040 0.33605
At1g70650 0.3248
At1g48570 0.27699
plastid
mitochondrion
B
BAD AVG GOOD
VAR3 : 84
At1g55040 : 83
At1g70650 : 87
At1g48570 : 84
cons : 84

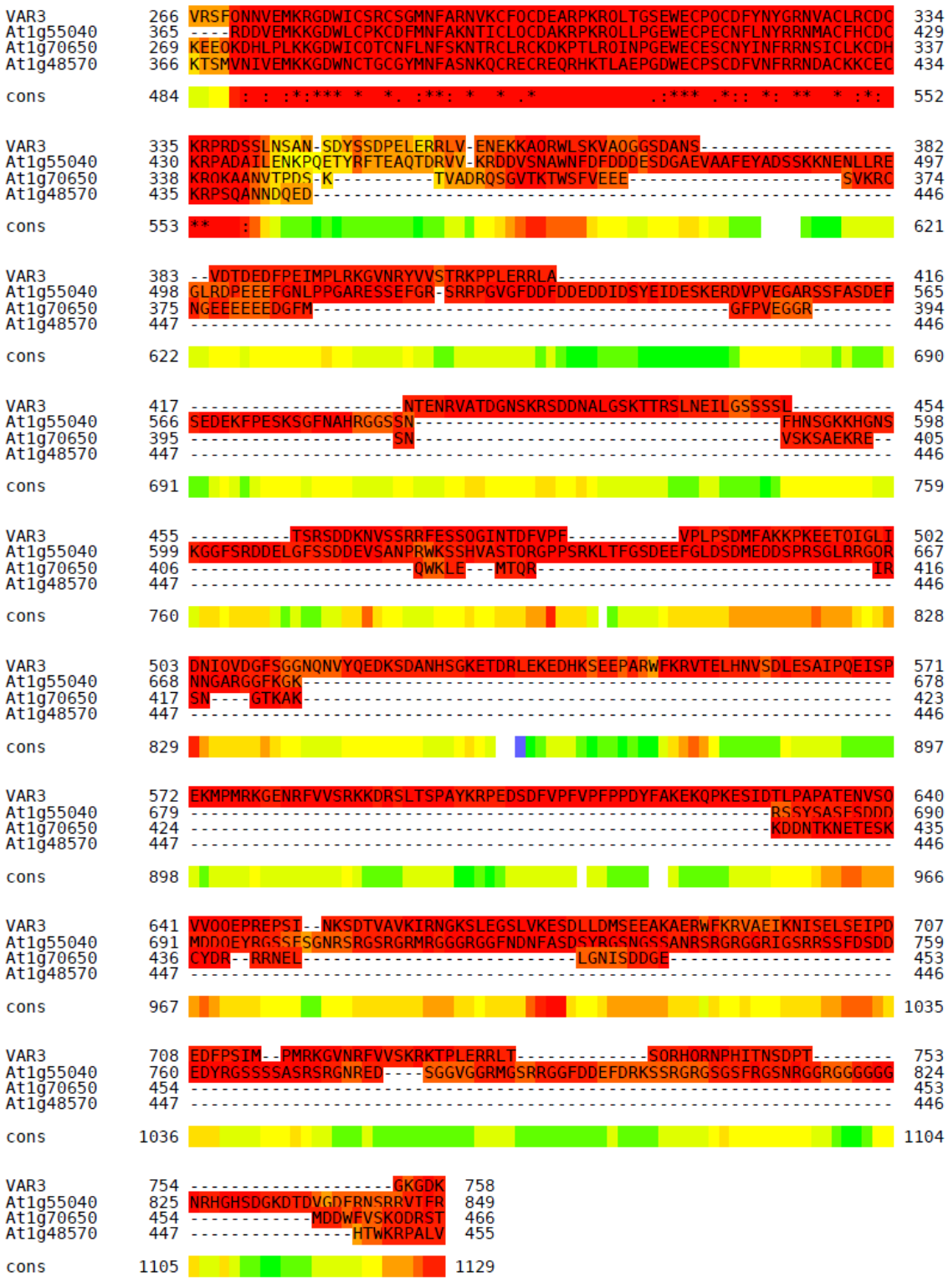


Figure 4.11. VAR3 belongs to a small family in Arabidopsis. A, phylogenetic tree of the family members and their predicted subcellular localization. The neighbour-

joining tree was built without distance corrections by Clustal Omega. Prediction of subcellular localization is from Target P. B, protein sequence of VAR3 family members are aligned by T-coffee.

Protein sequence alignment by T-coffee shows the presence of multiple highly conserved regions at the N-parts and variable regions at the C-parts (Figure 4.11A). These four proteins are very close in a phylogenetic tree indicating they belong to the same protein family (Figure 4.11B). Target P predicts all of the VAR3 family members to be organelle targeted, one in mitochondria and three in plastids.

Except for the zinc finger motif, no other annotated domain or motif was found in the VAR3 family. In order to find the hidden uncharacterized motifs, I performed motif scanning using MEME against all four members to look for motifs between 15aa to 70aa. 7 motifs were returned (Figure 4.12). Motif 1 is the zinc finger domain with 4 characteristic cysteine residues.

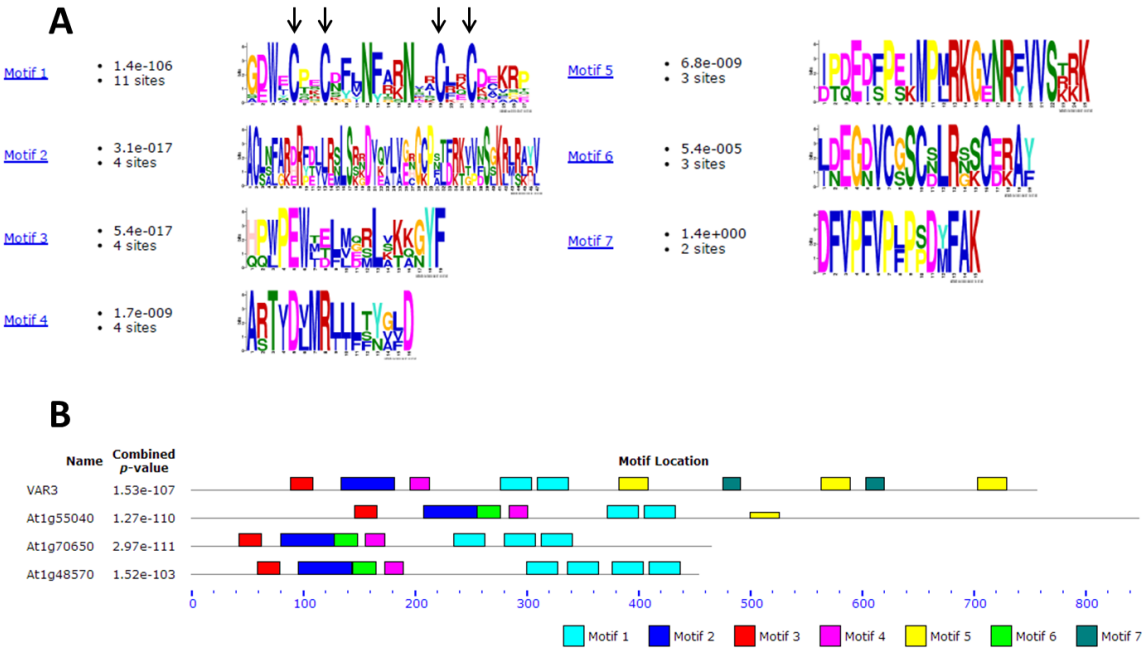


Figure 4.12. VAR3 and its family members contain multiple domains. A, motifs detected by MEME prediction for VAR3 family members. Motif1 is the RanBP2 type Zinc finger domain. Arrows indicate the Cysteines characteristic for Zinc finger domain. B, motif locations in VAR3 family.

As shown in Figure 4.12B, the zinc finger motif is shared by all four members, however, the number of repeats varies across the family. VAR3 and At1g55040 contain two zinc finger motifs, while At1g70650 has three and At1g48570 has four. Regions before the zinc finger motifs are relatively highly conserved, briefly spanning 4 distinct domains. However, one short domain, motif 6 is missing from VAR3. The sequences downstream of the zinc finger domains are very variable. VAR3 has three motif 5 repeats and 2 motif 7 repeats. These motifs are either missing or poorly conserved in other members.

Discussion

Discovery of RIP/MORF family and ORRM1 revealed the unexpected complexity of the plant RNA editing mechanism. In order to identify the unknown components of the editing apparatus, we employed a fishing strategy using epitope tagged ORRM1 as bait. Position of the epitope tag relative to ORRM1 has a huge effect on the protein's function. A 3xFlag-strepII tag adjacent to C terminus of ORRM1 disrupts its function in editing. This is not surprising, given that RRM domain at the C terminus of ORRM1 actually carries the editing activity. A tag in close proximity might change the RRM conformation or impose a hindrance for its potential binding partners. In contrast, 3FS-ORRM1 with a same tag at the N terminus

of ORRM1 is fully functional. Introduction of an N terminal tagged ORRM1 into mutant protoplasts could increase the editing levels which would be defective otherwise. Stable integration of this chimeric gene into mutant fully restored the editing, which was then used for co-immunoprecipitation of ORRM1.

Although both Flag and strepII tags were present in the hybrid protein, only Flag turned out to have high affinity to the corresponding resins. Loss of affinity for strepII is probably due to the internal position of this tag which may keep the tag from being exposed. The flexible linker VGAG and GGSG did not seem to maintain the affinity of strepII. The 3 tandem Flag tag was fully competent for the immunoprecipitation of ORRM1. One promising candidate interactor was found in MS/MS 12 times, the most frequently detected besides the bait ORRM1 (100 times). This candidate protein is encoded by *VAR3*, which was previously found to cause Arabidopsis variegated phenotype when the gene is disrupted by a *Ds* transposon (17). However, a T-DNA mutant which has an insert at the first exon showed a yellow instead of a variegated phenotype. Most identified variegated mutants are caused by unstable transposable elements that affect chloroplast function (21). So it is highly possible that the *Ds* insertion was lost in some cells in the previously reported mutant due to instability of transposon, which resulted in reverted phenotype in the green sectors. Although the authors detected no RNA expression of VAR3 in the mutant by the RT-PCR, it is probably due to the relatively low expression level of the endogenous VAR3 (17).

Interaction between VAR3 and ORRM1 was confirmed by a yeast two-hybrid assay which shows that VAR3 is bona fide interacting partner for ORRM1 and hence

a potential editing factor. Indeed editing of multiple plastid sites is severely affected by *var3* mutation. 14 sites have major loss of editing and 12 other sites' editing are significantly reduced. Most PPR editing factor mutants show no macroscopic phenotypic difference with wild type plants. Two exceptions are *clb19* and *ys1* which lose editing for *rpoA-200 clpP-559* and *rpoB-338* respectively (22, 23). *rpoA* and *rpoB* encode two different subunits of the plastid-encoded plastid RNA polymerase (PEP). *clpP* encodes a subunit of the plastid Clp protease complex. In *clb19* and *ys1*, proteins translated from unedited transcripts probably have comprised functions which consequently cause the drastic mutant phenotype. Notably, in *var3* editing of *rpoA-200* and *clpP-559* is completely abolished and *rpoB-338* editing efficiency is reduced. Thus the yellow-green phenotype of *var3* can be largely attributed to editing defects of *rpoA-200 clpP-559* and *rpoB-338*. *var3* survives sucrose media and older leaves turn light green, which is also seen in *clb19 ys1* mutant. A nuclear-encoded plastid RNA polymerase (NEP) also exists in plastids, which might have rescued the defect caused by defective PEP in the late stage of development (24).

Editing efficiency can be affected by other RNA processing steps. A reduced editing level might be caused by up-regulation of the transcripts which has saturated the editing machinery, as previously reported (25, 26). However, it is not likely the case for *var3*. First, many sites have no editing at all, which is highly unlikely to be caused by other types of RNA processing. Second, C targets that reside on the same transcripts are differently affected. *ndhD-878* has major loss of editing, while *ndhD-383* is barely affected. Third, *var3* effect on editing is tightly correlated with other editing factors. Editing sites that share PPR recognition factor are similarly affected in

var3 mutant. In addition, all 14 severely affected sites are also affected in the *orrm1* mutant. Therefore, VAR3 is a novel major plastid editing factor.

Since no second mutant is available, we performed both virus induced gene silencing and protoplast complementation to verify *VAR3* function in editing. In VIGS experiment, editing defects were confirmed in *VAR3* silenced plants. Given that VIGS only knocks down the gene expression, it is not surprising to see the editing level of the *VAR3* dependent sites reduced but not totally abolished. Introduction of a 35S::VAR3 into mutant protoplasts greatly increased the extent of editing of the editing defective sites such as *rpoA-200*. Both pieces of evidence show that the editing defect seen in *var3* mutant can be attributed to loss of *VAR3*.

My yeast two-hybrid results showed that VAR3 interacts with many other editing factors. VAR3 binds to the RIP-RIP part of ORRM1 but not the RRM part. VAR3 binds to CRR28 and OTP82, which have genetic interactions with VAR3 as well, based on the editing events in which they are involved. VAR3 interacts with RIP1, though the interaction is not as strong as that with ORRM1. VAR3 does not seem to directly interact with RIP2 or RIP9. However, ORRM1 can bind to RIP1 and RIP2. Previously we reported interaction between RIP-RIP domain of ORRM1 and CRR28 and OTP82(15). So interactions exist among ORRM1, VAR3, CRR28/OTP82 and RIPs, which strongly suggests a multi-component editing complex. VAR3 also dimerizes in yeast which might also occur *in vivo* in the apparatus. In addition, another interacting protein, NCED4, for VAR3 has been identified through a yeast two-hybrid screen(17). Although NCED4 is annotated as a putative polyene chain or carotenoid dioxygenase, it is unknown yet whether it also participates in plastid RNA editing.

Three proteins share high similarity with VAR3 which comprise a small family in Arabidopsis. The only well documented domain found in this family is the Ran binding protein 2 type zinc finger motif comprising a conserved 30 amino acid residue consensus (X_2GDWICX_2CX_3NFARRX_2CXRCX_2-PRPEX_2; pFAM00641) characterized in RAN binding protein 2 (RanBP2) and other nucleoporins. VAR3 contains two tandem RanBP2 zinc-finger domains while other members of the VAR3 family have various numbers of this domain. Plant RNA editing has been shown to be Zn^{2+} dependent (27). A zinc ion binding site is also a characteristic of cytidine deaminase which is believed to be the catalytic activity for plant RNA editing. Recently, two PPR-DYW proteins, ELI1 and DYW1, have been shown to have two zinc atoms bound to the DYW region, provoking suspicions about the elusive editing enzyme (28). It is intriguing that VAR3, a plastid editing factor has zinc binding potential. Given that VAR3 itself does not contain any deaminase-like domain, it is possible that it provides Zn^{2+} for the active site of the catalytic center in the editosome where it associates with other components. The N terminal part of the VAR3 family is more conserved with 3 or 4 highly similar motifs. Besides the zinc finger domain, VAR3 also contains three long repeats and two short repeats at the C-terminus, which is less conserved in other family members. However, the function of these domains is still unknown.

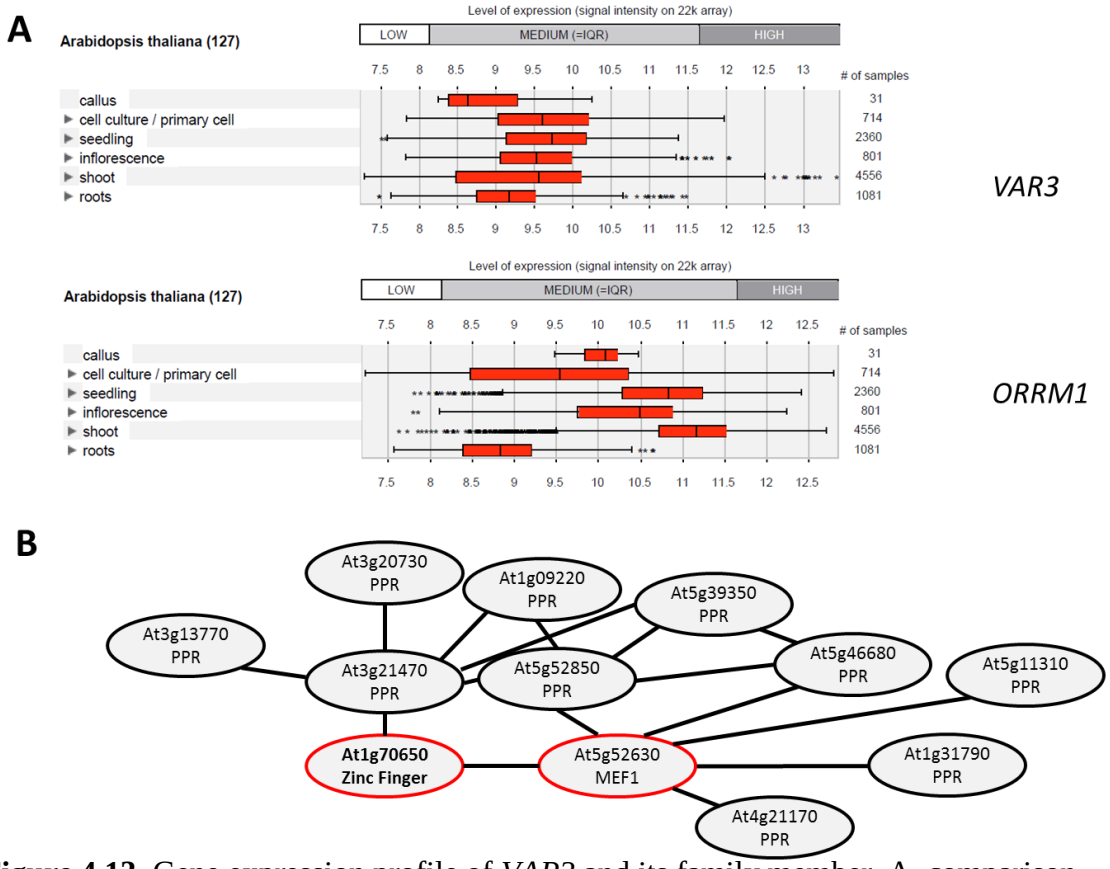


Figure 4.13. Gene expression profile of *VAR3* and its family member. A, comparison of *VAR3* and *ORRM1* expression levels in different tissues. Data acquired from Genevestigator. Expression levels have been converted to log values. B, one member of *VAR3* family is tightly co-expressed with Mitochondria Editing Factor1 and other PPR proteins. Figure is modified from co-expression database (Atted.jp)

VAR3 is a medium –level expressed protein in Arabidopsis according to the data from Genevestigator, which is comparable to the expression level of its interacting partner ORRM1. Proteins involved in the same pathway including RNA editing are usually co-expressed. Two PPR editing factors, CREF3 and CREF7, were discovered through co-expression analysis of known editing protein CRR21(29). Some of VAR3 family members are tightly co-expressed with other editing factors such as PPR proteins. The most outstanding example is the VAR3 family member

At1g70650, whose expression is closely co-regulated with Mitochondrial Editing Factor1 (MEF1) shown in Figure 4.13 (30). Besides MEF1, this gene is related to another PPR protein encoding gene At3g21470. Both MEF1 and At3g21470 share an expression pattern with many other PPR proteins which form a co-expression network. It is highly possible that additional members of the VAR3 family are also involved plant RNA editing.

VAR3 orthologs can be found in many other flowering species, including maize, rice, pea, etc. This suggests a common role in RNA editing across the flowering plants. So far, four kinds of core components of plant editosome have been identified: PPR protein, RIP/MORF, ORRM, VAR3 (possibly the whole family). The combined molecular mass of four factors is approximately the expected size of a fully competent editosome (200kD). So the most interesting question is whether any component is still missing. An *in vitro* editing assay with purified recombinant editing proteins would be one way to examine this. What is interesting, although A-to-I editing enzyme has been characterized in humans and other metazoan, the regulatory mechanism is totally unknown (31). Editing defects in Drosophila often cause abnormal locomotive behaviors. A blast of Ran-BP Zinc finger protein identified many proteins in animal kingdom. Although none of them share similarity with other domains of VAR3, many of them contain RRM motifs and zinc finger motif in the same protein. Mutation of a Drosophila zinc finger RRM protein leads to locomotive abnormality. It will be of great interest to test if the Drosophila RRM-zinc finger protein is actually involved in ADAR mediated RNA editing.

Material and Methods

Table 4.2. oligonucleotides used in chapter 4	
RecA_F	ATGGATTCACAGCTAGTCTTGTCTCTG
RecA_R	GTC GCG ATC GAA TTC AGA ACT GAT TTT GTG GGA G
3FS_F	ATC AGT TCT GAA TTC GAT CGC GAC GAC TAC AAA GAC CAT GAC GGT GAT
3FS_R	ATA GGA TCC TGA ACC CCC ACT TCC ACC TTT TTC AAA TTG AGG ATG AGA CCA ACC
ORRM1_163F	GGA AGT GGG GGT TCA GGA TCC TAT ATA GTC GAC TCT TCT GCA ATT TCC GCA CCG
ORRM1_R_WO	GAG CCC GAA ACT TGG TTG ACT TCT
ORRM1_R	CTA GAG CCC GAA ACT TGG TTG
ORRM1_F	ATG GAA GCT CTT ATT GCT TCC ACT TC
VAR3_F	ATG AAC AAC TCC ACC AGA CTC ATC TCC
VAR3_R	TCA TTT ATC TCC TTT ACC AGT GGG ATC
VAR3_VIGS_F	CCTCTTTGACTTCTCGTTCTGATG
VAR3_VIGS_R	ACT ACA AAC CGA TTC TCT CCT TTT CGC
VAR3_100F	CGT TTC CAC CGC CGT GCG TTT
RIP2_133F	TTCTCAATTCGATGTGGAGCTAAC
RIP2_R	TCA TCT TGT GTT TTC TCT GCG GCG
RIP9_175F	GCTGCGACGGTGGATTCGGAT
RIP9_R	TTAAGAGGAATCAGAGGCTGCTGG

Plasmid constructs

The ORRM1 coding sequence was cloned using primer pair ORRM1_1F and ORRM1_R_WO from previous constructs. The sequence was first TA cloned into PCR8/GW/TOPO (Invitrogen), and then shuttled into a modified PBI121 vector with a 3XFlag-strepII C-terminal tag. Alternatively, coding sequence of ORRM1 mature form (without predicted 54aa transit peptide) was fused to a N terminal 3XFlag-strepII tag sequence and an artificial transit peptide sequence from RecA in an overlapping PCR using primer pairs RecA_F, RecA_R, 3FS_F, 3FS_R, ORRM1_163F and ORRM1_R. This chimeric gene was cloned into PCR8 vector first and then PBI121 vector using LR ClonaseII (Invitrogen). The C-terminal tag on the vector was eliminated by the endogenous stop codon of ORRM1 in the sequence. RecA-RRM construct was from previous study, and used as a positive control. VAR3

coding sequence was cloned using primer pair VAR3_F and VAR3_R from Arabidopsis thaliana cDNA. PCR product was first ligated to PCR8/GW/TOPO and then the destination vectors pSAT4a and pAUL13 to create a transient expression construct and a stable complementation construct respectively.

Protoplast complementation

Protoplasts of *orrm1* or *var3* were prepared following the protocol from Jen Sheen's lab. Specially, light green leaves instead of yellow leaves from the *var3* mutant were used for protoplast. 10 µg plasmid DNA was used to transfect 2x10^4 cells. The transfected protoplasts were incubated in dark at room temperature for 3 days (*orrm1* complementation) or 1 day (*var3* complementation) before harvest.

Generation of transgenic plants

The 3XFlag-strepII-ORRM1ΔCTP constructs in both PBI121 and PAUL5 were used to transform Agrobacterium GV3101 strain. Standard root transformation protocol was followed to transform mutant *orrm1* and *var3* roots (32). Basically, the roots were first induced on Callus Inducing Medium (CIM) for 2 days and then infected with Agrobacteria in liquid media. Roots were incubated on CIM for another 2 days until they were overgrown by Agrobacteria and then bacteria were removed by several washing steps with liquid CIM containing Timentin and Carbenicillin. Roots were then cut into 0.5mm pieces and put onto Shoot Inducing Medium (SIM) containing

100mg/L Basta for selection. After the shoots grew out, they were removed from the calli and transferred onto a Root Inducing Medium (RIM). Fully grown transgenic plants with healthy roots were then transferred into soil.

Co-immunoprecipitation

About 10g leaves of each line were grinded in liquid nitrogen into fine powder. Total leaf protein was extracted using grinding buffer (150mM NaCl, 50mM Tris-HCl pH7.4, 1mM EDTA, 0.2%NP-40 and 1x Cocktail protease inhibitor (Sigma)). The extract was cleared by 13,000rpm centrifugation, 0.45μm filtration and then 30 minutes incubation with unconjugated agarose beads (Vector Laborotaries, CA) to minimize non-specific binding. 200μl anti-Flag agarose resins (Sigma) were first blocked with 4%BSA before 2 hours incubation with around 30μg pre-cleared protein extract. The washing step was done using washing buffer (150mM NaCl, 50mM Tris-HCl pH7.4, 1mM EDTA, 0.2% NP-40). IP was eluted with elution buffer (2M $MgCl_2$, 50mM Tris-HCl ph8.0, 150mM NaCl, 0.5%CHAPS). Final sample was prepared using SDS-PAGE sample prep kit (Thermo Scientific).

Immunoblotting and silver staining

Standard protocol for western blotting was followed. 1% of the IP samples were loaded onto an Any kD MINI-PROTEAN TGX precast gel (Bio-rad, Hercules, CA) followed by standard procedures of western blotting. α-Flag-M2-HRP (Sigma) was

used to detect Flag tagged proteins. 10% of the IP samples were separated on a 10% polyacrylamide gel before subject to silver staining.

Mutant growth and phenotyping

T-DNA insertional mutant *var3* SAIL_358_H03 was obtained from ABRC. After 3 days stratification, the seeds were placed onto MS plates under normal long day condition. Mutant plants and the wild type siblings were then transferred into soil after 7 weeks on MS media. Leaves from 4 weeks old and 8 weeks old plants were collected for further analysis.

RNA editing extent measurement

DNA contaminants were removed from RNA samples by TURBO DNase (Life Technology). cDNAs were reverse transcribed from the RNAs using the pooled reverse primers as in before (14). PCR products harboring the editing sites were either bulk sequenced or subject to Poisoned Primer Extension (PPE) assay (14).

Virus-induced gene silencing

VAR3 gene-specific region was picked from CATMA database and amplified using primer pair VAR3_VIGS_F and VAR3_VIGS_R. The fragment was first integrated into PCR8/GW/TOPO and then the silencing vector PTRV2/GW/GFP by a LR

reaction. Agrobacteria harboring the silencing construct were used to infiltrate 2 weeks old Arabidopsis seedlings that expressed GFP driven by 35S promoter. 5 weeks after infiltration, silencing efficiency was monitored by the expression of the co-silenced GFP in each individual. Successfully silenced plants which showed dark red from stem to leaf under UV light were collected for further analysis.

Yeast two-hybrid assay

Mature VAR3 coding sequence (without the N-terminal 33aa transit peptide) was amplified using primer pair VAR3_100F and VAR3_R from cDNA and cloned into PCR8/GW/TOPO (invitrogen). Mature RIP2 RIP9 coding sequences were amplified using primer pairs RIP2_133F and RIP2_R, RIP9_175F and RIP9_R from Arabidopsis thaliana cDNA respectively. PCR products were first cloned into PCR8/GW/TOPO and then pGADT7GW and pGBKT7GW destination vectors through homologous recombination by LR clonaseII(invitrogen). RIP1, RARE1, CRR28, OTP82, ORRM1 and its derived constructs were from previous works(14, 15). Empty vectors were used as negative controls. Two mating types of PJ69-4 yeast strain- a and α were used. Single transformants were obtained by transformation while double transformants through mating. Yeast harboring testing pairs were grown in leucine and tryptophan deficient media overnight before they were diluted with water into OD_{600} 0.5, 0.05, 0.005. 10 µl of each dilution was spotted onto leucine, tryptophan, histidine, adenine deficient media plates. Growth results were collected after 3 days incubation at 30 ℃.

References

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

2. Gray MW, Covello PS (1993) RNA editing in plant mitochondria and chloroplasts. *FASEB J* 7:64–71.

3. Hiesel R, Wissinger B, Schuster W, Brennicke A (1989) RNA editing in plant mitochondria. *Science* 246:1632–1634.

4. Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA metabolism. *Annu Rev Plant Biol* 61:125–155.

5. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281:37661–37667.

6. Tasaki E, Hattori M, Sugita M (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmFc transcript. *Plant J* 62:560–570.

7. Yin P et al. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504:168–171.

8. Barkan A et al. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8:e1002910.

9. Salone V et al. (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581:4132–4138.

10. Okuda K et al. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell Online* 21:146–156.

11. Nakamura T, Sugita M (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* 582:4163–4168.

12. Boussardon C et al. (2012) Two interacting proteins are necessary for the editing of the *ndhD-1* site in Arabidopsis plastids. *Plant Cell Online* 24:3684–3694.

13. Takenaka M et al. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci* 109:5104–5109.

14. Bentolila S et al. (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci* 109:E1453–E1461.

15. Sun T et al. (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc Natl Acad Sci* 110:E1169–E1178.

16. Yoo S-D, Cho Y-H, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* 2:1565–1572.

17. Næsted H et al. (2004) Arabidopsis *VARIEGATED 3* encodes a chloroplast-targeted, zinc-finger protein required for chloroplast and palisade cell development. *J Cell Sci* 117:4807–4818.

18. Bentolila S, Oh J, Hanson MR, Bukowski R (2013) Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet* 9:e1003584.

19. Robbins JC, Heller WP, Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the Arabidopsis chloroplast *accD* transcript. *RNA* 15:1142–1153.

20. Avis JM, Clarke PR (1996) Ran, a GTPase involved in nuclear processes: its regulators and effectors. *J Cell Sci* 109 (Pt 10):2423–2427.

21. Sakamoto W (2003) Leaf-variegated mutations and their responsible genes in Arabidopsis thaliana. *Genes Genet Syst* 78:1–9.

22. Chateigner-Boutin A-L et al. (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56:590–602.

23. Zhou W et al. (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. *Plant J* 58:82–96.

24. Zhelyazkova P et al. (2012) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24:123–136.

25. Kupsch C et al. (2012) Arabidopsis chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell Online* 24:4266–4280.

26. Tillich M et al. (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci* 106:6002–6007.

27. Hegeman CE, Hayes ML, Hanson MR (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in Arabidopsis in vitro. *Plant J* 42:124–132.

28. Hayes ML, Giang K, Berhane B, Mulligan RM (2013) Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem* 288:36519–36529.

29. Yagi Y et al. (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol* 10:1419–1425.

30. Zehrmann A, Verbitskiy D, Härtel B, Brennicke A, Takenaka M (2010) RNA editing competence of trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain. *FEBS Lett* 584:4181–4186.

31. Sommer B, Köhler M, Sprengel R, Seeburg PH (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* 67:11–19.

32. Weigel D, Glazebrook J (2006) In planta transformation of Arabidopsis. *Cold Spring Harb Protoc* 2006:pdb.prot4668.

Conclusion

In higher plants, RNA editing is an important RNA processing step that is believed to correct the defective organellar transcripts via a C-to-U conversion. Plant RNA editing is carried out by protein complexs named editosomes. However, except for the recognition factor—members of the pentatricopeptide repeat protein family, the remaining components of the editosome were largely unknown. My thesis project was to identify the unknown editing factors that are essential for plant organelle RNA editing. In this dissertation, I have described three new editing factors—RIP1, ORRM1 and VAR3.

RIP1 was found in a proteomics study of the RARE1 co-immunoprecipitates. RARE1 is a chloroplast PPR protein that specifies editing of *accD-794*. RIP1 interacts with the PPR motifs of the RARE1 protein. While RARE1 is chloroplast protein, RIP1 is dual targeted to both chloroplasts and mitochondria. In a *RIP1* knockdown mutant, editing of 14 chloroplast sites and 266 mitochondrial sites (out of 368 sites tested) is affected, among which 107 mitochondrial sites have a major loss of editing. Thus RIP1 is an editing factor for both mitochondria and chloroplasts. RIP1 belongs to a small protein family which contains 10 family members. RIP2 and RIP9 were later shown to be major chloroplast RNA editing factors, while RIP3 and RIP8 were shown to be major mitochondrial editing factors. Other members of the RIP family are not directly involved in RNA editing. The RIP proteins obviously play a different role than the PPR specificity factors do. While one PPR editing factor controls only one to several editing events, a RIP editing factor affects a large number of editing events. The RIP proteins can selectively bind to the PPR proteins and also can form

homodimers and heterodimers. The RIP proteins are the first non-PPR *trans*-acting factors identified for plant organelle RNA editing.

ORRM1 was found by a homology search with the RIP proteins. ORRM1 possesses two truncated RIP domains at its N-terminus and an RNA recognition motif at its C-terminus. Loss of *ORRM1* results in editing defects at 24 chloroplast sites in Arabidopsis. Surprisingly, the editing activity of ORRM1 is carried by the RRM domain, not the RIP-RIP domain, which indicates that ORRM1 belongs to a different family than the RIP proteins. ORRM1 can bind to two PPR editing factors, CRR28 and OTP82. The recombinant ORRM1 protein can also bind near some editing sites *in vitro*. However, it is unlikely that *in vivo* ORRM1 can recognize every *orrm1*-dependent site via specific interactions with individual RNAs. It is possible that the RRM domain of the ORRM1 protein can bind to other protein factors in the editosomes. The Arabidopsis genome encodes 196 RRM-containing proteins. A small clade of RRM-containing proteins was found based on the high similarity to ORRM1. Additional members of this family could also be editing factors.

VAR3 was found to co-immunoprecipitate with epitope-tagged ORRM1. In a *VAR3* null mutant, editing of 30 chloroplast sites was affected, 21 of which were also *orrm1*-dependent. VAR3 interacts with the RIP-RIP domain of the ORRM1 protein. In addition, VAR3 can also bind to two PPR editing factors, CRR28 and OTP82. VAR3 has two Ran2 binding protein type zinc finger motifs. Three highly similar proteins were found by homology search, all of which are targeted to organelles. One of the VAR3 related proteins is tightly co-expressed with a mitochondrial editing PPR protein—MEF1. It is highly possible that other members of the VAR3 family are also

plant organelle editing factors. The fact that VAR3 and its related proteins can potentially bind zinc ions is very intriguing given that the plant C-to-U editing is known to be zinc-dependent.

So far, three additional editing factors have been identified, which reveals an unexpected complexity of the plant editing machinery. One particular editing event requires one PPR recognition factor and probably more than one RIP protein. ORRM1 participates in editing of 24 chloroplast sites. The remaining 13 chloroplast sites and mitochondrial sites probably require other proteins of the ORRM family. It is also possible that ORRM proteins can substitute one another for some sites. 30 chloroplast editing events require VAR3 while other VAR3-related proteins might be involved in editing of the mitochondrial and the remaining plastid sites. Both VAR3 and RIP proteins can dimerize. In addition, a second PPR protein might supply a DYW domain as the deaminase activity. Taken together, editosomes for different sites could have different components and different compositions. Hundreds of different editosomes possibly exist for the over 600 organelle editing sites. Further investigation is needed to understand the molecular mechanism of these apparatuses, which will be of great help to future engineering of the editosome.

Appendix I

Expression of the RIP1 protein in the *rip1* T-DNA insertional mutant

In chapter 2, one *RIP1* T-DNA insertional mutant line, FLAG_150D11, was reported. The mutant has a severe dwarf phenotype and editing defects in both mitochondria and plastids. The editing defects can be partially rescued by a transient expression of *RIP1* under a 35S promoter. Plants that have RIP1 silenced by VIGS (Virus induced gene silencing) show similar editing defects with the T-DNA mutant. Surprisingly, the expression of RIP1 mRNA is increased four to six fold in the mutant compared to the wild type siblings (Figure 2.6E). The T-DNA is inserted 140bp upstream of the start codon of RIP1 in the FLAG_150D11 line (Figure 2.6A). It is possible that insertion of the T-DNA knocks down the expression of RIP1 at the translational level rather than the transcriptional level. To test this hypothesis, RIP1 specific polyclonal antibodies were raised to examine the expression of the RIP1 protein in the T-DNA mutant.

Three regions of the RIP1 protein were selected for antigenic peptides (Figure Apx. 1). The highly conserved region was selected so that the antibody against this region will be able to recognize other members of the RIP1 family. Antigenic index was used to predict the antigenic index, and a well predicted region was selected. RIP1 has a proline-rich C terminus which is absent in other family members. Around 100 amino acids were selected from each of the three regions to make antigenic peptides.

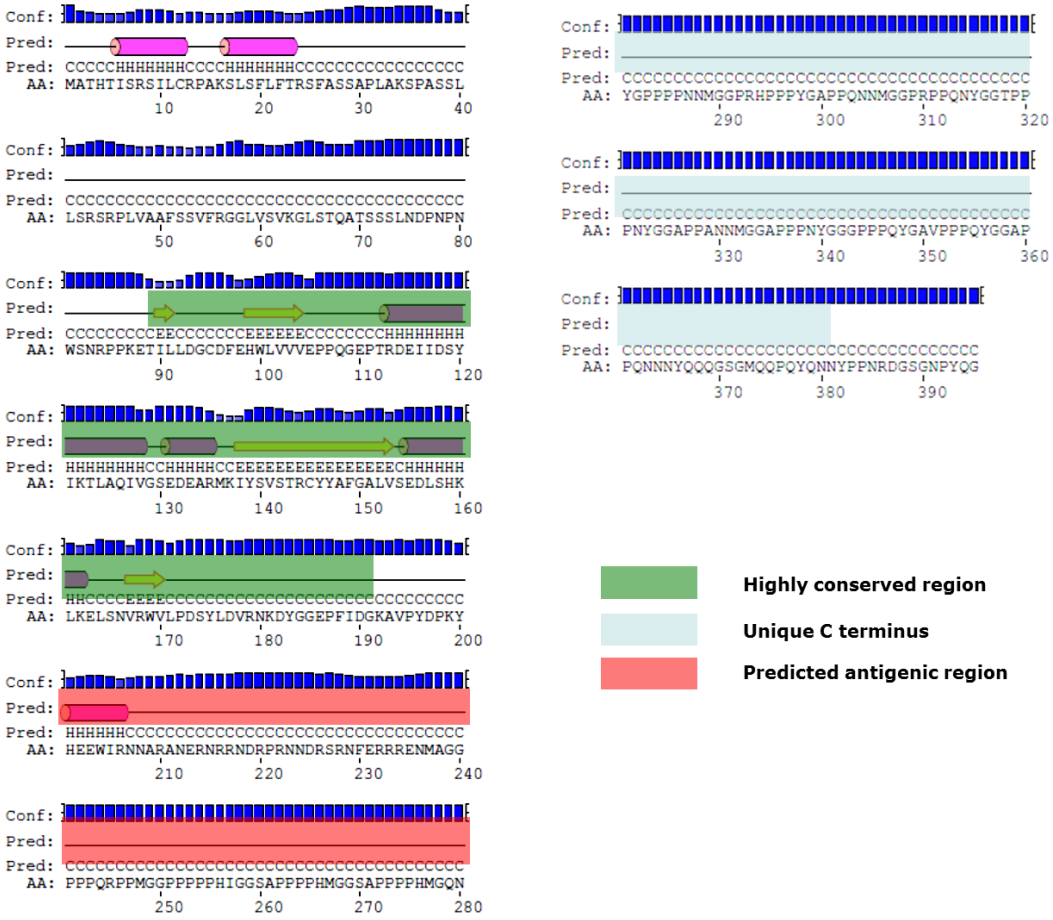


Figure Apx.1. RIP1 antigenic region selection. The protein secondary structure was predicted using software Jred. Highly conserved region is the region shared by the RIP family. Unique C terminus is the C-terminal region of RIP1 which is not found in other RIP proteins. Predicted antigenic region is according to the antigenic index.

Three regions were cloned and expressed in *E.coli* and the purified recombinant proteins were sent to Pocono Rabbit Farm and Laboratory to prepare rabbit antisera.

Figures Apx2-4 describe the RIP1 protein expression level in *rip1* T-DNA mutants and its wild type siblings detected by the antibodies that were raised against our antigenic peptides.

As shown in Figure Apx.2, the antibody against the RIP1 conserved region recognized a protein which has similar size with mature RIP1 (37kD). This protein was much less expressed in the *rip1* mutant. Besides, this antibody also recognizes another major band between 25kD and35kD, which may be another RIP protein. Intriguingly, in the *rip1* mutant that band is also much lighter compared to the wild type. It is possible that in the *rip1* mutant, expression of the other RIP proteins ais also negatively affected.

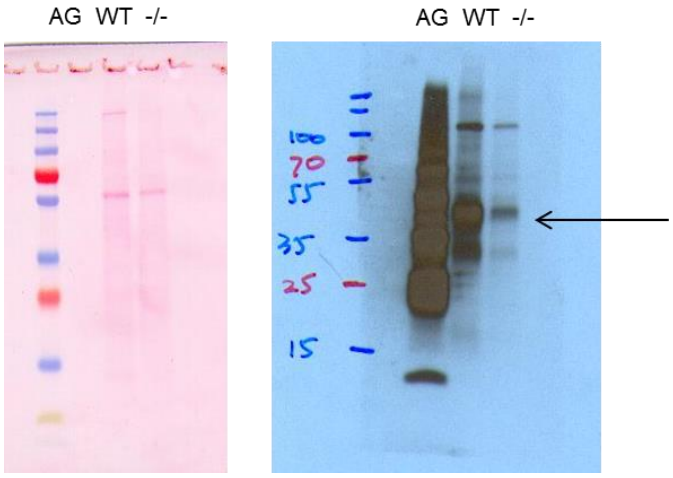


Figure Apx.2. RIP1 expression detected by anti-RIP1conserved region antibody. Ag, antigenic peptide. 10μg total leaf protein from the wild type or the *rip1* mutant was loaded. Arrow indicates the band of RIP1. Left, ponceau staining; right, immuno blot.

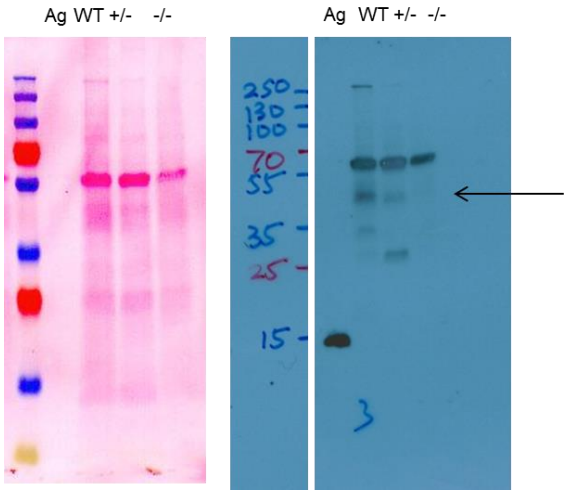


Figure Apx.3. RIP1 expression detected by anti-RIP1 predicted antigenic region antibody. Ag, antigenic peptide. 10 µg total leaf protein from the wild type or the *rip1* mutant was loaded. Arrow indicates the band of RIP1. Left, ponceau staining; right, immuno blot.

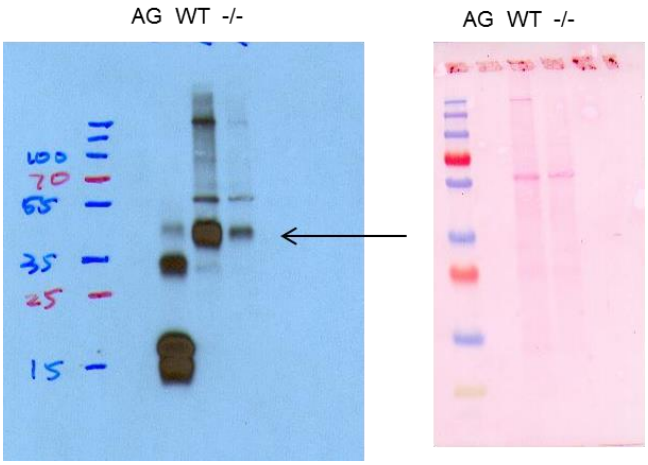


Figure Apx. 4. RIP1 expression detected by anti-RIP1unique C-terminus antibody. Ag, antigenic peptide. 10 µg total leaf protein from the wild type or the *rip1* mutant was loaded. Arrow indicates the band of RIP1. Left, immuno blot; right, ponceau staining.

In Figure Apx.3, the antibody against the predicted antigenic region also recognizes two major bands. One band is around 55kD and exists in every sample, and is likely to be the Rubisco large subunit. The other band is the size of RIP1 (37kD). Only the wild type and the heterozygous plants have detectable expression of this protein.

In Figure Apx.4, the antibody against the unique C-terminus of the RIP1 protein recognizes one major protein which is the right size for the mature RIP1. Expression of the RIP protein in the mutant plant is greatly reduced compared to the wild type plant.

Taken together, we generated three RIP1 antibodies for various purposes. Immunoblot results indicate that RIP1 protein expression is greatly reduced in the T-DNA insertional mutant line compared to the wild type. The insertion of the T-DNA probably inhibits the translation of *RIP1*.

Appendix II

Additional ORRM1/RIP1 interacting protein candidates from co-immunoprecipitation

Introduction

Plant RNA editing is carried out by a protein complex, for which three types of components have been reported—PPR proteins, RIP/MORF proteins and ORRM1. However, the editing enzyme is still elusive and whether other unknown components exist is not clear. In an effort to identify additional editing factors, I performed co-immunoprecipitation coupled with mass spectrometry for two known factors—RIP1 and ORRM1. In Chapter 4, I described a new factor—VAR3 found in ORRM1 co-IP. In this appendix, I will include other ORRM1 interacting candidates as well as RIP1 interacting candidates.

Result

C-terminal tagged RIP1 can greatly increase editing efficiency in *rip1* mutant

In order to immunoprecipitate RIP1 and ORRM1 and their interacting proteins, we epitope-tagged both proteins (Figure Apx.5). As mentioned in Chapter 4, C-terminal tag disrupts ORRM1 function. An N-terminal tagging strategy was applied to ORRM1. Conversely, the highly conserved region of RIP1 is on its N terminus so the tag was added to its C terminus. Two different constructs were made for this purpose.

RIP1-twin strepII has two strepII tags (WSHPQFEK(GGGS)$_2$GGSAWSHPQFEK) and HCF136 promoter from the pAUL vector which has similar expression level with RIP1, according to Genevestigator (1). RIP1-3Flag-strepII (designated RIP1-3FS in the remainder of this appendix) has 3 tandem Flag tags and a strepII tag and

Figure Apx.5. Schematic diagram of epitope tagged RIP1 and ORRM1. The RIP1 native promoter was cloned from -700bp to -1bp upstream of At3g15000. HCF136 promoter was from pAUL vector. CTP was cloned from RecA and mORRM1 is ORRM1 coding sequence without the transit peptide (54aa).

a native RIP1 promoter cloned from minus 700bp to minus 1bp upstream of RIP1 coding sequence. Both constructs were used to transform *rip1*+/- mutants via floral dipping. The transformants selected from hygromycin-containing media were then genotyped to find *rip1* homozygotes carrying a transgene. RNA was extracted from these transformants and editing efficiency was examined in poisoned primer extension for the defective sites in *rip1*. As shown in Figure Apx.6, editing extents of *cob-325* and *nad6-161* in *rip1* were severely reduced compared to wild type. Introducing a RIP1-twin strepII greatly increased the editing efficiency of both sites in the mutant from 15% to 60% and 10% to 60% respectively. Well expressing lines were picked for RIP1 co-immunoprecipitation.

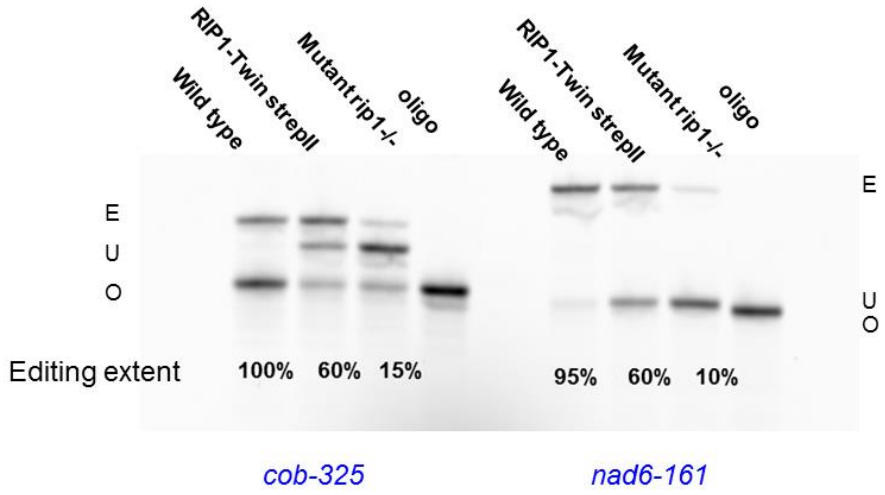


Figure Apx.6. Introduction of a C-terminal tagged RIP1 greatly increases editing extent of *rip1* mutant. RIP1-Twin strepII was in a *rip1* mutant background. Poisoned primer extension was used to examine editing efficiency of *cob-325* and *nad6-161*. E, edited. U, unedited. O, oligo.

Epitope tagged RIP1 can be efficiently immunoprecipitated

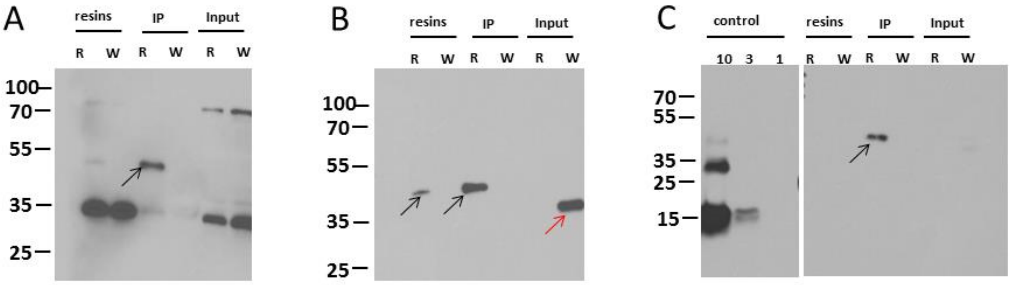


Figure Apx.7. Immunoblotting of RIP1-twin strepII co-IP. A, deteced by Strep-tactin-HRP. RIP1-twin strepII is low-expressed, only seen in the immunoprecipitate. B, detected by RIP1 antibody. Antibody detected one specific band in wild type plant which is endogenous RIP1. Transgenic line RIP1-twin strepII is in the *rip1* mutant background and has very low RIP1 expression level that is not detectable on the gel. The RIP1 antibody also recognizes tagged RIP1, which is slighter larger than RIP1. C, quantification of RIP1 in the IP. RIP1 antigenic region was used for quantification control. 1ng, 3ng, 10ng were loaded on the left. 10% of the total IP sample was loaded on the right of the gel. Black arrow indicates the tagged RIP1 ~40kD, red arrow indicates the endogenous RIP1~36kD. R, RIP1-twin strepII transformant; W, wild type *Col*.

Total leaf protein was extracted from twin strepII-tagged RIP1 transformants and wild-type plants. Strep-tactin resins were used to retain strepII-tagged protein and its interacting proteins. HCF136::RIP1-twin strepII was not well expressed, although HCF136 is documented to be a highly abundant protein and the transgenic line with the highest expression had been selected for this assay. In Figure Apx.7A, strep-tactin-HRP could not detect any expression of RIP1-twin strepII (around 40kD) in the crude extract. However, after the co-immunoprecipitation, tagged RIP1 was highly enriched in the IP sample (Figure Apx. 7A). Elution of RIP1 off the resin was quite effective, since it was barely detectable on the leftover resins. Notably, strep-tacin-HRP also recognizes two major non-specific proteins, 35kD and 70kD respectively. It is very likely they are biotin-containing proteins (2). However, neither of them seems to interfere with our assay. The 70kD protein was probably washed off during the washing step while the 35kD protein was retained on the resins even after the elution (Figure Apx.7A).

A RIP1 specific antibody was raised against the C-terminal 100 amino acids of RIP1. In order to confirm that the strep tag indeed pulled down tagged RIP1, RIP1 antibody was also used for immunoblotting the same sample with Figure Apx.3A. RIP1 is around 36kD while a twin strepII tagged RIP1 is around 40kD. The RIP1 antibody recognizes a specific band in a crude extract of wild-type control at around 36kD, but not in the transgenic plant (Figure Apx.7B). This again confirmed that RIP1-twin strepII was not well expressed in the transgenic plant. In the IP sample, no RIP1 was detected in the wild-type control indicating the binding was very specific.

For the transgenic line, a specific 40kD band was detected which is twin strepII tagged RIP1. Thus, RIP1 was effectively immunoprecipitated by the twin strepII tag.

In order to quantify the RIP1 amount in the IP, the RIP1 antigenic peptide was used as a positive control in a Western blot (Figure Apx.7C). 10% of the total IP was equivalent to around 3ng positive control. This is a fairly low amount of bait protein for mass spectrometry so it was not used in later analysis.

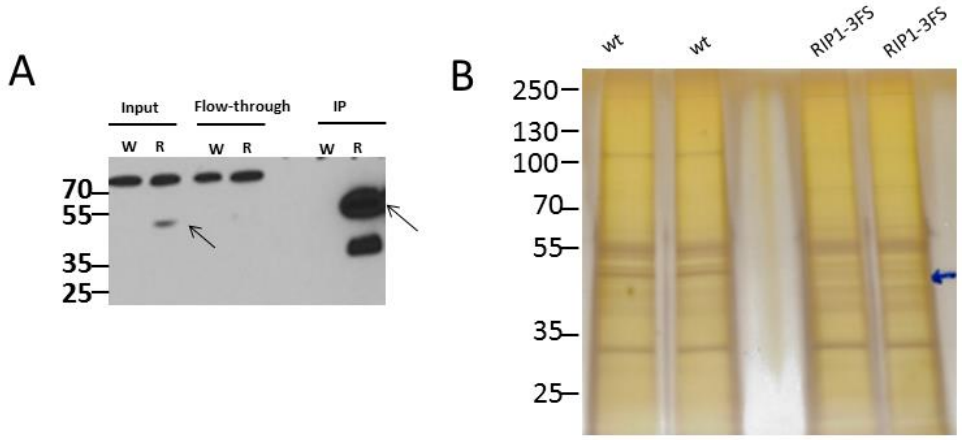


Figure Apx.8. RIP1-3Flag-strepII co-immunoprecipitation by strep-tactin. A, immunoblotting of RIP1-3Flag-strepII co-IP. Anti-Flag-HRP was used. W, wild type control; R, RIP1-3Flag-StrepII transformants. Input, protein extracts from each group; Flow-through, flow-through from Strep-Tactin resins; IP, elution from Strep-Tactin resin. Arrow indicates RIP1-3Flag-strepII which is around 42kD. B, silver staining of RIP1-3Flag-strepII co-IP. 10% of two replicates from each group were loaded to 10% Tris-Glyicine SDS-PAGE gel. Arrow indicates the approximate location of RIP1-3Flag-strepII.

In order to overcome the low expression of RIP1, a RIP1-3FS line with greater expression with a native RIP1 promoter was used later. Attempts to perform tandem purification using dual Flag and StrepII tag were not successful. It required a huge amount of protein to start with due to loss of protein during the procedures. Only

StrepII was used since it has higher specificity compared to Flag. Immunoprecipitates were first separated on SDS-PAGE and tagged RIP1 was detected by a strep-tactin-HRP antibody. RIP1-3FS is about 42kD. As shown in Figure Apx.8A, the Flag antibody recognizes a band in the transgenic plants which was missing in wild type control. One non-specific band around 80kD was observed in both the control and the experimental sample. The IP sample was greatly enriched for RIP1-3FS, as shown in Figure Apx.8A. The 80kD band disappeared after elution but a 35kD band was present. The IP samples from both wild-type control and RIP1-3FS transgenic plants were then separated on a SDS-PAGE gel and silver stained before subject to mass spectrometry (Figure Apx. 8B). Clear differences can be seen between two samples especially between 35kD and 55kD.

Mass spectrometry identified several candidates in ORRM1 and RIP1 co-IP

In chapter 4, the ORRM1 co-IP has been described. Both RIP1 and ORRM1 IP samples were subject to mass spectrometry to identify candidate interactors. A mock IP from wild type plants was also included to eliminate the non-specific interactions. The peptides found were then matched to the Arabidopsis protein database. Proteins that were enriched in the experimental sample compared to mock sample were selected as candidates.

Table Apx.1. Candidate proteins found in 3FS-ORRM1 immunoprecipitates using Flag tag

Accession No.	Number of peptides	Annotation	localization
AT3G20930.1	107	ORRM1-RNA-binding (RRM/RBD/RNP motifs) family protein	plastid
AT5G17790.1	12	variegated 3 - (var3) - interacts with NCED4/CCD4	plastid
AT3G46780.1	9	unknown protein (pTAC16)	thylakoid, plastid nucleoid

Table Apx.1 lists the three most abundant proteins found in ORRM1 IP that were not found in mock IP. Peptides from ORRM1 were found 107 times in our experiments, which makes it at the top of the list. This confirmed that tagged ORRM1 was efficiently pulled down by the Flag resins. The second protein is VAR3, which has been described in Chapter 4. The third protein is pTAC16, which has been found in thylakoids and nucleoids in some proteomics studies (PPDB, ppdb.tc.cornell.edu). pTAC proteins were first reported as Transcriptionally Active Chromosome proteins (3). pTAC proteins associate with PEP (plastid encoded plastid RNA polymerase), playing critical roles in plastid gene expression regulation. Functions of many pTAC proteins are still unknown, including pTAC16. Therefore, further investigation is needed to know whether pTAC16 is also involved plastid RNA editing.

Table Apx.2. Candidate proteins found in RIP1-3FS co-immunoprecipitates using strepII tag

Accession No.	number of peptides	Annotation	Note	localizaiton
AT3G15000.1	**10**	**RIP1 - mitochondria & plastid**		**mitochondria; plastid**
AT3G07660.1	4	Kinase-related protein of unknown function (DUF1296)	low abundant, phosphoprotein (since it was found in several p-proteome studies) without predicted mTP or cTP. Here we identified only 1 peptide (4 times). The single rice homolog has predicted mTP.	
AT1G26680.1	2	transcriptional factor B3 family protein	identified with a single peptide CSEIVLKNEQGVK (2x) in one of the 2 replicates. Chloroplast predicted. We found it multiple times in the chloroplast with this peptide (but never in total leaf extracts). This suggest that it could be a bonafide plastid protein. B3 DNA biding repeat protein	plastid

Table Apx.2 lists candidate proteins from the RIP1 IP. RIP1 peptides were found 10 times, which is a low number compared to the ORRM1 experiment. At3g07660 which encodes a protein of unknown function (DUF1296) was found 4 times. Since it is a low abundance protein in plants, it is unlikely to be a non-specific binding protein. The homologous gene in rice has a well-predicted mitochondria

presequence, which implies this might be a mitochondrial protein. At1g26680 was found 2 times, encoding a transcription factor B3 family protein. Neither of these candidates has a known function, so whether they are editing factors still needs further study.

Table Apx.3. candidate proteins found in 3FS-ORRM1 immunoprecipitates using Flag tag--complete list					
Accession	**adjusted abundance**	**Annotation**	**Curated Location**	**TargetP**	**ORMM1/wt**
AT3G20930.1	107	RNA-binding (RRM/RBD/RNP motifs) family protein	plastid	C	only in ORMM1
AT5G17790.1	12	variegated 3 - (var3) - interacts with NCED4/CCD4	plastid	C	only in ORMM1
AT5G02500.2	22.5	HSP70-1 (HSC70-1) (not plastid) - very abundant - dual localized nucleus&cytosol	nucleus; cytosol	_	only in ORMM1
AT5G17920.1	9	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase (ATCIMS) - abundant	cytosol	_	only in ORMM1
AT5G17920.2	9	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase (ATCIMS) - abundant	cytosol	_	only in ORMM1
AT3G46780.1	9	unknown protein (pTAC16)	thylakoid	C	only in ORMM1
AT5G09810.1	6	actin 7 (ACT7) / actin 2	not plastid	_	only in ORMM1
AT4G20850.1	5	Tripeptidyl-peptidase (TTP2)	plastid stroma	C	only in ORMM1
AT4G35090.1	6	catalase 2 (CAT2)	peroxisome	_	only in ORMM1
AT4G35090.2	6	catalase 2 (CAT2)	peroxisome	_	only in ORMM1
AT4G27440.1	4	PORB - constitutive expression - main protein in barley	thylakoid-peripheral-stromal-side	C	only in ORMM1
AT4G27440.2	4	PORB - constitutive expression - main protein in barley	thylakoid-peripheral-stromal-side	C	only in ORMM1
AT1G07920.1	3.3	elongation factor 1-alpha / EF-1-alpha	cytosol	_	only in ORMM1
AT1G07930.1	3.3	elongation factor 1-alpha (E-Tu)	not plastid	_	only in ORMM1
AT1G07930.2	3.3	elongation factor 1-alpha / EF-1-alpha	not plastid	_	only in ORMM1
AT1G07940.1	3.3	elongation factor 1-alpha / EF-1-alpha	not plastid	_	only in ORMM1
AT1G07940.2	3.3	elongation factor 1-alpha / EF-1-alpha		_	only in ORMM1
AT5G60390.1	3.3	elongation factor 1-alpha / EF-1-alpha - dual location cytosol & Nucleus	nucleus; cytosol	_	only in ORMM1
AT5G60390.2	3.3	elongation factor 1-alpha / EF-1-alpha - dual location cytosol & Nucleus	nucleus; cytosol	_	only in ORMM1
AT5G60390.3	3.3	elongation factor 1-alpha / EF-1-alpha - dual location cytosol & Nucleus	nucleus; cytosol	_	only in ORMM1

AT1G49240.1	4.6	actin 8 (ACT8)	not plastid	_	only in ORMM1
AT3G18780.1	4.6	actin 2 (ACT2)	vacuole	_	only in ORMM1
AT3G18780.2	4.6	actin 2 (ACT2)	vacuole	_	only in ORMM1
AT1G09640.2	3	elongation factor 1B-gamma, eEF-1B gamma	cytosol	_	only in ORMM1
AT1G48120.1	3	hydrolases;protein serine/threonine phosphatases		_	only in ORMM1
AT1G30230.1	3	elongation factor 1-beta / EF-1-beta	not plastid	_	only in ORMM1
AT1G30230.2	3	elongation factor 1-beta / EF-1-beta	not plastid	_	only in ORMM1
AT5G65730.1	2.6	xyloglucan endotransglucosylase/hydrolase 6		S	only in ORMM1
AT5G38410.1	4.8	Rubisco small subunit 3b (RBCS-3B)	plastid stroma	C	only in ORMM1
AT5G38410.2	4.8	Rubisco small subunit 3b (RBCS-3B)	plastid stroma	C	only in ORMM1
AT5G38410.3	4.8	Rubisco small subunit 3b (RBCS-3B)	plastid stroma	C	only in ORMM1
AT5G38420.1	4.8	Rubisco small subunit 2b (RBCS-2b)	plastid stroma	C	only in ORMM1
AT1G56190.1	2	phosphoglycerate kinase-1 (PGK-2)	plastid stroma	C	only in ORMM1
AT1G56190.2	2	phosphoglycerate kinase-1 (PGK-2)	plastid stroma	_	only in ORMM1
AT3G12780.1	2	phosphoglycerate kinase-1 (PGK-1)	plastid stroma	C	only in ORMM1
AT3G45140.1	2	lipoxygenase AtLOX2, plastid	plastid stroma	C	only in ORMM1
AT4G34200.1	2	D-3-phosphoglycerate dehydrogenase (3-PGDH) - EDA9 (embryo sac development arrest 9) -very abundant	plastid	C	only in ORMM1
AT5G38430.1	1.3	Rubisco small subunit 1b (RBCS-1b)	plastid stroma	C	only in ORMM1
AT3G25230.1	2	peptidyl-prolyl cis-trans isomerase / rotamase FK506-binding protein (ROF1)		_	only in ORMM1
AT3G25230.2	2	peptidyl-prolyl cis-trans isomerase / rotamase FK506-binding protein (ROF1)		_	only in ORMM1
AT1G20010.1	1	tubulin beta-5 chain (TUB5)	not plastid	_	only in ORMM1
AT1G71500.1	1	Rieske [2Fe-2S] domain (TEF5-Merchant)	thylakoid	C	only in ORMM1
AT2G13360.1	1	alanine-glyoxylate aminotransferase (AGT1)	peroxisome	_	only in ORMM1
AT2G13360.2	1	alanine-glyoxylate aminotransferase (AGT1)	peroxisome	_	only in ORMM1
AT2G47730.1	1	glutathione transferase GSTF8 (GST6) - cytosol and plastid	stroma; cytosol	C	only in ORMM1

AT3G13490.1	1	tRNA synthetase class II (D, K and N) protein (dual targeted cTP/mTP)	plastid stroma; mitochondria	M	only in ORMM1
AT4G20890.1	1	tubulin beta-9 chain (TUB9)	nucleus	_	only in ORMM1
AT5G44340.1	1	tubulin beta-4 chain (TUB4)	not plastid	_	only in ORMM1
AT1G72150.1	1	PATELLIN 1 cell plate frafficking (PATL1)	not plastid	_	only in ORMM1
AT2G39800.1	1	delta 1-pyrroline-5-carboxylate synthetase A / P5CS A (P5CS1)	not plastid	_	only in ORMM1
AT2G39800.2	1	delta 1-pyrroline-5-carboxylate synthetase A / P5CS A (P5CS1)	not plastid	_	only in ORMM1
AT2G39800.3	1	delta 1-pyrroline-5-carboxylate synthetase A / P5CS A (P5CS1)	not plastid	_	only in ORMM1
AT2G39800.4	1	delta 1-pyrroline-5-carboxylate synthetase A / P5CS A (P5CS1)	not plastid	_	only in ORMM1
AT3G55610.1	1	delta 1-pyrroline-5-carboxylate synthetase B / P5CS B (P5CS2)	cytosol	_	only in ORMM1
ATCG00470.1	1	CF1e - atpE	thylakoid-peripheral-stromal-side	_	only in ORMM1
AT1G55800.1	1	similar to ATP binding / nucleotide binding / phenylalanine-tRNA ligase		_	only in ORMM1
AT3G09880.1	1	Protein phosphatase 2A regulatory B subunit		C	only in ORMM1
AT3G44310.1	1082	nitrilase 1 (NIT1 or NITl) - dual localized in cytosol and nucleus	nucleus; cytosol	_	1.22
AT3G44310.3	1082	nitrilase 1 (NIT1 or NITl) - dual localized in cytosol and nucleus	nucleus; cytosol	_	1.22
ATCG00490.1	361	Rubisco large subunit (RBCL)	plastid stroma	_	1.71
AT4G10320.1	277	isoleucyl-tRNA synthetase, putative		_	1.54
AT2G01210.1	201	leucine-rich repeat transmembrane protein kinase, putative	plasma membrane	S	1.34
AT5G35200.1	138	ENTH/ANTH/VHS superfamily protein		_	0.82
AT5G26710.1	134	glutamate-tRNA ligase, putative / glutamyl-tRNA synthetase, putatuve / GluRS		C	0.79
ATCG00120.1	123	CF1a - atpA	thylakoid-peripheral-stromal-side	_	1.02

AT5G43940.1	106	alcohol dehydrogenase class III (ADHIII or ADH2) / glutathione-dependent formaldehyde dehydrogenase (GSH-FDH)	not plastid	_	0.74
AT5G43940.2	85	alcohol dehydrogenase class III (ADHIII or ADH2) / glutathione-dependent formaldehyde dehydrogenase (GSH-FDH)	not plastid	_	0.39
AT4G35630.1	128.2	phosphoserine aminotransferase - abundant	plastid stroma	C	1.18
AT3G08590.1	97	2,3-biphosphoglycerate-independent phosphoglycerate mutase	not plastid	_	0.94
AT3G08590.2	97	2,3-biphosphoglycerate-independent phosphoglycerate mutase	not plastid	_	0.94
AT2G17630.1	95.8	phosphoserine aminotransferase, putative		C	0.95
ATCG00480.1	89	CF1b - atpB	thylakoid-peripheral-stromal-side	_	0.82
AT2G40660.1	63	tRNA-binding region domain-containing protein		_	0.47
AT3G28220.1	66	MATH domain-containing protein	plastid	_	1.06
AT3G52930.1	59.4	fructose-bisphosphate aldolase	cytosol	_	1.18
AT2G39730.1	79	Rubisco activase	plastid stroma	C	1.93
AT2G39730.2	58	Rubisco activase	plastid stroma	C	1.15
AT2G39730.3	58	Rubisco activase	plastid stroma	C	1.15
AT1G57720.2	56.7	elongation factor 1B-gamma, eEF-1B gamma		_	1.13
AT1G57720.1	56.7	elongation factor 1B-gamma, eEF-1B gamma	not plastid	_	1.13
AT1G42970.1	49.4	glyceraldehyde-3-phosphate dehydrogenase B (GAPB)	plastid stroma	C	1.09
AT3G26650.1	62.8	glyceraldehyde 3-phosphate dehydrogenase A-1 (GAPA-1)	plastid stroma	C	1.95
AT5G02500.1	28.5	HSP70-1 (HSC70-1) (not plastid) - very abundant - dual localized nucleus&cytosol	nucleus; cytosol	_	0.50
AT4G37800.1	40.4	xyloglucan:xyloglucosyl transferase, putative / xyloglucan endotransglycosylase	not plastid	S	1.38
AT3G11770.1	24	unknown protein		_	0.71
AT1G09640.1	26.3	elongation factor 1B-gamma, eEF-1B gamma	cytosol	_	1.12
AT5G19510.1	30	elongation factor 1B alpha-subunit 2 (eEF1Balpha2)	not plastid	_	1.50
AT4G20360.1	42	elongation factor Tu (EF-Tu-1), plastid	plastid stroma	C	2.82

AT1G19480.1	24			C	1.40
AT1G19480.2	24			C	1.40
AT3G57560.1	18	N-acetyl glutamate kinase - interacts with PII	plastid stroma	C	1.25
AT4G29270.1	13			S	0.63
AT1G12900.1	21.8	glyceraldehyde-3-phosphate dehydrogenase A-2 (GAPA-2)	plastid stroma	C	2.07
AT1G12900.3	21.8	glyceraldehyde-3-phosphate dehydrogenase A-2 (GAPA-2)	plastid stroma	_	2.07
AT1G12900.4	21.8	glyceraldehyde-3-phosphate dehydrogenase A-2 (GAPA-2)	plastid stroma	_	2.07
AT1G67090.1	12.3	Rubisco small subunit-4 (RBCS-4)	plastid stroma	C	0.76
AT1G20620.1	11	catalase 3 (CAT3)	peroxisome	_	0.69
AT1G20620.2	11	catalase 3 (CAT3)	peroxisome	_	0.69
AT1G20620.5	11	catalase 3 (CAT3)	peroxisome	_	0.69
AT2G36460.1	21.9	Aldolase	cytosol	_	2.42
AT4G26530.1	14.4	fructose-bisphosphate aldolase, cytoplasmic	cytosol	_	1.32
AT4G26530.2	14.4	fructose-bisphosphate aldolase, cytoplasmic	cytosol	_	1.32
AT1G58270.1	9	meprin and TRAF homology domain-containing protein / MATH domain-containing protein		S	0.50
AT3G09440.1	8.5	HSP70 (HSC70-3) (not plastid)	not plastid	_	0.42
AT3G09440.2	8.5	HSP70 (HSC70-3) (not plastid)	not plastid	_	0.42
AT2G23390.1	22	unknown protein	plastid	M	3.40
AT1G61520.1	7	LHCI-3 - LHCI-680A CAB4	thylakoid-integral	C	0.40
AT1G61520.2	7	LHCI-3 - LHCI-680A CAB4	thylakoid-integral	_	0.40
AT1G61520.3	7	LHCI-3 - LHCI-680A CAB4	thylakoid-integral	C	0.40
AT5G26000.1	7	thioglucoside glucohydrolase 1 (TGG1) (myrosinase)	not plastid	S	0.40
AT5G26000.2	7	thioglucoside glucohydrolase 1 (TGG1) (myrosinase)	not plastid	S	0.40
AT1G55490.1	9.1	Cpn60-beta-2	plastid stroma	C	1.12
AT1G55490.2	9.1	Cpn60-beta-2	plastid stroma	C	1.12
AT3G13470.1	9.1	Cpn60-beta-1	plastid stroma	C	1.12
AT5G56500.1	5.6	Cpn60-beta-3 - model .2 has higher score than model .1	plastid stroma	C	0.30
AT5G56500.2	5.6	Cpn60-beta-3 - model .2 has higher score than model .1	plastid stroma	C	0.30
AT1G13440.1	19.5	glyceraldehyde-3-phosphate dehydrogenase C-2 (GAPC-2)	nucleus	_	3.88
AT3G04120.1	19.5	glyceraldehyde-3-phosphate dehydrogenase C-1 (GapC-1)	cytosol	_	3.88
AT1G13440.2	16.5	glyceraldehyde-3-phosphate dehydrogenase C-2 (GAPC-2)	nucleus	_	3.13
AT5G64430.1	5	octicosapeptide/Phox/Bem1p (PB1)		_	0.25

AT5G24280.1	3			_	only in wt
AT4G04640.1	17	CF1y - atpC	thylakoid-peripheral-stromal-side	C	7.50
AT5G35630.1	5	glutamate-ammonia ligase (GS2), chloroplast	plastid stroma	C	1.50
AT5G35630.2	5	glutamate-ammonia ligase (GS2), chloroplast	plastid stroma	C	1.50
AT5G35630.3	5	glutamate-ammonia ligase (GS2), chloroplast	plastid stroma	C	1.50
AT1G05910.1	2	cell division cycle protein 48 (CDC48)		_	only in wt
AT2G29560.1	2	Enolase (ENO3)	cytosol	_	only in wt
AT1G12900.2	6.5	glyceraldehyde-3-phosphate dehydrogenase A-2 (GAPA-2)	plastid stroma	S	3.33
AT5G14740.1	4.5	beta carbonic anhydrase 2 (CA2) - gene model with cTP	plastid stroma	_	2.00
AT5G14740.2	4.5	beta carbonic anhydrase 2 (CA2)- N-term truncated gene model without cTP	plastid stroma	_	2.00
AT3G01500.1	1.5	beta carbonic anhydrase-1 (beta CA1) - this model lacks the cTP	plastid stroma	_	only in wt
AT3G01500.2	1.5	beta-carbonic anhydrase-1 (beta CA1) - model has CTP	plastid stroma	C	only in wt
AT3G01500.3	1.5	beta-carbonic anhydrase-1 (beta CA1) - model with cTP but C-term truncated	plastid stroma	C	only in wt
AT5G14740.3	1.5	beta-carbonic anhydrase 2 (CA2) (little support for this gene model)	plastid stroma	_	only in wt
AT5G14740.4	1.5	beta-carbonic anhydrase 2 (CA2) (little support for this gene model)	plastid stroma	_	only in wt
AT5G14740.5	1.5	beta-carbonic anhydrase 2 (CA2) (little support for this gene model)	plastid stroma	_	only in wt
AT3G60750.1	2	transketolase-1 (TKL-1)	plastid stroma	C	1.00
AT3G60750.2	2	transketolase-1 (TKL-1)	plastid stroma	C	1.00
AT1G04820.1	1	tubulin alpha-2/alpha-4 chain (TUA4)	not plastid	_	only in wt
AT1G31330.1	1	psaF- subunit III - LTP - hydrophobic	thylakoid-integral	C	only in wt
AT1G50010.1	1	unknown protein	plastid	_	only in wt
AT1G71080.1	1	RNA polymerase II transcription elongation factor		_	only in wt
AT3G24503.1	1	aldehyde dehydrogenase (ALDH1a)	not plastid	_	only in wt
AT4G14960.1	1	tubulin alpha-6 chain (TUA6)	cytoskeleton	_	only in wt
AT4G14960.2	1	tubulin alpha-6 chain (TUA6)	cytoskeleton	_	only in wt
AT4G38630.1	1	26S proteasome regulatory subunit S5A (RPN10) (MBP1)	cytosol	_	only in wt
AT5G06450.1	1	unknown protein		_	only in wt

Table Apx.4.Candidate proteins found in RIP1-3FS co-immunoprecipitates using strepII tag--complete list

Accession	total adjusted abundance	Annotation	Curated Location	TargetP	RIP-Strep/wt
AT3G14415.1	16	glycolate oxidase-1 (GOX-1) - peroxisome	peroxisome	_	only in RIP-Strep
AT4G10340.1	13	LHCII-5 - CP26	thylakoid-integral	C	only in RIP-Strep
AT1G36180.1	10.7	acetyl-CoA carboxylase - ACC2	plastid	C	only in RIP-Strep
AT3G15000.1	**10**	**RIP1 - mitochondria & plastid**	**mitochondria; plastid**	**C**	**only in RIP-Strep**
AT1G55490.1	5.5	Cpn60-beta-2	plastid stroma	C	only in RIP-Strep
AT3G13470.1	5.5	Cpn60-beta-1	plastid stroma	C	only in RIP-Strep
AT3G09440.1	5	HSP70 (HSC70-3) (not plastid)	not plastid	_	only in RIP-Strep
AT3G52960.1	5	Peroxiredoxin IIE (PrxII E)	plastid stroma	C	only in RIP-Strep
AT4G09010.1	5	TL29	thylakoid-peripheral-lumenal-side	C	only in RIP-Strep
AT3G07660.1	4	Kinase-related protein of unknown function (DUF1296)		_	only in RIP-Strep
AT1G07890.1	3	L-ascorbate peroxidase 1, cytosolic (APX1)	cytosol	_	only in RIP-Strep
AT4G01310.1	3	50S ribosomal protein L5	plastid ribosome	C	only in RIP-Strep
AT4G27440.1	3	PORB - constitutive expression - main protein in barley	thylakoid-peripheral-stromal-side	C	only in RIP-Strep
AT5G13510.1	3	50S ribosomal protein L10	plastid ribosome	C	only in RIP-Strep
AT5G54190.1	3	PORA - down regulated in light; high in Prolamellar body	thylakoid-peripheral-stromal-side	C	only in RIP-Strep
AT5G64430.1	3	octicosapeptide/Phox/Bem1p (PB1)		_	only in RIP-Strep
AT1G26680.1	2	transcriptional factor B3 family protein		C	only in RIP-Strep
AT1G32060.1	2	phosphoribulokinase-2 (PRK-2)	plastid stroma	C	only in RIP-Strep
AT2G03440.1	2	nodulin-related, similar to Early nodulin 12B precursor (N-12B)		_	only in RIP-Strep
AT3G08740.1	2	elongation factor P (EF-P)	plastid stroma	C	only in RIP-Strep
AT3G50820.1	2	psbO OEC33-like	thylakoid-peripheral-lumenal-side	C	only in RIP-Strep
AT4G34620.1	2	30S ribosomal protein S16B (n-encoded homologue of c-enc S16A) - dual targeted mito & plastid	mitochondria; plastid ribosome	M	only in RIP-Strep
AT5G24490.1	2	PSRP-1 associates with 30S (translation factor? Interact with RRF?)	plastid ribosome	C	only in RIP-Strep
AT1G11860.1	1	Glycine cleavage T-protein	not plastid	M	only in RIP-Strep
AT1G48830.1	1	40S ribosomal protein S7 (RPS7A)	cytosol	_	only in RIP-Strep
AT1G56070.1	1	elongation factor 2, EF-2	not plastid	_	only in RIP-Strep
AT1G67700.1	1	unknown protein	plastid	M	only in RIP-Strep
AT2G36530.1	1	Enolase (ENO2 also LOS2) dual targeted nucleus & cytosol	cytosol; nucleus	_	only in RIP-Strep
AT2G40560.1	1	Protein kinase superfamily protein		_	only in RIP-Strep
AT2G47730.1	1	glutathione transferase GSTF8 (GST6) - cytosol and plastid	cytosol; plastid	C	only in RIP-Strep

AT2G47940.1	1	DegP2 - HhoA homologue or DegQ	plastid stroma	C	only in RIP-Strep
AT3G04790.1	1	ribose 5-phosphate isomerase (PRI)	plastid stroma; thylakoid-peripheral-stromal-side	C	only in RIP-Strep
AT3G13930.1	1	Dihydrolipoamide Acetyltransferase, pyruvate DH complex	mitochondria	M	only in RIP-Strep
AT3G18890.1	1	Tic62 -interacts with FNR (dual localized thylakoid & envelope)	envelope-inner-peripheral-stromal-side; thylakoid	C	only in RIP-Strep
AT4G26530.1	1	fructose-bisphosphate aldolase, cytoplasmic	cytosol	_	only in RIP-Strep
AT4G34030.1	1	3-methylcrotonyl-CoA carboxylase non-biotinylated subunit (MCCB)	mitochondria	M	only in RIP-Strep
AT5G16285.1	1	F-box family protein		_	only in RIP-Strep
AT5G66190.1	1	FNR-1	thylakoid-peripheral-stromal-side	C	only in RIP-Strep
ATCG00680.1	1	psbB CP47	thylakoid-integral	M	only in RIP-Strep
ATCG00750.1	1	30S ribosomal protein S11	plastid ribosome	C	only in RIP-Strep
AT2G28000.1	18	Cpn60-alpha-1	plastid stroma	C	17.00
AT1G32990.1	6	50S ribosomal protein L11	plastid ribosome	C	5.00
AT3G62030.1	6	peptidylprolyl isomerase ROC4 (CYP20-3)	plastid stroma	C	5.00
AT3G11630.1	5	2-Cys Peroxiredoxin A (Prx A; formerly named BAS1)	plastid stroma; thylakoid-peripheral-stromal-side	C	4.00
AT3G55800.1	5	sedoheptulose-bisphosphatase (SBPase)	plastid stroma	C	4.00
AT1G01090.1	40	Plastidial Pyruvate Dehydrogenase E1alpha subunit	envelope-inner-peripheral-stromal-side	C	3.00
AT1G09340.1	4	HIP1.3 (cytosol - heteroglycan-interaction) or Rap38/CSP41B (plastid-RNA binding)	cytosol; plastid stroma	_	3.00
AT3G48870.1	5.5	ClpC2 (also named HSP93-III) - highly similar to ClpC1	plastid stroma	C	2.67
AT5G50920.1	5.5	ClpC1 (also named HSP93-V) - highly similar to ClpC2	envelope-inner-peripheral-stromal-side; plastid stroma	C	2.67
AT3G46780.1	35	unknown protein (pTAC16)	thylakoid	C	2.50
AT5G15530.1	13.2	biotin carboxyl carrier protein BCCP-2 - part of ACCase complex (found only in reproductive organs)	envelope-inner-peripheral-stromal-side	C	2.38
AT1G30120.1	20.5	Plastidial Pyruvate Dehydrogenase E1 - beta subunit	plastid	C	2.15
AT5G49910.1	25.1	cpHSP70-2 (Dnak homologue)	plastid stroma	C	2.10
AT4G28750.1	6	psaE-1 subunit IV - stromal side	thylakoid-peripheral-stromal-side	C	2.00
AT1G44575.1	3	psbS (NPQ4 - null mutant)	thylakoid-integral	C	2.00
AT3G26520.1	3	gamma tonoplast intrinsic protein1;2 (gamma-TIP2)	vacuole-tonoplast	_	2.00
AT3G44310.1	3	nitrilase 1 (NIT1 or NITl) - dual localized in cytosol and nucleus	cytosol; nucleus	_	2.00
AT3G01500.1	50.3	beta carbonic anhydrase-1 (beta CA1) - this model lacks the cTP	plastid stroma	_	1.96

AT3G16950.1	35.1	E3 -dihydrolipoamide dehydrogenase 1, plastidic (ptlpd1)	plastid	C	1.88
AT4G24280.1	26.9	cpHSP70-1 (DnaK homologue)	plastid stroma	C	1.72
AT3G60750.1	40	transketolase-1 (TKL-1)	plastid stroma	C	1.67
ATCG00490.1	345	Rubisco large subunit (RBCL)	plastid stroma	_	1.61
ATCG00540.1	13	petA - cytochrome f (cleavable ss of 35 aa)	thylakoid-integral	_	1.60
AT3G27830.1	6.5	50S ribosomal protein L12-A (homologous to L12-C)	plastid ribosome	C	1.60
AT3G27850.1	6.5	50S ribosomal protein L12-C (homologous to L12-A)	plastid ribosome	C	1.60
AT5G14740.1	43.7	beta carbonic anhydrase 2 (CA2) - gene model with cTP	plastid stroma	_	1.57
ATCG00480.1	133	CF1b - atpB	thylakoid-peripheral-stromal-side	_	1.51
AT1G74730.1	5	unknown protein	thylakoid	C	1.50
AT2G21330.1	10.4	fructose-bisphosphate aldolase-1 (SFBA-1)	plastid stroma	C	1.42
AT4G16155.1	84	E3 - dihydrolipoamide dehydrogenase 2 (ptlpd2), plastid- subunit of pyruvate dehydrogenase complex	plastid	_	1.41
AT5G66570.1	12	psbO OEC33	thylakoid-peripheral-lumenal-side	C	1.40
AT4G09320.1	7	NDPK1 - very abundant & multiple localizations	cytosol; nucleus; peroxisome	_	1.33
AT1G42970.1	83.7	glyceraldehyde-3-phosphate dehydrogenase B (GAPB)	plastid stroma	C	1.32
AT4G20360.1	71	elongation factor Tu (EF-Tu-1), plastid	plastid stroma	C	1.29
AT5G16390.1	95.9	biotin carboxyl carrier protein BCCP-1 - part of ACCase complex	envelope-inner-peripheral-stromal-side; plastid stroma	C	1.28
ATCG00120.1	200	CF1a - atpA	thylakoid-peripheral-stromal-side	_	1.27
AT3G54050.1	9	fructose-bisphosphatase (FBPA), high CEF1 (hcef1)	plastid stroma	C	1.25
AT1G36160.1	453.3	acetyl-CoA carboxylase - ACC1	cytosol	_	1.24
AT1G15820.1	11	LHCII-6 - CP24	thylakoid-integral	M	1.20
AT1G12900.1	54.1	glyceraldehyde-3-phosphate dehydrogenase A-2 (GAPA-2)	plastid stroma	C	1.18
AT1G06680.1	26	PsbP-1 OEC23 Tat lTP (model with cTP)	thylakoid-peripheral-lumenal-side	C	1.17
AT1G72610.1	15	germin-like protein (GER1)	extracellular	S	1.14
AT1G13440.1	15.5	glyceraldehyde-3-phosphate dehydrogenase C-2 (GAPC-2)	nucleus	_	1.07
AT3G04120.1	15.5	glyceraldehyde-3-phosphate dehydrogenase C-1 (GapC-1)	cytosol	_	1.07
AT2G39730.1	115	Rubisco activase	plastid stroma	C	1.02
AT4G38970.1	23.6	fructose-bisphosphate aldolase-2 (SFBA-2)	plastid stroma; plastoglobules	C	1.02
AT4G05180.1	71.4	psbQ OEC16-like Tat lTP	thylakoid-peripheral-lumenal-side	C	1.02

AT3G08940.1	24	LHCII-4.2 - CP29 - model .2 is better than model .1	thylakoid-integral	C	1.00
AT4G09650.1	6	CF1d - atpD	thylakoid-peripheral-stromal-side	C	1.00
AT4G32470.1	6	ubiquinol-cytochrome C reductase complex 14 kDa protein	mitochondria	M	1.00
AT1G71500.1	4	Rieske [2Fe-2S] domain (TEF5-Merchant)	thylakoid	C	1.00
AT3G15190.1	4	30S ribosomal protein S20	plastid ribosome	C	1.00
ATCG00770.1	4	30S ribosomal protein S8	plastid ribosome	_	1.00
AT1G13930.1	2	abundant unknown protein - weak similarity to 60S ribosomal su	not plastid	_	1.00
AT5G14220.1	2	protoporphyrinogen oxidase (PPO) (PPOX)		_	1.00
AT1G07920.1	1.6	elongation factor 1-alpha / EF-1-alpha	cytosol	_	1.00
AT1G07930.1	1.6	elongation factor 1-alpha (E-Tu)	not plastid	_	1.00
AT1G07940.1	1.6	elongation factor 1-alpha / EF-1-alpha	not plastid	_	1.00
AT5G60390.1	1.6	elongation factor 1-alpha / EF-1-alpha - dual location cytosol & Nucleus	cytosol; nucleus	_	1.00
AT3G26650.1	40.3	glyceraldehyde 3-phosphate dehydrogenase A-1 (GAPA-1)	plastid stroma	C	1.00
AT5G35360.1	286	biotin carboxylase (BC) - part of ACCase complex	envelope-inner-peripheral-stromal-side	C	0.99
AT3G25860.1	230.1	E2 - dihydrolipoamide acetyltransferase, plastid - subunit of pyruvate dehydrogenase complex	plastid	C	0.96
AT1G03090.2	15	3-methylcrotonyl-CoA carboxylase (MCCA)	mitochondria	M	0.88
AT4G21280.1	81.6	psbQ OEC16 Tat ltp	thylakoid-peripheral-lumenal-side	C	0.87
AT5G35630.1	13	glutamate-ammonia ligase (GS2), chloroplast	plastid stroma	C	0.86
AT3G12780.1	31	phosphoglycerate kinase-1 (PGK-1)	plastid stroma	C	0.82
AT1G31330.1	20	psaF- subunit III - LTP - hydrophobic	thylakoid-integral	C	0.82
AT5G17920.1	20	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase (ATCIMS) - abundant	cytosol	_	0.82
AT5G02500.1	37	HSP70-1 (HSC70-1) (not plastid) - very abundant - dual localized nucleus&cytosol	cytosol; nucleus	_	0.76
AT5G38430.1	5.6	Rubisco small subunit 1b (RBCS-1b)	plastid stroma	C	0.70
AT5G64040.1	15	psaN - TAT LTP	thylakoid-peripheral-lumenal-side	C	0.67
AT3G14420.1	10	glycolate oxidase-1 (GOX-1) - peroxisome & plastid	peroxisome; plastid	_	0.67
AT5G04140.1	5	ferredoxin-dependent glutamate synthase/glu1/Fd-GOGAT 1 (dual cTP and mTP)	plastid stroma	C	0.67
AT2G34590.1	10.5	E1 beta pyruvate dehydrogenase complex	plastid	C	0.62
AT2G01210.1	61	leucine-rich repeat transmembrane protein kinase, putative	plasma membrane	S	0.56
AT5G08050.1	14	unknown protein	thylakoid	C	0.56
AT5G38410.1	15.2	Rubisco small subunit 3b (RBCS-3B)	plastid stroma	C	0.55

AT5G38420.1	15.2	Rubisco small subunit 2b (RBCS-2b)	plastid stroma	C	0.55
AT4G26910.1	5.8	dihydrolipoamide succinyltransferase - part of OGDC	mitochondria	M	0.53
AT4G04640.1	9	CF1y - atpC	thylakoid-peripheral-stromal-side	C	0.50
AT5G26000.1	6	thioglucoside glucohydrolase 1 (TGG1) (myrosinase)	not plastid	S	0.50
ATCG00780.1	6	50S ribosomal protein L14	plastid ribosome	_	0.50
AT4G03280.1	3	petC - Rieske Fe-S protein	thylakoid-peripheral-lumenal-side	C	0.50
ATCG00130.1	13	CFO-I - atpF	thylakoid-integral	_	0.44
AT1G34430.1	110.9	Plastidial Dihydrolipoamide Acetyltransferase, pyruvate DH complex	envelope-inner-peripheral-stromal-side	C	0.43
ATCG00640.1	7	50S ribosomal protein L33	plastid ribosome	_	0.40
AT1G67090.1	31.9	Rubisco small subunit-4 (RBCS-4)	plastid stroma	C	0.39
AT4G01150.1	8	unknown protein	thylakoid-integral	C	0.33
AT3G45140.1	11	lipoxygenase AtLOX2, plastid	plastid stroma	C	0.22
ATCG00270.1	6	psbD D2	thylakoid-integral	_	0.20
AT5G01530.1	7	LHCII-4.1-CP29	thylakoid-integral	C	0.17
AT5G54770.1	9	THI1 -involved in thiamine synthesis (vitamine B) (ARA6) - dual localized mitos & plastid	mitochondria; plastid stroma	C	0.13
AT3G55410.1	139.3	2-oxoglutarate dehydrogenase, E1 subunit	mitochondria	M	0.07
AT5G55070.1	34.2	2-oxoglutarate dehydrogenase E2 subunit - part of OGDC	mitochondria	M	0.06

Materials and methods

Generation of transgenic plants

The RIP1 coding sequence was amplified from cDNA clone U67651 obtained from ABRC, followed by TA TOPO cloning into PCR8/GW/TOPO vector (Invitrogen). The coding sequence was then shuttled into the PAUL15 destination vector by LR Clonase II (Invitrogen) to create a RIP1-strepIII expression cassette driven by a HCF136 promoter. A RIP1-3xFlag-strepII expression construct was made using a modified PBI121-Gateway-3Flag-strepII vector. The native promoter of RIP1 was cloned from *A. thaliana* genomic DNA using primer pair RIP1_promoter_F(AAG CTT TTT CAA ACA ATG AAA GTA TGA GAG TGG C) and RIP1_promoter_R (ACT AGT GGA AGC TCT AGA TTG GGC TTC G). The hygromycin resistance cassette was cloned from the PH7RWG2.0 vector using primer pair HYR_F(GTTT

AAAC ATT ATC AGC TTG CAT GCC GGT) and HYG_R (GGG CCC ATC ATA CAT GAG AAT TAA GGG AGT CAC G). The RIP1 promoter fragment was digested with HindIII and SpeI and then ligated into the HindIII and XbaI sites of PBI121-RIP1-3Flag-strepII to replace the 35S promoter. The plant selective marker in this construct was then replaced by the hygromycin fragment, which was digested and then ligated into the PmeI and ApaI sites. PAUL15-RIP1-strepIII and Native promoter-RIP1-3Flag-strepII were then transformed into Agrobacterium tumefaciens GV3101 by electroporation. Standard floral dipping was performed to transform *rip1+/-* flowers. The transformants were selected on MS plates containing 25mg/L hygromycin before being transferred into soil.

Co-immunoprecipitation

Total leaf protein was extracted by grinding ~10g leaves from transgenic plants or wild-type plants grown in the same conditions using extraction buffer (100mM Tris-HCl pH8.0, 150mM NaCl, 0.5mM EDTA, 0.2% NP-40). The protein extract was cleared by centrifugation at 13,000rpm 20min followed by passage through 0.45μm filter (Corning). 600μl Strep-Tactin resin suspension (IBA, Germany) was packed into a gravity column. 30mg total protein was used for each immunoprecipitation, which was performed following the manufacture's manual. 100ul column bed anti-Flag agarose resins (Sigma) were used for Flag-tag immunoprecipitation. The elution was done by incubating the resins with Flag-elution buffer (2M $MgCl_2$, 50mM Tris-HCl

ph8.0, 150mM NaCl, 0.5%CHAPS). Elution was then concentrated using an SDS sample prep kit (Thermo Scientific).

Immunoblotting and silver staining

Protein samples were concentrated using an SDS sample prep kit (Thermo Scientific). 1% of the IP samples were loaded onto an Any kD MINI-PROTEAN TGX precast gel (Bio-rad, Hercules, CA) followed by standard procedures of western blotting. α-Flag-M2-HRP (Sigma) was used to detect Flag-tagged proteins. Strep-tactin-HRP (IBA) was used to detect strepII-tagged proteins. 10% of the IP samples were separated on a 10% polyacrylamide gel before being subjected to silver staining.

References

1. Lyska D, Engelmann K, Meierhoff K, Westhoff P (2013) pAUL: A gateway based vector system for adaptive expression and flexible tagging of proteins in Arabidopsis. *PLoS ONE* 8:e53787.

2. Nikolau BJ, Wurtele ES, Stumpf PK (1985) Use of streptavidin to detect biotin containing proteins in plants. *Anal Biochem* 149:448–453.

3. Pfalz J, Liere K, Kandlbinder A, Dietz K-J, Oelmüller R (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell Online* 18:176–197.