Molecular Analysis of Sequence Determinanats and Protein Factors Involved in Chloroplast RNA Editing

by Wade Patrick Heller

MOLECULAR ANALYSIS OF SEQUENCE DETERMINANTS AND PROTEIN FACTORS INVOLVED IN CHLOROPLAST 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

Wade Patrick Heller

August 2010

MOLECULAR ANALYSIS OF SEQUENCE DETERMINANTS AND PROTEIN FACTORS INVOLVED IN CHLOROPLAST RNA EDITING

Wade Patrick Heller, Ph. D.

Cornell University 2010

In land plants, RNA editing converts 25-40 cytidine nucleotides within chloroplast transcripts to uridines and approximately 500 such conversions occur in mitochondrial transcripts. Shared or similar mechanisms and machinery are believed to underlie RNA editing in both organelles. In order for a particular C-target to be correctly recognized and edited, two distinct components are required: the *cis*-element, a sequence component contained within the transcript itself, and *trans*-factors, protein components which are recruited to the *cis*-element. Prior studies that utilized editing assays *in planta* and *in vitro* allowed characterization of *cis*-elements for a number of C-targets and revealed that subsets of C-targets are related. Although a high level of sequence identity near related C-targets has not been observed, mapping of a competition element to a 5 nt block, located 20 nt upstream of the maize C-targets rpoB C467 and rps14 C80, was achieved using an *in vitro* assay. This region is likely to be a binding site for an as of yet unidentified common *trans*-factor. A number of *trans*-factors, all members of the pentatricopeptide repeat (PPR) protein family, that affect other C-targets have begun to be identified in recent years. One such protein, RARE1, was identified by a reverse genetic screen of chloroplast PPR proteins. Comparative genomics was utilized to predict orthologous pairs of PPR proteins encoded by the genomes of

Arabidopsis and rice. A subset of Arabidopsis-specific chloroplast PPR proteins were studied as possible editing factors for Arabidopsis-specific RNA C-targets, leading to the identification of RARE1. PPR proteins apparently do not have catalytic editing activity themselves and are postulated to form complexes with another protein containing this activity. To discover additional *trans*-factors participating in RNA editing, immunoprecipitation of RARE1 was performed and the constituents of the co-immunoprecipitate were identified by mass spectrometry. One candidate from this analysis, RIP1, was confirmed to interact with RARE1 by yeast two-hybrid analysis, although RIP1 does not contain any characterized domains and its function is unknown.

BIOGRAPHICAL SKETCH

Wade Patrick Heller was born to Seth and Deborah Heller of Bainbridge, New York in 1983. He attended Bainbridge-Guilford Central High School and graduated in 2001. Afterward, Wade attended Rochester Institute of Technology, and earned a Bachelor of Science degree in Biotechnology with Highest Honors in May 2005. While attending RIT, Wade met his future wife, Lindsay, who moved with him to Ithaca in the fall of 2005 when he began graduate school at Cornell University in the Field of Biochemistry, Molecular and Cell Biology. Wade and Lindsay were married in 2006, and in 2007 their first son, Maddox, arrived. Three years later, another boy, Lucas, was born. Upon receipt of Wade's Ph.D. in August 2010, the family is bound for Hilo, Hawaii, where Wade has accepted a postdoctoral research position with the USDA Agricultural Research Service.

This dissertation is dedicated to my loving wife, Lindsay, and our children who have endured this journey along with me. I hope that my children will someday know the thrill of scientific discovery for themselves, and that they will have spouses as supportive as mine to guide them along their paths.

ACKNOWLEDGMENTS

First, I would like to thank my wife, Lindsay, for standing by my side and supporting me, especially during the time we've been at Cornell. I would also like to thank her, and our children, for keeping me in-touch with reality and providing the necessary life distractions that have enriched the journey. I would be remiss for not thanking my parents and grandparents who have instilled in me the value of education from an early age, and for their support of my career decisions, even when that means we will soon be separated by 5,000 miles of land and sea.

I would like to thank my advisor, Maureen Hanson, for the opportunity to have been a member of her lab for the past five years and for her guidance. I am grateful to have shared in experimental planning and critique with past and present members of the Hanson lab with whom I've interacted, especially Michael Hayes and John Robbins, and for technical assistance by Lin Lin. Lastly, I would also like to thank the National Science Foundation and the National Institutes of Health for providing the funding that has made possible this experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa(s), amino acid(s)

ACCase, acetyl co-A carboxylase

bp(s), base pair(s)

GC/MS, gas chromatography coupled mass spectrometry

Het, heterozygote

HM, homozygous mutant

IPTG, Isopropyl β-D-1-thiogalactopyranoside

nt(s), nucleotide(s)

MES, (*N*-morpholino)ethanesulfonic acid

MW, molecular weight (in kDa)

MS, mass spectrometry

MS/MS, tandem mass spectrometry

PCR, polymerase chain reaction

PDS, phytoene desaturase

PPE, poisoned primer extension

PPR, pentatricopeptide repeat

RT-PCR, reverse transcription polymerase chain reaction

SD, standard deviation

SDS, sodium dodecyl sulfate

T-DNA, transfer DNA (*Agrobacterium*)

UTR, untranslated region

VIGS, virus-induced gene silencing

WT, wild-type

X-Gal, 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

CHAPTER 1

INTRODUCTION

RNA editing is a biological phenomenon exhibited by diverse organisms in which mRNA sequences are altered from their genomically encoded forms. There are two major forms of mRNA editing: insertion/deletion editing and nucleotide conversion editing. Insertion/deletion editing occurs in trypanosome mitochondria, causing non-encoded uridine nucleotides to be inserted or nucleotides to be deleted from transcripts (1). *Physarum polycephalum* exhibits a similar form of insertional editing in which non-templated G nucleotides are introduced into transcripts (2). Nucleotide conversion editing usually is the result of deamination reactions and causes pre-existing cytidine or adenosine nucleotdides within transcripts to be changed to uridine or inosine nucleotides, respectively. In animals, adenosine to inosine RNA editing alters signaling components of the nervous system (3), and cytidine to uridine (C-to-U) editing of the *apoB* mRNA of mammals is necessary for production of two alternate protein forms with different stability properties (4). The type of editing RNA editing discussed in this dissertation concerns C-to-U editing of transcripts encoded by the chloroplast and mitochondrial genomes of land plants.

In a typical land plant, around 30 particular Cs in chloroplast transcripts are edited to Us, and over 500 such C-targets are present in transcripts of the mitochondrial genome. Despite the disparity in number of C-targets, the mechanism of RNA editing is believed to be common to both organelles, involving shared or highly similar nucleus-encoded protein components. This

work solely focuses on RNA editing of chloroplast transcripts, but implications of the findings are likely applicable to mitochondrial editing as well.

The RNA editing apparatus is divided into *cis*-elements, the local sequence and structure of the mRNA transcript itself sufficient to specify a given C-target, and *trans*-factors, protein components which bind *cis*-elements and carry out the editing activity. Studies both *in vivo* and *in vitro* have mapped *cis*-elements of various C-targets to between 20 nt upstream of C-targets and a few nt downstream of the C-target (5-8). Involvement of protein *trans*-factors in RNA editing was first indirectly shown by UV-crosslinking experiments in which proteins of particular molecular weights in chloroplast extracts bound putative *cis*-element sequences within synthetic RNAs (9,10). Analysis of the barley *albostrians* and maize *iojap* mutants have shown that *trans*-factors of editing are nuclear gene products, as these mutants lack chloroplast ribosomes and therefore are unable to translate plastome-encoded proteins, yet retain the ability to carry out RNA editing of plastid transcripts (11,12). To date, almost all experimentally identified *trans*-factors belong to the pentatricopeptide repeat (PPR) protein family, which are believed to act as sequence recognition factors binding the *cis*-elements of a number of C-targets, a concept that is later discussed in more detail.

Historically, the first discoveries of chloroplast RNA editing involved two C-targets within translation initiation codons, maize rbcL C2 and tobacco psbL C2, editing of which creates ATG codons from genomically encoded ACG codons (13,14). Another example of start codon formation as a result of RNA editing is the case of Arabidopsis ndhD C2; however, editing of all C-targets does not cause such dramatic changes. Nevertheless, almost all known chloroplast C-targets of editing are within the coding regions of transcripts, and

the consequence of an editing event is usually a change in an amino acid residue of the encoded protein. In Arabidopsis editing of 32 out of the 34 known C-targets causes a codon to specify a different amino acid, while the function of editing the remaining two C-targets, one in the 3' UTR of *accD* and the other in the first intron of *rps12*, is unclear. In general, the outcome of amino acid codon changes incurred by editing is the restoration of evolutionarily conserved residues in the encoded proteins.

Two substantial breakthroughs in characterization of the chloroplast RNA editing machinery both involved editing of synthetic templates. The first was analysis *in planta* of transplastomic plants containing transgenes encoding RNA editing substrates, first described in (6). The other was analysis of RNA editing *in vitro* by incubation of editing substrates prepared by *in vitro* transcription with extracts of purified chloroplasts, first shown in (9). Analysis *in planta* requires the time-consuming and expensive generation of transgenic plants, whereas the *in vitro* editing assay allows many different synthetic editing substrates, often mimicking 'mutant' *cis*-elements, to be tested within a relatively short time. Chapter 2 of this dissertation is an extensive *in vitro* characterization of *cis*-elements for a subset of related RNA editing C-targets whose relationship was uncovered by *in planta* analysis. In this case, transplastomic tobacco overexpressing a region of the maize *rpoB* transcript containing the ZMrpoB C467[1] C-target had been found to exhibit reduced editing of the orthologous NTrpoB C473 C-target, as well as at two additional C-targets, rps14 C80 and psbL C2 (5). This was the first evidence that *cis*-elements compete for *trans*-factors *in vivo*, and alignment of the putative *cis*-elements of these three C-targets revealed three small (3-4 nt)

[1]Editing C-target nomenclature: (ZM)rpoB C467, C nt 467 of the (*Zea mays*) *rpoB* transcript.

blocks of sequence homology that were observed if small gaps were introduced into one or more of the sequences. The experiments in Chapter 2 identified a single 5 nt sequence predicted to be the binding site of a common *trans*-factor mediating cross-competition in editing of the C-targets in *rpoB* and *rps14*.

A longstanding model for the existence of RNA editing postulates that C-targets of editing and the machinery of editing have co-evolved (15). The catalytic component is believed to have been derived from a pre-existing activity, and the ability of this catalytic component to 'correct' particular defects in transcripts derived from mutated loci is believed to have allowed such mutations to become fixed over time. As components of the RNA editing complex evolved to recognize different positions of transcripts, T-to-C mutations arising at these positions were tolerated, and over evolutionary time became fixed; similarly, C-to-T mutations at positions previously affected by RNA editing obviates the need for the editing machinery to recognize these positions and allows the machinery to diverge. Thus, as editing C-targets have been gained or lost by particular lineages throughout evolution, components of the editing complex recognizing them have likely also become divergent between different plant species. For example, Arabidopsis and rice share only eight common chloroplast C-targets, whereas Arabidopsis has 26 non-shared C-targets and rice has 13 non-shared C-targets (16,17). It expected that *trans*-factors affecting the non-shared C-targets are likewise divergent between the two species.

A major breakthrough in the field of RNA editing came with the identification of the first *trans*-factor, CRR4 (18). CRR4 was identified by a genetic screen of mutagenized Arabidopsis seedlings with a *chlororespiratory*

*r*eduction (*crr*) phenotype indicative of a defect in the NAD(P)H dehydrogenase complex in plastids. The molecular defect in *crr4* is the inability to edit the ndhD C2 C-target, resulting in the absence of the *ndhD* start codon in the transcript, and the mutation causing the defect maps to a locus encoding a PPR protein. Additionally, CRR4 was shown to bind the *cis*-element region of its target (19,20), and this binding data is the basis for the idea that PPR proteins are site-specificity factors of RNA editing. Subsequently, other PPR protein editing factors have been identified by forward genetic analyses, in which the molecular cause of defective chloroplast function or biogenesis was found to be an editing defect. Factors identified in this manner include CLB19, CRR21, CRR22, CRR28, and LPA66 (19,21-23). However, these identifications have been dependent on phenotypes that are both detectable and viable. As it became evident that PPR proteins are likely to be involved in all chloroplast (and, more recently, mitochondrial) RNA editing, reverse genetic screens became favorable as these do not rely on visible phenotypes and can utilize gene silencing in cases which a null genetic mutation causes plant or embryo lethality.

PPR proteins represent a family of over 450 members, and this family is highly expanded in plants relative to non-photosynthetic eukaryotes (24,25). In addition to RNA editing, PPR proteins have been found to be involved in a variety of regulation events within chloroplasts, including polycistron cleavage, splicing, transcript stabilization and translation (26-29). PPR proteins are made up of tandem repeats of a degenerate 35 aa motif. The P-L-S class of PPR proteins has PPR motifs of slightly variable length, with the addition of an E (extended) domain; some members additionally carry the DYW domain, named for its conserved terminal tripeptide. With one exception, all PPR

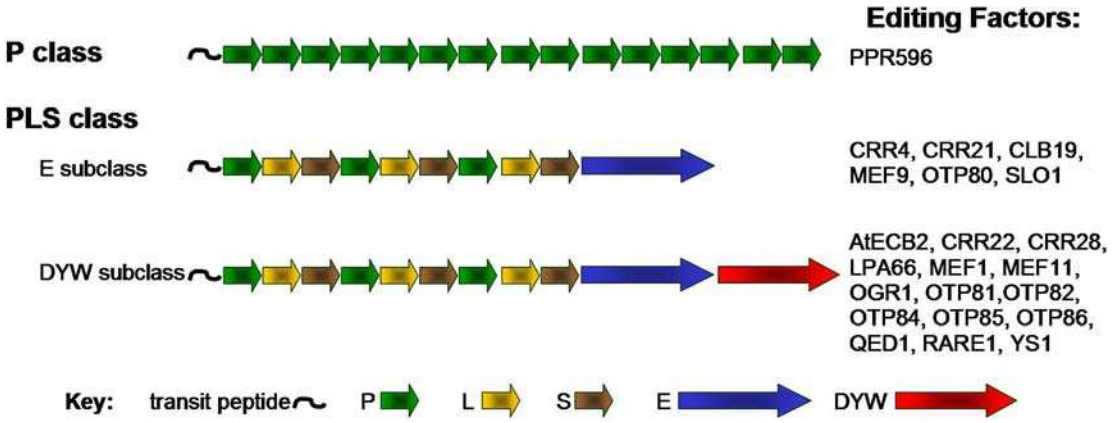


Figure 1.1. Diagrams showing motif structure of various PPR protein classes. Known RNA editing *trans*-factors of each class are listed at right; data derived from (18,19,21-23,30-38) except for QED1 (John Robbins, unpublished data). Represented motifs/domains include: P, PPR motif; L, long (generally 36 aa) PPR motif; S, short (generally 31 aa) PPR motif; E, extended domain; DYW, DYW domain. Figure modified from (25).

proteins identified as a chloroplast RNA editing factors are of the P-L-S class, and 60% (15 of 25) of these known to be involved in chloroplast or mitochondrial editing carry the DYW domain (Figure 1.1). Recently, a P-class PPR protein, PPR596, was reported to affect mitochondrial editing of one C-target (39); however, unlike the knockout of editing of particular C-targets observed in mutants of the P-L-S editing factors, *ppr596* mutants exhibit an increase in editing efficiency of a C-target. A more detailed discussion of PPR protein domains can be found in Chapter 4.

A reverse genetic screen was used to identify RARE1, a PPR-DYW editing factor essential for editing of the *accD* transcript in Arabidopsis, and the experiments leading to its identification comprise Chapter 3 of this dissertation. The screen took advantage of (1) the observation that RNA editing C-targets are divergent between distantly related species, in this case Arabidopsis and rice, and (2) a prediction that *trans*-factors required for editing of the non-conserved C-targets would likewise be divergent. Targeting prediction for

Arabidopsis genes encoding PPR proteins was performed, and of those strongly predicted to be plastid localized, only 8 were identified as Arabidopsis genes not having rice orthologs by reciprocal best hit analysis. One of the candidates, RARE1, was determined to be essential for RNA editing of the accD C794 C-target, and this was the first identification of a *trans*-factor based on a reverse genetics screen.

Aside from PPR proteins, the only other class of proteins implicated in RNA editing are the chloroplast ribonucleoproteins. These proteins are highly abundant and involved in the stabilization of chloroplast transcripts, including those not known to be affected by RNA editing. Originally, it was shown that immunodepletion of one of these proteins, CP31, reduced editing activity in an *in vitro* assay (9). Subsequently, Arabidopsis *cp31a* and *cp31b* mutants were shown to have reduced editing of partially overlapping sets of C-targets, but a total loss of editing was not observed for any C-target, even in double mutants (40). It is therefore difficult to determine whether CP31A and CP31B participate directly in RNA editing, such as in a complex with PPR proteins, or indirectly affect editing by causing altered transcript levels. For example, increasing or decreasing transcript abundance alters the ratio of editing substrates to editing factors, and assuming such factors are present in limited quantities, the fraction of substrates edited would change correspondingly.

Although the DYW domain of PPR proteins has been implicated as possibly containing the catalytic activity of editing based on phylogenetic and bioinformatic analyses (41), the idea was negated by experiments showing that editing defects in *crr22* and *crr28* mutants could be restored by introduction of transgenes encoding CRR22 or CRR28 lacking their respective DYW domains (23). As catalytic activity has not been ascribed to PPR

proteins themselves, it is instead proposed that the PPR proteins associate with a separate catalytic factor within RNA editing complexes (Figure 1.2). Support for the existence of RNA editing complexes comes from estimation of the native molecular weight of protein complexes containing the PPR editing factors CRR4 and RARE1. In both cases, the PPR protein-containing complexes are estimated to be approximately 200 kDa (Charles Bullerwell, unpublished data, and Figure 4.8D), which is substantially larger than the mass of the PPR proteins themselves. However, reverse genetic screens of deaminases have failed to identify the critical enzymatic component of RNA editing to date.

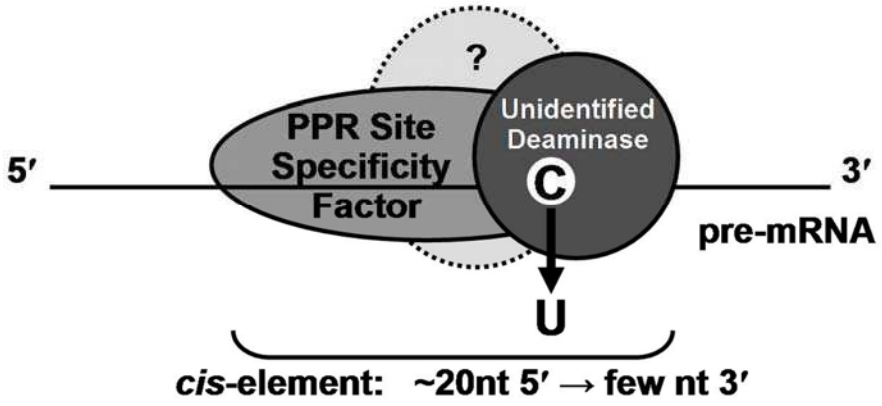


Figure 1.2. Model for plant organelle RNA editing. A PPR protein acting as a site specificity factor is recruited to a pre-mRNA containing an editing C-target by the *cis*-element sequence. The PPR protein is thought to associate with a cytidine deaminase, which catalyzes the editing reaction. The identity of the catalytic subunit is unknown and additional unknown constituents (represented by '?') may also be needed. Figure modified from (20).

Although the protein-protein interaction partners of editing factor PPR proteins remain completely unknown, complex immunoprecipitation and subsequent identification of constituents by mass spectrometry is one obvious route to identify additional *trans*-factors of editing. Of course, proteins with cytidine deaminase activity, and possibly chloroplast ribonucleoproteins, are likely

candidates that might arise from such an analysis. Chapter 4 of this dissertation describes the experimental identification of proteins interacting with the RARE1 editing factor. Transgenic plants expressing affinity-tagged RARE1 were made in order to perform the above described experiment. Although no putative deaminase has yet been identified from this data, the identification of a RARE1-interacting protein, RIP1 (RARE1-INTERACTING PROTEIN 1), has been confirmed by yeast two-hybrid analysis. This protein belongs to the 8-member DIFFERENTIATION AND GREENING (DAG) protein family, one member of which, DAL1, has been shown to be essential for processing of the chloroplast *rrn* operon (42,43). Targeting of proteins of this family to chloroplasts or mitochondria is predicted, and it is therefore tempting to imagine they play a role in RNA editing within both organelles. No functional domains have been identified within these proteins and their molecular function is completely unknown. Analysis of RNA editing defects in *rip1* mutants and *RIP1*-silenced plants have shown partial editing defects of several C-targets, strengthening the argument that RIP1 plays a role in RNA editing. Further analyses will be needed to determine the role that RIP1 and possibly other DAG-family proteins are performing in RNA editing complexes.

REFERENCES

1. Hajduk, S.L., Harris, M.E. and Pollard, V.W. (1993) RNA editing in kinetoplastid mitochondria. *FASEB J*, **7**, 54-63.

2. Cattaneo, R. (1991) Different types of messenger RNA editing. *Annu Rev Genet*, **25**, 71-88.

3. Stapleton, M., Carlson, J.W. and Celniker, S.E. (2006) RNA editing in Drosophila melanogaster: New targets and functional consequences. *RNA*, **12**, 1922-1932.

4. Chen, S.H., Habib, G., Yang, C.Y., Gu, Z.W., Lee, B.R., Weng, S.A., Silberman, S.R., Cai, S.J., Deslypere, J.P., Rosseneu, M. *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.

5. Chateigner-Boutin, A.L. and Hanson, M.R. (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol Cell Biol*, **22**, 8448-8456.

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

7. Hayes, M.L. and Hanson, M.R. (2007) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA*, **13**, 281-288.

8. Reed, M.L., Peeters, N.M. and Hanson, M.R. (2001) A single alteration 20 nt 5' to an editing target inhibits chloroplast RNA editing in vivo. *Nucleic Acids Res*, **29**, 1507-1513.

9. Hirose, T. and 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.

10. Miyamoto, T., Obokata, J. and 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.

11. Halter, C.P., Peeters, N.M. and Hanson, M.R. (2004) RNA editing in ribosome-less plastids of iojap maize. *Curr Genet*, **45**, 331-337.

12. Zeltz, P., Hess, W.R., Neckermann, K., Borner, T. and 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.

13. Hoch, B., Maier, R.M., Appel, K., Igloi, G.L. and Kossel, H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, **353**, 178-180.

14. Kudla, J., Igloi, G.L., Metzlaff, M., Hagemann, R. and Kossel, H. (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable psbL mRNA by a C to U substitution within the initiation codon. *EMBO J*, **11**, 1099-1103.

15. Covello, P.S. and Gray, M.W. (1993) On the evolution of RNA editing. *Trends Genet*, **9**, 265-268.

16. Chateigner-Boutin, A.L. and 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.

17. Tsudzuki, T., Wakasugi, T. and Sugiura, M. (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol*, **53**, 327-332.

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

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

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

21. Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C. and Zhang, L. (2009) LPA66 is required for editing psbF chloroplast transcripts in Arabidopsis. *Plant Physiol*, **150**, 1260-1271.

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

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

24. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

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

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

27. Nakamura, T., Meierhoff, K., Westhoff, P. and Schuster, G. (2003) RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. *Eur J Biochem*, **270**, 4070-4081.

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

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

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

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

32. Robbins, J.C., Heller, W.P. and Hanson, M.R. (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.

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

34. 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.

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

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

37. Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I. and Huang, J. (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*.

38. Yu, Q.B., Jiang, Y., Chong, K. and 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.

39. Doniwa, Y., Ueda, M., Ueta, M., Wada, A., Kadowaki, K. and Tsutsumi, N. (2010) The involvement of a PPR protein of the P subfamily in partial RNA editing of an Arabidopsis mitochondrial transcript. *Gene*, **454**, 39-46.

40. Tillich, M., Hardel, S.L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J.M., Lehwark, P., Leister, D., Small, I.D. and Schmitz-Linneweber, C. (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A*, **106**, 6002-6007.

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

42. Babiychuk, E., Fuangthong, M., Van Montagu, M., Inze, D. and Kushnir, S. (1997) Efficient gene tagging in Arabidopsis thaliana using a gene trap approach. *Proc Natl Acad Sci U S A*, **94**, 12722-12727.

43. Bisanz, C., Begot, L., Carol, P., Perez, P., Bligny, M., Pesey, H., Gallois, J.L., Lerbs-Mache, S. and Mache, R. (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.

CHAPTER 2

CROSS-COMPETITION IN EDITING OF CHLOROPLAST RNA TRANSCRIPTS *IN VITRO* IMPLICATES SHARING OF TRANS-FACTORS BETWEEN DIFFERENT C TARGETS[2]

Wade P. Heller, Michael L. Hayes and Maureen R. Hanson

ABSTRACT

C-to-U plant organellar RNA editing is required for the translation of evolutionarily conserved and functional proteins. 28 different C targets of RNA editing have been identified in maize chloroplasts, and hundreds of Cs are edited in mitochondria. Mutant analysis in Arabidopsis has indicated that absence of a single site-specific recognition protein can result in loss of editing of a single C target, raising the possibility that each C target requires a recognition protein. Here we show that transcripts encompassing two editing sites, ZMrpoB C467 and ZMrps14 C80 can compete editing activity from each other *in vitro*, despite limited sequence similarity. The signal causing competition overlaps a 5′ *cis*-element required for editing efficiency. A single five-nucleotide mutation spanning the region from -20 to -16 relative to the edited C of rpoB C467 is sufficient to eliminate its substrate editing as well as its ability to compete editing activity from rps14 C80 substrates. A corresponding mutation in an rps14 C80 competitor likewise eliminated its

[2]Originally published in J Biol Chem. 2008 Jan;283(12):7314-9
WPH contributed to experimental design, conducted all experiments, and contributed all figures. MLH contributed to experimental design. WPH and MRH wrote the manuscript text.

ability to compete editing activity from rpoB C467 substrates. Taken together, our results indicate that the RNA sequences mediating both editing efficiency and cross-competition are highly similar and that a common protein is involved in their editing. Sharing of trans-factors can facilitate editing of the large number of different C targets in plant organelles so that a different protein factor would not be required for every editing site.

INTRODUCTION

Post-transcriptional modification of plant organellar mRNAs by RNA editing is required for maintenance of functional protein sequences (1-3), and also for the introduction of translation initiation codons in particular transcripts (4,5). Typically, chloroplast genomes of higher land plants have on the order of 30 to 40 editing sites, while mitochondria generally have greater than 400 (6-12). To date, 28 cytidine-to-uridine editing sites have been identified in 15 chloroplast transcripts in maize (12,13), all of which alter the encoded amino acid, except one site in the 5′UTR of *ndhG*.

It is currently believed that the plant organellar RNA editing machinery consists of two distinct components: the *cis*-element which uniquely identifies a given editing site by its sequence and structure within the transcript itself, and the *trans*-acting factors, which are likely to be proteins that recognize the *cis*-element and catalyze the editing reaction (14). The sequences surrounding all editing sites in a given organism do not show obvious similarity to each other either by direct sequence aligment or secondary structure prediction. However, transplastomic tobacco that overexpress a fragment of maize *rpoB* or tobacco *ndhF* transcripts spanning the rpoB C467 or ndhF

C290 editing sites, respectively, showed reduced editing at the cognate tobacco sites, as well as at additional sites (15), indicating that at least some *cis*-elements are related. These three editing sites therefore form a “cluster” affected by overexpression of transcripts carrying only one C target (Figure 2.1). Furthermore, three 2-3 nt regions of sequence identity exist between the sites of the rpoB C473 cluster within 20 nt 5′ of the edited C, when gaps are introduced in the sequences (15).

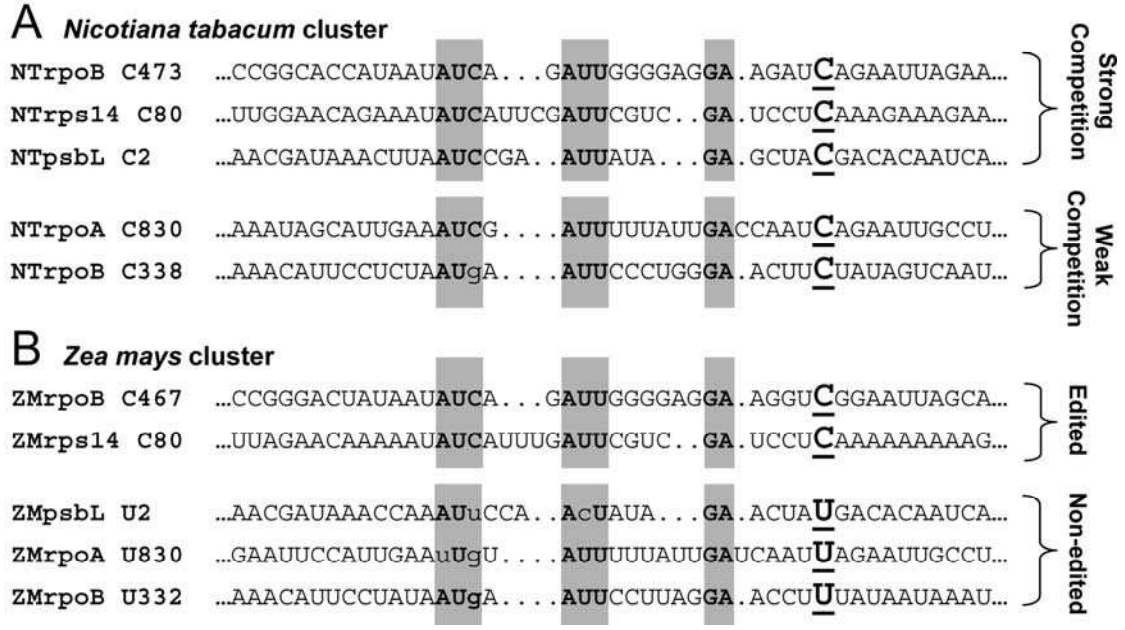


Figure 2.1. Sequence identity in members of the NTrpoB C473 editing site cluster. A, Alignment of sequences from tobacco editing sites affected by overexpression of ZMrpoB C467 transgene, as reported in (15). Shaded boxes indicate regions of sequence identity when the alignment allows gaps B, Orthologous maize editing sites aligned using the tobacco model.

Two nuclear-encoded protein factors have been identified that are believed to be responsible for sequence recognition of editing sites in the *ndhD* transcript of Arabidopsis: CRR4 which is critical for ATndhD C2 editing, and CRR21 which is required for ATndhD C383 editing (16,17). Both of these proteins are members of the pentatricopeptide (PPR) class of proteins, which

consists of >450 members in Arabidopsis that are largely targeted to chloroplasts and/or mitochondria (18). PPR motif-containing proteins have been implicated in additional organellar RNA processing or maturation events as well as RNA editing (19-21). As *crr4* or *crr21* knockouts specifically affect editing at one C target only, they are believed to be sequence recognition factors for the sites they affect. While CRR4 and CRR21 evidently do not affect editing of multiple C targets, *in vivo* competition data regarding other editing sites as well as our data from *in vitro* analysis presented here suggests that trans-factors required for editing of multiple C targets will be described in the future. Possibly genes encoding such trans-factors rarely emerge in mutant screens because their loss of function would often have lethal consequences.

An *in vitro* editing assay has been developed to study *cis*-elements near C targets of editing. Editing of RNA substrates, transcribed *in vitro*, occurs in extracts prepared from isolated chloroplasts (22,23). A major benefit of this strategy for studying the editing machinery is that many mutant substrates can be studied for editing efficiency at the same time, under controlled conditions. Alternatively, *cis*-elements can be studied *in vivo* in transgenic plants; however, this technique is very limited in terms of the number of substrates that can be tested by the amount of time and expense required to generate such transgenic plants, as well as variability due to transformation and regeneration. The *in vitro* strategy has previously been used in our laboratory to identify *cis*-elements of the tobacco rpoB C473 and psbE C214 editing sites (23-25).

We have previously studied the tobacco rpoB C473 site extensively both *in vivo* and *in vitro*. A transgene containing the sequence from 20 nt 5′ to

6 nt 3′ (-20/+6) of the NTrpoB C473 site is sufficient for tobacco editing *in vivo* (26), and a -31/+22 transgene is edited more efficiently *in vivo* (24). Furthermore, a synthetic substrate containing the -31/+22 region of rpoB C473 was shown to be sufficient for *in vitro* editing in both tobacco and maize extracts (24). Here we report our findings on the relationship of the *cis*-element of maize rpoB C467 to that of its cluster member, rps14 C80. The existence of a five-member rpoB C473 cluster has been functionally proven in tobacco using an *in vivo* approach (15); however, in the putative orthologous maize cluster, editing occurs only at rpoB C467 and rps14 C80, because the remaining three members of the maize cluster have a genomically encoded U at the position of the C target in tobacco (Figure 2.1). We have found that synthetic maize rpoB C467 and rps14 C80 RNAs can both be edited by maize chloroplast extracts and both are capable of reducing the editing extent of transcripts carrying either the rpoB C467 or rps14 C80 sites. We have taken advantage of our competition assay to localize the *cis*-elements of each editing site that are responsible for the competition effect. Previously, sequences responsible for cross-competition *in vivo* could not be studied by *in vitro* editing assays because substrates representing two or more cluster members could not be edited *in vitro*. We observe that the cis-elements causing the competition effect co-localize with those that determine editing efficiency, indicating that the same trans-factor is likely to mediate both responses.

MATERIALS AND METHODS

Synthesis of editing substrates in vitro. DNA templates for RNA

Table 2.1. Oligonucleotides used in the experiments (IDT, Coralville, IA)

Name	Sequence 5′ to 3′	Purpose
SKRpoB54_F	CGCTCTAGAACTAGTGGATCGGGACTATAATAT CAGATTG	PCR ZMrpoB54WT
KSRpoB54_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATT	PCR ZMrpoB54WT
T7SK	TAATACGACTCACTATAGGGCGCTCTAGAACTA GTGGATC	template construction
SK	CGCTCTAGAACTAGTGGATC	RT-PCR
KS	TCGAGGTCGACGGTATC	RT-PCR
RpoB54m1_F	CTCTAGAACTAGTGGATCCCCTGAATAATATCA GATTGGGGAGG	Pcr rpoB m1
RpoB54m2_F	CTCTAGAACTAGTGGATCGGGACTTATTAATCA GATTGGGGAGGAAGGT	Pcr rpoB m2
RpoB54m3_F	CTCTAGAACTAGTGGATCGGGACTATAATTAGT CATTGGGGAGGAAGGTCGG	Pcr rpoB m3
RpoB54m4_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GTAACCGGAGGAAGGTCGGAAT	Pcr rpoB m4
RpoB54m4_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATTCCGACCTTCCTCCGGTT	Pcr rpoB m4
RpoB54m5_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATTGGCCTCCAAGGTCGGAAT	Pcr rpoB m5
RpoB54m5_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATTCCGACCTTGGAGGCCAA	Pcr rpoB m5
RpoB54m6_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATTGGGGAGGTTCCACGGAAT	Pcr rpoB m6
RpoB54m6_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATTCCGTGGAACCTCCCCAA	Pcr rpoB m6
RpoB54m7_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATTGGGGAGGAAGGTCCCTTA	Pcr rpoB m7
RpoB54m7_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA TAAGGGACCTTCCTCCCCAA	Pcr rpoB m7
RpoB54m8_R	TCGAGGTCGACGGTATCTTTTTTATCAATACGAT ATTCCGACCTTCCTCCCC	Pcr rpoB m8
RpoB54m9_R	TCGAGGTCGACGGTATCTTTTTTTAGTTATGCTA ATTCCGACCTTCCTC	Pcr rpoB m9
RpoB54m10_R	TCGAGGTCGACGGTATCAAAAAAATCAATTGCT AATTCCGACC	Pcr rpoB m10
RpoB54-20m_F	CTCTAGAACTAGTGGATCGGGACTATAATTTCA GATTGGG	Pcr rpoB -20m
RpoB54-19m_F	CTCTAGAACTAGTGGATCGGGACTATAATAACA GATTGGGG	Pcr rpoB -19m
RpoB54-18m_F	CTCTAGAACTAGTGGATCGGGACTATAATATGA GATTGGGGA	Pcr rpoB -18m
RpoB54-17m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCT GATTGGGGAG	Pcr rpoB -17m
RpoB54-16m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA CATTGGGGAGG	Pcr rpoB -16m
RpoB54-15m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GTTTGGGGAGGA	Pcr rpoB -15m
RpoB54-14m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GAATGGGGAGGAA	Pcr rpoB -14m

Table 2.1 (Continued)

RpoB54-13m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATAGGGGAGGAAG	Pcr rpoB -13m
RpoB54-12m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATTCGGGAGGAAGG	Pcr rpoB -12m
RpoB54-11m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATTGCGGAGGAAGGT	Pcr rpoB -11m
RpoB54-10m_F	CTCTAGAACTAGTGGATCGGGACTATAATATCA GATTGGCGAGGAAGGTC	Pcr rpoB -10m
RpoB54-9m_F	GGGACTATAATATCAGATTGGGCAGGAAGGTCG GAATTAGCAAT	Pcr rpoB -9m
RpoB54-8m_F	GGGACTATAATATCAGATTGGGGTGGAAGGTCG GAATTAGCAAT	Pcr rpoB -8m
RpoB54-7m_F	GGGACTATAATATCAGATTGGGGACGAAGGTCG GAATTAGCAAT	Pcr rpoB -7m
RpoB54-6m_F	GGGACTATAATATCAGATTGGGGAGCAAGGTCG GAATTAGCAAT	Pcr rpoB -6m
RpoB54-5m_F	GGGACTATAATATCAGATTGGGGAGGTAGGTCG GAATTAGCAAT	Pcr rpoB -5m
RpoB54-4m_F	GGGACTATAATATCAGATTGGGGAGGATGGTCG GAATTAGCAAT	Pcr rpoB -4m
RpoB54-3m_F	GGGACTATAATATCAGATTGGGGAGGATCGTCG GAATTAGCAAT	Pcr rpoB -3m
RpoB54-2m_F	GGGACTATAATATCAGATTGGGGAGGATGCTCG GAATTAGCAAT	Pcr rpoB -2m
RpoB54-1m_F	GGGACTATAATATCAGATTGGGGAGGATGGAC GGAATTAGCAAT	Pcr rpoB -1m
RpoB54+1m_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATTCGGACCTTCCTC	Pcr rpoB +1m
RpoB54+2m_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATTGCGACCTTCCT	Pcr rpoB +2m
RpoB54+3m_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA ATACCGACCTTCC	Pcr rpoB +3m
RpoB54+4m_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA AATCCGACCTTC	Pcr rpoB +4m
RpoB54+5m_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTA TTTCCGACCTT	Pcr rpoB +5m
RpoB54+6m_R	TCGAGGTCGACGGTATCTTTTTTATCAATTGCTT ATTCCGACCT	Pcr rpoB +6m
Rps14-100_F	CGCTCTAGAACTAGTGGATCTAATAATAGGAAC GGCACAT	Pcr rps14 -100nt
Rps14-24_F	CGCTCTAGAACTAGTGGATCAATATCATTTGATT CGTCGATCC	Pcr rps14 -24nt
Rps14-20_F	CGCTCTAGAACTAGTGGATCTCATTTGATTCGT CGATCCT	Pcr rps14 -20nt
Rps14-16_F	CGCTCTAGAACTAGTGGATCTTGATTCGTCGAT CCTCAAA	Pcr rps14 -16nt
Rps14-12_F	CGCTCTAGAACTAGTGGATCTTCGTCGATCCTC AAA	Pcr rps14 -12nt
Rps14+5_R	TCGAGGTCGACGGTATCTTTTTGAGGATCGACG AATCAAATG	Pcr rps14 +5nt
Rps14+10_R	TCGAGGTCGACGGTATCCTTTTTTTTTGAGGAT CGACGAATCAAATG	Pcr rps14 +10nt

Table 2.1 (Continued)

Rps14+30_R	TCGAGGTCGACGGTATCGGGGAAACTTTGCTTCTTAT	Pcr rps14 30nt, PPE rps14 5′mutants
Rps14-20m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATCAAATGTTATTTTTGTTCT	Pcr rps14 -20m
Rps14-19m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATCAAATCATATTTTTGTTCT	Pcr rps14 -19m
Rps14-18m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATCAAAAGATATTTTTGTTCT	Pcr rps14 -18m
Rps14-17m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATCAATTGATATTTTTGTTCT	Pcr rps14 -17m
Rps14-16m_R	GGAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATCATATGATATTTTTGTTC	Pcr rps14 -16m
Rps14-15m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATCTAATGATATTTTTGTTC	Pcr rps14 -15m
Rps14-14m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAATGAAATGATATTTTTG	Pcr rps14 -14m
Rps14-13m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGAAACAAATGATATTTTTG	Pcr rps14 -13m
Rps14-12m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGATTCAAATGATATTTTTG	Pcr rps14 -12m
Rps14-11m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACGTATCAAATGATATTTTTG	Pcr rps14 -11m
Rps14-10m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGACCAATCAAATGATATTTTTG	Pcr rps14 -10m
Rps14-9m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGAGGAATCAAATGATATTTTTG	Pcr rps14 -9m
Rps14-8m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCGTCGAATCAAATGATATTTTTG	Pcr rps14 -8m
Rps14-7m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATCCACGAATCAAATGATATTTTTG	Pcr rps14 -7m
Rps14-6m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATGGACGAATCAAATGATATTTTTG	Pcr rps14 -6m
Rps14-5m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGAACGACGAATCAAATGATATTTTTG	Pcr rps14 -5m
Rps14-4m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGTTCGACGAATCAAATGATATTTTTG	Pcr rps14 -4m
Rps14-3m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGCATCGACGAATCAAATGATATTTTTG	Pcr rps14 -3m
Rps14-2m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGACGATCGACGAATCAAATGATATTTTTG	Pcr rps14 -2m
Rps14-1m_R	GAAACTTTGCTTCTTATCTTTTTTTTTGTGGATCGACGAATCAAATGATATTTTTG	Pcr rps14 -1m
Rps14+1m_R	GAAACTTTGCTTCTTATCTTTTTTTTAGAGGATCGACGAATCAAATGATATTTTTG	Pcr rps14 +1m
Rps14+2m_R	GAAACTTTGCTTCTTATCTTTTTTTATGAGGATCGACGAATCAAATGATATTTTTG	Pcr rps14 +2m
Rps14+3m_R	GAAACTTTGCTTCTTATCTTTTTTATTGAGGATCGACGAATCAAATGATATTTTTG	Pcr rps14 +3m
Rps14+4m_R	GAAACTTTGCTTCTTATCTTTTTATTTGAGGATCGACGAATCAAATGATATTTTTG	Pcr rps14 +4m

Table 2.1 (Continued)

Rps14+5m_R	GAAACTTTGCTTCTTATCTTTTATTTTGAGGATC GACGAATCAAATGATATTTTTG	Pcr rps14 +5m
Rps14+6m_R	GAAACTTTGCTTCTTATCTTTATTTTTGAGGATC GACGAATCAAATGATATTTTTG	Pcr rps14 +6m
Rps14m3′_R	GAAACTTTGCTTCTTATCTTTTTTTTTGAGGATC GACGAATCATTACTTATTTTTGTTC	PCR rps14 m20-′
ZMpsbL 165_C_F	CTCTAGAACTAGTGGATCCCCCCAATGACCAT AGATCG	PCR ZMpsbL ′C2′ 165
ZMpsbL 165_C_R	TCGAGGTCGACGGTATCGGGTTTGATTGTGTC GTAGTTCTATA	PCR ZMpsbL ′C2′ 165
ZMpsbL_-16m_R	TAGTTCTATAGTTGGGATTTGGTTTATCG	PCR ZMpsbL ′C2′ -16m
ZMpsbL_-11m_R	TAGTTCTATAATTGGAATTTGGTTTATCG	PCR ZMpsbL ′C2′ -11m
ZMpsbL-16/-11m_R	TAGTTCTATAATTGGGATTTGGTTTATCG	PCR ZMpsbL ′C2′ -16/-11m
ZMpsbL165_T_R	TCGAGGTCGACGGTATCGGGTTTGATTGTGTC ATAGT	PCR ZMpsbL T2 165
NTpsbL 165_F	CGCTCTAGAACTAGTGGATCTGCCCTAATGAC TATAGATC	PCR NTpsbL C2 165
NTpsbL 165_R	TCGAGGTCGACGGTATCGGGTTTGATTGTGTC GTAGCTCT	PCR NTpsbL C2 165
NTpsbLm18-_R	TTTGATTGTGTCGTAGCTCTATAATTGCCTATA AGTTTATCG	PCR NTpsbL -18to-14m
T7KS(-)RpoB_F	TAATACGACTCACTATAGGGGATACCGTCGAC CTCGAGGGACTATAATATCAGATTG	PCR rpoB54 competition
SK(-)RpoB_R	GATCCACTAGTTCTAGAGCGTTTTTTATCAATT GCTAATTCCG	PCR rpoB54 competition
T7KS(-)Rps14_F	TAATACGACTCACTATAGGGGATACGGTCGAC CTCGATAATAATAGGAACGGCACAT	PCR rps14 competition
SK(-)Rps14_R	GATCCACTAGTTCTAGAGCGGGGGAAACTTTG CTTCTTAT	PCR rps14 competition
KS(-)	GATACCGTCGACCTCGA	RT-PCR
SK(-)	GATCCACTAGTTCTAGAGCG	RT-PCR
RpoB_PPE_C	ATAATATCAGATTGGGGAGG	PPE rpoB 3′mutants
RpoB_PPE_G	TCGAGGTCGACGGTATCTTTTTTATCAATTG	PPE rpoB 5′mutants
Rps14_PPE_C	ATCAATTGATTCGTCGATCC	PPE rps14 3′mutants

substrates were made by PCR amplification (Taq MasterMix Kit, Qiagen, Valencia, CA) from maize genomic DNA using gene specific primers (IDT, Coralville, IA) containing overhanging bacterial fragments SK and KS (Table 2.1). The T7 promoter sequence was then added by a subsequent PCR step to the 5′ end of the templates. Mutant templates were made by incorporation of mismatches in primers used for PCR. RNA substrates were then

transcribed *in vitro* from their DNA templates and purified as described previously (25).

Preparation of editing competent maize extracts. Maize chloroplast extracts were prepared from leaves of 7 to 10 day-old plants as previously described (24). Leaf tissue was homogenized and intact chloroplasts were isolated by gradient sedimentation using Percoll (Amersham Biosciences). Buffers and conditions for chloroplast isolation, extraction, and dialysis as previously described (23). Extracts contained 2-4 μg/μl protein.

In vitro editing reaction. Editing reactions for substrate editing without competitor RNAs were as previously described (23), using 0.1 fmol of RNA and 4 μl of extract. For competition experiments, competitor RNAs were added to the editing reaction mixture prior to the addition of 10fmol of RNA substrate. Following incubation to allow editing, 1 μl of the editing reaction mixture was used for cDNA synthesis and subsequent PCR amplification as in (23). All substrates, with the exception of the two competition substrates, used the KS primer for RT and SK and KS primers for PCR. For competition experiments, RT of the substrate RNA used the SK(-s) primer, and the subsequent PCR used KS(-s) and SK(-s) primers. The pTri competitor is a 128-nt control transcript containing a fragment of conserved human 18S rRNA sequence unrelated to any known editing site, and is transcribed from a template included with the T7MEGAshortscript kit (Ambion).

Poisoned primer extension. To determine the editing efficiency in a given reaction, poisoned primer extension was performed as previously described (23). Different oligonucleotides were used for extension of substrates from each site, depending on the presence of mutations either 5′ or 3′ of the edited C, and are listed in Table 2.1.

RESULTS

A 54-nt ZMrpoB C467 substrate, containing 31 nt upstream and 22 nt downstream of the target C, and flanked by SK and KS sequences on the 5′ and 3′ ends, respectively, was found to edit 60±4% in replicate experiments. Substrates were made with blocks of 5 or 6 nts at a time mutated to the complementary nucleotides along the length of the sequence (Figure 2.2A), in order to evaluate the significance of each sequence block for C467 editing (Figure 2.2B). Six of the mutant substrates had reduced editing efficiency relative to the wild-type substrate, and three of these six were virtually unedited using standard assay conditions. The three five-nt blocks that

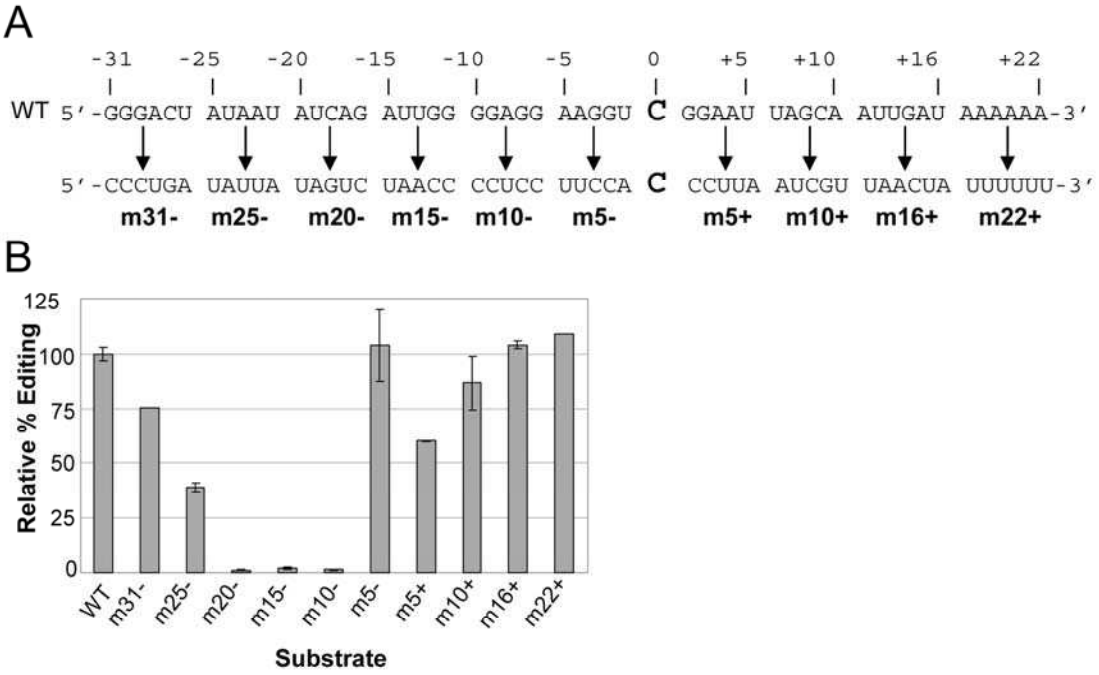


Figure 2.2. Editing efficiency *in vitro* of ZMrpoB C467 substrates containing multiple mutations. A, Each substrate consisted of WT sequence (upper line), with the exception of the 5-6 nts indicated in the lower line. Substrates also contained SK and KS sequences on the 5′ and 3′ ends, respectively, for RT-PCR amplification. B, Relative editing efficiencies of the 10 substrates indicated in (A), as compared to a WT substrate, which was 63% edited under the reaction conditions. Error bars represent 1 SD from the mean in replicate samples.

contain these critical sequences were m20-, m15-, and m10-, which collectively span the region from -20 to -6 (Figure 2.2).

A 131-nt ZMrps14 C80 substrate (100 nt upstream and 30 nt downstream) was edited *in vitro* to 28±5%. Substrates containing 5′ and 3′ truncations of the 131-nt substrate were assayed for editing efficiency, in order to determine the minimal sequence required for editing to proceed (Figure 2.3). Truncation substrates containing at least 20-nt 5′ and at least five-nts 3′ retained editing efficiency, but substrates with less than 20-nt 5′ had a marked decrease in editing efficiency.

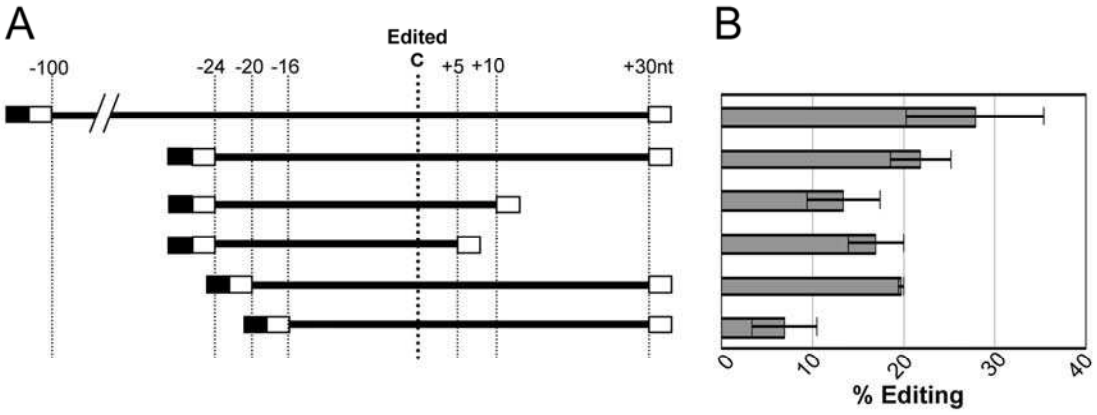


Figure 2.3. Editing efficiency *in vitro* of rps14 C80 substrates with varying amounts of sequence around the edited C. A, Substrate lengths tested. Open boxes indicate sequences used for universal amplification of substrates during RT-PCR, SK on 5′ ends of substrates, and KS on 3′ ends. Black boxes refer to the T7 promoter sequence for *in vitro* transcription. B, Editing efficiency of each substrate shown in (A). Error bars represent 1 SD from the mean in replicate samples.

As determined above, the critical *cis*-elements directing editing of both rpoB C467 and rps14 C80 lie within 20-nt 5′ of the target C, and some sequence further 5′ or on the 3′ side of the target may also be involved in recognition. To further explore the sequence requirements for editing of each substrate, we made single nucleotide mutations at each position within the -20/+6 windows for each site by altering the wild-type sequence to the

complementary nucleotide, and each was assayed for editability (Figure 2.4). The relative editing efficiency for each substrate was calculated. For rpoB C467, 35% of the mutated substrates had reduced editing efficiency, 35% had enhanced editing efficiency, and 30% had no effect. Of the rps14 C80 mutated substrates, 58% had reduced editing efficiency, 8% had enhanced editing efficiency, and 35% had no effect.

The raw (unaligned) comparison between the two critical windows reveals 6 common positions that negatively affect editing when mutated, and only 3 of these are the same nucleotide in both sequences. Using gaps to align the rpoB C467 and rps14 C80 editing sites, as reported by (15), reveals 6 positions in common that negatively affect editing when mutated, and 5 of these 6 are the same nucleotide in both sequences (Figure 2.4C).
To further investigate *cis*-elements affecting editing and the similarity between the rpoB C467 and rps14 C80 sites, we performed *in vitro* competition experiments. First we established that 100-fold self-competitor was sufficient to reduce editing of rpoB C467 substrate to virtually undetectable levels, while inclusion of pTri 18S RNA, the 128-nt control transcript of unrelated sequence, did not reduce editing of C467. Cross-competition was observed; a 100-fold amount of rpoB C467 substrate reduced rps14 C80 substrate editing to a similarly low level as self-competitor, and rps14 C80 reciprocally reduced rpoB C467 editing.

Figure 2.5A shows the results of self-competition experiments using rpoB C467 substrate and rpoB C467 wild-type and multiple-nt mutated competitors. Three 5-nt mutated competitors had reduced competition relative to wild-type, and in the case of competition with the m20- sequence, editing efficiency of the substrate was >80% relative to the no-competitor and pTri

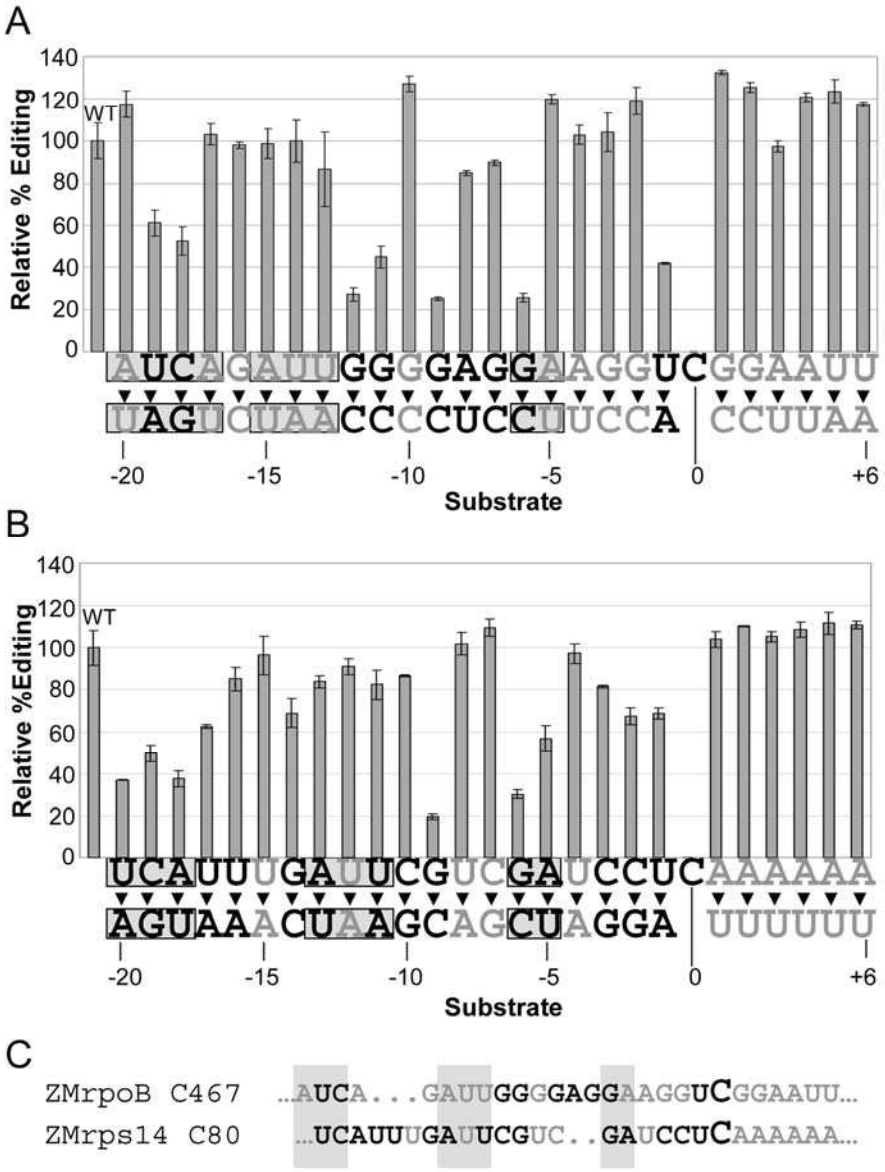


Figure 2.4. Effect of single nucleotide mutations within the critical editing windows of ZMrpoB C467 and ZMrps14 C80 on *in vitro* editing efficiency. A, Relative editing efficiencies of rpoB C467 54-nt substrate containing mutations to the complementary base at the specified positions, as compared to WT substrate, which was edited 60-74% in replicate experiments. Shaded boxes indicate regions of sequence identity between the two sites and critical nucleotides (positions which, when mutated, cause a reduction of editing efficiency below WT level) are shown in bold. Error bars represent 1 SD from the mean. B, Same as (A), except for the rps14 C80 131-nt substrate. Actual editing efficiency of WT substrate was 47-61% under the reaction conditions. C, Alignment as in Figure 2.1, with critical single nucleotides from (A) and (B) shown in bold; sequence identity in the cluster is marked by shaded boxes.

controls. In the case of rpoB cross-competition of rps14 C80 substrate editing, four of the 5-nt mutated competitors exhibited reduced competition relative to wild-type competitors (Figure 2.5B). Three of these also showed reduced competition of the rpoB C467 substrate (Figure 2.5A). The same 5-nt mutation had the largest effect on competition, and completely eliminated the ability of the rpoB C467 competitor to reduce editing of rps14 C80. Furthermore, when the -20 to -16 region of the rps14 C80 competitor was mutated, cross-competition was likewise eliminated (Figure 2.6).

In tobacco, overexpression of the maize rpoB C467 site *in vivo* reduces editing of psbL C2 as well as rps14 C80. The nucleotide at the editing site in the maize *psbL* transcript is a genomically encoded U, and thus drift could

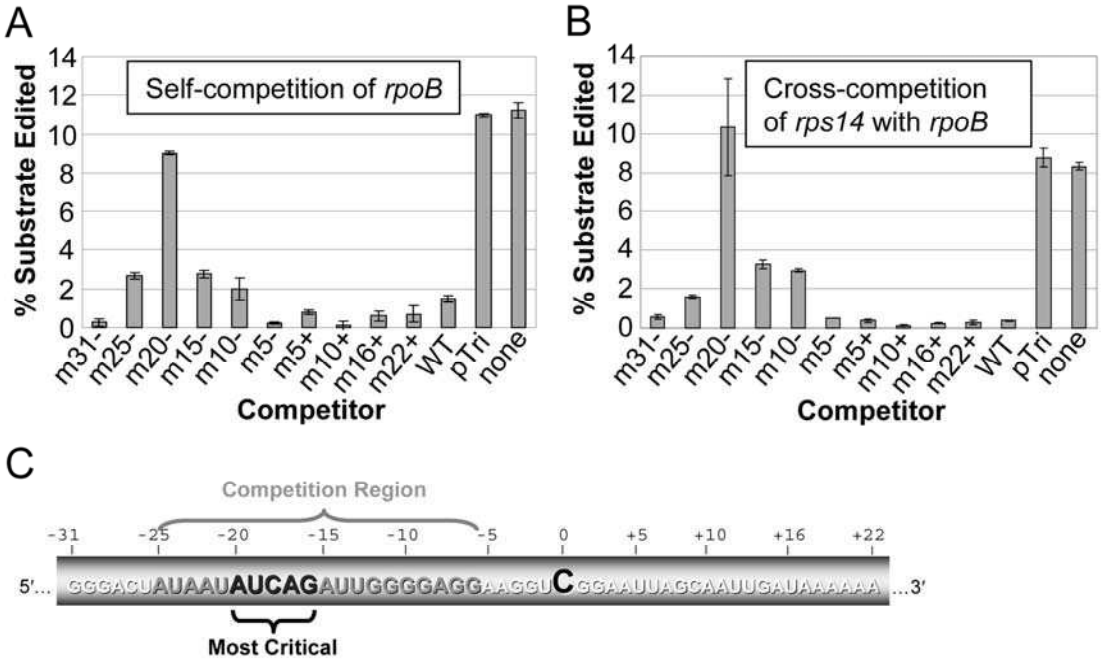


Figure 2.5. Effect of ZMrpoB substrate mutations on *in vitro* self- and cross-competition A, Actual percent edited of WT ZMrpoB C467 54-nt substrate when competitors were added to *in vitro* editing reactions at a ratio of 100:1. Substrates m31- to m22+ refer to those indicated in Figure 2.2. Error bars represent 1 SD from the mean from replicate reactions. B, Same as (A) except testing *rpoB* competitors for cross-competition with the *rps14* 131-nt substrate. C, Model of rpoB C467 competition region.

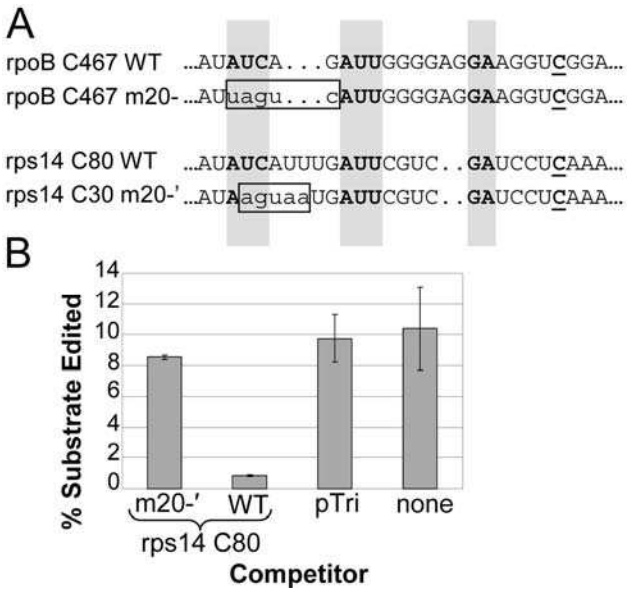


Figure 2.6. Effect of *rps14* competitor mutations on *in vitro* cross-competition with rpoB C467 substrate. A, Sequence alignments of WT and mutant rpoB C467 and rps14 C80 competitor RNAs. Boxed regions show the 5 nts altered in the mutated competitors, and shaded positions represent sequence identity from Figure 2.1. B, Editing efficiency of rpoB C467 substrate under competition from the RNAs indicated.

have occurred in a cis-element that might have been present in a progenitor of maize in which psbL C2 was edited. However, we noted that some of the nucleotides found to be critical for editing of maize rpoB C467 and rps14 C80 are present at comparable positions in maize upstream of psbL U2. We therefore wondered whether a maize *psbL* RNA might reduce editing of rpoB C467 and rps14 C80. Indeed, we found that maize *psbL* RNA was able to reduce editing of rpoB C467 and rsp14 C80 substrates if a higher competitor-to-substrate ratio was used. Whereas rpoB C467 and rps14 C80 could compete editing activity at 100-fold competitor-to-substrate, maize psbL U2 competitor or an artificial psbL ′C2′ competitor did not cause an appreciable decrease in substrate editing efficiency unless a 1000-fold competitor-to-substrate ratio was used (Figure 2.7).

Comparison of the *psbL* transcripts from maize and tobacco reveals

several sequence differences upstream of the editing site, in addition to the C/U editing site difference itself (Figure 2.7A). At the 1000-fold competitor-to-substrate ratio needed to observe competition by *psbL* substrates, a

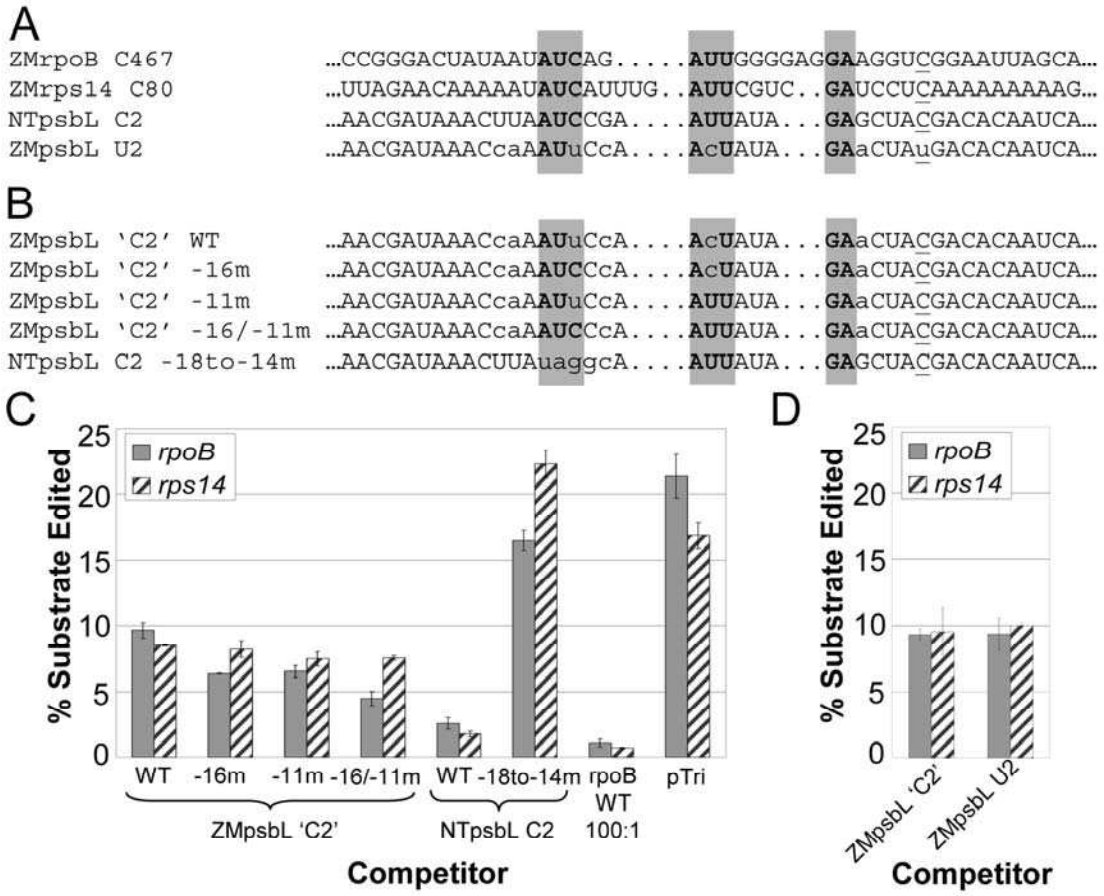


Figure 2.7. Effect of psbL ′C2′ competitors on editing efficiency of rpoB C467 and rps14 C80 substrates. A, rpoB C467 editing site cluster as previously described (15), showing nucleotide differences between the tobacco and maize sequences. Conserved elements shown in shaded boxes, divergence of the maize gene from the tobacco gene is shown by the use of lowercase letters. B, psbL ′C2′ RNAs used for *in vitro* competition experiments, including conserved element restoration mutants. U2 in the WT sequence was mutated to a C in all competitor RNAs to correspond to the tobacco psbL C2 editing site. C, Shaded bars indicate editing extent of the rpoB C467 substrate, and dashed bars indicate editing extent of rps14 C80 substrate. Ratio of competitor to substrate was 1000 to 1, except in the *rpoB* 100:1 column, in which the ratio was 100:1. D, Competition as in (C), except using ZMpsbL ′C2′ and ZMpsbL U2 competitor RNAs.

competitor containing the tobacco sequence was able to reduce editing of rpoB C467 and rps14 C80 substrates more efficiently than one containing the maize sequence (Figure 2.7C). Two of the differences between the tobacco and maize *psbL* sequences, at the -16 and -11 positions of the maize transcript, relative to the editing site, are changes in the conserved sequence elements identified in the tobacco editing site cluster. Restoration of these conserved positions in maize psbL ′C2′ competitors (see Figure 2.7B) enhanced rpoB C467 competition slightly, but not to the level of competition exerted by the tobacco psbL C2 competitor (Figure 2.7C). The improvement in competition of by the mutated maize competitors was higher for rpoB C467 substrate than for rps14 C80 substrate. Mutating the tobacco *psbL* competitor in the region corresponding to the rpoB m20- mutant competitor likewise eliminated its ability to compete editing activity from either rpoB C467 or rps14 C80 substrates (Figure 2.7C).

Competitors carrying the genomically-encoded U at the editing site of maize *psbL* RNAs did not compete differently than those with a C at this position. When these maize psbL U2 and psbL ′C2′ competitors were added to reactions containing rpoB C467 substrate, the editing efficiencies were reduced to 9%. There also was no significant difference between the psbL U2 and psbL ′C2′ competitors on rps14 C80 substrate editing; with either competitor, the editing efficiencies were reduced to 10%.

DISCUSSION

Self-competition of editing of endogenous *psbL* transcript with *psbL* transgene transcripts was first observed *in vivo* by Chaudhuri et al. (27) and

provided the initial indirect evidence for the existence of site-specific recognition factors for chloroplast RNA editing. Subsequently, when Hirose and Sugiura (28) developed editing-competent tobacco chloroplast extracts, they observed that oligoribonucleotides carrying sequences 5′ to psbL C2 or a *ndhB* C target of editing were effective self-competitors for editing, but that no cross-competition occurred when either *psbL* or *ndhB* RNAs were added to radiolabeled oligoribonucleotide substrates specified by the other gene. Likewise, Miyamoto et al. (22) showed that *psbE* and *petB* editing substrates will undergo self-competition but not cross-competition *in vitro*. Mutations in either *psbE* or *petB* RNAs between -5 and -1 upstream of the C target of editing did not affect self-competition extent (29). These prior reports utilized a thin-layer chromatography separation of radiolabeled C and U to assay editing and competition extent, which is not easily quantified. We have used a sensitive and precise poisoned primer extension assay (25) in order to quantify the effect of mutation of substrates and competitors on the extent of editing in maize chloroplast extracts. We show that swapping the SK/KS flanking sequences of substrate and competitor RNAs is a convenient method to assay self-competition *in vitro*.

Our report is the first to demonstrate cross-competition between transcripts of two different editing sites *in vitro*. The cross-competition of maize rpoB C467 and rps14 C80 substrates *in vitro* is consistent with our prior finding of cross-competition of the orthologous editing sites *in vivo* by over-expression of *rpoB* transgene transcripts in tobacco transplastomic plants (15).

We have demonstrated that *psbL* competitors carrying either a C or U at the location of the edited nucleotide are equally effective in reducing editing of rpoB C467 and rps14 C 80 in maize extracts *in vitro*. The effectiveness of

either C- or U-containing competitors was not unexpected, given that *ndhF* transgene transcripts carrying either C or U at the editing site were effective competitors of endogenous *ndhF* transcript editing in transplastomic tobacco plants *in vivo* (30). Evidently both unedited and edited transcripts can be recognized by trans-factors.

Cross-competition between editing sites *in vivo* and *in vitro* reveals similarities in *cis*-elements between sequences surrounding different C targets of editing. The simplest explanation for cross-competition would be the existence of a single trans-factor that recognizes these similar editing sites. This hypothesis is an attractive explanation for the capability of the plant to recognize hundreds of different Cs with a high degree of specificity. Imperfect selectivity of a trans-factor in recognition of C targets could explain how a new T-to-C mutation could be corrected at the transcript level in the plant in which it first arises, leading to evolutionary improvement of editing efficiency at new targets through modification of trans-factors to recognize multiple targets. Alternatively, there could be different factors, perhaps evolutionarily related, that recognize each editing site, perhaps with different efficiencies. Transcripts carrying one editing site could possibly bind both factors, resulting in reduced editing extent of two different C targets when the competitor transcript is in great excess.

A study of an Arabidopsis editing mutant has shown that the lack of a single trans-factor, CRR4, can prevent editing of the ATndhD C2 target (16) and that CRR4 specifically binds to an RNA fragment containing ndhD C2 (17). In tobacco, editing of ndhD C2 is affected by overexpression of an *ndhF* transgene, and sequences 5′ to both C targets exhibit some similarity, suggesting that the two sites may share the same or related trans-factors.

The Arabidopsis ndhF C290 and ndhD C2 editing sites also exhibit 5′ sequence similarity, but no effect on ndhF C290 editing in the ndhD C2 editing-deficient mutant was detected (16). Several hypotheses can be created to explain these apparently contradictory findings. First, possibly an ortholog of CRR4 in tobacco can bind to both ndhD C2 and to the ndhF C290 site and thus ndhF C290 editing would be reduced in the presence of excess *ndhF* transcript. Second, there may be a trans-factor that is shared between ndhD C2 and ndhF C290 editing sites that is not an ortholog of CRR4, but is a factor that is remains to be discovered. If such a factor is in limiting quantities and is bound by *ndhF* transcript, editing of ndhD C2 could be reduced even if a CRR4 ortholog only binds to *ndhD* transcript. Future identification of all of the components of editing complexes that act on C targets with related 5′ sequence should reveal the nature of the relationships between the editing complexes responsible for converting different plant organelle Cs to Us.

REFERENCES

1. Bock, R., Kossel, H. and Maliga, P. (1994) Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *EMBO J*, **13**, 4623-4628.

2. Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H. and Nagano, Y. (2001) Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem*, **276**, 3937-3940.

3. Zito, F., Kuras, R., Choquet, Y., Kossel, H. and Wollman, F.A. (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.

4. Hoch, B., Maier, R.M., Appel, K., Igloi, G.L. and Kossel, H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, **353**, 178-180.

5. Kudla, J., Igloi, G.L., Metzlaff, M., Hagemann, R. and Kossel, H. (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable psbL mRNA by a C to U substitution within the initiation codon. *EMBO J*, **11**, 1099-1103.

6. Corneille, S., Lutz, K. and Maliga, P. (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol Gen Genet*, **264**, 419-424.

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

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

9. Inada, M., Sasaki, T., Yukawa, M., Tsudzuki, T. and Sugiura, M. (2004) A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant Cell Physiol*, **45**, 1615-1622.

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

11. Tillich, M., Funk, H.T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M. and Maier, R.M. (2005) Editing of plastid RNA in Arabidopsis thaliana ecotypes. *Plant J*, **43**, 708-715.

12. Tsudzuki, T., Wakasugi, T. and Sugiura, M. (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol*, **53**, 327-332.

13. Tillich, M., Schmitz-Linneweber, C., Herrmann, R., and R. Maier. (2001) The plastid chromosome of maize (Zea mays): Update of the complete sequence and transcript editing sites. *Maize Gen. Coop. Newsl*, **75**, 42-44.

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

15. Chateigner-Boutin, A.L. and Hanson, M.R. (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol Cell Biol*, **22**, 8448-8456.

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

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

18. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

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

20. Hashimoto, M., Endo, T., Peltier, G., Tasaka, M. and Shikanai, T. (2003) A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. *Plant J*, **36**, 541-549.

21. Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N. and Schuster, G. (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat

protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell*, **15**, 1480-1495.

22. Miyamoto, T., Obokata, J. and 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.

23. Hegeman, C.E., Hayes, M.L. and Hanson, M.R. (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in Arabidopsis in vitro. *Plant J*, **42**, 124-132.

24. Hayes, M.L., Reed, M.L., Hegeman, C.E. and Hanson, M.R. (2006) Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript in vivo and in vitro. *Nucleic Acids Res*, **34**, 3742-3754.

25. Hayes, M.L. and Hanson, M.R. (2007) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA*, **13**, 281-288.

26. Reed, M.L., Peeters, N.M. and Hanson, M.R. (2001) A single alteration 20 nt 5' to an editing target inhibits chloroplast RNA editing in vivo. *Nucleic Acids Res*, **29**, 1507-1513.

27. Chaudhuri, S., Carrer, H. and Maliga, P. (1995) Site-specific factor involved in the editing of the psbL mRNA in tobacco plastids. *EMBO J*, **14**, 2951-2957.

28. Hirose, T. and 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.

29. Miyamoto, T., Obokata, J. and Sugiura, M. (2004) A site-specific factor interacts directly with its cognate RNA editing site in chloroplast transcripts. *Proc Natl Acad Sci U S A*, **101**, 48-52.

30. Reed, M.L., Lyi, S.M. and Hanson, M.R. (2001) Edited transcripts compete with unedited mRNAs for trans-acting editing factors in higher plant chloroplasts. *Gene*, **272**, 165-171.

CHAPTER 3

A COMPARATIVE GENOMICS APPROACH IDENTIFIES A PPR-DYW PROTEIN THAT IS ESSENTIAL FOR C-TO-U EDITING OF THE ARABIDOPSIS CHLOROPLAST *accD* TRANSCRIPT[3]

John C. Robbins[4], Wade P. Heller[4] and Maureen R Hanson

ABSTRACT

Several nuclear-encoded proteins containing pentatricopeptide repeat (PPR) motifs have previously been identified to be trans-factors essential for particular chloroplast RNA editing events through analysis of mutants affected in chloroplast biogenesis or function. Other PPR genes are known to encode proteins involved in other aspects of organelle RNA metabolism. A function has not been assigned to most members of the large plant PPR gene family. Arabidopsis and rice each contain over 400 PPR genes, of which about a fifth exhibit a C-terminal DYW domain. We describe here a comparative genomics approach that will facilitate identification of the role of RNA-binding proteins in organelle RNA metabolism. We have implemented this strategy to identify an Arabidopsis nuclear-encoded gene *RARE1* that is required for editing of the chloroplast *accD* transcript. RARE1 carries 15 PPR motifs, an E/E+ and a DYW domain, whereas previously reported editing factors CRR4, CRR21, and CLB19 lack a DYW domain. The *accD* gene encodes the β Carboxyl-

[3]Modified from: RNA. 2009 Jun;15(6):1142-53.
[4]These authors contributed equally to this work. Specifically, JCR contributed to experimental design, and Figures 3.8, Supplemental Figures 3.S2-4. WPH conducted all experiments and contributed all other figures. All authors contributed to manuscript text.

transferase subunit of acetyl coA carboxylase, which catalyzes the first step in fatty acid biosynthesis in chloroplasts. Despite a lack of accD C794 editing and lack of restoration of an evolutionarily conserved leucine residue in the β carboxyltransferase protein, *rare1* mutants are unexpectedly robust and reproduce under growth room conditions. Previously the serine-to-leucine alteration encoded by editing was deemed essential in the light of the finding that a recombinantly expressed "unedited" form of the pea acetyl coA carboxylase was catalytically inactive.

INTRODUCTION

Vascular plant organelle transcripts undergo C-to-U RNA editing (reviewed in [1-4]). In Arabidopsis, 34 editing events are known to occur in chloroplast transcripts (4,5), while 508 Cs are known to be modified to Us in Arabidopsis mitochondria (6). The amino acid encoded by edited transcripts often differs from the one predicted from unedited transcripts, usually resulting in increased evolutionary conservation of the amino acid sequence from the one predicted from genomic sequence (7), although start and stop codons in organelle transcripts are also sometimes created by C-to-U editing (8,9). The residues modified by RNA editing are often important for the three-dimensional structure of the protein (10). RNA editing appears to be a mechanism to correct defective organelle genes at the transcript level.

The suite of particular C-to-U editing events varies from one plant species to another, even though RNA editing probably arose in an ancestor common to the land plants (11). Between divergent species, such as between dicots and monocots, RNA editing C targets vary considerably. In species that

do not contain a particular C-target of editing, a T is almost always genomically encoded, resulting in the equivalent of a C-to-U edited transcript. Of the 34 known Arabidopsis chloroplast editing events, 26 do not occur in rice (5,12) (Table 3.1). Between two grasses, rice and maize, only 8 differences in the C targets of editing were reported (13). Within 3 Solanaceous species, there are 11 C targets that are not conserved among tomato, tobacco and deadly nightshade (14). No differences in the editing targets in the plastid (4) or mitochondrion (6) has been detected in ecotypes of *Arabidopsis thaliana*, although variation in the efficiency of editing of particular mitochondrial Cs has been observed (15,16).

The cis-elements required for RNA editing of particular Cs in chloroplasts and mitochondria are typically within about 30 nt 5' and 10 nt 3' of the C target. Sequences surrounding editing sites have been analyzed either by introduction of altered cis-elements into transgenic tobacco plastids (17,18), or by assaying RNA variants in chloroplast or mitochondrial extracts competent for editing *in vitro* (19-21). High-level expression of an RNA carrying an *rpoB* editing site or an *ndhF* editing site led to the development of the "cluster hypothesis" of cis-elements and trans-factors (22). Transplastomic plants overexpressing the *rpoB* transcript exhibited reduced editing at 4 other C targets whose sequences immediately 5' exhibited some similarity. Likewise, editing extents of two different sites in plants overexpressing the *ndhF* transcripts were reduced in efficiency. In both cases, there was some sequence similarity immediately 5' to the C target in the *rpoB* or *ndhF* transcripts in the sites whose editing efficiency was reduced (22). This finding can best be explained as the presence of trans-factors that are shared among the *rpoB*-related sites or the *ndhF*-related sites. Similar results have been

Table 3.1. Differences in editing targets between Arabidopsis and rice. Data derived from (5,12). *Actual editing has not been assayed; editing is assumed by orthology to maize (11). [1]Position in 3' UTR from A of ATG. [2]Gene not present in rice. [3]Position in 5' UTR from A of ATG. [4]Position with intron 1, where *i1*1 is first nucleotide after splice site. [NE]Not edited.

	Arabidopsis		Rice	
Gene	Edited Nucleotide	Δ Amino Acid	Edited Nucleotide	Δ Amino Acid
accD	C794	S265→L	n/a[2]	n/a
	C1568[1]	3'UTR	n/a[2]	n/a
atpA	T	L383	C1148	S383→L
atpF	C92	P31→L	T	L31
clpP	C559	H187→Y	T	Y187
matK	C640	H214→Y	T	Y236
	T	Y415	C1351*	H420→Y
ndhA	T	L159	C473	S158→L
	T	F355	C1070	S357→F
ndhB	C149	S50→L	T	L50
	C467	P156→L	C467	P156→L
	C586	H196→Y	C586	H196→Y
	C[NE]	S204	C611	S204→L
	T	F235	C704	S235→F
	T	L246	C737	P246→L
	C746	S249→F	T	F249
	C830	S277→L	C830	S277→L
	C836	S279→L	C836	S279→L
	C872	S291→L	C[NE]	S291
	C1255	H419→Y	T	Y419
	C1481	P494→L	C1481	P494→L
ndhD	C2	T1→M	T	M1
	C383	S128→L	T	L128
	C674	S225→L	T	L225
	C878	S293→L	C878	S293→L
	C887	P296→L	T	L296
ndhF	T	L21	C62	S21→L
	C290	S97→L	T	L97
ndhG	T	5'UTR	C(-10)[3]	5'UTR
	C50	S17→F	T	F17
	T	L116	C347*	P116→L
petL	C5	P2→L	T	L2
psbE	C214	P72→S	T	S72
psbF	C77	S26→F	T	F26
psbZ	C50	S17→L	C[NE]	S17
rpl2	T	M1	C2	T1→M
rpl23	C89	S30→L	T	F30
rpoA	C200	S67→F	T	F67
rpoB	C338	S113→L	T	F111
	T	L158	C467	S156→L
	C551	S184→L	C545	S182→L
	T	L189	C560	S187→L
	C2432	S811→L	T	L811
rpoC1	C488	S163→L	T	L163
rps8	T	L61	C182	S61S→L
rps12	C(*i1* 58)[4]	intron	C(*i1* 54)[NE]	intron
rps14	C80	S27→L	C80	S27→L
	C149	P50→L	T	L47
ycf3	T	M20	C185	T62→M

obtained when RNA editing in chloroplast and mitochondrial extracts has been assayed. Addition of excess amounts of one editing substrate results in reduced editing of additional editing substrates exhibiting weak similarity (20,23).

The involvement of a trans-factor in editing of more than one site suggests how plants may survive the creation of new C targets for editing. If an organelle genome acquires a T-to-C mutation that becomes fixed, the plant may not survive if an encoded amino acid altered by the affected codon is important for protein function. However, if a pre-existing trans-factor can edit a new site sufficiently to produce at least some functional protein, the plant may be able to survive (24). Selection can then occur either on the cis-elements or the trans-factor(s) in order to improve the efficiency of editing. The higher conservation of a 5' element to a potential C target in *psbE* transcripts in land plants than in those carrying a genomically encoded T is consistent with selection for a sequence efficiently recognized by a trans-factor (25).

Trans-factors required for editing of particular C targets in chloroplasts have been identified, and all have been pentatricopeptide repeat (PPR)-containing proteins, members of a large gene family in plants (26). The P class of PPR proteins carries 35 amino-acid repeats, while the PLS family carries "long" 35-36 amino acid repeats as well as "short" 31 amino acid repeats. Members of the PLS class also have E and E+ "extended" domains absent from the P class. Some members of the PLS class additionally carry a DYW domain, named for the characteristic final tripeptide (27,28). The first three reported chloroplast trans-factors are all members of the PLS E/E+ class, and all were identified in Arabidopsis mutants with defects in chloroplast function. The first two known chloroplast editing trans-factors, CRR4 and

CRR21, were found in Arabidopsis mutants with defective NAD(P)H dehydrogenase (29,30), and each is required for editing of a different C target in the *ndhD* transcript. Additionally, an editing factor named CLB19 was identified through analysis of an Arabidopsis chlorophyll-deficient mutant, and is required for editing of C targets in *rpoA* and *clpP* transcripts (31). The membership of these first three editing factors in the PLS E/E+ subclass lacking a DYW domain was surprising, as phylogenetic considerations suggest that the DYW subclass is associated with the development and retention of chloroplast RNA editing (32). However, a PPR-DYW gene has been shown to be required for intergenic processing of a chloroplast transcript (33) and there is evidence for function of DYW domains as endoribonucleases (34). While this paper was under review, two reports appeared that described the identification of three PPR-DYW proteins as chloroplast site-specific editing factors (35,36).

To identify additional editing trans-factors, we have developed an alternative approach to forward genetic screens of chloroplast biogenesis mutants. If an editing site is lost by a C to T mutation in the genome, trans-factors previously recognizing this site will no longer be selectively maintained, allowing for their divergence or loss over evolutionary time. Likewise, if a T to C change occurs in one lineage but not another, a trans-factor is likely to evolve to allow efficient editing in that lineage but will not be needed in other lineages.

To design an efficient reverse genetic screen for RNA editing trans-factors, we examined the Arabidopsis genome sequence to identify putative chloroplast-targeted PPR proteins that lack identifiable orthologs in the complete rice sequence. By narrowing our candidate list to those genes

whose proteins are predicted to be chloroplast targeted by both the TargetP and Predotar algorithms, we identified a list of 8 candidate genes likely to carry out functions in Arabidopsis that are not required in rice chloroplasts. Our analysis of one of these 8 Arabidopsis genes encoding PPR proteins revealed a factor essential for editing of the *accD* transcript, which encodes a subunit of the heteromeric acetyl-CoA carboxylase (ACCase) that is found in plastids of dicots but is not present in the rice and other Gramineae (37).

MATERIALS AND METHODS

Plant Growth. Arabidopsis thaliana ecotype Columbia was grown in 16 hr light/8 hr dark under full-spectrum fluorescent lights in a growth room at 26^{o} C for use in VIGS experiments. For all other experiments, plants were grown in 15 hr light/9 hr dark under metal halide lamps at 26^{o} C.

Virus-induced gene silencing (VIGS). Silencing fragments were amplified by PCR using the gene sequence tag (GST) primers (Table 3.S1) designated by the Complete Arabidopsis Transcriptome Microarray (CATMA) database (38), using Taq PCR Mastermix (Qiagen, Valencia, CA). PCR products were cloned using pCR8/GW/TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA), and plasmid DNA was isolated from spectinomycin-resistant transformants using the PureLink Quick Plasmid Miniprep Kit (Invitrogen). To generate the silencing vector pTRV2GFPGW, we first inserted Gateway Reading Frame A (Gateway Conversion System, Invitrogen) at the *Sma*I site of pYL170 (39), and subsequently inserted a 200bp fragment of the mGFP4 (40) gene between *Sac*I and *Xba*I. The mGFP4 fragment was amplified using the SacI-mGFP4 and XbaI-mGFP4 primers listed in Table 3.S1, cloned into

pCR8/GW/TOPO, cut from vector backbone and gel purified prior to ligation to the pYL170-Gateway derivative. Recombination of the pCR8 entry clones with pTRV2GFPGW using LR Clonase II enzyme mix (Invitrogen) produced the final silencing constructs which were electrotransformed into *E.coli* DHα, then purified plasmid DNA was subcloned into *Agrobacterium tumefaciens* GV3101. *Agrobacterium* strains were maintained on LB agar containing 50μg/ml kanamycin and 100μg/ml rifampicin.

Ecotype Col-0 mGFP5-expressing Arabidopsis kindly provided by Dominique Roberston, NC State University, Raleigh, NC (41) were used for visual screening of GFP expression or silencing under long wave UV irradiation using a BLAK-RAY Model B 100 AP lamp (UVP, Upland, CA). Plant growth conditions and Agro-inoculation of Arabidopsis leaves were as described in (39).

RNA isolation and cDNA synthesis. Two weeks after inoculation, plants were screened for co-silencing of GFP, and individual silenced rosette leaves were harvested and flash-frozen in liquid nitrogen. Total leaf RNA was isolated using Trizol reagent (Invitrogen), and cDNA was synthesized using the Sensiscript Reverse Transcriptase Kit (Qiagen) and random nonamer primers. Fragments of 18 chloroplast transcripts containing all known Arabidopsis editing sites were amplified using the primers listed in Table 3.S1. For semi-quantiative RT-PCR analysis, RNA was primed with random nonamers or At5g13270_RT_R for production of *ACTIN* or *RARE1* cDNA; PCR was performed with Actin_F and Actin_R to generate *ACTIN* product, and At5g13270_RTPCR_F and At5g13270_RTPCR_R to generate *RARE1* product. Aliquots of PCR reactions were removed at cycles indicated (Figure 3.5), and electrophoresed in 1% agarose gels.

Poisoned primer extension (PPE). RNA was isolated from Arabidopsis tissues and analyzed by PPE as previously described (42). Primers used for PPE are listed in Table 3.S1, as either PPE_C or PPE_G, denoting whether the sequencing reactions contained ddCTP or ddGTP, respectively. PPE reactions were performed as in (43), except for PPE oligonucleotides with the 5' Hexachlorofluorecein (HEX) tag modification as in (44), purchased from IDT (Coralville, IA), in which case 0.5 pmol of fluorescently labeled primer was used in place of radiolabeled primer. HEX tagged extension products were detected by a Typhoon 9400 imager (GE Healthcare, Piscataway, NJ), using the 555 nm bandpass-20 filter, excitation at 532nm, and the photomultiplier tube voltage set at 600.

Identification of T-DNA insertional mutants. Line WiscDsLox330H10 (stock CS851454) with a T-DNA insertion mapped within the coding region of At5g13270, was obtained from the Arabidopsis Biological Resource Center (ABRC, The Ohio State University, Columbus, OH). A segregating population was grown on soil and genotyped using the primers listed in Table 3.S1, where primers At5g13270_5'_+1726 and At5g13270_3'_+2328 amplify the wild-type allele, and At5g13270_5'_+1726 and WiscLB amplify the mutant allele. WS-4 Line FLAG_424E06 was obtained from the INRA Versaille T-DNA collection (45). The insertional allele was verified by primers At5g13270_5'_-468 and LB4 and primers At5g13270_5'_-468 and At5g13270_3'_+123 amplified the wild-type allele (Table 3.S1). PCR was performed with BioMix Red (Bioline, Taunton, MA).

Confocal microscopy. A Leica DMRE-7 (SDK) microscope with a TCS-SP2 confocal scanning head (Leica Microsystems Inc., Bannockburn, IL) and a 63x water immersion objective was used to collect confocal laser scanning

images of chloroplasts in wild-type and mutant plants. Chlorophyll fluorescence was excited with 633 nm light and collected between 660 and 700 nm in order to make projections of Z-series.

Bioinformatic analysis. The PlantRBP (http://plantrbp.uoregon.edu/) was consulted to identify putative orthologous groups. Prediction of chloroplast transit sequences was performed with Predotar v. 1.03 (http://urgi.versailles.inra.fr/predotar/predotar.html [46]) and TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/ [47]). Sequences were translated and aligned using T Coffee Version_7.44 (http://www.tcoffee.org [48]). Alignments and residue characteristics were displayed using GeneDoc (http://www.psc.edu/biomed/genedoc). Information about protein motifs was obtained from Pfam (http://pfam.sanger.ac.uk/).

RESULTS

Identification of candidate PPR genes affecting Arabidopsis chloroplast RNA metabolism

Examination of the POGS/Plant RBP RNA-binding protein database (http://plantrbp.uoregon.edu/), which integrates data from Arabidopsis, maize, and rice genome sequencing (49), revealed the existence of Arabidopsis models for genes encoding PPR proteins predicted to be targeted to the chloroplast by both the Predotar and TargetP databases (Table 3.2). Eight such Arabidopsis genes had no obvious orthologs in rice according to PlantRBP, and Reciprocal Best Hit analysis using the Oryza sativa nuclear genome sequence confirmed the absence of putative rice orthologs.

Table 3.2. Known and candidate genes affecting chloroplast RNA metabolism. Data for are derived from the PlantRBP database, (29-31,35,36) and this paper. C targets are numbered relative to the A of the ATG predicted translation initiation codon, where A = 1.

Arabidopsis Locus	Gene Name	Putative Rice Ortholog Exists?	Predotar/TargetP Plastid Targeted (one/both/neither)	Location of RNA Editing Event
At1g05750	*CLB19*	Yes	Both	rpoA C200, clpP C559
At1g11290	*CRR22*	Yes	Both	ndhD C887, ndhB C746, rpoB C551
At1g59720	*CRR28*	Yes	Both	ndhB C467, ndhD C878
At1g79080		No	Both	
At2g45350	*CRR4*	No	Neither	ndhD C2
At3g04760		No	Both	
At3g26630		No	Both	
At3g22690		No	Both	
At4g14190		No	Both	
At5g13270	*RARE1*	No	Both	accD C794
At5g24830		No	Both	

Identification of a candidate editing factor by virus-induced gene silencing

Preliminary experiments resulted in less than 50% of Arabidopsis bleached seedlings among those we agroinfiltrated with a tobacco rattle virus-induced gene silencing (VIGS) vector containing a phytoene desaturase (PDS) gene sequence (39). We also found that bleached tissue resulting from silencing of PDS exhibited altered RNA editing levels at some Arabidopsis chloroplast sites (Figure 3.S1). Because we needed to identify silenced tissue for our studies, we introduced a GFP sequence into the silencing vector so that we could visually screen for silencing of GFP in transgenic Arabidopsis expressing GFP (39,41). RNA could then be isolated from leaves exhibiting reduced GFP expression to assay RNA editing efficiency.

We introduced sequences from the genes encoding the known editing factors CRR4 and CRR21 into the VIGS vector. Gene-specific sequences were selected from the Complete Arabidopsis Transcriptome Microarray (CATMA) database (38) gene-sequence tag database.

Following agroinfiltration and visual screening for reduced GFP fluorescence, we assayed editing of chloroplast Cs using a sensitive poisoned primer extension (PPE) assay (42,50). As expected, the efficiency of editing of the start codon of *ndhD* was specifically reduced in plants inoculated with a *CRR4*-containing VIGS vector, and likewise, editing of the ndhD C383 site was reduced in plants silenced for *CRR21* (Figure 3.1A). We applied this same strategy to At5g13270 and discovered that it exhibited reduced editing efficiency of the accD C794 site relative to plants that were uninoculated or silenced for CRR4 (Figure 3.1B). No other editing defects were detected in the At5g13270-silenced plants.

Phenotype of insertional mutants in At5g13270

We obtained an Arabidopsis ecotype Columbia line carrying a coding region T-DNA insertion in At5g13270 (WiscDsLox330H10) from the Arabidopsis Biological Resource Center. A second T-DNA allele (FLAG_424E06), carrying a promoter insertion in At5g13270, was obtained from the FLAGdB collection made by INRA (Figure 3.2).

Segregating populations of WiscDsLox330H10 were genotyped, and wild-type, heterozygous, and homozygous mutant plants were identified. Homozygous plants were vigorous and set seed under our growth room conditions (Figure 3.3A). No obvious differences in size, morphology, or number of chloroplasts between homozygous mutants and wild-type were

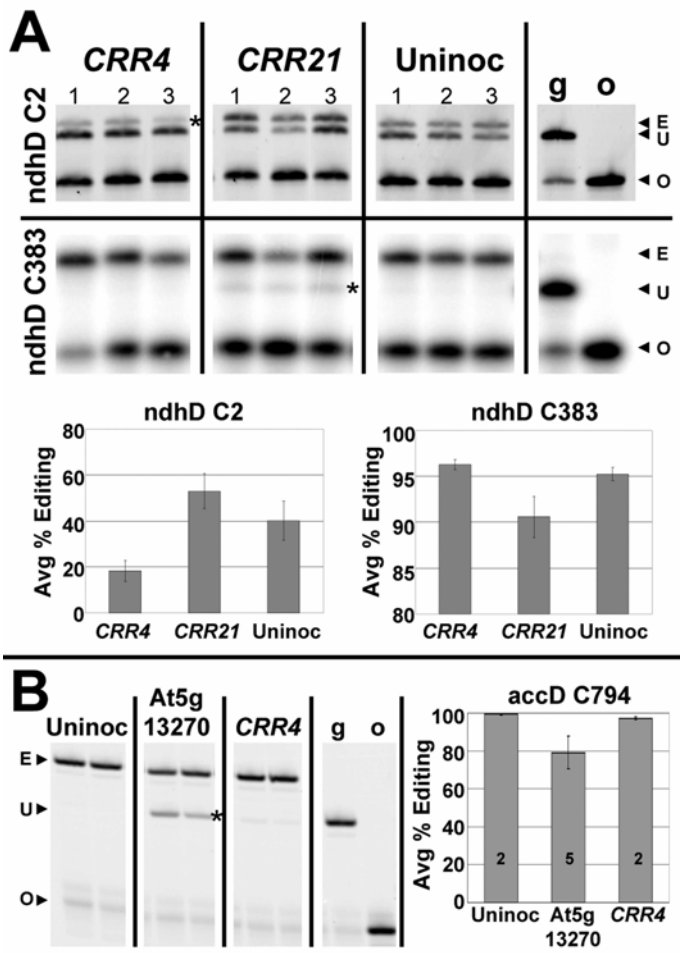


Figure 3.1. VIGS of Arabidopsis genes affects editing of particular C targets. A, Assays of RNA editing of two targets in *ndhD* transcripts following inoculation of Arabidopsis plants with gene silencing vectors designed to silence the known editing factor genes *CRR4* and *CRR21*. PPE products for the ndhD C2 and ndhD C383 editing sites from uninoculated leaves and leaves exhibiting GFP co-silencing of *CRR4* or *CRR21*. Extension products are E, edited; U, unedited; and O, oligo; corresponding to 45, 42, and 35 nt, respectively for ndhD C2, and 41, 35, and 25 nt for ndhD C383. Bottom panel: Bar chart showing average percent editing of *ndhD* sites from 3 replicate plants inoculated with the same silencing constructs as in the top panel. Error bars represent one standard deviation from the mean. B, Assays of RNA editing at accD C794 following inoculation of Arabidopsis plants with vectors designed to silence At5g13270 and *CRR4*. Left panel: PPE products for the accD 794 site from uninoculated leaves and leaves exhibiting GFP co-silencing. Extension products are E, edited, 34nt; U, unedited, 30nt; and O, oligo, 22nt. Bar chart showing average percent editing of accD C794 from 2, 5, and 2 replicate plants, respectively, inoculated with GFP co-silencing constructs as in the left panel. Error bars represent one standard deviation from the mean.

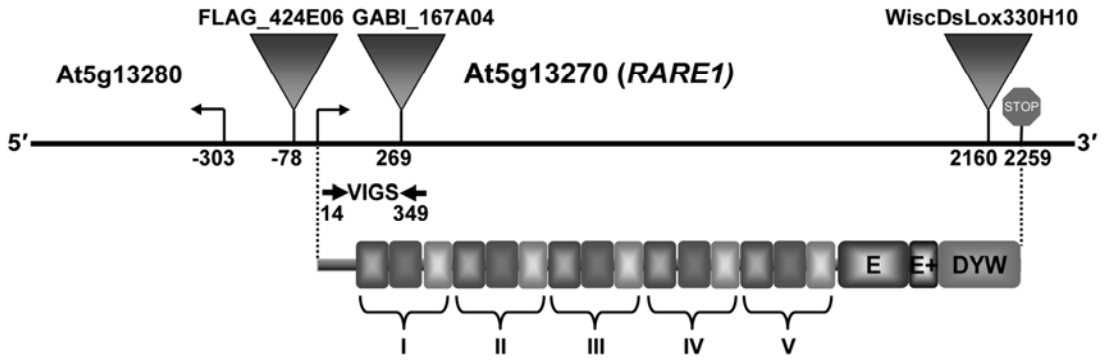


Figure 3.2. Structure of the At5g13270 locus and the predicted RARE1 protein. The locations of T-DNA insertions associated with decreased accD C794 editing are indicated. The 335 nt region of At5g13270 RNA targeted by VIGS is delimited by facing arrows. Depicted below the gene model is the modular organization of the predicted RARE1 protein. Starting at residue 47, five tandem PLS blocks (I –V), of either 101 or 102 residues are shown followed by the E, E+ and DYW domains (27). RARE1 motif coordinates are defined in Figure 3.S2. The P, L, and S forms of the PPR motifs are indicated left to right by three different shadings of gray boxes. Figure is drawn to scale.

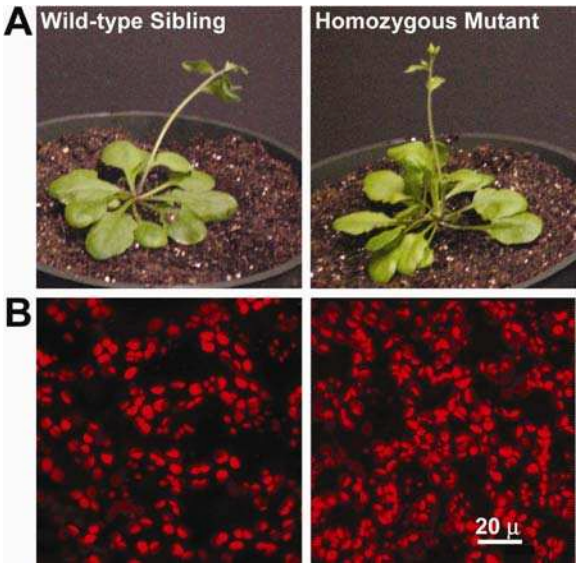


Figure 3.3. A mutant lacking editing of accD C794 exhibits robust growth and chloroplasts of normal appearance. A, Plant growth phenotype of a wild-type Columbia plant and homozygous mutant WiscDsLox330H10 B, Confocal microscopy of mesophyll chloroplasts from rosette leaves of wild-type Columbia and the homozygous WiscDsLox330H10 mutant. Images false-colored red to correspond to chlorophyll autofluorescence.

detected when autofluorescent mesophyll chloroplasts were examined by confocal microscopy (Figure 3.3B).

RNA was prepared from leaves of wild-type, heterozygous, and homozygous WiscDsLox330H10 mutant plants and RNA editing extent was assayed by PPE at all 34 known chloroplast editing sites. The poisoned primer extension assay we used is known to be able to detect edited transcripts representing as little as 1% of the RNA population (42). All sites exhibited wild-type extent of RNA editing in the wild-type and heterozygous plants. In the homozygous WiscDsLox33H10 mutants, no editing was detected at accD C794 (Figure 3.4A). FLAG_424E06 plants exhibit an editing defect at the accD C794 site, though partial editing does occur (Figure 3.4B). Because the FLAG insertion is in the promoter region, possibly some transcript and protein is produced that allows a low level of editing. GABI_167A04 plants do not edit accD C794 (Figure 3.4C). These findings led us to name At5g13270 as *RARE1* (first gene Required for *AccD* RNA Editing). Because mutations of *RARE1* do not affect editing of the second C target in *accD* (Figure 3.4A), it is likely that another gene will be found that is required for editing of accD C1568.

To determine whether or not the WiscDSLox330H10 line expressed the *RARE1* RNA despite the T-DNA insertion and complete lack of editing of accD C794, we performed semi-quantiative RT-PCR on leaves of wild-type and the mutant. Transcripts of At5g13270 were detected in both wild-type and the mutant lines (Figure 3.5). Transcripts of At5g13270 were also observed in the FLAG_424E06 line (data not shown).

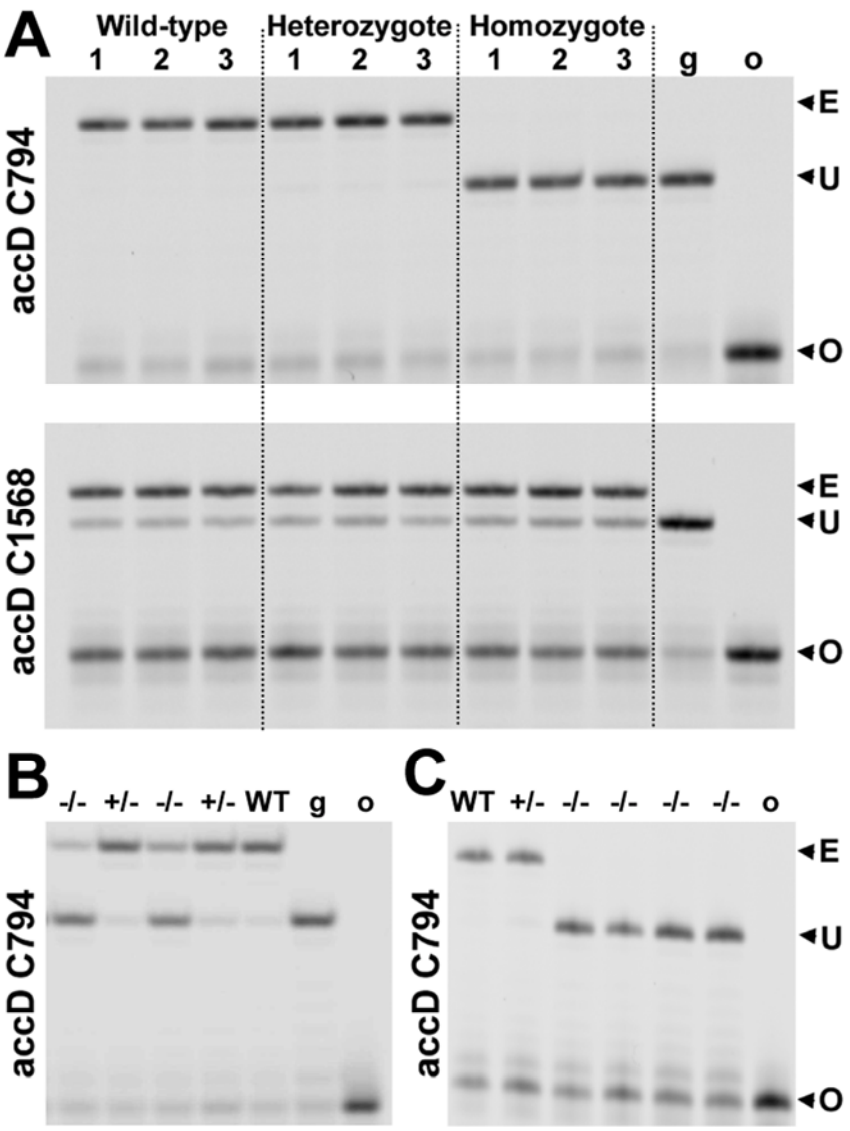


Figure 3.4. Homozygous mutant lines exhibit editing defects in accD C794. RNA assayed in each lane was prepared separately from a single leaf of a different individual plant. E, U, and O indicate edited extension product, unedited extension product, and oligonucleotide, respectively, and represent 34, 30, and 22 nt for the C794 site. A, PPE of RNA from Columbia (wt), heterozygous, and homozygous WiscDsLox330H10 plants. The edited and unedited extension products and oligonucleotide for the C1568 site are 33, 31, and 24 nt in size. B, PPE assay of accD C794 editing in progeny of a heterozygous plant carrying the FLAG_424E06 insertion allele. WT (ecotype WS-4), heterozygotes (+/−) and homozygotes (−/−). Control lanes g and o correspond to genomic DNA template and oligonucleotide only, respectively. C, PPE assay of accD C794 editing in Columbia (WT), a heterozygote (+/−) and homozygous mutant (−/−) plants carrying the GABI_167A04 insertion allele.

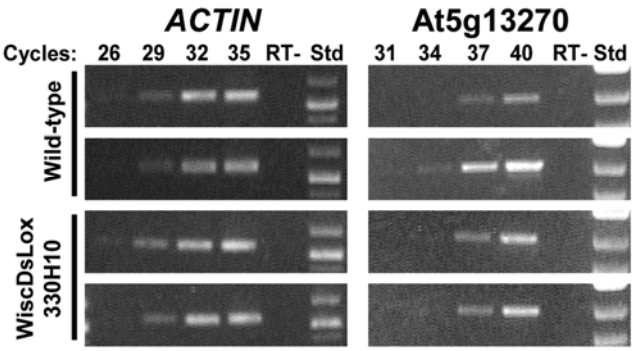


Figure 3.5. Transcripts of At5g13270 are present in both wild-type and WiscDsLox330H10 homozygous mutant plants. Semi-quantitative RT-PCR of *ACTIN* and *RARE1* transcripts in two wild-type and two mutant plants. PCR product sizes are 571bp and 990bp for *ACTIN* and *RARE1*, respectively. Number of PCR cycles is indicated above each lane.

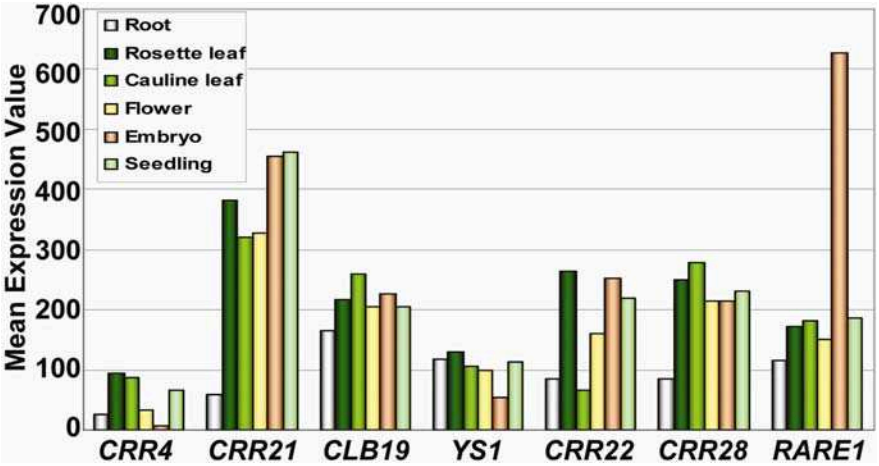


FIGURE 3.6. Expression levels of the seven known *trans*-factors of RNA editing according to online microarray data (https://www.genevestigator.ethz.ch). Accession numbers for editing factors are shown in Table 3.2.

Expression levels of *RARE1* transcripts and extent of editing in different tissues

Most PPR protein-encoding genes that have been examined are expressed at low levels, though one class of mitochondrial PPR genes are expressed at higher levels (51). To determine expression levels of *RARE1* in various tissues, we examined the GENEVESTIGATOR Arabidopsis microarray gene expression database (https://www.genevistigator.ethz.ch [52]). We also

compared the expression of *RARE1* with that of six additional chloroplast editing factors (Figure 3.6). Transcripts of all six of these editing factors are expressed at relatively low levels. *CLB19* and *RARE1* edit transcripts encoding proteins expected to be active in both non-green and green tissues in the plant, and transcript abundance does not vary much except for a tripling of expression of *RARE1* in embryo tissue. Because CRR4, CRR21, CRR22, and CRR28 each edit a transcript encoding a NAD(P)H dehydrogenase subunit not needed in non-green tissues, it is not surprising that expression of both genes in root is very low (Figure 3.6).

To determine whether editing of accD C794 occurs in both green and non-green tissues, PPE was performed on WiscDsLox330H10 RNA isolated from roots, rosette and cauline leaves, and mature flowers of homozygous mutant and wild-type plants. No tissue-specific variation in editing extent was observed and no mutant tissues exhibited any editing of accD C794 (Figure 3.7).

RARE1 is a PPR-DYW protein required for editing of *accD* transcripts

The protein encoded by At5g13270 contains 15 PPR motifs organized into 5 PLS blocks, E/E+ and DYW domains (Figure 3.2). The location of the two T- DNA insertions within the gene and the sequence utilized for virus-induced gene silencing is also indicated in Figure 3.2. Taken together, the reduced editing of *accD* transcripts associated with two mutant alleles of *RARE1* and in plants inoculated with a VIGS vector reveals the importance of this PPR-DYW protein in editing of a serine to leucine codon in the β carboxyltransferase subunit of the heteromeric acetyl-coA carboxylase.

One of the two T-DNA insertion alleles we have analyzed, WiscDsLox330H10, exhibits an insertion within the DYW domain. The mutant

protein predicted from the transcripts we detected (Figure 3.5) would contain only the N-terminal portion of the DYW domain, followed by 17 amino acids encoded by the T-DNA insertion (Figure 3.2) While most of the predicted

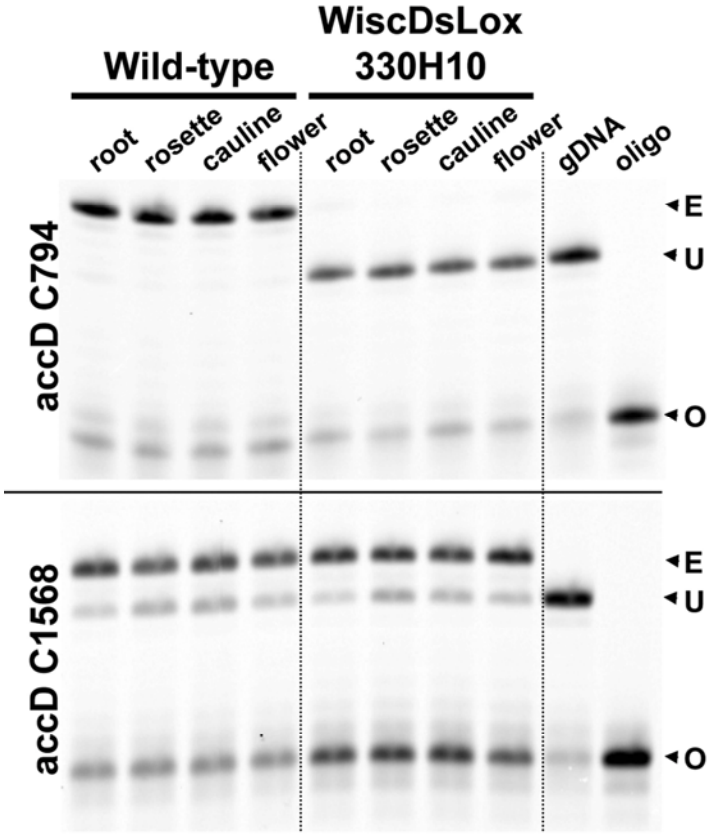


Figure 3.7. No tissue-specific variation in editing extent in *accD* sites was detected in four tissues of wild-type and mutant plants. Editing of accD C794 and C1568 sites was assayed in RNA extracted from four different tissues of wild-type Columbia and WiscDsLox330H10 homozygous mutant plants. Extension product sizes as in Figure 3.4.

mutant protein, including the PPR motifs, E/E+, and part of the DYW domains, does not differ from wild-type RARE1, it is evidently unable to function in editing of *accD* transcripts either due to a lack of protein stability or a necessity for the entire DYW domain for functionality

Both the sequence and organization are conserved among all seven editing factors (29,31,35,36,53), especially in the latter part of PLS block 4,

PLS block 5, and the C-terminal E/E+ domains (Figure 3.8 and Figure 3.S2). In contrast, the N-terminal halves are less conserved, as a whole, which is primarily a function of variability in the number of PLS blocks. RARE1 and the other three PPR-DYW editing factors also exhibit high sequence similarity in the DYW domain, but the region bridging the end of the E/E+ domain and the beginning of the DYW domain (the E-DYW bridge: commencing with residue 657 of RARE1) is much less conserved than what is found in other regions of

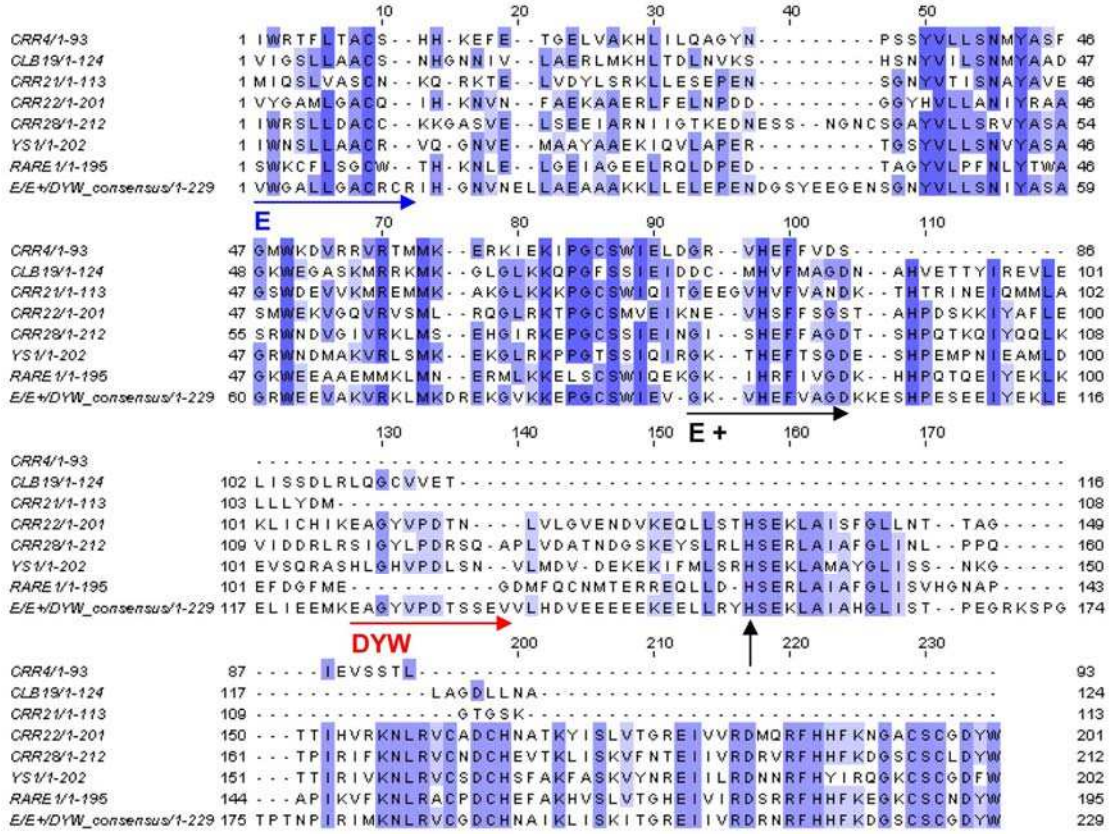


Figure 3.8. Comparative alignments of the C-terminal regions (E-to-DYW domain) of seven PPR-containing proteins (CRR4, CRR21, CLB19, CRR22, CRR28, YS1 and RARE1) that are known to influence chloroplast RNA C-to-U editing. The C-terminal domains (E-to-DYW) of published protein model sequences and the consensus E-to-DYW sequence (27) were aligned using T-Coffee v. 5.05 and are presented using Jalview v. 2 (54). Intensity of coloring reflects the percentage of sequence identity. The arrows indicate the beginning of the consensus of the E, E+ and DYW domains. The vertical arrow indicates the beginning of the highly conserved region, which continues to the C-terminus, of the DYW domain. Figure 3.8 contributed by JCR.

the C-termini (Figure 3.8 and Figure 3.S2). An alignment of the 5 PLS blocks with each other and with a highly conserved PPR-related motif in the E domain is shown in Figure 3.S3.

Further details about the relationships between the PLS blocks in RARE1 are shown in Figure 3.S3. As is the case for nearly all PPR-motif-containing genes ending with a DYW domain, At5g13270 is intronless. Genes exhibiting high similarity to *A. thaliana RARE1* can be identified in the genome sequence data for grape and poplar (Figure 3.S4). A cDNA from grape (Genbank accession CB976854) indicates that this species edits *accD* transcripts at the same codon as in *A. thaliana*. Physicochemical conservation between the three putative orthologous sequences is also quite high (data not shown). Comparison of the RARE1 orthologs indicate a high degree of conservation throughout the protein sequences; however, the region that corresponds to the E-DYW bridge in Figure 3.8 shows a lesser degree of similarity (Figure 3.S4).

The N-terminus of mature RARE1 has not been determined, but TargetP predicts a transit sequence 80 amino acids in length. If cleavage does occur at this location, then it would affect a PPR motif that begins at the 47th amino acid from the N-terminus. A block conserved between the predicted Arabidopsis, grape, and poplar proteins begins at the 21st N-terminal amino acid. Previously, fusion of 100 amino acids of N-terminal sequence of RARE1 with DsRed2 was reported to result in plastid targeting in electroporated tobacco protoplasts (Lurin et al. 2004). Before any information was available about its function, RARE1 happened to be used as a representative DYW protein in assays of four PPR proteins for binding to

homoribopolymers, and was observed to bind to polyG, but not to polyA, polyC, polyT nor to single- or double-stranded DNA (Lurin et al., 2004).

DISCUSSION

We have shown the effectiveness of a novel comparative genomics strategy that takes into consideration the co-evolution of nuclear-encoded ligands and their corresponding organellar substrates for identification of the function of proteins involved in plant organelle RNA metabolism. Candidate genes affecting species-specific aspects of organelle metabolism can be selected by finding proteins carrying RNA-binding domains that are present in one species but lack identifiable orthologs in a second species. Subsequently, plants silenced or mutated in these genes can be tested for loss of species-specific events affecting transcription, RNA processing, or translation.

We have shown that two mutant alleles of an Arabidopsis nuclear gene affect RNA editing of a chloroplast transcript that is not encoded by the rice genome. Not only does the editing event at accD C794 not occur in the rice genome, but also the *accD* gene itself does not exist in the rice chloroplast genome. Thus other RNA metabolism events in Arabidopsis that involve the *accD* gene may require additional nuclear genes that are not needed in rice and other Graminae that lack the *accD* gene. The *RARE1 accD* editing factor gene was successfully detected among a group of Arabidopsis PPR genes predicted to carry a chloroplast transit sequence and lacking identifiable rice orthologs. Future analysis of the remaining 7 Arabidopsis PPR genes in Table 3.2 could focus on RNA editing sites that occur in Arabidopsis but not rice, as well as other events in RNA metabolism that are specific to Arabidopsis.

Furthermore, the converse experiment can be performed to identify the role of RNA-binding proteins in rice genes by identification of rice genes for chloroplast RNA-binding proteins that lack Arabidopsis orthologs, for example.

In other model and crop species for which sufficient genome sequence and mutant collections are available, this strategy can be applied for identification of RNA-binding proteins responsible for 5' and 3' RNA processing, RNA turnover, intron splicing, and translational activation or repression, in addition to RNA editing events. Because CRR4, one of the four verified chloroplast editing factors listed in Table 3.1, was not predicted to be targeted to the chloroplast by either Predotar or TargetP, candidate nuclear genes for proteins affecting organelle RNA metabolism should obviously not be eliminated by a lack of predicted organelle targeting until prediction algorithms become more reliable. In fact, *CRR4* is an Arabidopsis gene lacking an ortholog in rice and would have been on the list of candidate genes in Table 3.2 except for the lack of a predicted chloroplast transit sequence. Editing factors CLB19 and CRR21 would also not have been identified as candidates using our strategy because of the presence of putative orthologs in the rice nuclear genome. As rice does not edit the C targets for which CLB19 and CRR21 are required in Arabidopsis, the putative rice orthologs identified bioinformatically are clearly not performing identical functions in the dicot and monocot species. Some RNA metabolism genes will not be readily identified by our strategy in cases where evolution of the ancestral gene has resulted in divergent functions in recognizably similar genes that remain present in distantly related species.

We have identified only one editing event for which the RARE1 protein is required: accD C794. Like CRR4 and CRR21, RARE1 appears to be

essential for only one C-to-U modification. Therefore the cis-elements surrounding accD C794 are unlikely to be present in a cluster of editing sites (22) that would cross-compete for RARE1 if one transcript were in excess. We have not observed any strong similarity of the sequences 5' to accD C794 in close proximity to other C targets of editing in Arabidopsis.

The robustness of the plants that lack the editing required to produce a leucine codon 265 in Arabidopsis *accD* transcripts is surprising, as this residue was thought to be important for heteromeric ACCase enzyme activity. The leucine residue is genomically encoded in a number of plant species and is

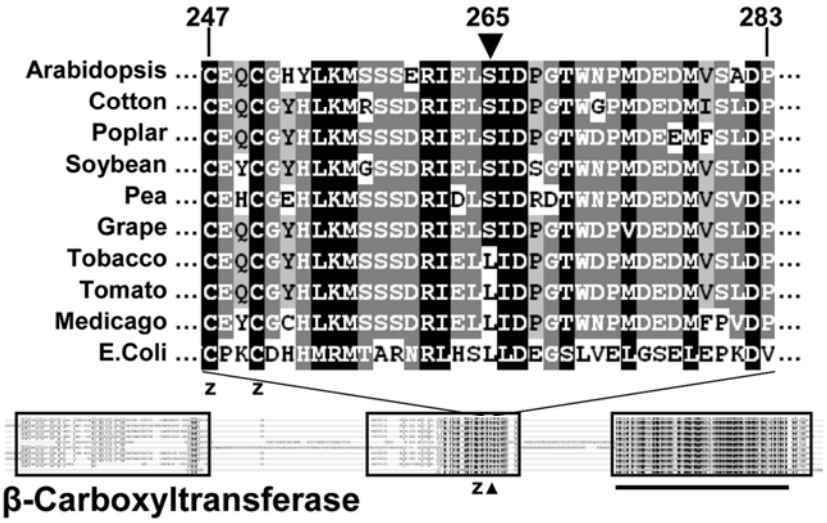


Figure 3.9. Conservation of protein sequences near the amino acid in the β carboxyltransferase subunit of ACCase affected by editing of C794. Alignment of β carboxyltransferase protein sequences from nine plant and one bacterial species. Conserved regions are boxed. Alignment was performed by T-COFFEE and viewed with GeneDoc. Genbank accessions containing the *accD* gene from each species used were: AP000423, DQ345959, EF489041, DQ317523, X56315, DQ424856, Z00044, AM087200, AC093544, NC_000913. Upper panel: Amino acid sequences of the central conserved region containing the residue affected by RNA editing (arrow). Conserved cysteinyl zinc ligands (55) are indicated by (z). Numbers indicate residues in the Arabidopsis sequence. Lower panel: Graphical depiction of ACCase β carboxyltransferase subunit protein alignment. An enlarged version of the complete alignment can be viewed in Figure 3.S3. The conserved carboxyl transferase domain is underlined, arrow and z symbols as in upper panel.

also present in *E. coli* (Figure 3.9). In the three-dimensional structure of the *E. coli* subunit, the leucine codon is present at the subunit interface (10,55) and thus may affect multimeric enzyme integrity. When tobacco plants were transformed with a construct designed to disrupt the *accD* gene, homoplasmic plastid transformants could not be obtained, evidently because loss of all wild-type copies of *accD* would be lethal. Tobacco plants selected to carry some defective *accD* genes exhibited severe leaf development defects (56). Furthermore, *accD* transcription and protein accumulation is evidently important in the specialized type of chloroplasts found in tomato fruit, where photosynthetic genes are downregulated but *accD* expression is upregulated (57).

When the wild-type forms of the two subunits of the pea carboxyltransferase were expressed in *E. coli* with a His tag on the β subunit and purified on a nickel column, the protein complex exhibited carboxyltransferase and ACCase activity (58). Expression in *E. coli* of a version of the *accD* gene that would encode a β subunit corresponding to the unedited transcript resulted in neither carboxyltransferase nor ACCase activity. Furthermore, the mutant subunit exhibited different solubility properties (59). In fact, the pea ACCase study is one of the few that are often cited to demonstrate an important biological role of chloroplast RNA editing. Most evidence for the importance of RNA editing in plant organelle protein function does not result from direct biochemical studies, but instead from observations that a particular amino acid is highly conserved in comparison to the amino acid that would be encoded by an unedited transcript. The predicted pea and Arabidopsis proteins differ considerably in the N-terminal portion of the β carboxyltransferase prior to the zinc-binding domain, and beginning at 24

codons after the edited codon, pea exhibits a 112-codon insertion relative to Arabidopsis (Figure 3.S5.). The Arabidopsis protein therefore varies considerably from its pea ortholog that was tested by heterologous expression by Sasaki et al. (59). It is possible that the diverged sequence allows some function of the Arabidopsis protein despite the presence of a serine at residue 265.

Another possibility is that the Arabidopsis plastid ACCase is not required for plant development because of the presence of a homomeric eukaryotic-type ACCase in the plastid, as is the case in rice and other grasses that lack the heteromeric bacterial-type ACCase (37). The Arabidopsis nuclear genome exhibits two genes expected to encode eukaryotic-type homomeric ACCases, *ACC1* and *ACC2*. Disruption of the *ACC1* gene results in embryo lethality (60). While the ACC1 protein has been shown to function in the cytosol, both Predotar and TargetP predict a plastid transit sequence at the N-terminus of Arabidopsis ACC2. In the related species *Brassica napus*, the N-terminus of the ACC2 protein was shown to target GFP to tobacco chloroplasts (61).

While the current report is focused on RNA metabolism, it will be worthwhile in future studies to study the activity and accumulation of the heteromeric ACCase in *rare1* mutants, as well as more rigorously examine the growth and development of *rare1* mutant plants under a variety of environmental conditions. If the unedited *accD* transcript results in a dysfunctional protein, fatty acid content of vegetative and reproductive tissues of mutants may be altered. Assays of the Arabidopsis enzyme *in vivo* and *in vitro* will reveal whether or not Arabidopsis can tolerate the presence of a S265 in the β carboxyltransferase subunit of ACCase, unlike the pea enzyme.

Future work will reveal whether the homomeric ACCase(s) or other enzymes that may be present in Arabidopsis may be able to compensate for a heteromeric ACCase containing β carboxyltransferase translated from unedited transcripts.

REFERENCES

1. Bock, R. (2000) Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochimie*, **82**, 549-557.

2. Hanson, M.R., Sutton, C.A. and Lu, B. (1996) Plant organelle gene expression: altered by RNA editing *Trends in Plant Science*, **1**, 57-64.

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

4. Tillich, M., Funk, H.T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M. and Maier, R.M. (2005) Editing of plastid RNA in Arabidopsis thaliana ecotypes. *Plant J*, **43**, 708-715.

5. Chateigner-Boutin, A.L. and 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.

6. Bentolila, S., Elliott, L.E. and Hanson, M.R. (2008) Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics*, **178**, 1693-1708.

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

8. Kudla, J., Igloi, G.L., Metzlaff, M., Hagemann, R. and Kossel, H. (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable *psbL* mRNA by a C to U substitution within the initiation codon. *EMBO J*, **11**, 1099-1103.

9. Wintz, H. and Hanson, M.R. (1991) A termination codon is created by RNA editing in the petunia atp9 transcript. *Curr Genet*, **19**, 61-64.

10. Yura, K. and Go, M. (2008) Correlation between amino acid residues converted by RNA editing and functional residues in protein three-dimensional structures in plant organelles. *BMC Plant Biol*, **8**, 79.

11. Tillich, M., Lehwark, P., Morton, B.R. and Maier, U.G. (2006) The evolution of chloroplast RNA editing. *Mol Biol Evol*, **23**, 1912-1921.

12. Tsudzuki, T., Wakasugi, T. and Sugiura, M. (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol*, **53**, 327-332.

13. Corneille, S., Lutz, K. and Maliga, P. (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol Gen Genet*, **264**, 419-424.

14. Kahlau, S., Aspinall, S., Gray, J.C. and Bock, R. (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J Mol Evol*, **63**, 194-207.

15. Bentolila, S., Chateigner-Boutin, A.L. and Hanson, M.R. (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.

16. Zehrmann, A., van der Merwe, J.A., Verbitskiy, D., Brennicke, A. and Takenaka, M. (2008) Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*. *Mitochondrion*, **8**, 319-327.

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

18. Reed, M.L., Peeters, N.M. and Hanson, M.R. (2001) A single alteration 20 nt 5' to an editing target inhibits chloroplast RNA editing in vivo. *Nucleic Acids Res*, **29**, 1507-1513.

19. Hayes, M.L. and Hanson, M.R. (2007) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA*, **13**, 281-288.

20. van der Merwe, J.A., Takenaka, M., Neuwirt, J., Verbitskiy, D. and Brennicke, A. (2006) RNA editing sites in plant mitochondria can share cis-elements. *FEBS Lett*, **580**, 268-272.

21. Hirose, T. and 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.

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

23. Heller, W.P., Hayes, M.L. and Hanson, M.R. (2008) Cross-competition in editing of chloroplast RNA transcripts in vitro implicates sharing of trans-factors between different C targets. *J Biol Chem*, **283**, 7314-7319.

24. Covello, P.S. and Gray, M.W. (1993) On the evolution of RNA editing. *Trends Genet*, **9**, 265-268.

25. Hayes, M.L. and Hanson, M.R. (2008) High conservation of a 5' element required for RNA editing of a C target in chloroplast psbE transcripts. *J Mol Evol*, **67**, 233-245.

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

27. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

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

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

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

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

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

33. Hashimoto, M., Endo, T., Peltier, G., Tasaka, M. and Shikanai, T. (2003) A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in Arabidopsis. *Plant J*, **36**, 541-549.

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

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

36. Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I. and Huang, J. (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*.

37. Konishi, T., Shinohara, K., Yamada, K. and Sasaki, Y. (1996) Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. *Plant Cell Physiol*, **37**, 117-122.

38. Crowe, M.L., Serizet, C., Thareau, V., Aubourg, S., Rouze, P., Hilson, P., Beynon, J., Weisbeek, P., van Hummelen, P., Reymond, P. *et al.* (2003) CATMA: a complete Arabidopsis GST database. *Nucleic Acids Res*, **31**, 156-158.

39. Burch-Smith, T.M., Schiff, M., Liu, Y. and Dinesh-Kumar, S.P. (2006) Efficient virus-induced gene silencing in Arabidopsis. *Plant Physiol*, **142**, 21-27.

40. Haseloff, J., Siemering, K.R., Prasher, D.C. and Hodge, S. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc Natl Acad Sci U S A*, **94**, 2122-2127.

41. Turnage, M.A., Muangsan, N., Peele, C.G. and Robertson, D. (2002) Geminivirus-based vectors for gene silencing in Arabidopsis. *Plant J*, **30**, 107-114.

42. Hayes, M.L. and Hanson, M.R. (2007) Assay of editing of exogenous RNAs in chloroplast extracts of Arabidopsis, maize, pea, and tobacco. *Methods Enzymol*, **424**, 459-482.

43. Hegeman, C.E., Hayes, M.L. and Hanson, M.R. (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in Arabidopsis in vitro. *Plant J*, **42**, 124-132.

44. Roberson, L.M. and Rosenthal, J.J. (2006) An accurate fluorescent assay for quantifying the extent of RNA editing. *RNA*, **12**, 1907-1912.

45. Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M. and Lecharny, A. (2002) FLAGdb/FST: a

database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Res*, **30**, 94-97.

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

47. Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, **2**, 953-971.

48. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, **302**, 205-217.

49. Walker, N.S., Stiffler, N. and Barkan, A. (2007) POGs/PlantRBP: a resource for comparative genomics in plants. *Nucleic Acids Res*, **35**, D852-856.

50. Peeters, N.M. and Hanson, M.R. (2002) Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA*, **8**, 497-511.

51. Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M. and Giege, P. (2008) PPR336 is associated with polysomes in plant mitochondria. *J Mol Biol*, **375**, 626-636.

52. Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol*, **136**, 2621-2632.

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

54. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**, 1189-1191.

55. Bilder, P., Lightle, S., Bainbridge, G., Ohren, J., Finzel, B., Sun, F., Holley, S., Al-Kassim, L., Spessard, C., Melnick, M. *et al.* (2006) The structure of the carboxyltransferase component of acetyl-coA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme. *Biochemistry*, **45**, 1712-1722.

56. Kode, V., Mudd, E.A., Iamtham, S. and Day, A. (2005) The tobacco plastid accD gene is essential and is required for leaf development. *Plant J*, **44**, 237-244.

57. Kahlau, S. and Bock, R. (2008) Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chromoplast gene expression largely serves the production of a single protein. *Plant Cell*, **20**, 856-874.

58. Kozaki, A., Kamada, K., Nagano, Y., Iguchi, H. and Sasaki, Y. (2000) Recombinant carboxyltransferase responsive to redox of pea plastidic acetyl-CoA carboxylase. *J Biol Chem*, **275**, 10702-10708.

59. Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H. and Nagano, Y. (2001) Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem*, **276**, 3937-3940.

60. Baud, S., Guyon, V., Kronenberger, J., Wuilleme, S., Miquel, M., Caboche, M., Lepiniec, L. and Rochat, C. (2003) Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in Arabidopsis. *Plant J*, **33**, 75-86.

61. Schulte, W., Topfer, R., Stracke, R., Schell, J. and Martini, N. (1997) Multi-functional acetyl-CoA carboxylase from *Brassica napus* is encoded by a multi-gene family: indication for plastidic localization of at least one isoform. *Proc Natl Acad Sci U S A*, **94**, 3465-3470.

CHAPTER 4

A DAG-FAMILY PROTEIN INTERACTS WITH ARABIDOPSIS CHLOROPLAST RNA EDITING FACTOR RARE1

ABSTRACT

Cytidine to uridine RNA editing is an important post-transcriptional modification affecting expression of chloroplast and mitochondrial genomes. A group of 25 proteins, exclusively belonging to the pentatricopeptide repeat (PPR) protein family, have been identified as *trans*-factors of RNA editing by forward and reverse genetic screens. However, none of these PPR proteins have been found to contain the enzymatic activity mediating C-to-U nucleotide conversion. Although a protein containing such catalytic activity seems likely to associate with PPR site recognition factors, the editing enzyme as well as other proteins that may interact with PPR proteins within RNA editing complexes are completely unknown. In an attempt to identify additional RNA editing complex components, immunoprecipitation of a RARE1-containing chloroplast RNA editing complex was performed and identification of putative constituents has been attempted by mass spectrometry. One of the proteins identified within the immunoprecipitate, RIP1 (RARE1-INTERACTING PROTEIN 1), was found to interact with RARE1 *in vivo* by yeast two-hybrid analysis. In Arabidopsis, RIP1 belongs to a largely uncharacterized family of eight proteins for which mitochondrial or chloroplast localization is predicted. *rip1* mutants exhibit a dwarf seedling phenotype and display an altered RNA editing profile for various chloroplast C-targets.

INTRODUCTION

RNA editing converts 34 particular cytidine nucleotides of Arabidopsis chloroplast transcripts to uridine (1,2). A model for recognition of targets has been derived from analysis of sequences of these transcripts containing the RNA editing C-targets (*cis*-elements) and the discovery that certain pentatricopeptide repeat (PPR) proteins are essential for editing of one or more C-targets. Current evidence indicates that a *cis*-element of roughly 20 nt upstream of a C-target is recognized by a PPR protein (3-7), which recruits a catalytic component to the RNA editing complex (8). However, the protein factor(s) providing the catalytic deaminase activity have not been identified.

The first proteins shown to affect RNA editing were a class of abundant mRNA binding proteins containing RRMs (RNA recognition motifs). Immunodepletion of one such protein, CP31, from tobacco leaf extracts reduced editing ability *in vitro* (9). CP31 has also been shown to bind the *psbA* transcript (10), which is not known to contain C-targets of RNA editing. Subsequent analysis of the *CP31A* and *CP31B* loci of Arabidopsis showed an involvement of the encoded proteins in RNA editing of partially overlapping subsets of some chloroplast C-targets (11). However, absolute knockout of editing of any one C-target was not observed even in double mutants, increasing the doubt that cpRNPs have a specific function in RNA editing outside of their accepted role in transcript stabilization. Little else on the topic of RRM-containing proteins related to RNA editing is known; instead, all current evidence indicates that proteins of the PPR family act as site specificity factors by binding to *cis*-element sequences. Of course, a function for RNA binding proteins aside from C-target recognition are possible, e.g., generation

of a conformational change in the structure of the RNA such that the *cis*-elements or C-targets themselves become accessible to other proteins of RNA editing complexes.

PPR proteins consist of tandem arrays of a degenerate 35 aa motif, although variation in the exact length of repeats is observed in the P-L-S subclass, and also some carry E (extended) and DYW domains (12). Attempts to produce recombinant PPR proteins from prokaryotic expression systems have had limited success. Only one RNA editing factor PPR protein has been successfully purified by such an approach. Direct binding *in vitro* of CRR4 to the *cis*-element of its C-target, ndhD C2, has been demonstrated (8,13). A few other successful purifications have been accomplished for PPR proteins involved in organelle RNA maturation events outside of RNA editing. Arabidopsis HCF152, which is involved in polycistron cleavage rather than editing, has been shown to bind its target RNA (14,15), and three maize PPR proteins also not known to be involved in editing, CRP1, PPR5 and PPR10, have been shown to bind RNA *in vitro* (16,17).

Among the RNA editing factors of chloroplast and mitochondrial C-targets that have been identified, 11 of 15 of the chloroplast factors and 3 of 5 of the mitochondrial factors belong to the DYW subclass of PPR proteins (13,18-30). Although the DYW domain has been implicated as a possible catalytic component of RNA editing by bioinformatic analyses (31), null mutants of two PPR-DYW proteins, CRR22 and CRR28, can be complemented with transgenes having DYW truncations, indicating that the domain is non-essential in these two cases (23).

As in plant organelle transcripts, C-to-U editing occurs in the mammalian *apoB* mRNA. In mammalian C-to-U editing, the deaminase

enzyme, APOBEC-1, has been identified (32), and recognition of the correct target occurs by a protein factor, ACF, which binds a *cis*-element and the enzyme (33), fulfilling the same role as PPR proteins are postulated to have in plant organellar editing. Reverse genetic analyses of plant cytidine deaminases has failed to date to identify the catalytic factor of C-to-U editing, although a plastid targeted deaminase specifically affecting A-to-I RNA editing of tRNA-Arg has recently been identified (34).

Difficulty with reverse genetic identification of organellar RNA editing factors other than PPR proteins has made the determination of protein-protein interaction partners of the PPR proteins an obvious approach for uncovering other classes of proteins involved in RNA editing, including the possible identification of the catalytic component of editing. To that end, we report here immunoprecipitation of the PPR-DYW editing protein RARE1, which is essential for editing of the accD C794 chloroplast C-target, and the identification of putative interaction protein by mass spectrometry based proteomics. Candidate proteins were examined further by yeast two-hybrid analysis.

MATERIALS AND METHODS

Generation of α-RARE1 antibody. A 159 aa polypeptide spanning the E, E+ and the beginning of the DYW domain of RARE1 (24), 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 4.1) were used to amplify the fragment by PCR, which was cloned into vector pCR2.1/TOPO (Invitrogen, Carlsbad, CA), before subcloning the EcoRI-SalI

Table 4.1. Oligonucleotides used (IDT, Coralville, IA)

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	TAAAGATCATGACATCGATTACAAGGATGACGATGA CAAGGTCGGCGCCGGTT	PCR1 F
3XFLAG-StrepIIF2	CCCGGGGACTACAAAGACCATGACGGTGATTATAA AGATCATGACATCGATT	PCR2 F
3XFLAG-StrepIIR	TGACAAGGTCGGCGCCGGTTGGTCTCATCCTCAAT TTGAAAAATAAGAGCTC	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_+1992R	TTCCATGAAACCATCAAACTCCTTAAGC	ΔDYW
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_+1188R	TTAACCCTGGTAGGGGTTGCC	Rip1 R
At4g28750_+1F	ATGGCGATGACGACAGCATCTACG	cloning
At4g28750_+432R	TTAAGCTGCAACTTCTTCGACCTC	cloning
At1g53280_+1F	ATGGCGTCGTCGTCGTTGT	cloning
At1g53280_+1317R	TTACACAAGTGTTGCCTTTGAGAGC	cloning
At2g42220_+1F	ATGGCGGGGATCATAAGCCC	cloning
At2g42220_+705R	TTAGCTTGTTGGAGGAAAGAGCTTC	cloning
At3g62820_+1F	ATGAAAACTCCCATGAGTTCTTCTATCACG	cloning
At3g62820_+585R	TTACAAACCATTAGCCGCTAGCTTG	cloning

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. 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 (35) was modified to contain a 3XFLAG-StrepII affinity tag C-terminally fused to a Gateway cassette in place of the GusA gene. The affinity tag (see Figure 4.3) contains the 3XFLAG epitope (Sigma) N-terminally fused to the StrepII epitope (IBA, St. Louis, MO) with a 4 aa V-G-A-G linker (36). 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. 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 (37). For native promoter constructs, 311 bp 5' of the start codon was amplified using Rare1_-311F in combination with the primers above. For the DYW deletion (ΔDYW), primer Rare1_+1992R was

used. 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 (38). Transgenic plants were selected on MS agar plus 50 µg/ml kanamycin and 100 µg/ml carbenicillin.

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 (39). 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 (40), goat anti-Rabbit IgG-HRP (GE Healthcare) secondary antibody was used for detection; otherwise, anti-FLAG M2-HRP (Sigma-Aldrich) was used according to manufacturer's protocol.

Size exclusion chromatography. Stromal protein (0.5mg) was prepared as in (41), dialyzed against KEX buffer (30 mM HEPES-KOH, pH 8.0, 200 mM KOAc, 10 mM $Mg(OAc)_2$, and 5 mM DTT) (42), 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 (39), 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 (43), except Anti-FLAG M2 Magnetic Agarose (Sigma-Aldrich) was used, 10 mg total leaf extract prepared as above (without 2-mercaptoethanol) was used for each immunoprecipitation, 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 [17]). $MgCl_2$ 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. 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 (44).

Processing of the MS data, database searches, quantification of identified proteins and data submission to PPDB and PRIDE. 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 (44). 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 (44).

Protein-protein interaction verification in vivo. Yeast two-hybrid analysis was performed with the ProQuest Two-Hybrid System (Invitrogen), using Gateway-ready cDNA clones for the candidate interacting proteins listed in Table 4.2 obtained from the Arabidopsis Biological Resource Center (ABRC, The Ohio State University), including: Gateway clones G67651, G13301, U24617, U15553, G18824, GC00161 and GC105048 corresponding to candidates AT3G15000, AT1G67700, AT3G57620, AT5G40370, AT5G47890, AT1G15010 and AT2G44920, respectively; as well as SSP pUni cDNA clones C105357 and U11217 corresponding to candidates AT4G28750 and

AT1G53280. For the latter two clones, Gateway versions were made by TOPO cloning into pCR8/GW/TOPO using primer pairs At4g28750_+1F with At4g28750_+432R and At1g53280_+1F with At1g53280_+1317R. The remaining two candidates, AT2G42220 and AT3G62820, were cloned by RT-PCR using primers At2g42220_+1F with At2g42220_+705R and At3g62820_+1F with At3g62820_+585R. Additionally AT3G15000 was cloned without a putative transit peptide of 56 aa, using primers At3g15000_+169F and At3g15000_+1188R. 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 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.

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 wild-type allele and T-DNA alleles were amplified with primer pairs At3g15000_-442F with At3g15000_+99R, or At3g15000_-442F with FLAG_LB4, respectively. 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 (45).

Virus-induced gene silencing (VIGS). VIGS using a GFP co-silencing marker as in (24,46) of At3g15000 was done as using CATMA primers (47) At3g15000_VF and At3g15000-VR. Tissue was collected 18 days post inoculation.

Analysis of RNA editing by poisoned primer extension (PPE). All 34 known Arabidopsis chloroplast RNA editing C-targets (1,2) were assayed as in (24).

RESULTS

Anti-RARE1 antibody (α-RARE1) was found to interact with a protein of approximately 75 kDa in wild-type leaves, but absent in *rare1* mutants (Figure 4.1A,C). Use of purified chloroplast stroma (as prepared for RNA editing *in vitro* assays in [41]) for immunoblots eliminated many cross-reacting protein bands (Figure 4.1C). For size estimation of a RARE1 containing protein complex, size exclusion chromatography of stroma (dialyzed against KEX buffer for maintenance of chloroplast protein complexes as formulated in [42]) was performed, and individual fractions were subjected to SDS-PAGE and immunoblotted with α-RARE1 antibody (Figure 4.1D, upper). The peak RARE1 fraction (number 23) was also the peak fraction for a 200 kDa size standard, β-Amylase, and thus the native molecular weight of the RARE1 protein complex was estimated to be approximately 200 kDa. Notably, ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) as detected by an α-Rubisco LSU (large subunit) antibody, was effectively separated from RARE1-containing protein complexes by this method (Figure 4.1D, lower). RARE1 protein was immunoprecipitated using magnetic Dynabeads-Protein A

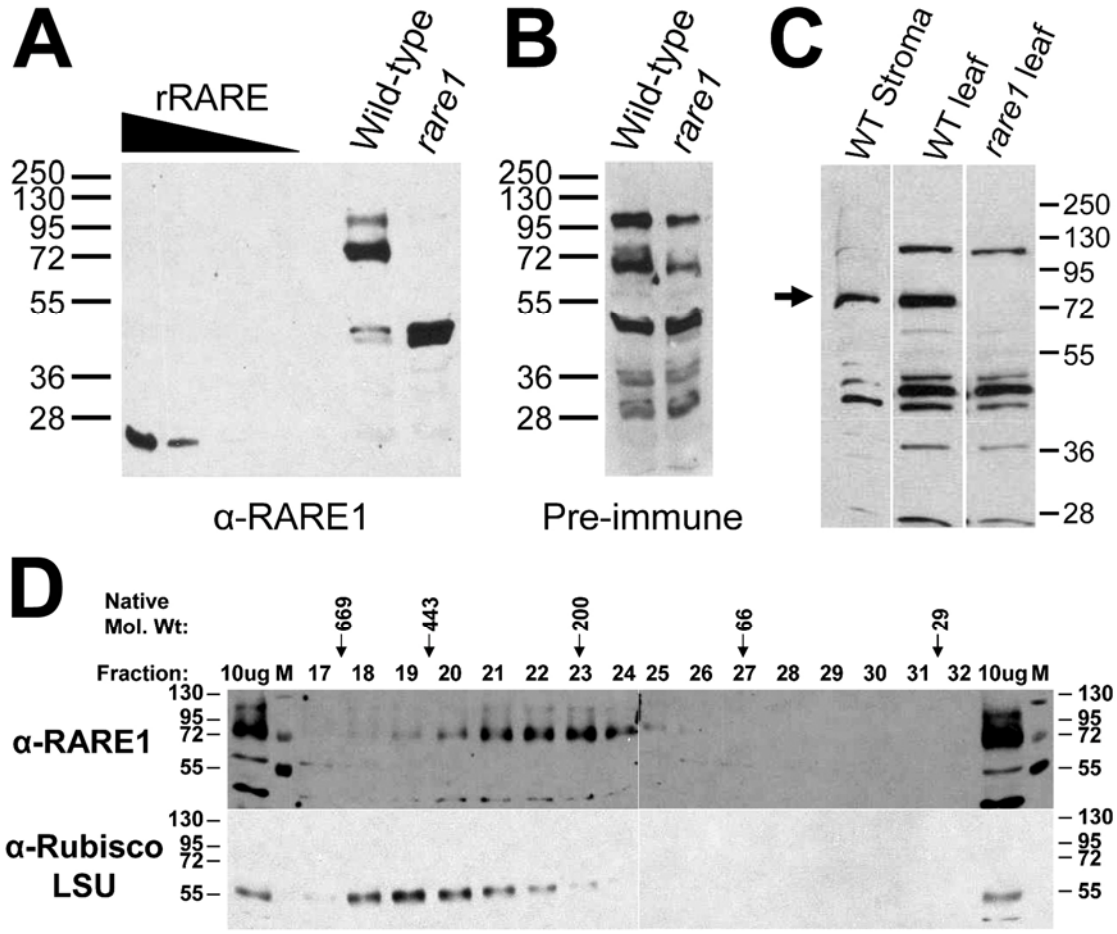


Figure 4.1. Immunoblots of wild-type and *rare1* protein extracts using α-RARE1 antibody. A, α-RARE1 antibody reacts with a 75 kDa protein in wild-type stroma, which is absent in *rare1* (WiscDsLox330H10) stroma. rRARE1 is a 194 aa recombinant protein used for antibody purification and is loaded from 2.5 ng to 0.2 ng (5-fold serial dilutions, left to right). B, Pre-immune serum reactivity against wild-type and *rare1* stroma. C, Total leaf protein extracts contain a number of additional cross-reacting proteins compared to stroma. Arrow indicates RARE1 protein. Loading for all plant protein samples A-C is 20 ug/lane. D, Size exclusion chromatography fractions of wild-type stroma probed with α-RARE1 antibody or α-Rubisco LSU antibody, with peak fraction elution pattern for size standards indicated.

crosslinked to α-RARE1 antibody, and the immunoprecipitate was found to be free of protein contaminants commonly detected by α-RARE1 antibody under denaturing conditions (immunoblots), as shown in Figure 4.2A. Unfortunately, even an as much as 30 µg α-RARE1 antibody bound to the Dynabeads was

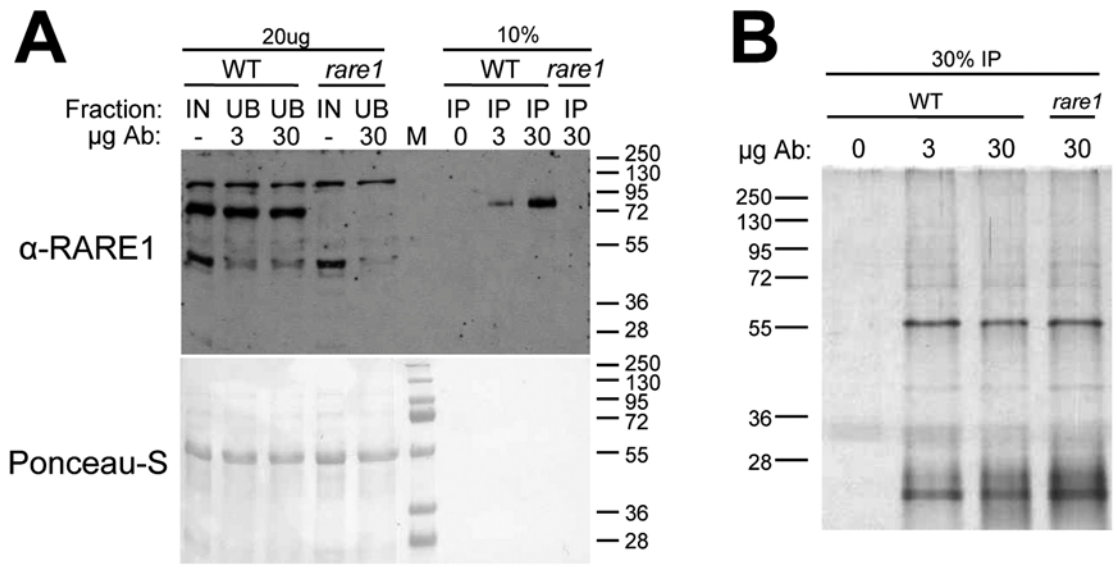


Figure 4.2. Immunoprecipation of RARE1 using α-RARE1 antibody. A, Immunoblot of input (IN), unbound (UB) or immunoprecipate (IP) fractions from α-RARE1 immunoprecipitation. Loading indicated in µg or % of immunoprecipitate. 0, 3, or 30 µg of antibody bound to 1.5mg Dynabeads-Protein A per immunoprecipitation. B, Silver stain gel of 30% each IP fraction.

unable to immunodeplete RARE1 from 2 mg total leaf protein extract, and no 75 kDa band corresponding RARE1 in the SDS-PAGE separated silver-stained gel could be detected. Additionally, RARE1 could not be eluted from Dynabeads using the gentle (low pH) protocol; instead the immunoprecipitate was eluted with sample buffer, resulting in the high background seen in Figure 4.2B.

As an alternative to immunoprecipitation with α-RARE1 antibody, transgenic plants were produced that express RARE1 protein tagged with epitopes to commercially available antibodies/purification reagents. The 3XFLAG and StrepII tags were chosen because of prior successes with plant protein purification (36,43), and a novel tandem affinity tag incorporating the two sequences is shown in Figure 4.3B. However, despite sequence verification upon initial insertion of the DNA fragment into the plant

transformation plasmid, transgenic plants were found to carry a 4 bp deletion within the StrepII tag (see Figure 4.3B, red highlighting), thereby inactivating the StrepII epitope and resulting in a frameshift change and translational fusion to a C-terminal 16 aa polypeptide derived from the T-DNA, as shown in Figure 4.3C. For this reason, only the 3XFLAG epitope was used for immunoblots and immunoprecipitation.

Transgenic plants expressing the various RARE1-fusion or deletion constructs shown in Figure 4.3A in a *rare1* background were found to have varying degrees of accD C794 RNA editing activity (Figure 4.4, upper). Primary 35S::RARE1 transformants were found to have a wide range of editing activity (10-100 percent, data not shown); however, at least one homozygous T2 plant gave rise to a stable line having fully restored accD C794 RNA editing activity. A line carrying a construct with RARE1-3FS expressed from the native promoter had approximately 60% accD C794 RNA editing and was used for the 3XFLAG immunoprecipitation described here. Other transgenic lines, 35S::RARE1-3FS, and 35S::RARE1ΔDYW exhibited less RNA editing, with the other most highly edited lines being 40 and 12% edited, respectively. An immunoblot using α-RARE1 antibody on total leaf protein from each line indicates the relative abundance of RARE1 protein from each line (Figure 4.4, middle panel). The protein gel used for this particular experiment is a NuPAGE Bis-Tris, 4-12% gradient gel and shows a migration for RARE1 at a larger molecular weight (95 kDa compared to 75 kDa in Tris-Glycine gels). RARE1-3FS and RARE1ΔDWY have a calculated MW of 92 and 74 kDa, respectively. Abundance of the RARE1ΔDYW protein is difficult to gauge by the immunoblot for two reasons. First, the cross-reacting protein of ~80 kDa present in all of the samples could be comigrating with the

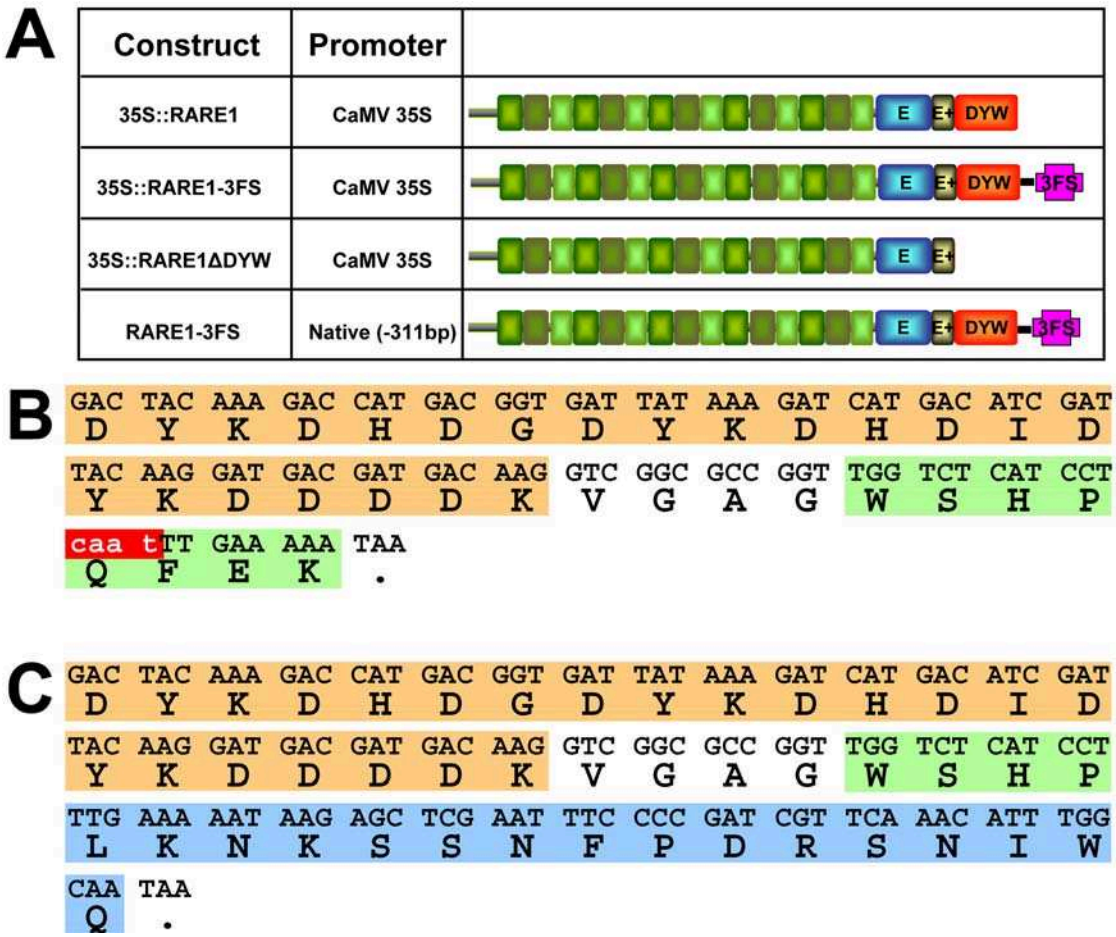


Figure 4.3. Constructs used for floral dip transformation. A, Construct name, promoter, and cartoon of protein to be expressed. Green boxes indicate five P-L-S blocks of PPR motif repeats; E (extended), E+ (extended+) and DYW refer to the additional C-terminal domains of the RARE1 protein. 3FS indicates the addition of the 3xFLAG-StrepII tandem affinity tag. B, Intended DNA and protein sequence of the 3xFLAG-StrepII tag used for transformation. Orange and green highlighting indicate the 3xFLAG and StrepII epitopes, respectively. Red highlighting indicates the four nt deletion present in the actual transformants obtained. C, Nucleotide and protein sequence of actual tag used. Highlighting as in B, except blue, indicating the 16 aa C-terminal addition caused by the translation of the T-DNA after deletion of the nucleotides indicated in B.

RARE1ΔDYW protein. Second, the epitope used to generate the α-RARE1 antibody is partially absent in the transgene, so the lack of protein detection does not necessarily indicate that it does not accumulate in these plants.

Size exclusion chromatography confirmed that RARE1-3FS is in a

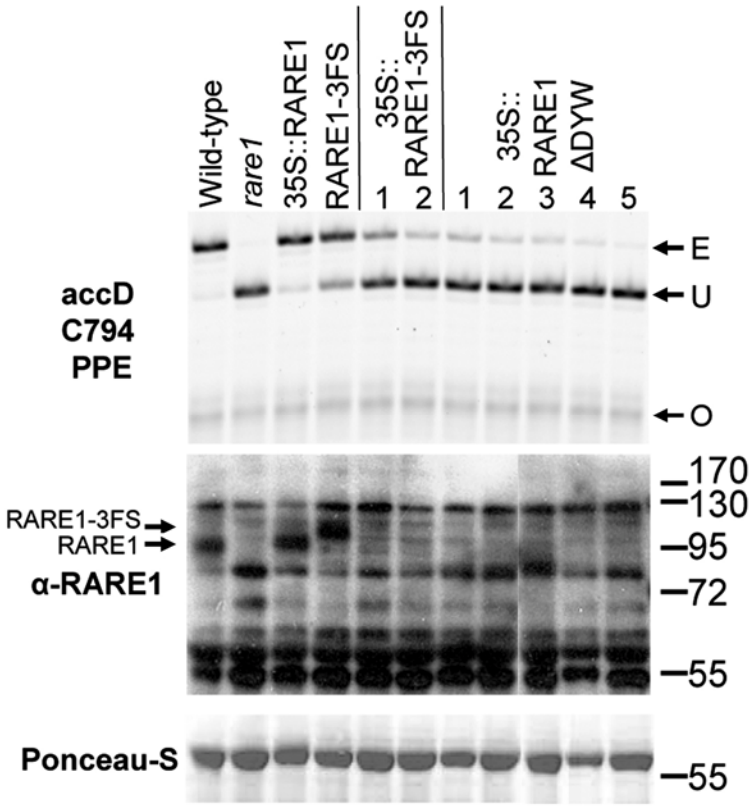


Figure 4.4. accD C794 RNA editing and RARE1 protein production in transgenic lines. Upper panel, poisoned primer extension products E (edited), U (unedited) and O (unincorporated oligo) are 34, 30 and 22 nt, respectively. *rare1* line is WiscDsLox330HI0 homozygous mutant, all transgenic samples labeled as in Figure 4.3A. 35S::RARE1 is from a stable homozygous line, RARE1-3FS heterozygous for the transgene, and transgene copy number in the remaining samples is unknown (≥1 copies). Middle panel, 20 ug total leaf protein from each sample probed with α-RARE1 to show relative abundance of RARE1 protein in the individual lines. Bottom panel, Ponceau-S stain of Rubisco large subunit to show approximately equal loading.

protein complex of approximately the same MW (~200 kDa) in RIPA buffer as the native protein is in KEX buffer (Figure 4.5). Immunoprecipitation with anti-FLAG agarose was performed. Elution of the immunoprecipitate with 3XFLAG peptide, as recommended by the manufacturer and reported in (43), was found to be ineffective (data not shown); instead, elution with 2 M $MgCl_2$ and 0.5% CHAPS detergent was used. Wild-type control, *rare1* control, and RARE1-3FS immunoprecipitates were resolved by SDS-PAGE, and tryptic

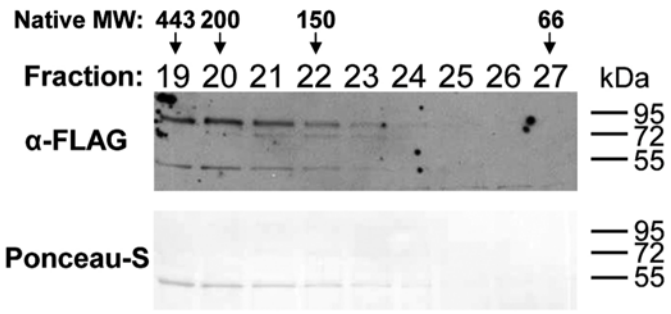


Figure 4.5. Size exclusion chromatography shows RARE1-3FS complex in is similar molecular weight range as the previously measured RARE1 complex in Figure 4.1D. Due to buffer differences, the particular fraction(s) in which size standards and RARE1 complexes eluted are not identical to Figure 4.1D. Labeling as in Figure 4.1D.

peptides extracted from bands of the wild-type and RARE1-3FS lanes were analyzed by mass spectrometry (MS) for protein identification. A protein band with an apparent MW of 85 kDa (arrow, Figure 4.6B) and specific to the RARE1-3FS sample was postulated to be the RARE1-3FS protein (bait). This tentative identification was supported by the MS results, as RARE1 exhibited the highest number of spectral queries for its respective gel section. Proteins comprising other bands unique to the RARE-3FS immunoprecipitate (12, 23, and 24 kDa) have not yet been identified.

In total, peptides from proteins encoded by 524 and 194 genetic loci were identified in the wild-type control and RARE1-3FS samples, respectively. Of the 18 unique proteins identified in the RARE1-3FS sample, 6 were excluded as candidate interacting proteins as they were identified in other partial MS proteomic analyses of wild-type bands from this immunoprecipitation (data not shown). The protein and peptide data used to identify RARE1 and the remaining 11 candidate interacting proteins is shown in Table 4.2. None of the candidates identified by MS are annotated as cytidine deaminases or known to be cpRNPs, though both types of proteins

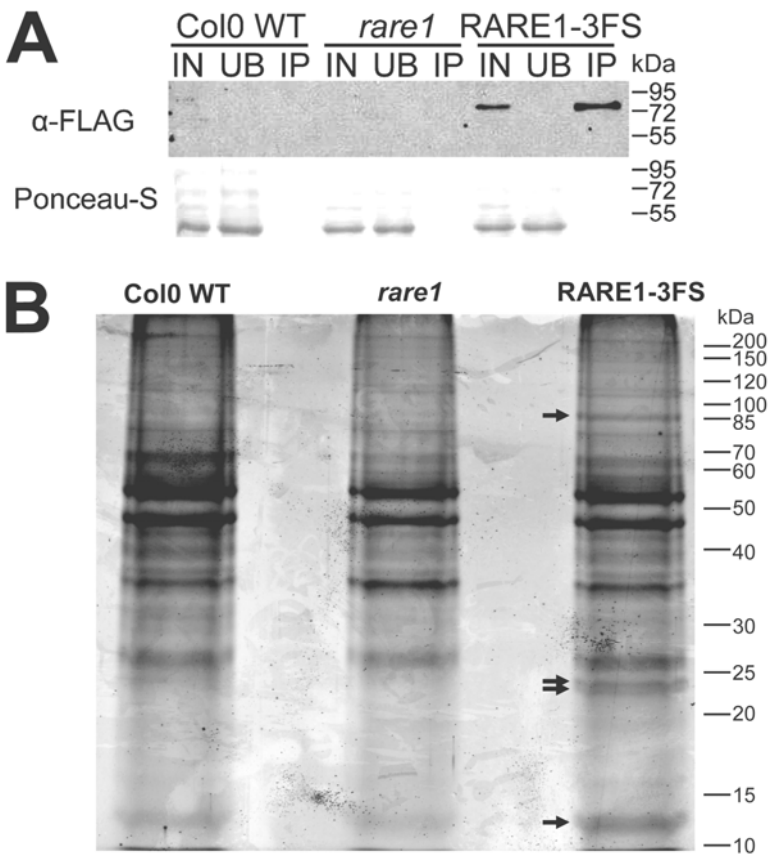


Figure 4.6. RARE1 protein complex co-immunoprecipitation using 3XFLAG antibody against RARE1-3FS protein extracts. A, Immunoblot showing RARE-3FS protein in the input (IN) and immunoprecipitate (IP) fractions from the transgenic line immunoprecipitation, and depletion in the unbound (UB) fraction. Wild-type and *rare1* (background of transgenic line) are included; Ponceau-S stain shows equal loading of control and transgenic samples. B, SyproRuby stain of 50% of each immunoprecipitate. Arrows indicate protein bands specific to RARE1-3FS sample.

were anticipated as likely candidates for RARE1-interacting proteins.

Yeast two-hybrid analysis confirmed protein-protein interaction *in vivo* between RARE1 and AT3G15000.1, the top candidate from the MS data by number of spectral queries (Figure 4.7). This protein was named RIP1 (RARE1-Interacting Protein 1) and a map of the *RIP1* locus is shown in Figure 4.8A. The RARE1-RIP1 interaction was observed only when the RIP1 fusion protein had the putative 56 aa transit peptide removed from its N-terminus.

Table 4.2. MS/MS data of RARE1 and candidate interacting proteins from FLAG immunoprecipitation of RARE1-3FS. 'Queries' indicates the number of MS/MS spectra representing peptide(s) derived from a given protein. 'Peptides' indicates the number of unique peptides identified by one or more MS/MS spectra.

Protein	Annotation	TargetP	MW (-cTP)	Queries	Peptides	Peptide Sequence(s)
AT5G13270.1	RARE1 (Required for accD RNA Editing 1)	cTP	84 (75) kDa	69	17	AVGLFSGMLASGDKPPSSMYTTLLK LKEFDGFMEGDMFQCNMTER AGVSVSSYSYQCLFEACR AGLCSNTSIETGIVNMYVK KPVACTGLMVGYTQAGR LAIAFGLISVHGNAPAPIK SLIGSQYGESALITMYSK FIVGDKHHPQTQEIYEK KLNEAFEFLQEMDK TTMISAYAEQGILDK LNEAFEFLQEMDK LFDEMSELNAVSR NLELGEIAGEELR ACASLEELNLGK HVSLVTGHEIVIR ELSCSWIQEK SGLLDEALK
AT3G15000.1	Similarity to DAG protein	cTP	43 (37)	10	1	TLAQIVGSEDEAR
AT4G28750.1	Subunit E of Photosystem I	cTP	15 (10)	3	2	DSPAAAAAPDGATATKPKPPPIGPK VNYANISTNNYALDEVEEVAA
AT2G42220.1	Rhodanese-like Domain Containing Protein	cTP	26 (20)	3	1	LLLVCQEGLR
AT5G40370.1	Glutaredoxin	SP	12	2	2	AKEIVNSESVVVFSK LVPLLTEAGAIAGK
AT3G57260.1	Beta 1,3-Glucanase 2	SP	37	2	2	VSTAIATDTTTDTSPPSQGR SFLEPVIGFLASK

Table 4.2 (Continued)

Protein	Annotation	TargetP	MW (-cTP)	Queries	Peptides	Peptide Sequence(s)
AT3G62820.1	Invertase/Pectin Methylesterase Inhibitor Family Protein	SP	21	2	1	AGVSVAITDNKDVLR
AT1G67700.1	Auxin-regulated Protein	mTP	26	2	1	AALSTSDVIELPTQDQLK
AT5G47890.1	NADH-Ubiquinone Oxidoreductase B8 Subunit	cTP	11 (9)	2	1	ILLCQSSPASAPTR
AT1G15010.1	Unknown Protein	cTP	16 (12)	1	1	IATGTNR
AT1G53280.1	DJ-1 Family Protein	cTP	47 (42)	1	1	VLVDGNLITSR
AT2G44920.1	Thylakoid Lumen Pentapeptide Repeat	cTP	22 (19)	1	1	VNLTNANLEGATVTGNTSFK

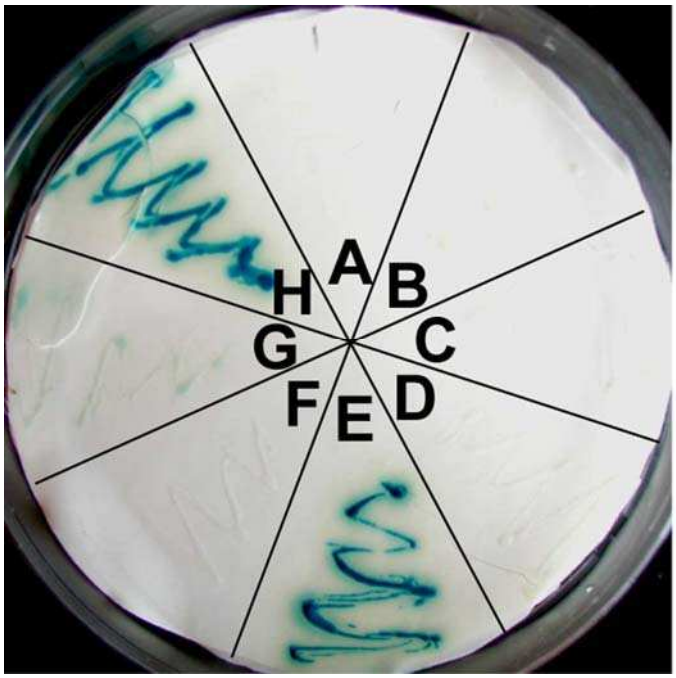


Figure 4.7. Yeast two-hybrid analysis reveals RARE1-RIP1 interaction *in vivo*. X-gal reporter assay of *lacZ* transcriptional activation is shown. All strains are double transformants containing plasmids: A, pDEST32 and pDEST22-RIP1FL; B, pDEST32 and pDEST22-RIP1ΔcTP; C, pDEST32-RARE1 and pDEST22; D, pDEST32-RARE1 and pDEST22-RIP1FL; E, pDEST32-RARE1 and pDEST22-RIPΔcTP; F, pEXP32-Krev1 and pEXP22-RalGDS-m2; G, pEXP32-Krev1 and pEXP22-RalGDS-m1; H, pEXP32-Krev1 and pEXP22-RalGDS-wt. 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 show no autoactivation of *lacZ* expression occurs if only RARE1- or RIP1-fusion proteins are expressed. All pDEST32-RARE1 constructs have deletion of a 33 aa transit peptide, RIP1FL denotes full-length RIP1 without transit peptide removal and RIP1ΔcTP indicates removal of a 56 aa transit peptide.

Thorough analysis of the RARE1 binding capacity of all remaining candidate proteins in Table 4.2 awaits generation of clones lacking transit peptides for the 7 candidates predicted to carry them.

Homozygous FLAG_150D11 mutants, which have a T-DNA inserted 140 bp upstream of the *RIP1* CDS exhibit a dwarf phenotype (Figure 4.8D), also exhibit an altered chloroplast RNA editing profile relative to wild type siblings (Table 4.3). Due to time constraints, only limited analysis of the *rip1*

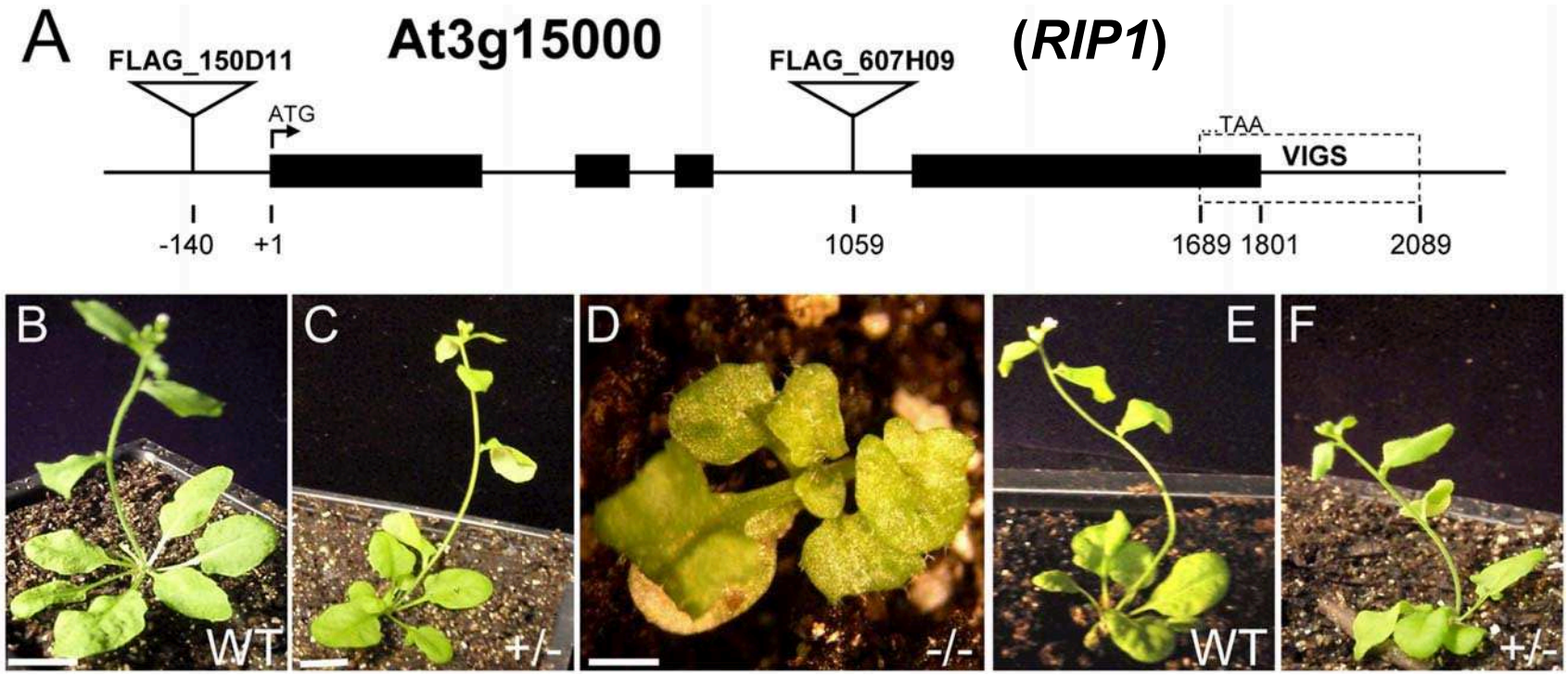


Figure 4.8. The *RIP1* locus. A, Map of At3g15000 (*RIP1*) showing exons in black rectangles, T-DNA insertions described in this text, and region used for VIGS. B-D, Images of segregating progeny of wild-type, heterozygote, and homozygous mutant genotypes for the FLAG_150D11 insertion. Plants are 32 days old and scale bars represent 10 mm (B and C) and 1 mm (D). E-F, Wild-type and heterozygote segregating progeny for the FLAG_607H09 insertion taken at the same plant age as in B-D.

lines has been performed. Due to its promoter location, some functional *RIP1* transcript could be produced from the FLAG_150D11 line which is yet be be assayed by RT-PCR. Regarding the FLAG_607H09 line, no homozygous mutant plants have yet been identified due to a low allele frequency in the population grown from stock center seeds.

PPE data for accD C794 in a segregating *rip1* FLAG_150D11 population is shown in Figure 4.9A; editing in the homozygous mutant is reduced relative to wild-type (90% in mutant compared to 98% in wild-type). *RIP1*-VIGS (Figure 4.9B) did not result in a significant reduction in accD C794 editing, in contrast to VIGS of *RARE1* (as shown in Figure 4.9B and in [24]). However, other chloroplast RNA editing C-targets are differentially edited in *rip1* FLAG_150D11 homozygous and heterozygous mutants, and in *RIP1*-VIGS plants. In particular, petL C5 RNA editing is decreased to 23% in a single *rip1* homozygous mutant compared to 87% average in three wild-type plants; three heterozygotes exhibited an average of 74% editing of this C-target (Figure 4.9C). Notably, *RIP1*-VIGS plants likewise exhibit a reduction of petL C5 editing (74% in *RIP1*-VIGS and 89% in empty vector control, Figure 4.9C). Considering all 34 known chloroplast C-targets of editing present in Arabidopsis, 11 exhibited changes in RNA editing >10%, either increased or decreased (Table 4.3), and in general, heterozygous mutants exhibited an intermediate level of editing.

DISCUSSION

Although plant organellar RNA editing is believed to be modulated by a multi-protein complex, formal estimation of the molecular weight of such a

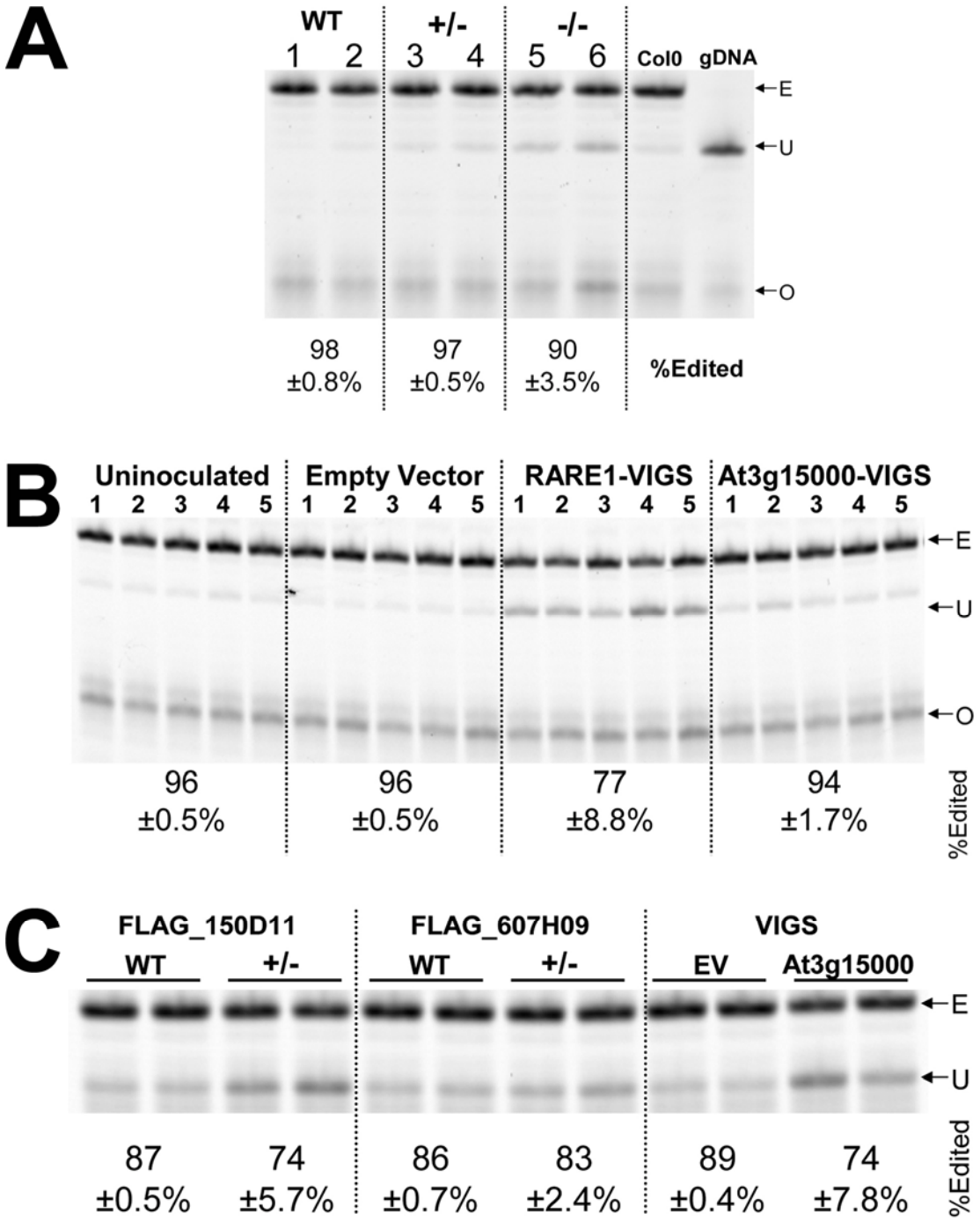


Figure 4.9. RNA editing in *rip1* plants. A, accD C794 editing in wild-type, heterozygote, or homozygous mutants for the FLAG_150D11 insertions. B, accD C794 editing in uninoculated or VIGS-treated plants. C, petL C5 editing in wild-type and heterozygous plants for FLAG_150D11 and FLAG_607H09 insertions, as well as empty vector control and At3g15000 (RIP1) VIGS. Extension product lengths for accD C794 as in Figure 4.5; lengths of petL C5 extension products E, U and O, are 30, 29 and 27 nt, respectively.

Table 4.3. 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 data derived from (13,18-20,22-24,30) except for QED1 (John Robbins, unpublished data).

Trans-factor (if known)	**Chloroplast C-target**	**Genotype**			**Δ Editing HM::WT**
		WT	**Het**	**HM**	
	petL C5	87%	77%	23%	64%↓
QED1	rps12 C(i1)58	18	29	59	41↑
	rpoB C2432	66	75	91	25↑
	accD C1568	43	37	55	12↑
	matK C640	88	93	95	7↑
	ndhB C872	90	90	84	6↓
	rpoC1 C488	44	40	81	37↑
CRR28	ndhD C878	94	91	69	25↓
	ndhB C467	84	80	69	15↓
CRR22	ndhD C887	83	75	50	23↓
	rpoB C551	93	94	86	7↓
	ndhB C746	97	97	95	2↓
	rps14 C80	93	83	59	34↓
CLB19	rpoA C200	84	78	70	14↓
	clpP C559	87	84	91	4↑
CRR4	ndhD C2	49	35	37	12↓
OTP85	ndhD C674	96	94	89	7↓
OTP86	rps14 C149	77	82	83	6↑
RARE1	accD C794	98	97	90	8↓
	ndhB C586	90	89	86	4↓
	ndhB C830	97	98	93	4↓
	ndhB C1255	95	95	91	4↓
	atpF C92	97	97	94	3↓
CRR21	ndhD C383	98	98	95	3↓
OTP80	rpl23 C89	82	81	85	3↑
	ndhB C149	96	97	94	2↓
OTP82	ndhB C836	96	97	94	2↓
	ndhG C50	81	80	79	2↓
OTP84	ndhB C1481	98	98	96	2↓
	ndhF C290	97	98	95	2↓
	psbZ C50	94	94	95	1↑
LPA66	psbF C77	78	82	80	2↑
	psbE C214	93	90	93	none
YS1	rpoB C338	72	67	72	none

complex has not been published. Here, we find RARE1-containing protein complexes to be in the molecular weight range of roughly 200 kDa (Figure 4.1D), which is consistent with the size of a CRR4-containing protein complex (Charles Bullerwell, unpublished data). This is approximately 125 kDa larger than the apparent molecular weight of RARE1 alone, and thus we sought to identify the constituent proteins of the complex using a co-immunoprecipitation approach. Although α-RARE1 antibody purified from rabbit polyclonal antisera could not effectively immunoprecipitate enough RARE1 to attempt protein identification by mass spectrometry based proteomics, 3XFLAG-tagged RARE1 and anti-FLAG agarose immunoprecipitated a protein band of the appropriate molecular weight of RARE1-3FS (apparent MW is 85 kDa) and its identity was verified by MS analysis. Eleven candidate RARE1-intereacting proteins were specifically identified in the RARE1-3FS sample (absent in wild-type) and each was tested for protein-protein interaction with RARE1, of which RIP1 was found to associate with RARE1 in yeast (Figure 4.7).

RIP1's annotation, "similarity to DAG protein," refers to an *Antirhinnum majus* protein, DIFFERENTIATION AND GREENING (DAG) which was shown affect expression of the plastid genome (*rpoB* expression), accumulation of nuclear gene products targeted to plastids, and arrests chloroplast development in the proplastid stage (48). RIP1 belongs to an eight member protein family in Arabidopsis; alignment of the *A. majus* DAG and Arabidopsis DAG-family protein sequences are shown in Figure 4.10. The only member of the Arabidopsis protein family characterized to date is DAL1 (DAG-LIKE 1), AT2G33430.1, mutants of which have an albino phenotype and have been shown to have defects in chloroplast *rrn* operon processing (49,50). Pfam (http://pfam.sanger.ac.uk) does not identify any functional domains within

Figure 4.10. Alignment of Arabidopsis DAG family protein member sequences and *A. majus* DAG. All protein sequences from TAIR9 (51), except for *A. majus* DAG, which is accession CAA65064.1 from EMBL-Bank. Related proteins were identified by BlastP and examination of POGs/PlantRBP (pogs.uoregon.edu [52]). Alignment was made with T-Coffee v.744 (www.t-coffee.org, [53]), and viewed with GeneDoc (http://www.psc.edu/biomed/genedoc). Matrix shows percent identity and similarity pairwise comparisons for the nine protein sequences.

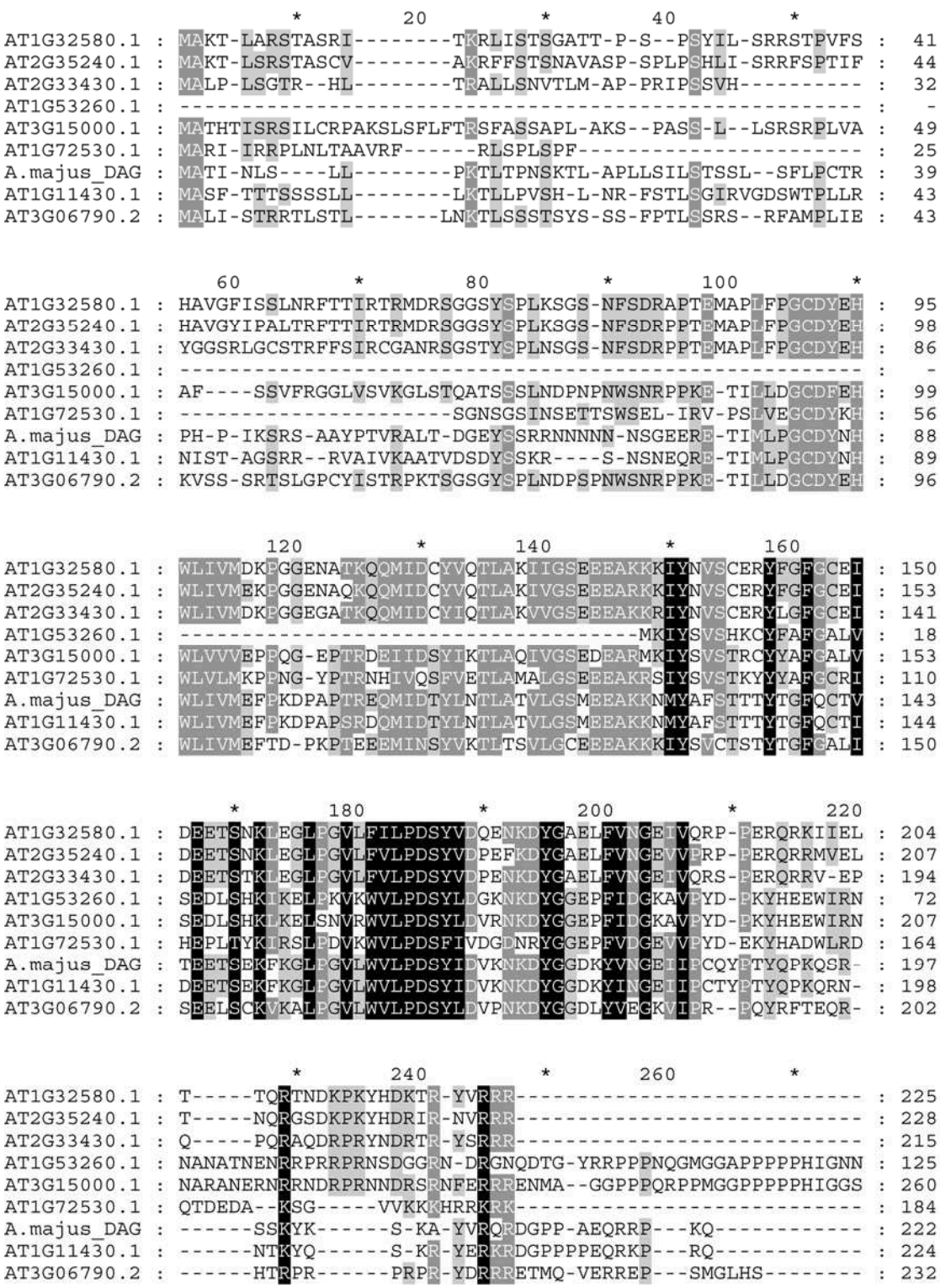

AT1G32580.1 : MAKT-LARSTASRI--------TKRLISTSGATT-P-S--PSYIL-SRRSTPVFS : 41
AT2G35240.1 : MAKT-LSRSTASCV--------AKRFFSTSNAVASP-SPLPSHLI-SRRFSPTIF : 44
AT2G33430.1 : MALP-LSGTR--HL--------TRALLSNVTLM-AP-PRIPSSVH---------- : 32
AT1G53260.1 : ---------- : -
AT3G15000.1 : MATHTISRSILCRPAKSLSFLFTRSFASSAPL-AKS--PASS-L--LSRSRPLVA : 49
AT1G72530.1 : MARI-IRRPLNLTAAVRF------RLSPLSPF---------- : 25
A.majus_DAG : MATI-NLS----LL--------PKTLTPNSKTL-APLLSILSTSSL--SFLPCTR : 39
AT1G11430.1 : MASF-TTTSSSSLL--------LKTLLPVSH-L-NR-FSTLSGIRVGDSWTPLLR : 43
AT3G06790.2 : MALI-STRRTLSTL-------LNKTLSSSTSYS-SS-FPTLSSRS--RFAMPLIE : 43
AT1G32580.1 : HAVGFISSLNRFTTIRTRMDRSGGSYSPLKSGS-NFSDRAPTEMAPLFPGCDYEH : 95
AT2G35240.1 : HAVGYIPALTRFTTIRTRMDRSGGSYSPLKSGS-NFSDRPPTEMAPLFPGCDYEH : 98
AT2G33430.1 : YGGSRLGCSTRFFSIRCGANRSGSTYSPLNSGS-NFSDRPPTEMAPLFPGCDYEH : 86
AT1G53260.1 : ---------- : -
AT3G15000.1 : AF----SSVFRGGLVSVKGLSTQATSSSLNDPNPNWSNRPPKE-TILLDGCDFEH : 99
AT1G72530.1 : ----------SGNSGSINSETTSWSEL-IRV-PSLVEGCDYKH : 56
A.majus_DAG : PH-P-IKSRS-AAYPTVRALT-DGEYSSRRNNNNN-NSGEERE-TIMLPGCDYNH : 88
AT1G11430.1 : NIST-AGSRR--RVAIVKAATVDSDYSSKR----S-NSNEQRE-TIMLPGCDYNH : 89
AT3G06790.2 : KVSS-SRTSLGPCYISTRPKTSGSGYSPLNDPSPNWSNRPPKE-TILLDGCDYEH : 96
AT1G32580.1 : WLIVMDKPGGENATKQQMIDCYVQTLAKIIGSEEEAKKKIYNVSCERYFGFGCEI : 150
AT2G35240.1 : WLIVMEKPGGENAQKQQMIDCYVQTLAKIVGSEEEARKKIYNVSCERYFGFGCEI : 153
AT2G33430.1 : WLIVMDKPGGEGATKQQMIDCYIQTLAKVVGSEEEAKKRIYNVSCERYLGFGCEI : 141
AT1G53260.1 : ----------MKIYSVSHKCYFAFGALV : 18
AT3G15000.1 : WLVVVEPPQG-EPTRDEIIDSYIKTLAQIVGSEDEARMKIYSVSTRCYYAFGALV : 153
AT1G72530.1 : WLVLMKPPNG-YPTRNHIVQSFVETLAMALGSEEEAKRSIYSVSTKYYYAFGCRI : 110
A.majus_DAG : WLIVMEFPKDPAPTREQMIDTYLNTLATVLGSMEEAKKNMYAFSTTTYTGFQCTV : 143
AT1G11430.1 : WLIVMEFPKDPAPSRDQMIDTYLNTLATVLGSMEEAKKNMYAFSTTTYTGFQCTI : 144
AT3G06790.2 : WLIVMEFTD-PKPTEEEMINSYVKTLTSVLGCEEEAKKKIYSVCTSTYTGFGALI : 150
AT1G32580.1 : DEETSNKLEGLPGVLFILPDSYVDQENKDYGAELFVNGEIVQRP-PERQRKIIEL : 204
AT2G35240.1 : DEETSNKLEGLPGVLFVLPDSYVDPEFKDYGAELFVNGEVVERP-PERQRRMVEL : 207
AT2G33430.1 : DEETSTKLEGLPGVLFVLPDSYVDPENKDYGAELFVNGEIVQRS-PERQRRV-EP : 194
AT1G53260.1 : SEDLSHKLKELPKVKWVLPDSYLDGKNKDYGGEPFIDGKAVEYD-PKYHEEWIRN : 72
AT3G15000.1 : SEDLSHKLKELSNVRWVLPDSYLDVRNKDYGGEPFIDGKAVEYD-PKYHEEWIRN : 207
AT1G72530.1 : HEPLTYKLRSLPDVKWVLPDSFIVDGDNRYGGEPFVDGEVVEYD-EKYHADWLRD : 164
A.majus_DAG : TEETSEKFKGLPGVLWVLPDSYIDVKNKDYGGDKYVNGEIIECQYPTYQPKQSR- : 197
AT1G11430.1 : DEETSEKFKGLPGVLWVLPDSYIDVKNKDYGGDKYINGEIIECTYPTYQPKQRN- : 198
AT3G06790.2 : SEELSCKVKALPGVLWVLPDSYLDVPNKDYGGDLYVEGKVIER--PQYRFTEQR- : 202
AT1G32580.1 : T-----TQRTNDKPKYHDKTR-YVRRR---------- : 225
AT2G35240.1 : T-----NQRGSDKPKYHDRIR-NVRRR---------- : 228
AT2G33430.1 : Q-----PQRAQDRPRYNDRTR-YSRRR---------- : 215
AT1G53260.1 : NANATNENRRPRRPRNSDGGRN-DRGNQDTG-YRRPPPNQGMGGAPPPPPHIGNN : 125
AT3G15000.1 : NARANERNRRNDRPRNNDRSRNFERRRENMA--GGPPPQRPPMGGPPPPPPHIGGS : 260
AT1G72530.1 : QTDEDA--KSG-----VVKKKHRRKRK---------- : 184
A.majus_DAG : ------SSKYK------S-KA-YVRQRDGPP-AEQRRP---KQ---------- : 222
AT1G11430.1 : ------NTKYQ------S-KR-YERKRDGPPPPEQRKP---RQ---------- : 224
AT3G06790.2 : ------HTRPR------PRPR-YDRRRETMQ-VERREP---SMGLHS-------- : 232

Figure 4.10 (Continued)

```
                280         *       300         *       320         *
AT1G32580.1 : ------------------------------------------------------- :   -
AT2G35240.1 : ------------------------------------------------------- :   -
AT2G33430.1 : ------------------------------------------------------- :   -
AT1G53260.1 : PNMPPHIQ-----PPNMNQNYRGPPPPPNMNQNYQGPPAPNMNQNYQGPPPSNMG : 175
AT3G15000.1 : APPPPHMGGSAPPPPHMGQNY-GPPPPNNMGGPRHPPP-------YGAPPQNNMG : 307
AT1G72530.1 : ------------------------------------------------------- :   -
A.majus_DAG : ------------------------------------------------------- :   -
AT1G11430.1 : ------------------------------------------------------- :   -
AT3G06790.2 : ------------------------------------------------------- :   -

                340         *       360         *       380
AT1G32580.1 : ------------------------------------------------------- :   -
AT2G35240.1 : ------------------------------------------------------- :   -
AT2G33430.1 : ------------------------------------------------------- :   -
AT1G53260.1 : QNYQGPPPPNMNQSYQGPPPPNMNQSYQG-PPPSNMGQNYRGPSLPPPNMSQNYE : 229
AT3G15000.1 : ----GPRPP---QNYGGTPPPN----YGGAPPANNMGG------APPPN----YG : 341
AT1G72530.1 : ------------------------------------------------------- :   -
A.majus_DAG : ------------------------------------------------------- :   -
AT1G11430.1 : ------------------------------------------------------- :   -
AT3G06790.2 : ------------------------------------------------------- :   -

                *       400         *       420         *
AT1G32580.1 : ----------------------------------------------------ENMR : 229
AT2G35240.1 : ----------------------------------------------------ENMR : 232
AT2G33430.1 : ----------------------------------------------------ENTR : 219
AT1G53260.1 : GPPPPNMNG----GWSGNYQQNGGYQQQGQGGGMQ--QQPY------PPNRNQS : 271
AT3G15000.1 : GGPPPQYGAVPPPQYGGAPPQNNNYQQQGSGMQQPQYQNNYPPNRDGSGNPYQG : 395
AT1G72530.1 : ----------------------------------------------------KKLI : 188
A.majus_DAG : ---------------------------------------E------ATPESST : 230
AT1G11430.1 : ---------------------------------------E------PAASDSS : 232
AT3G06790.2 : ----------------------------------PV--NPG------EFNKPSA : 244
```

	AT1G 35280.1	AT2G 35240.1	AT3G 33430.1	AT1G 53260.1	AT3G 15000.1	AT1G 75230.1	A.majus DAG	AT1G 11430.1
AT2G 35240.1	81 (% Identity) 89 (% Similarity							
AT3G 33430.1	65 75	65 75						
AT1G 53260.1	9 14	9 13	9 13					
AT3G 15000.1	20 33	21 32	21 30	36 40				
AT1G 75230.1	25 42	24 41	24 40	11 16	18 30			
A.majus DAG	33 45	32 46	33 46	8 13	17 29	24 37		
AT1G 11430.1	33 46	31 47	32 45	9 13	17 29	22 36	67 76	
AT3G 06790.2	35 48	37 48	36 48	12 15	26 36	24 38	39 55	39 56

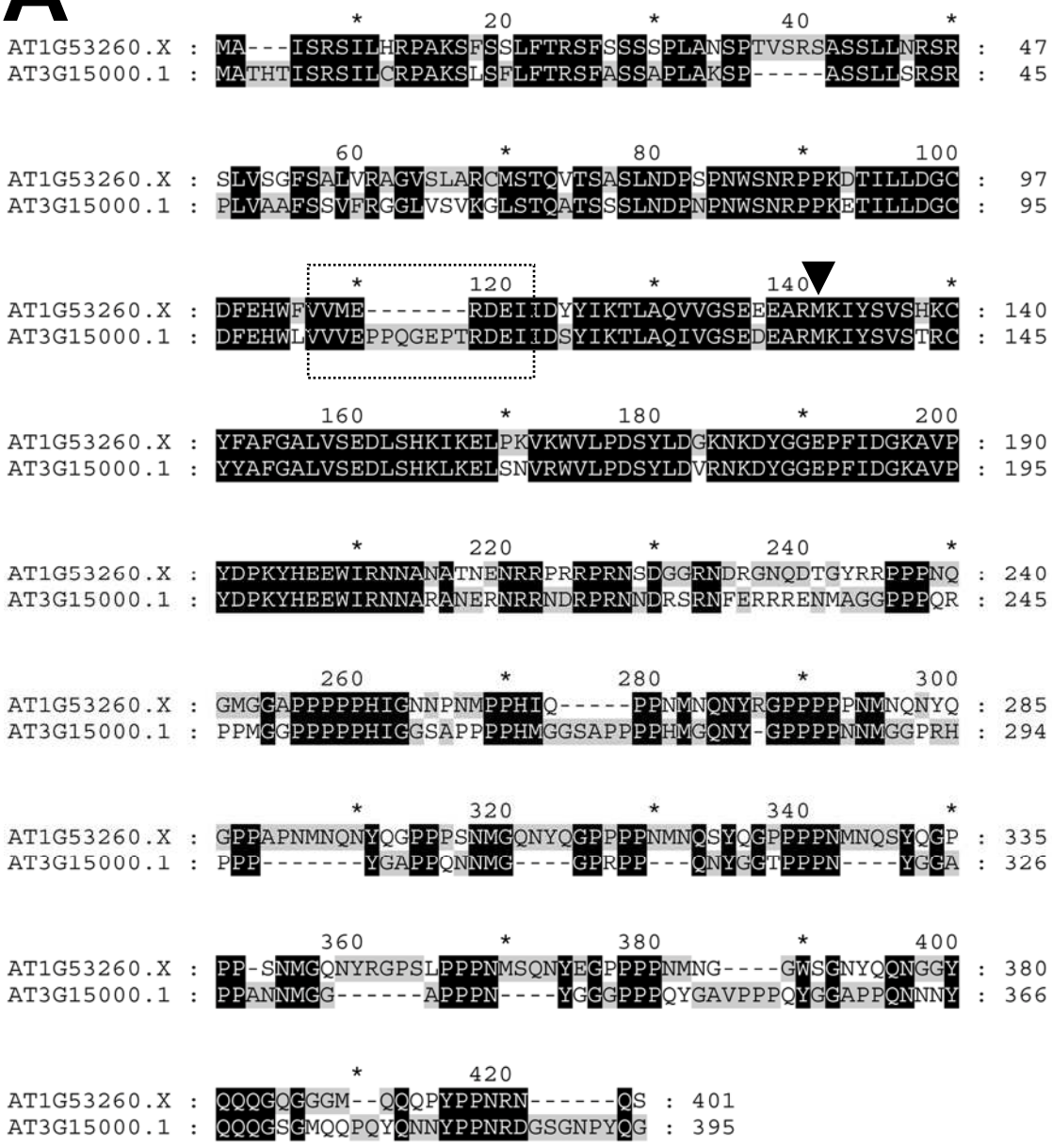
these proteins; however, amino acids 90-198 of RIP1 and the corresponding regions of other proteins in the family show a considerably higher degree of similarity, likely indicating a functional domain.

Relative to the other family members, only RIP1 and AT1G53260.1 have a roughly 150 aa C-terminal proline-rich extension, setting these two proteins apart from the other family members. Furthermore, the At1g53260 locus may have undergone mutation within a sequence that may have previously comprised its first exon. This locus carries a sequence which is highly similar to RIP1 except for 8 nts of divergent sequence that cause a frameshift relative to the remainder of the gene model and divergence from the RIP1 sequence (Figure 4.11). All cDNA sequences for At1g53260 in public databases are continuous across this frameshift region, indicating it does not likely contain an intron, thus functional protein is probably not translated. Therefore, I reason that RIP1 is likely the only functional DAG family protein containing the C-terminal proline-rich extension in the Arabidopsis proteome.

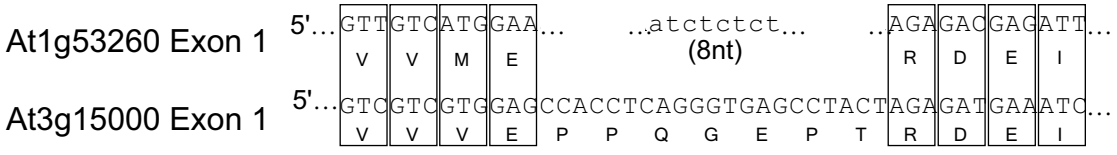
A cladogram derived from the alignment in Figure 4.10 shows the evolutionary relationship of the DAG-family proteins and their targeting predictions from TargetP v 1.1 (54) and Predotar v 1.03 (55) are shown in Figure 4.12. Targeting prediction for the protein family is split evenly split between chloroplasts and mitochondria (except for AT1G53260.1, which is predicted to be secreted, as it lacks the putative plastid transit peptide that would be encoded by the aforementioned alternate first exon). Uptake of DAL1 by pea chloroplasts confirmed the presence of a plastid transit peptide (54), but does not rule out that it could also be targeted to mitochondria. RIP1 was selected for localization analysis in (55), and found to co-localize with a mitochondrial marker in Arabidopsis epidermal cells, but according to the

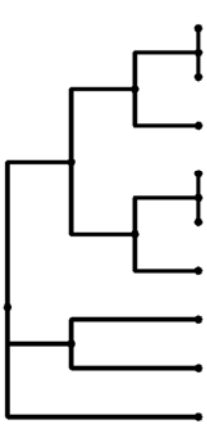
Figure 4.11. Alignment of RIP1 (At3g15000.1) and an alternative gene model for At1g53260. The At1g53260.X model is identical to At1g53260.1 at the C-terminus (after black triangle), but the N-terminus has been extended to place the A of the translation initiation codon at position 19860908 of chromosome 1 (GenBank accession NC_003070) and 8 nt removed after the 107th codon to correct a frameshift change. Start of the At1g53260.1 model indicated by black triangle. The region of the At1g53260.X model surrounding the frameshift change is boxed in A; the nucleotide and translated protein sequence for both *RIP1* and At1g53260.X gene models is shown in B. Percent identity and similarity for the protein sequences are 58 and 66, respectively.

A


AT1G53260.X : MA---ISRSILHRPAKSFSSLFTRSFSSSSPLANSPTVSRSASSLLNRSR : 47
AT3G15000.1 : MATHTISRSILCRPAKSLSFLFTRSFASSAPLAKSP-----ASSLLSRSR : 45
AT1G53260.X : SLVSGFSALVRAGVSLARCMSTQVTSASLNDPSPNWSNRPPKDTILLDGC : 97
AT3G15000.1 : PLVAAFSSVFRGGLVSVKGLSTQATSSSLNDPNPNWSNRPPKETILLDGC : 95
AT1G53260.X : DFEHWFVVME-------RDEIIDYYIKTLAQVVGSEEEARMKIYSVSHKC : 140
AT3G15000.1 : DFEHWLVVVEPPQGEPTRDEIIDSYIKTLAQIVGSEDEARMKIYSVSTRC : 145
AT1G53260.X : YFAFGALVSEDLSHKIKELPKVKWVLPDSYLDGKNKDYGGEPFIDGKAVP : 190
AT3G15000.1 : YYAFGALVSEDLSHKLKELSNVRWVLPDSYLDVRNKDYGGEPFIDGKAVP : 195
AT1G53260.X : YDPKYHEEWIRNNANATNENRRPRRPRNSDGGRNDRGNQDTGYRRPPPNQ : 240
AT3G15000.1 : YDPKYHEEWIRNNARANERNRRNDRPRNNDRSRNFERRRENMAGGPPPQR : 245
AT1G53260.X : GMGGAPPPPPHIGNNPNMPPHIQ-----PPNMNQNYRGPPPPPNMNQNYQ : 285
AT3G15000.1 : PPMGGPPPPPHIGGSAPPPPHMGGSAPPPPHMGQNY-GPPPPNNMGGPRH : 294
AT1G53260.X : GPPAPNMNQNYQGPPPSNMGQNYQGPPPPNMNQSYQGPPPPNMNQSYQGP : 335
AT3G15000.1 : PPP-------YGAPPQNNMG----GPRPP---QNYGGTPPPN----YGGA : 326
AT1G53260.X : PP-SNMGQNYRGPSLPPPNMSQNYEGPPPPNMNG----GWSGNYQQNGGY : 380
AT3G15000.1 : PPANNMGG------APPPN----YGGGPPPQYGAVPPPQYGGAPPQNNNY : 366
AT1G53260.X : QQQGQGGGM--QQQPYPPNRN------QS : 401
AT3G15000.1 : QQQGSGMQQPQYQNNYPPNRDGSGNPYQG : 395

B


At1g53260 Exon 1
5'...GTTGTCATGGAA...
...atctctct...
(8nt)
...AGAGACGAGATT...
V V M E
R D E I
At3g15000 Exon 1
5'...GTCGTCGTGGAGCCACCTCAGGGTGAGCCTACTAGAGATGAAATC...
V V V E P P Q G E P T R D E I



Protein Model	TargetP cTP	TargetP mTP	TargetP SP	TargetP Other	Predotar
AT1G32580.1	**0.869**	0.178	0.001	0.005	mTP
AT2G35240.1	**0.779**	0.428	0.011	0.006	mTP
AT2G33430.1 (DAL1)	0.461	**0.585**	0.002	0.017	cTP
AT1G53260.1	0.028	0.088	**0.544**	0.432	none
AT3G15000.1 (RIP1)	**0.914**	0.128	0.001	0.012	cTP
AT1G72530.1	0.320	**0.733**	0.021	0.017	mTP
A. majus DAG	**0.914**	0.042	0.022	0.054	cTP
AT1G11430.1	**0.808**	0.035	0.080	0.019	cTP
AT3G06790.2	**0.777**	0.649	0.002	0.010	mTP

Figure 4.12. Cladogram of Arabidopsis DAG family proteins and *A. majus* DAG as derived from alignment of these proteins (Figure 4.10), and targeting prediction of the proteins from TargetP and Predotar. cTP, mTP and SP indicate chloroplast transit peptide, mitochondrial transit peptide or secretory pathway signal peptide, respectively.

author's own admission, chloroplast targeting has not been ruled out. In order for RARE1 and RIP1 to interact *in planta*, the two proteins must be co-localized within the same subcellular compartment. A fusion protein consisting of 100 aa from the N-terminus of RARE1 fused to GFP localizes to punctuate spots within *Nicotiana benthamiana* chloroplasts (Appendix IV). RIP1's subcellular localization experiments should certainly be revisited, and in particular, biomolecular fluorescence complementation analysis could be used to determine whether or not the RIP1-RARE1 interaction occurs *in planta*.

Chloroplast RNA editing was surveyed in *rip1* mutants, and although RARE1 specifically affects accD C794 editing, editing efficiency of accD C794 is not the most affected among chloroplast C-targets in *rip1* plants (Table 4.3). Interestingly, *rip1* mutants exhibited higher or lower editing efficiency of a number of chloroplast C-targets, and the effect loosely correlates with particular PPR trans-factors affecting editing of one or C-targets. For

example, of the five C-targets affected by QED1 (AT2G29760), rps12 C(i1)58, rpoB C2432, accD C1568 and matK C872 were edited more efficiently in *rip1* mutants, while the fifth site, ndhB C872, was edited slightly less efficiently. The five C-targets collectively affected by CRR28 and CRR22 all exhibit reduced editing in *rip1* plants. The most affected C-target, petL C5, undergoes only 23% editing in one *rip1* plant, as compared to 87% average the wild-type plants.

As the *rip1* allele used for this study is upstream of the coding region and quantitative analysis of *RIP1* expression has not been done, it is not possible to determine whether RIP1 is essential for editing events, or only has an effect on efficiency of editing. Furthermore, given the strong dwarfism phenotype of *rip1* mutants having the promoter insertion, true knock-out plants may be inviable. The actual cause of the dwarf phenotype cannot be attributed to defective chloroplast RNA editing of a known C-target, in particular the most severe defect in petL C5 editing, as the *petL* gene has been shown to be non-essential in another vascular plant, tobacco (56,57).

Given RIP1's identification in proteomic analyses of mitochondrial proteins (58,59), and prior localization results (60), it is tempting to consider that RIP1 may be playing a role in both organelles. A defect within mitochondria—perhaps RNA editing of one or more C-targets, could be the cause of the strong phenotype exhibited in *rip1* plants. A very useful experiment for more accurately determining the localization of RIP1 would be to generate an antibody against RIP1 for use on an immunoblot against various subcellular protein fractions (especially chloroplast and mitochondrial fractions). Alternatively, transgenic lines expressing affinity-tagged RIP1 could be used for this experiment. Although not detected by microscopy in plants

expressing the N-terminus of RARE1 fused to GFP, RARE1 could be possibly be localized to mitochondria, in addition to chloroplasts. If this were the case, the RARE1-RIP interaction *in planta* could potentially be occurring mitochodria, as well. Indeed, the potential exists for association to occur in chloroplasts, only, mitochondria, only, or in both organelles. Another possibility is that RARE1 and RIP1 do not associate *in planta* due to compartmental separation; instead, RIP1 may bind other PPR protein(s) similar to RARE1 within mitochondria, such as mitochondrial RNA editing factor(s). If this is the case, there is perhaps another DAG-family protein that binds RARE1 in chloroplasts but failed to be identified in the immunoprecipitate either due to low abundance or poor ionization in the mass spectrometer.

I propose that PPR proteins interact with DAG-family proteins on the basis of these findings, specifically (1) RARE1, a PPR protein, interacts with RIP1, a DAG-family protein, (2) the effect on chloroplast C-target RNA editing in *rip1* mutants correlates with the PPR proteins that affect specific subsets of editing C-targets and (3) targeting to chloroplasts, mitochondria, or both, is predicted for these protein families (Figure 4.12 and [12]). Future analysis will undoubtedly determine whether RIP1 or other DAG-family proteins bind other PPR editing trans-factors, including those affecting mitochondrial C-targets. As the editing defects exhibited in *rip1* plants are not universal, RIP1 is unlikely to have catalytic editing activity; plausible functions for RIP1 other than binding PPR proteins directly for site-specific recognition include recruiting the actual catalytic component, stabilizing the complex, or improving the recognition of C targets.

Dissection of the RARE1-RIP1 interaction by yeast two-hybrid analysis may allow assignment of functional domains in these proteins. A likely scenario is that the E domain, universal to all P-L-S PPR proteins identified as editing factors, recruits the catalytic protein either directly or indirectly through protein-protein interactions. Strong biochemical evidence for a universal role of the E domain was shown by generation of chimeras of the CRR4 and CRR21 proteins. Although the primary protein sequences of the two proteins' E domains are divergent, functional complementation between the two domains was observed in *crr4* and *crr21* backgrounds (13). The E domain of CRR4 was also shown to be important for editing of its C-target (13), and deletion of the E and DYW domains from *CRR22* and *CRR28* transgenes abolished editing activity of all sites restored by the introduction of wild-type or DYW-only deletion transgenes into identical mutant backgrounds (23).

If indeed RIP1 binds to PPRs other than RARE1, perhaps the E domain is where the binding occurs. As editing of C-targets known to be affected by CRR22 and CRR28 are also among the targets most affected in *rip1* plants (see Table 4.3), and the DYW domain of CRR22 and CRR28 has been shown to be non-essential for editing, it is unlikely that RIP1 is binding PPR proteins within this domain. The dispensability of the DYW domain from CRR22 and CRR28 is in contrast to the observation that RARE1ΔDYW constructs could not restore editing in *rare1* plants to the same degree as the full-length version (Figure 4.4). Alternatively, RIP1 could bind particular PPR motifs within RARE1 and other PPR proteins with which it may interact.

Designation of functional domains to RIP1 and other DAG proteins could also be determined in the future. Specifically, by determining the location of the PPR-binding domain within RIP1 as either within the conserved

core or within the unique C-terminal extension, affinity of other DAG family proteins for PPR proteins could be predicted.

This discussion has considered RIP1 as the only RARE1-interacting protein identified by the MS analysis of the immunoprecipitate; however, exhaustive screens of RARE1's possible interaction with all 10 other candidates have not been performed. Specifically, recloning a set of 7 of the candidates predicted to have transit peptides for assay in the yeast two-hybrid screen has not yet been done. Indeed, the interaction of RIP1 with RARE1 was observed only when a predicted transit peptide of 56 aa was removed from RIP1. An attempt to determine the exact RARE1 transit peptide length empirically was unsuccessful; however, it has been noted that the TargetP predicted length of 80 aa may be inaccurate as it would disrupt a PPR motif and a region of sequence conservation among grape and poplar orthologs that continues N-terminal to the predicted cleavage site (24). A peptide mapping to amino acids 63-75 of RARE1, LNEAFEFLQEMDK, was identified by MS (Table 4.2), indicating that the actual transit sequence must be less than 63 aa. Examination of sequence preferences proximal to mapped transit peptide cleavage sites of chloroplast proteins observed by MS (61) indicated that a possible transit peptide length may be 33 aa, so codons encoding these 33aa were removed from RARE1 for the yeast two-hybrid studies.

The strength of data implicating the remaining candidates as interaction partners with RARE1 is weak, i.e., relatively few spectral queries support peptides from each candidate; however, an exhaustive analysis of the remaining 10 candidates with transit peptide truncations (if predicted) for interaction with RARE1 and RIP1, could result in the identification of additional proteins comprising the RARE1 protein complex.

Although it has not been included in Table 4.3, another PPR-DYW protein, AT1G15510, also known as AtECB2 (*Arabidopsis thaliana* EARLY CHLOROPLAST BIOGENESIS 2) and VAC1 (VANILLA CREAM 1), has been studied by two groups and *atecb2*/*vac1* plants have reduced editing of the accD C794 and ndhF C290 C-targets (28,62). This partial reduction of editing is in contrast to the complete knockout of editing caused by mutation of genes encoding PPR-DYW proteins in *rare1* and *otp84* mutants, which are defective in editing of the accD C794 and ndhF C290 C-targets, respectively (20,24). Additionally, *rare1* and *otp84* plants are able to grow photoauxotrophically, whereas mutants of the At1g15510 locus require growth on synthetic media supplemented with sucrose and have severe defects in the accumulation of photosynthetic pigments (62). For this reason, the nature of the relationship between AtECB2/VAC1 and RNA editing of the two C-targets whose complete loss did not result in such a strong phenotype is called into question. Indeed, multiple partial editing defects were observed in an unrelated albino mutant, *ispH*, whose molecular defect is in the plastid isoprenoid biosynthesis pathway, which is not known to involve RNA editing. While it is possible that the AtECB2/VAC1 protein may be somehow involved in RNA editing, its role is clearly different from that of RARE1 and OTP84. The editing defect in *atecb2*/*vac1* mutants could be an indirect effect of some other molecular defect. One way to test whether AtECB2/VAC1 is present in RNA editing complexes is to assay this interaction by yeast two-hybrid analysis with RARE1, RIP1 and any other protein(s) identified within the RARE1-containing RNA editing complex. As accD C794 editing is one of the two C-targets specifically reduced in the *atecb2*/*vac1* mutant, it is expected that if AtECB2/VAC1 is a genuine RNA editing factor, it would bind either the *accD*

transcript directly or indirectly by interaction with RARE1 or some other component of the editing complex.

REFERENCES

1. Chateigner-Boutin, A.L. and 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.

2. Tillich, M., Funk, H.T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M. and Maier, R.M. (2005) Editing of plastid RNA in Arabidopsis thaliana ecotypes. *Plant J*, **43**, 708-715.

3. Chateigner-Boutin, A.L. and Hanson, M.R. (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol Cell Biol*, **22**, 8448-8456.

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

5. Hayes, M.L. and Hanson, M.R. (2007) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA*, **13**, 281-288.

6. Heller, W.P., Hayes, M.L. and Hanson, M.R. (2008) Cross-competition in editing of chloroplast RNA transcripts in vitro implicates sharing of trans-factors between different C targets. *J Biol Chem*, **283**, 7314-7319.

7. Reed, M.L., Peeters, N.M. and Hanson, M.R. (2001) A single alteration 20 nt 5' to an editing target inhibits chloroplast RNA editing in vivo. *Nucleic Acids Res*, **29**, 1507-1513.

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

9. Hirose, T. and 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.

10. Nakamura, T., Ohta, M., Sugiura, M. and Sugita, M. (2001) Chloroplast ribonucleoproteins function as a stabilizing factor of ribosome-free mRNAs in the stroma. *J Biol Chem*, **276**, 147-152.

11. Tillich, M., Hardel, S.L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J.M., Lehwark, P., Leister, D., Small, I.D. and Schmitz-Linneweber, C. (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A*, **106**, 6002-6007.

12. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

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

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

15. Nakamura, T., Meierhoff, K., Westhoff, P. and Schuster, G. (2003) RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. *Eur J Biochem*, **270**, 4070-4081.

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

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

18. Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C. and Zhang, L. (2009) LPA66 is required for editing psbF chloroplast transcripts in Arabidopsis. *Plant Physiol*, **150**, 1260-1271.

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

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

21. Kim, S.R., Yang, J.I., Moon, S., Ryu, C.H., An, K., Kim, K.M., Yim, J. and An, G. (2009) Rice OGR1 encodes a pentatricopeptide repeat-

DYW protein and is essential for RNA editing in mitochondria. *Plant J*, **59**, 738-749.

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

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

24. Robbins, J.C., Heller, W.P. and Hanson, M.R. (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.

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

26. 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.

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

28. Yu, Q.B., Jiang, Y., Chong, K. and 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.

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

30. Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I. and Huang, J. (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*.

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

32. Teng, B., Burant, C.F. and Davidson, N.O. (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science*, **260**, 1816-1819.

33. Mehta, A., Kinter, M.T., Sherman, N.E. and Driscoll, D.M. (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**, 1846-1854.

34. Karcher, D. and Bock, R. (2009) Identification of the chloroplast adenosine-to-inosine tRNA editing enzyme. *RNA*, **15**, 1251-1257.

35. Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J*, **6**, 3901-3907.

36. Witte, C.P., Noel, L.D., Gielbert, J., Parker, J.E. and Romeis, T. (2004) Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. *Plant Mol Biol*, **55**, 135-147.

37. Kwok, E.Y. and Hanson, M.R. (2004) GFP-labelled Rubisco and aspartate aminotransferase are present in plastid stromules and traffic between plastids. *J Exp Bot*, **55**, 595-604.

38. Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W. and Chua, N.H. (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat Protoc*, **1**, 641-646.

39. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

40. Makino, A., Mae, T. and 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.

41. Hegeman, C.E., Hayes, M.L. and Hanson, M.R. (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in Arabidopsis in vitro. *Plant J*, **42**, 124-132.

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

43. Gillman, J.D., Bentolila, S. and Hanson, M.R. (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.

44. Friso, G., Majeran, W., Huang, M., Sun, Q. and van Wijk, K.J. (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.

45. Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M. and Lecharny, A. (2002) FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of Arabidopsis thaliana T-DNA transformants. *Nucleic Acids Res*, **30**, 94-97.

46. Burch-Smith, T.M., Schiff, M., Liu, Y. and Dinesh-Kumar, S.P. (2006) Efficient virus-induced gene silencing in Arabidopsis. *Plant Physiol*, **142**, 21-27.

47. Crowe, M.L., Serizet, C., Thareau, V., Aubourg, S., Rouze, P., Hilson, P., Beynon, J., Weisbeek, P., van Hummelen, P., Reymond, P. *et al.* (2003) CATMA: a complete Arabidopsis GST database. *Nucleic Acids Res*, **31**, 156-158.

48. Chatterjee, M., Sparvoli, S., Edmunds, C., Garosi, P., Findlay, K. and Martin, C. (1996) DAG, a gene required for chloroplast differentiation and palisade development in Antirrhinum majus. *EMBO J*, **15**, 4194-4207.

49. Babiychuk, E., Fuangthong, M., Van Montagu, M., Inze, D. and Kushnir, S. (1997) Efficient gene tagging in Arabidopsis thaliana using a gene trap approach. *Proc Natl Acad Sci U S A*, **94**, 12722-12727.

50. Bisanz, C., Begot, L., Carol, P., Perez, P., Bligny, M., Pesey, H., Gallois, J.L., Lerbs-Mache, S. and Mache, R. (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.

51. Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L. *et*

al. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res*, **36**, D1009-1014.

52. Walker, N.S., Stiffler, N. and Barkan, A. (2007) POGs/PlantRBP: a resource for comparative genomics in plants. *Nucleic Acids Res*, **35**, D852-856.

53. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, **302**, 205-217.

54. Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, **2**, 953-971.

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

56. Fiebig, A., Stegemann, S. and Bock, R. (2004) Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Res*, **32**, 3615-3622.

57. Schwenkert, S., Legen, J., Takami, T., Shikanai, T., Herrmann, R.G. and Meurer, J. (2007) Role of the low-molecular-weight subunits PetL, PetG, and PetN in assembly, stability, and dimerization of the cytochrome b6f complex in tobacco. *Plant Physiol*, **144**, 1924-1935.

58. Brugiere, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., Ravanel, S., d'Herin, P., Garin, J., Bourguignon, J. *et al.* (2004) The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. *Phytochemistry*, **65**, 1693-1707.

59. Heazlewood, J.L., Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J. and Millar, A.H. (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**, 241-256.

60. Nelson, B.K., Cai, X. and Nebenfuhr, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J*, **51**, 1126-1136.

61. Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. and van Wijk, K.J. (2008) Sorting signals, N-terminal

modifications and abundance of the chloroplast proteome. *PLoS One*, **3**, e1994.

62. Tseng, C.C., Sung, T.Y., Li, Y.C., Hsu, S.J., Lin, C.L. and Hsieh, M.H. Editing of accD and ndhF chloroplast transcripts is partially affected in the Arabidopsis vanilla cream1 mutant. *Plant Mol Biol*, **73**, 309-323.

APPENDIX I

Table 3.S1. Oligonucleotides used (IDT, Coralville, IA)

Name	Sequence 5′ to 3′	Purpose
SacI-mGFP4_F	GAGCTCAGGAGAGGACCATCTTCTT	pTRV2-GFPGW
XbaI-mGFP4_R	TCTAGACCATTCTTTTGTTTGTCTGC	
At1g79080_V_F	AATGTGCTGTTGACTGGTTT	Silencing
At1g79080_V_R	GCCGGTATATCATCTCGTCT	
At3g04760_V_F	TTGGAATCAATGGTGGGA	
At3g04760_V_R	AGTAAAGGGAAAGTCCTATGCAACC	
At3g22690_V_F	GCGATGTTGGGTAATGTT	
At3g22690_V_R	GATTCACTGTTCTCAAACAC	
At3g26630_V_F	TTTATCGTTAATCCATGGCG	
At3g26630_V_R	CTGGGATTGATGAGAGATCA	
At4g14190_V_F	GGACTGAACTTGCTTGGT	
At4g14190_V_R	GACTGGTTCTTCACTTCTTC	
At5g13270_V_F	CTGTACAGTCTTCGTTTAGCTCTT	
At5g13270_V_R	AAGGATTCTCGATACCCATCC	
At5g24830_V_F	GGAAGGGTCGATTGAAGA	
At5g24830__R	TACTGCTCTTCTGTATGCATGTG	
At5g52850_V_F	GTGCAAGATTGTTTTGGACTGCT	
At5g52850_V_R	GACACTCAGTGCCATTTCATGCT	
accD_+676_F	TGTGGATTCAATGCGACAAT	RNA Editing analysis
accD_+1638_R	GTTTGTCTAGTCTAATTTGAACTTCC	
accD_C794_PPE_G	/HEX/CCATAGGATTCCAAGTACCCGG	
accD_C1568_PPE_G	/HEX/CACTTTTTAGCTTGTTGATAGAGG	
atpF_+20_F	CTTTCGTTTACTTGGGTCACTGG	
atpF_+214_R	CACGCAGTTCTTCTGAATTTCGAATAG	
atpF_C92_PPE_G	/HEX/ATCAATACACCGAACACTACACTTAG	
clpP_+402_F	GGCACAAACGGGAGAATTTA	
clpP_+599_R	CCTATTTTTTATTGAACCGCTACAAG	
clpP_C559_PPE_G	/HEX/CCTATTTTTTATTGAACCGCTACAAG	
matK_+491_F	TGGTTCAAACCCTACGTTACC	
matK_+730_R	TTCGCTCAAAAAGGACTTCA	
matK_C640_PPE_G	/HEX/TGCGTAGAAAAAAAAAGATGGATTCG	
ndhB_+6_F	CTGGCATGTACAGAATGAAAATTTC	
ndhB_+730_R	GCTTGAACCCAATTCCTACAGTG	
ndhB_C149_PPE_C	/HEX/TGGCCTAATTCTTCTTCTGATGATC	
ndhB_C467_PPE_G	CCAGATAATAGGTAGGAGCATAAACTG	
ndhB_C586_PPE_C	TGGGTGGGGCAAGCTCTTCTATTC	
ndhB_+694_F	GCGCTTATATTCATCACTGTAGG	
ndhB_+1525_R	GTATCCTGAGCAATCGCAATAATC	
ndhB_C746_PPE_C	/HEX/CAAGCTTTCCCTAGCCCC	
ndhB_C830_PPE_G	AGGAATATCGAAAATTCGAGTGGCTGAAG	
ndhB_C836_PPE_G	AAGGAATATCGAAAATTCGAGTGG	
ndhB_C872_PPE_G	/HEX/CCAGAAGAAGATGCCATTCATTTG	
ndhB_C1255_PPE_G	/HEX/TGCCTGCCATCCACACCAG	
ndhB_C1481_PPE_G	AGCAATCGCAATAATCGGGTTCATTG	
ndhD_-40_F	GAGTACGCGTTCTTTGGACCTGGTG	
ndhD_+941_R	CCCATATGAGATACAGAAGAATAGGC	
ndhD_C2_PPE_G	/HEX/TTGGAAAAACTACAATTATTGTTAACCAAGG AAAA	

Table 3.S1 (Continued)

ndhD_C383_PPE_C	/HEX/CTGATGTTAGCAATGTACAGCGGTC	RNA Editing analysis
ndhD_C674_PPE_C	/HEX/TGGATTTCTTATTGCTTTTGCCGTC	
ndhD_C878_PPE_C ndhD_C887_PPE_C	AGTCGGTACAATCCAAATAATTTATGCAGC /HEX/CCAAATAATTTATGCAGCTTCAACATCTC	
ndhF_+30_F ndhF_+800_R ndhF_C290_PPE_C	GGATCATACCTTTCATTCCACTTC GCAGCATGTATAAGAGCCGAAATG /HEX/CGGTTACTTTATCGACCCACTTACTTC	
ndhG_+3_F ndhG_+365_R ndhG_C50_PPE_C	GGATTTGCCTGGACCAATAC CCCGTACCATGACGTATC /HEX/TGGACCAATACATGATTTTCTTTTAGTTTTTC	
psbE_+49_F psbE_+252_R psbE_C214_PPE_C	ATTCGATACTGGGTCATTCATAGC CTAAAACGATCTACTAAATTCATCGAG /HEX/CAAGGAATTCCATTAATAACTGGCC	
psbF_+10_F psbF_+117_R psbF_C77_PPE_C	GATAGGACCTATCCAATTTTTACAG CGTTGGATGAACTGCATTGC /HEX/TCATGGACTAGCTGTACCTACC	
petL_-28_F petL_+72_R petL_C5_PPE_C	AAAAAAACATATTTTATTGAGTCCCTTCATG GACCAATAAACAGAACTGAGGTTATAG /HEX/ACATATTTTATTGAGTCCCTTCATGC	
rpl23_+30F rpl23_+260_R rpl23_C89_PPE_G	GACAAAAGTATTCGGTTATTGGGG GGTGGAATAGAATAACCCGG /HEX/ACCCAATGCTTTATTTCTGTCCTAG	
rpoA_+40F rpoA_+538_R rpoA_C200_PPE_G	CAGTGGAAGTGTGTTGAATC AACAGGCATGAATACAGCATC /HEX/TAGAATAGTCATGTGGTACGTTCTC	
rpoB_+65_F rpoB_+792_R rpoB_C338_PPE_C rpoB_C551_PPE_G	GGTTTTATCGGTTTATTGATCAGGG CGGCGACCAATCCTTCCTAATTCAC /HEX/CATTCCTTTAATGAATTCCCTTGGAAC TCTCGTAGATTCAAACCCATAGCTG	
rpoB_+2310_F rpoB_+2732_R rpoB_C2432_PPE_C	TTCAGGTATCGACTTCAAAAGAAAC GGATACTCGGGTTCAAATACCC /HEX/TGGTTCAAGTTATAACCCAGAAATAATTC	
rpoC1_+465_F rpoC1_+1494_R rpoC1_C488_PPE_G	CCTACTTTCTTACGATTACGAGGTT GAACAGCCATTTGATCCCC /HEX/TGGGATGCTGTATTTCCAGGATTG	
rps14_+15_F rps14_+281_R rps14_C80_PPE_C rps14_C149_PPE_G	TTGATTTATAGGGAGAAGAAGAGGC CCTGGCAACAAACATGCCTGAAC ATCAATTGATTCGTCGATCC /HEX/TGTAGGTGCACTATTACGCGG	
rps12_+1_F rps12_+295_R rps12_Ci158_PPE_G	ATGCCAACCATTAAACAACTT GATTGGATTTGCACCAATGGAAACC /HEX/CATTCTATTTTATTCATAGGTACTGATCCG	
psbZ_+16_F psbZ_+165_R psbZ_C50_PPE_C	CAATTGGCAGTTTTTGCATTAATTATTACTTC CCCACCAAGAAGACTAATCCAA /HEX/CAATTGGCAGTTTTTGCATTAATTATTACTTC	
At5g13270_5'_+1726 At5g13270_3'_+2328 WiscLB	GAAATTGCTGGTGAGGAACT TGCAGACTCATTCTTGTTGG AACGTCCGCAATGTGTTATTAAGTTGTC	Geno-typing
At5g13270_5'_-468 At5g13270_3'_+123 LB4	TCTAATGGAGGAGCCATCTC AACAGACGGAGAAGAAGTGC CGTGTGCCAGGTGCCCACGGAATAGT	
GABI-LB	ATAATAACGCTGCGGACATCTACATTTT	

Table 3.S1 (Continued)

Actin_F Actin_R	TGGCATCACACTTTCTACAATGAGCT CGTGGATCCCTGCAGCTTCCATTCC	RT-PCR
At5g13270_RTPCR_F At5g13270_RTPCR_R	AGTCTCGGTTGGAACTCCACTTGTTG GTTTGTGGATGGTGTTTATCACC	

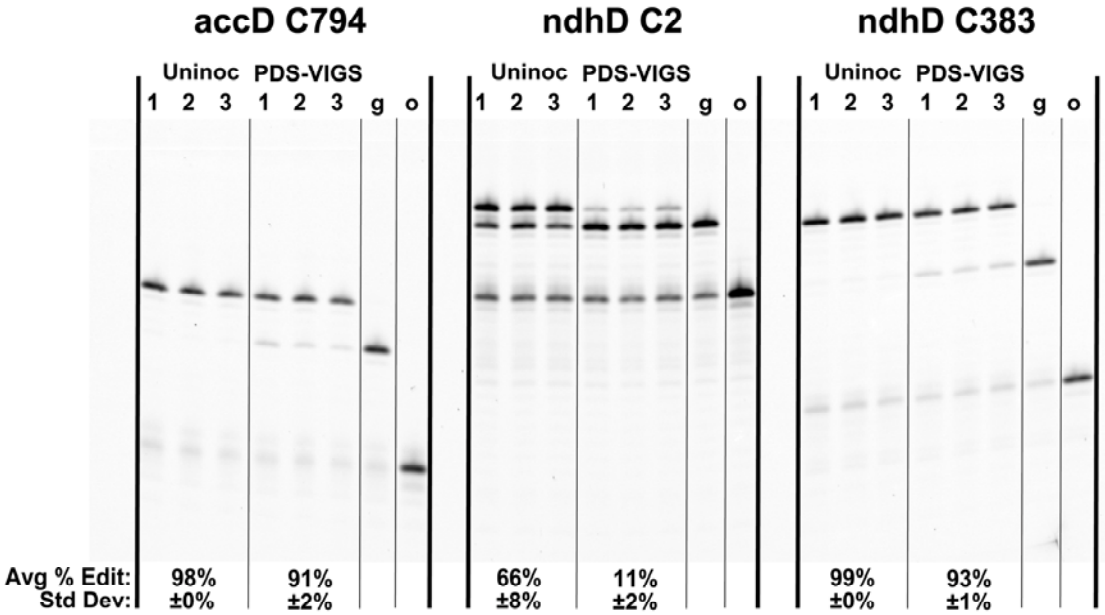
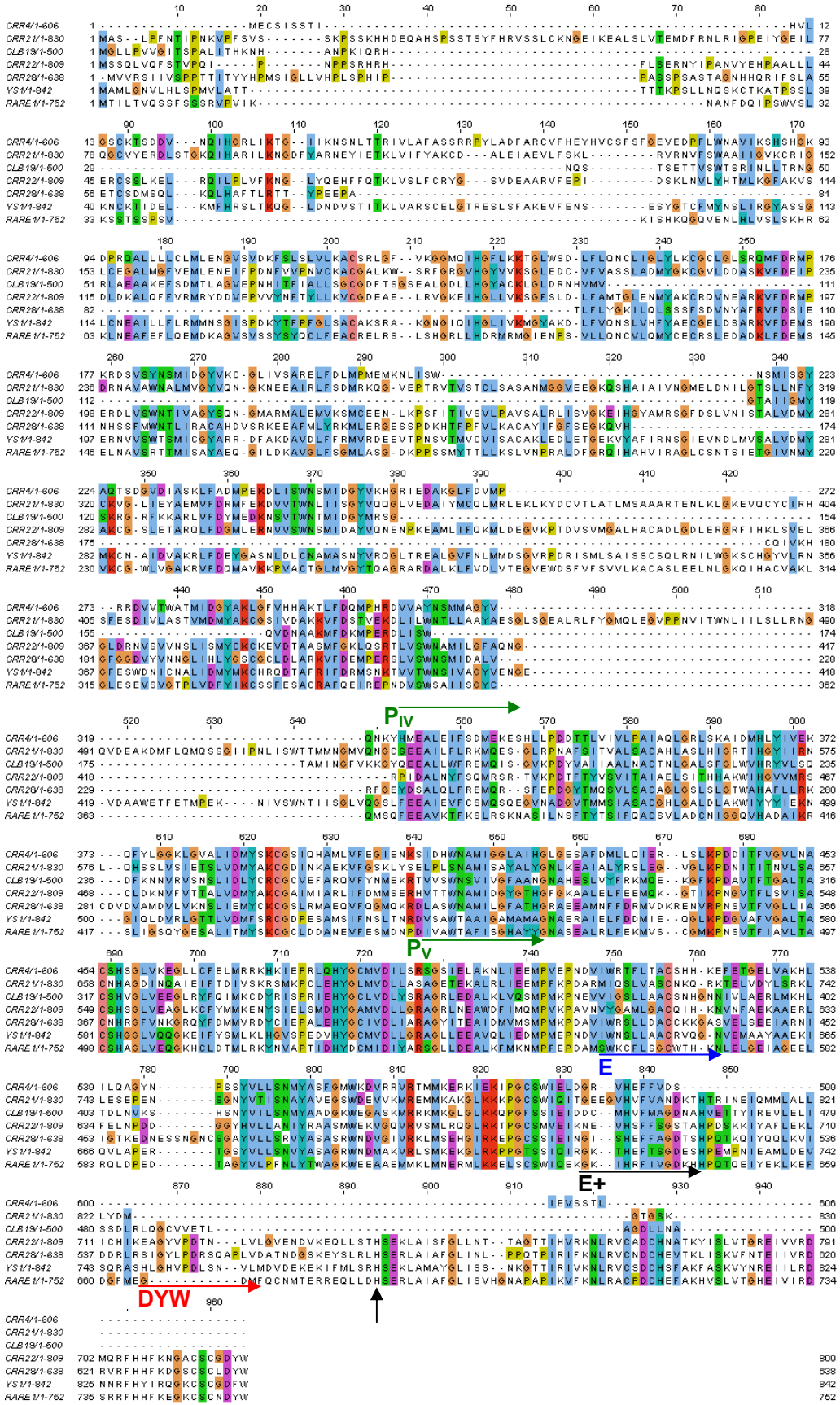
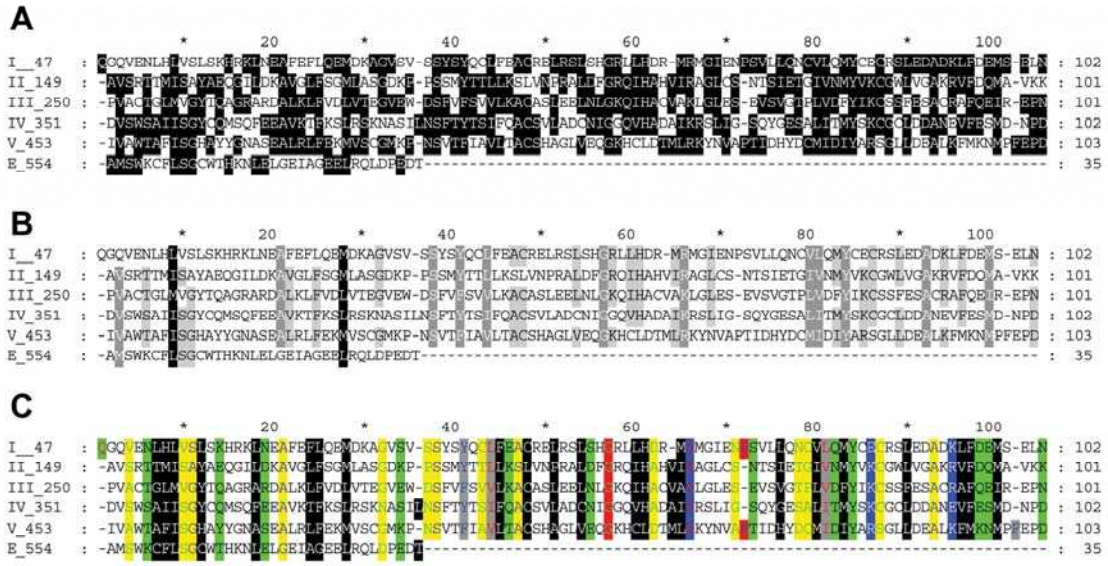


Figure 3.S1. Poisoned primer extension assay of three *Arabidopsis* chloroplast RNA editing events in uninoculated or in plants that have undergone silencing of the gene encoding phytoene desaturase (*PDS*-VIGS). Sizes of extension products for accD C794 are 22 nt (unincorporated oligo), 30 nt (unedited), and 34 nt (edited). Corresponding sizes for ndhD C2 and ndhD C383 editing sites are 35, 42, and 45 nt, and 25, 35, 41 nt, respectively. Control lanes g and o correspond to genomic DNA template and oligonucleotide only, respectively.

Figure 3.S2. Comparative alignments of seven PPR-containing proteins (CRR4, CRR21, CLB19, CRR22, CRR28, YS1 and RARE1) that are known to influence chloroplast RNA C-to-U editing. Published protein model sequences were aligned using T-Coffee v. 5.05 and are presented using Jalview v. 2. Hues reflect the conservation groups of amino acids. Green arrows indicate the first residues of PLS blocks P_I through P_V of RARE1. The blue, black and red arrows indicate the first residue of the E, E+ and DYW domains of RARE1, respectively. The vertical arrow indicates the beginning of the highly conserved region, which continues to the C-terminus, of the DYW domain. Figure contributed by JCR.


PIV
PV
E
E+
DYW



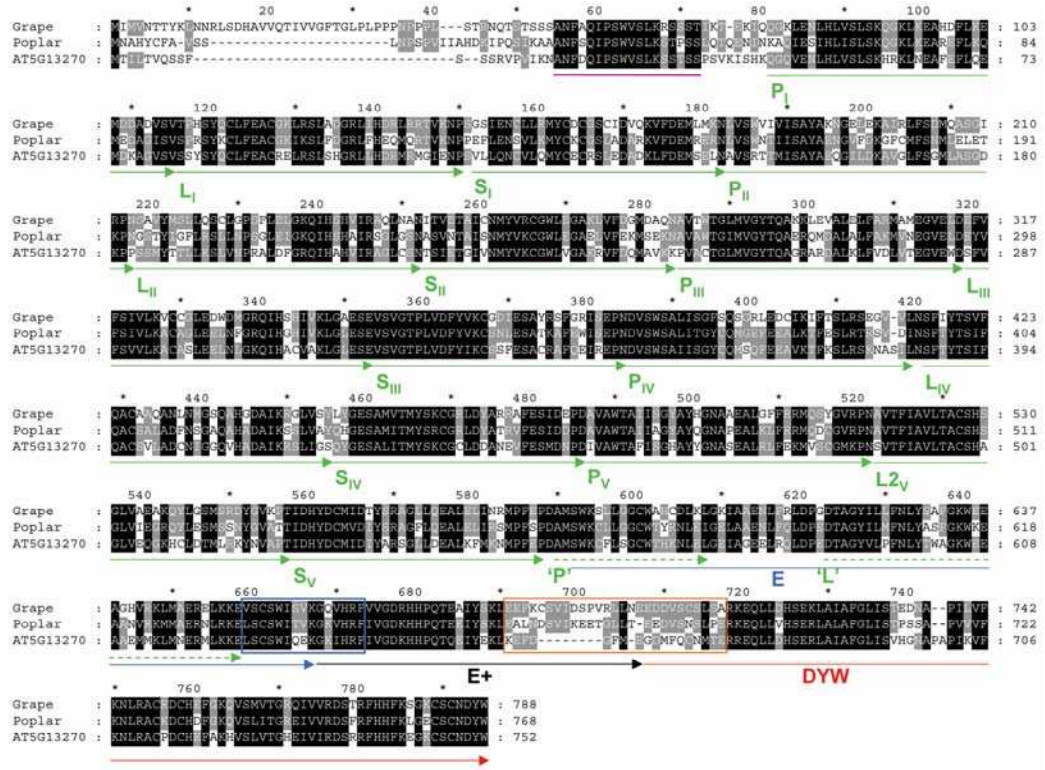
D

	II_149	III_250	IV_351	V_453	
I__47	21%	22%	19%	17%	Identity
	48%	51%	49%	45%	Similarity
II_149		29%	27%	21%	Identity
		54%	48%	43%	Similarity
III_250			26%	24%	Identity
			51%	49%	Similarity
IV_351				25%	Identity
				49%	Similarity

Figure 3.S3. Alignment of PLS blocks and the first 35 residues of the E motif of RARE1. The gene encoding *A. thaliana* RARE1 (gene ID: AT5G13270-TAIR-G) was translated and aligned using T-Coffee Version_7.44. Motifs were classified essentially as described in (1). The PLS block number (I - V) or the E motif and the position of their first residues are indicated on the left margin of each figure. A, Alignments are displayed using GeneDoc employing both the quantified mode and similarity groups. B, Alignments are displayed using GeneDoc employing both the conserved residue shading mode and similarity groups. C, Alignments are displayed using GeneDoc employing the physicochemical mode. Twelve chemical property codes are as follows: blue text on red background: Proline (P), green text on red background: Glycine (G), blue text on yellow background: tiny, green text on yellow background: small, red text on blue background: positive, green text on blue background: negative, white text on blue background: charged, red text on green background: amphoteric, black text on green background: polar, red text on gray background: aliphatic, blue text on gray background: aromatic, white text on black background: hydrophobic. D, Degrees of residue identity and similarity were established by pairwise comparisons of PLS block sequences. Figure contributed by JCR.

Figure 3.S4. Comparison of *A. thaliana* RARE1 with predicted protein sequences of two additional dicotyledons. Orthology was established using reciprocal best hit analysis of protein models. Accession numbers for poplar (*Populus trichocarpa* are AC208482.1 GI:15611965 and for grape (*Vitis vinifera*, Pinot Noir cultivar) are AM451645.2 GI:147770184). Sequences were translated, aligned using T-Coffee Version_7.44, and displayed using GeneDoc. A, Alignment and motif organization of the predicted *A. thaliana* RARE1 and orthologous sequences, displayed with the conserved residue shading mode and similarity groups. RARE1 is a member of the PLS subfamily of PPR proteins with a motif organization of 46-P-L-1-S-P-L-S-P-L-S-P-1-L-S-P-L2-S-4-E-E+-DYW, with gaps of 1-4 aa indicated (1). Commencing with residue 47, each of five tandem PLS blocks (I - V) are of 101 to 103 aa, and is comprised of three tandem repeats. Subscripts indicate a particular PLS block. PLS block motif lengths (underlined in green) are as follows: P: 35 aa; L: 35 aa; L2: 36 aa; S: 31 aa, as described in (1). C-terminal PLS subfamily motif lengths are as follows: E: 76 aa (underlined in blue); E+: 31 aa (underlined in black); DYW: 88 aa (underlined in red), as described in (1). A conserved region immediately upstream of the first PLS block (residues 21 - 46 of RARE1) is indicated (underlined in purple). A conserved 15 amino acid-motif in the E domain (2, 3) is boxed in blue. A relatively unconserved region (*A. thaliana* RARE1 has a 8 residue gap), bridging the E+ and DYW motifs, (residues 656 - 673) is indicated (orange block). Alignments of RARE1 with other *Arabidopsis* PPR proteins with DYW motifs shows this gap to be uncharacteristic of PPR-DYW proteins (data not shown). TargetP, but not Predotar, predicted the grape and poplar N-termini used herein, whereas both TargetP and Predotar predicted the *Arabidopsis* RARE1 N-terminus employed herein. B, Degrees of residue identity and similarity were established by pairwise comparisons of protein model sequences. Figure contributed by JCR.

A



B

	Poplar	AT5G13270	
Grape	64%	57%	Identity
	79%	72%	Similarity
Poplar		60%	Identity
		76%	Similarity

Figure 3.S5. Detailed alignment of protein sequences of the β carboxyltransferase protein sequences from 9 plant and one bacterial species, as described in abbreviated form in Figure 3.8. Cysteinyl zinc ligands (4) are indicated by blue highlighting. Red: position of serine altered to leucine by RNA editing. Green: genomically-encoded leucines. Carboxyltransferase domain according to Pfam is indicated by green underline.

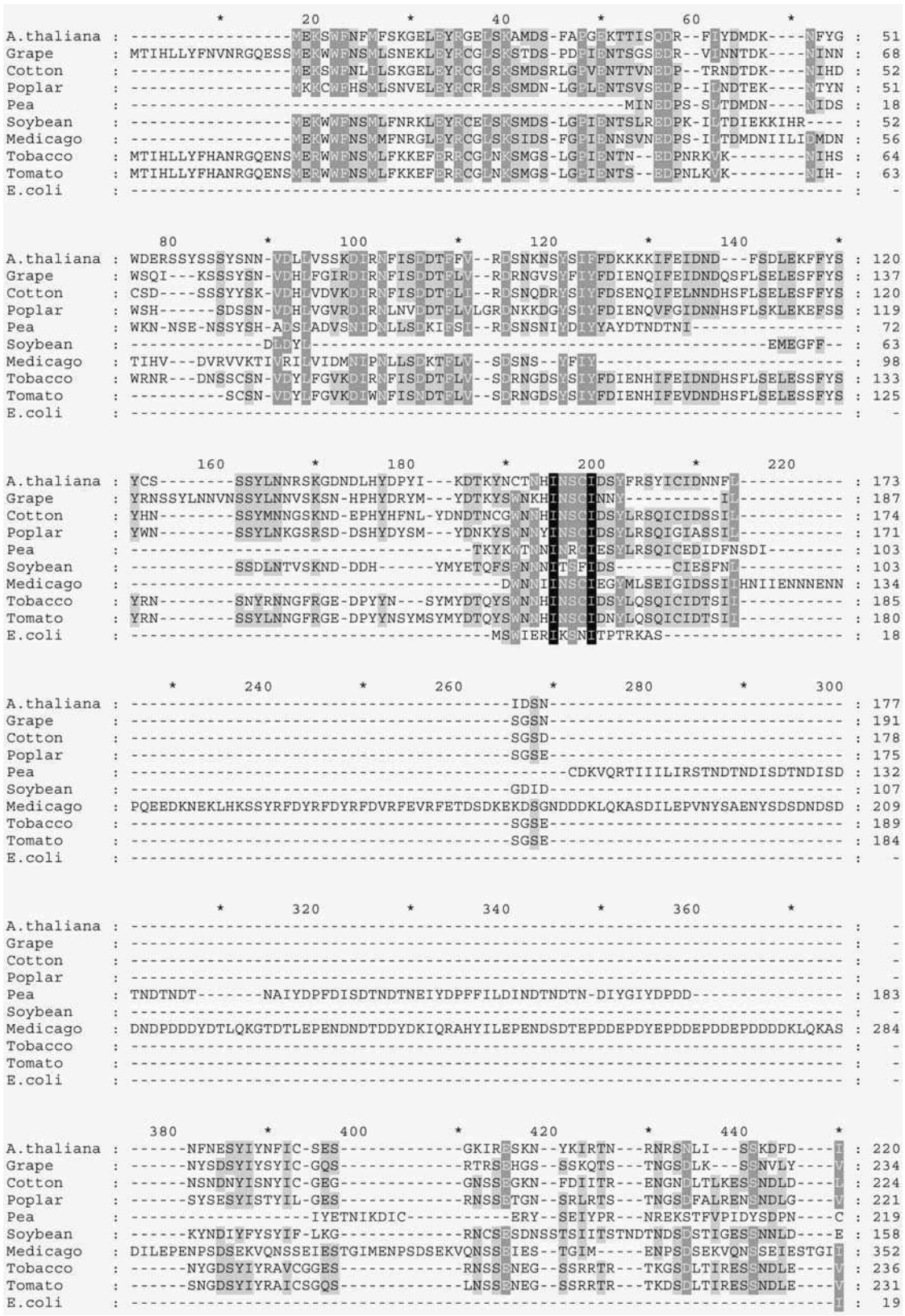
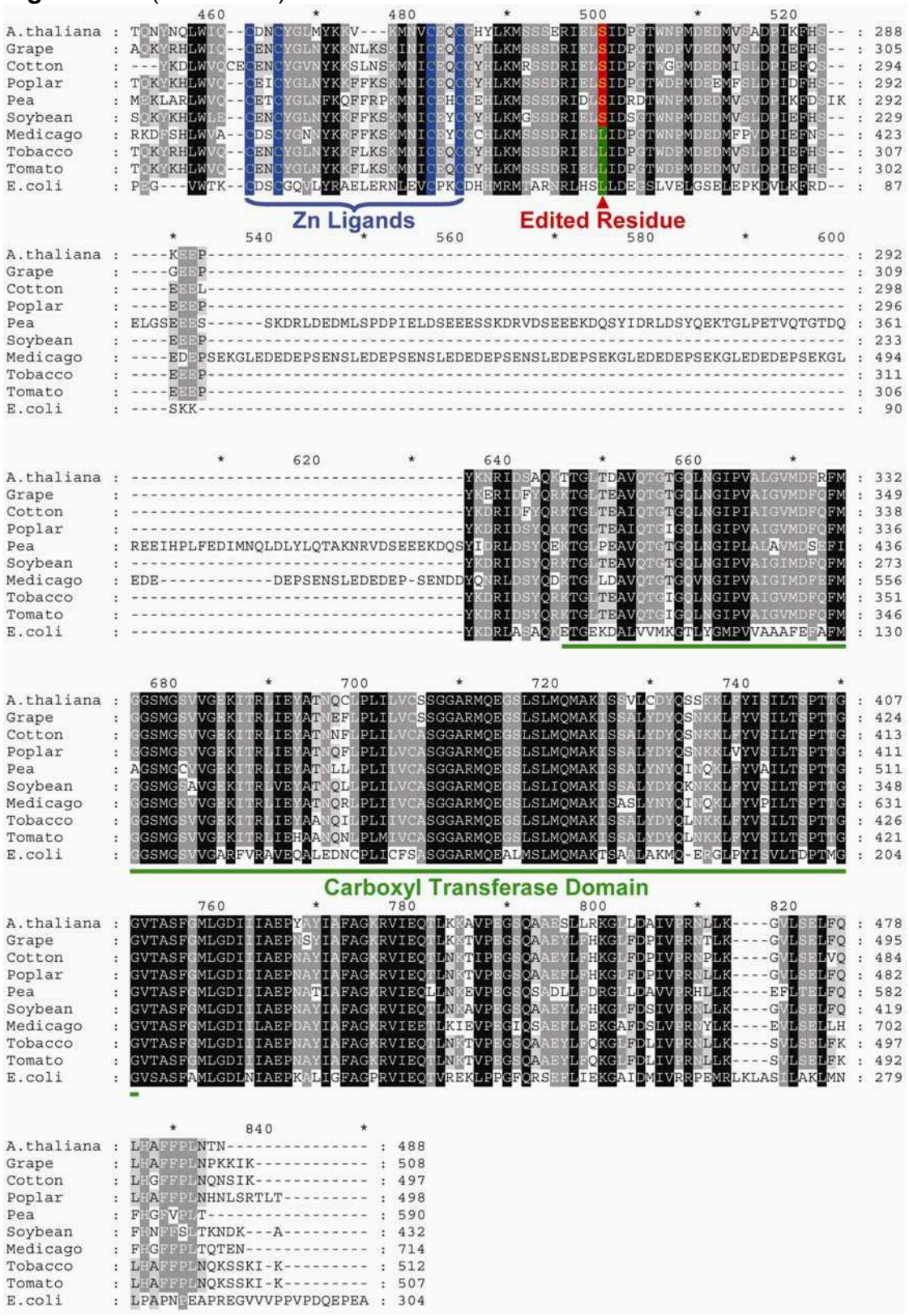

A.thaliana
Grape
Cotton
Poplar
Pea
Soybean
Medicago
Tobacco
Tomato
E.coli

Figure 3.S5 (Continued)



REFERENCES

1. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

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

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

4. Bilder, P., Lightle, S., Bainbridge, G., Ohren, J., Finzel, B., Sun, F., Holley, S., Al-Kassim, L., Spessard, C., Melnick, M. *et al.* (2006) The structure of the carboxyltransferase component of acetyl-coA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme. *Biochemistry*, **45**, 1712-1722.

APPENDIX II

ANALYSIS OF FATTY ACID CONTENT IN *rare1* SEEDS

Due to the lack of an overt phenotype in *rare1* mutant plants, seeds were submitted to Rob Last's Chloroplast 2010 project for a more detailed analysis. No photosynthetic impairment, morphological defect, or significant biochemical difference was noted in homozygous mutant seedlings containing any of the three alleles. However, reproducible differences in fatty acid content of seeds have been detected (Figure AII.1). For this analysis, plants were grown by me, genotyped by PCR (primers listed in Chapter 3), and mature seeds were collected and sent to Michigan State University where total seed fatty acid methyl esters were prepared for GC/MS and analyzed by Imad Ajjawi (Rob Last Lab). For the two *rare1* null editing alleles, WiscDsLox330H10 and GABI167A04, fatty acids 18:0, 18:1Δ9, and 18:2 were decreased relative to wild-type, and 18:3, 20:1, and 22:1 were increased (Figure AII.1A). Mutants with the weaker FLAG_424E06 promoter insertion allele display a weaker fatty acid profile alteration (Figure AII.1B). Only 18:1Δ9 is decreased and only 18:3 is increased relative to wild-type. This is not surprising, as the concentration of these two fatty acids were most affected in the two *rare1* null allele mutants.

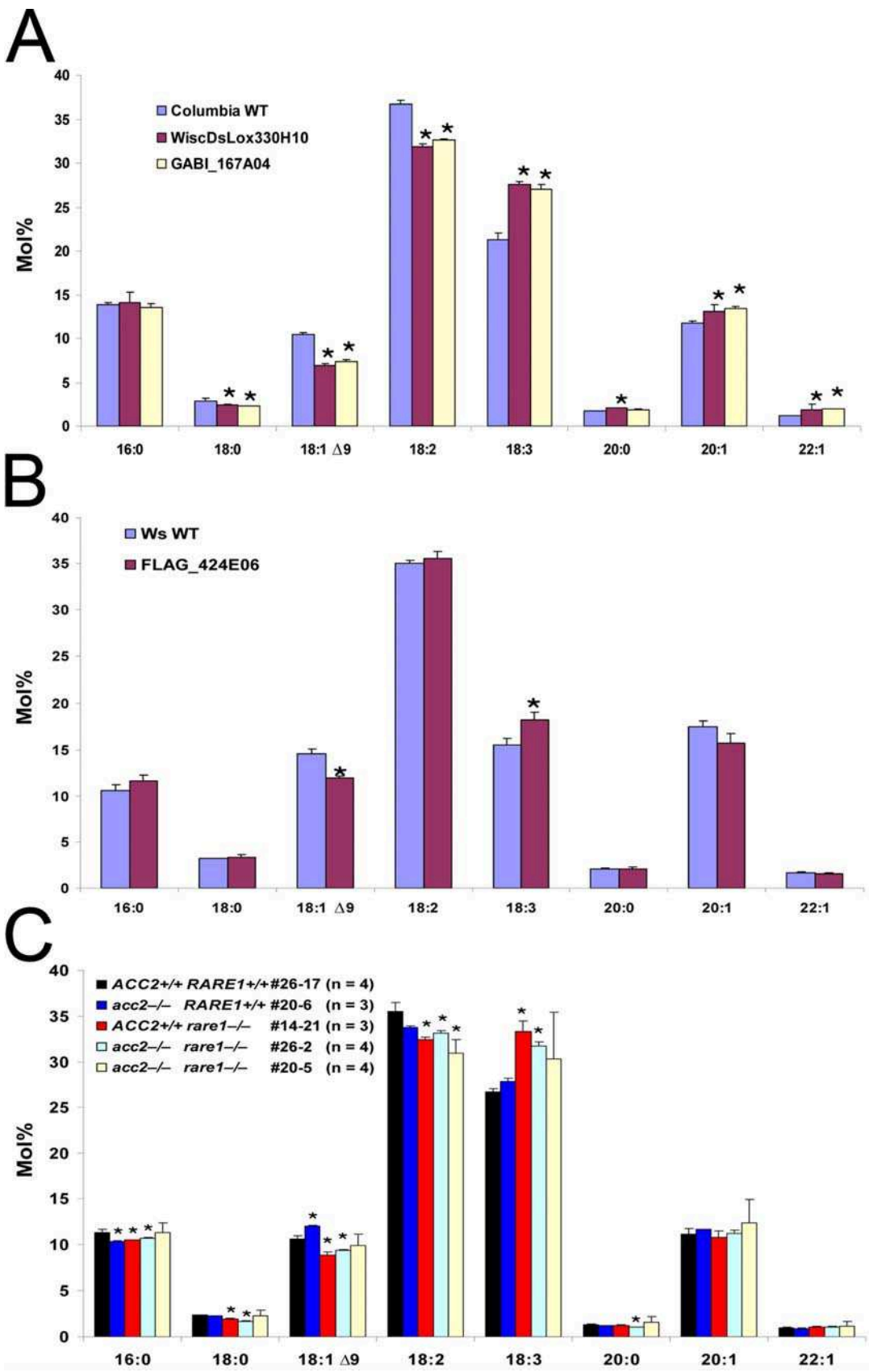
Sasaki et al. (1) have shown that pea carboxyltransferase requires RNA editing for activity *in vitro*; however, Arabidopsis *rare1* mutants evidently do not have an absolute requirement of editing in order to synthesize fatty acids *in vivo*. Given that RNA editing influences activity of the pea enzyme, it is unsurprising that *rare1* mutants display an altered fatty acid profile relative to

wild-type plants. However, it is difficult to correlate the actual differences observed directly to diminished acetyl-coA carboxylase (ACCase) activity alone. Since ACCase catalyzes the synthesis of malonyl-coA, an early intermediate in fatty acid biosynthesis, all fatty acids synthesized by chloroplasts should be affected in ACCase mutants. The fatty acid profiles of *rare1* mutants indicate that, although there are reproducible decreases or increases of particular fatty acids, fatty acid synthesis as a whole is not strongly affected in *rare1* mutants, and that the particular differences observed are possibly being regulated by some component other than ACCase itself. Nonetheless, the lack of leucine codon 265 restoration in *accD* is the only defect known to be exhibited by *rare1* mutants, and this lack of editing does influence Arabidopsis fatty acid biosynthesis to some degree.

An alternative possibility for robust growth of *rare1* mutants explored in (2) is the presence of the *ACC2* gene (At3g36180), which encodes a heteromeric-type ACCase with a putative plastid transit peptide and could possibly be providing ACCase activity in *rare1* mutants regardless of the *accD* editing defect. To test this, *acc2* (Salk_148966 allele) / *rare1* (WiscDsLox330H10) double mutants were generated and seed fatty acid content was analyzed as above (Figure AII.1C). Primers for amplifying *ACC2* wild type allele were 5'-ATCACACACTTCAAAGACGGG and 5'-CCATTCGCAAACACCCAATG, and the T-DNA was amplified with primers 5'-TTTTGCCGATTTCGGAAC and 5'-CCATTCGCAAACACCCAATG. *acc2* single mutants did not display a fatty acid profile alteration (the 18:1Δ9 change in plant 20-6 was not observed in a replicate plant of the same genotype, data not shown), and the *acc2*/*rare1* double mutants display a fatty acid phenotype similar to *rare1* single mutants. Taken together, these results indicates that

ACC2 does not compensate for heteromeric ACCase activity in *rare1* mutants, nor does it appear play a major role in chloroplast fatty acid biosynthesis in general.

Figure AII.1. Fatty acid methyl ester analysis of *rare1* and *rare1*/*acc2* seeds. A, Molar ratios of various fatty acid methyl esters in wild-type, WiscDsLox330H10 and GABI_167A04 *rare1* mutant seeds. Asterisks (*) indicate statistically significant change from wild type, $P<0.02$, n=4 for wild-type and WiscDsLox330H10, and n=3 for GABI_167A04. B, As in (A) except for FLAG_424E06 allele, $P<0.02$, and n=4 for both samples. C, As in (A) except for *acc2* (Salk_148966) / *rare1* (WiscDsLox330H10) crosses. $P<0.01$, and number of replicates is indicated in figure.


A
Columbia WT
WiscDsLox330H10
GABI_167A04
Mol%
16:0
18:0
18:1 Δ9
18:2
18:3
20:0
20:1
22:1
B
Ws WT
FLAG_424E06
Mol%
16:0
18:0
18:1 Δ9
18:2
18:3
20:0
20:1
22:1
C
ACC2+/+ RARE1+/+ #26-17 (n = 4)
acc2–/– RARE1+/+ #20-6 (n = 3)
ACC2+/+ rare1–/– #14-21 (n = 3)
acc2–/– rare1–/– #26-2 (n = 4)
acc2–/– rare1–/– #20-5 (n = 4)
Mol%
16:0
18:0
18:1 Δ9
18:2
18:3
20:0
20:1
22:1

REFERENCES

1. Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H. and Nagano, Y. (2001) Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem*, **276**, 3937-3940.

2. Robbins, J.C., Heller, W.P. and Hanson, M.R. (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.

APPENDIX III

EXPRESSION OF RECOMBINANT ARABIDOPSIS α- AND β-CARBOXYLTRANSFERASE

Previously, recombinant pea carboxyltransferase containing either edited or unedited versions of the β subunit has shown that only complexes containing the edited form have activity *in vitro* (1). Additionally, acetyl co-A carboxylase, of which carboxyltransferase is a component, is an essential enzyme in dicots, and copy number reduction of the chloroplast *accD* gene (i.e., generation of heteroplasmic mutants) causes severe developmental defects (2). Contrary to this result, Arabidopsis *rare1* mutant plants, which fail to edit accD C794 and produce unedited β-carboxyltransferase, are robust and display no visible phenotype (3), although minor fatty acid profile differences have been detected in seeds (Appendix II).

An alternative non-isotopic assay for carboxyltransferase activity has been developed, which monitors the reverse enzymatic reaction, i.e., the formation of acetyl-coA from malonyl co-A (4). In order to determine the enzymatic activity of Arabidopsis carboxyltransferase containing edited and unedited β subunits, this assay could be used with purified recombinant enzyme of both forms.

MATERIALS AND METHODS

The Arabidopsis genes encoding α- and β-carboxyltransferase, *accA* (At2g38040) and *accD* (AtCg00500), respectively, were cloned by PCR

(Accuprime Taq, Invitrogen). Unedited *accD* was amplified from genomic DNA, and edited *accD* was amplified from random primed cDNA prepared with the Sensiscript RT Kit (Qiagen). For both, the PCR primers were 5'-CATATGATGGAAAAATCGTGGTTCAA and 5'-CTCGAGTTAATTTGTGTTCAAAGGAA. *accA* cDNA was prepared using the Superscript III Kit (Invitrogen) and primer 5'-GGCGAAGCTGGGGTTA, followed by PCR with this primer and 5'-GGATCCATGTCTCGGCTCAAGAAAGGGAAG, which begins after the sequence encoding a TargetP predicted chloroplast transit peptide of 54 aa. Following sequence verification, *accA* and *accD* were subcloned into pDEST14 (no tag) or pDEST17 (N-terminal His tag) using LR Clonase II (Invitrogen). Coexpression vectors were made by ligating the NaeI-ScaI fragment of pDEST14-accD (both edited and unedited) to the 5.5 kb fragment of ScaI-EcoRV cut pDEST17-accA. The final constructs were transformed into *E.coli* Rosetta (DE3) cells and protein expression in log phase cultures was induced with 1mM IPTG for 16 hours at 25 degrees C. Proteins were purified with Ni-NTA Agarose (Qiagen) according to the manufacturers recommended protocol. Fractions were mixed with LDS-Sample Buffer and electrophoresed in NuPAGE 4-12% Bis-Tris gels (Invitrogen) using MES-SDS Running Buffer according the manufacturer's suggested protocol.

RESULTS AND DISCUSSION

Soluble His-tagged recombinant Arabidopsis α-carboxyltransferase was purified on Ni-NTA agarose (Figure AIII.1). Optimization of NaCl concentration in the wash buffer (500mM, Figure AIII.1B) resulted in a higher level of purity

compared to the standard recommended protocol (Figure AIII.1A). β-carboxyltransferase expressed alone was not soluble with N-terminal His-tag (Figure AIII.1A) or untagged (data not shown).

The expression and purification scheme used by Sasaki et al. (1) was to coexpress both subunits form a single operon in *E.coli* placing the His-tag sequence on the N-terminus of α-carboxyltransferase, and leaving β-carboxyltransferase untagged. A similar construct was made for expressing the Arabidopsis proteins, as described above; the expression and purification are shown in Figure AIII.1C. Unfortunately, no β-carboxyltransferase (apparent MW of around 60 kDa anticipated, see Figure AIII.1A) was produced, probably due to recombination within the co-expression vector. The promoter and terminator regions for the pDEST14-accD vectors and pDEST17-accA are identical, and in the *E. coli* host strain, recombination events between the duplicated regions are possible. A way to avoid such events is to generate new coexpression constructs that place both genes in the same operon so that duplication is the promoter and terminator is not necessary for expression.

Figure AIII.1. Attempts to express and purify recombinant Arabidopsis α- and β-carboxyltransferase from *E. coli*. A, Purification fractions from 6xHis-tagged α- and β-carboxyltransferase. E1-E4 indicate successive elution fractions. B, Purification of α- carboxyltransferase using optimized (500mM) NaCl concentration. C, Failed attempt to solubilize and purify untagged β-carboxyltransferase using 6xHis-tagged α-carboxyltransferase.

A
α-CT
β-CT
Lysate
Unbound
E1
E2
E3
E4
M
Lysate
E1
Pellet
α
β
250
130
95
72
55
36
28
17
B
M
Lysate
Wash 1
E1
E2
E3
E4
250
130
95
72
55
36
28
C
Lysate
Unbound
Wash 1
Wash 2
E1
E2
E3
E4
M
170
130
100
70
55
40
35
25

REFERENCES

1. Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H. and Nagano, Y. (2001) Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem*, **276**, 3937-3940.

2. Kode, V., Mudd, E.A., Iamtham, S. and Day, A. (2005) The tobacco plastid accD gene is essential and is required for leaf development. *Plant J*, **44**, 237-244.

3. Robbins, J.C., Heller, W.P. and Hanson, M.R. (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.

4. Liu, X., Fortin, P.D. and Walsh, C.T. (2008) Andrimid producers encode an acetyl-CoA carboxyltransferase subunit resistant to the action of the antibiotic. *Proc Natl Acad Sci U S A*, **105**, 13321-13326.

APPENDIX IV

RARE1 LOCALIZATION STUDY

In order to confirm the finding in Lurin et al. (1) that RARE1 contains a chloroplast transit peptide, the first 100 codons of *RARE1* were cloned by PCR using primers 5'-TCCATCAACTATGACGATTCTCACTGT and 5'-AGATAAAGACCTCAATTCTCTGCAAG into pCR8/GW/TOPO (Invitrogen). This sequence was recombined into vector pEARLEYGATE103 (2) to generate a C-terminal GFP fusion construct. The resulting plasmid was electrotransformed into *Agrobacterium tumefaciens* stain GV3101 and agro-infiltration of *Nicotiana benthamiana* seedlings was performed. GFP- and chlorophyll auto-fluorescence within mesophyll cells were observed three days post-infiltration. GFP signal was observed as punctuate spots within chloroplasts, with 1 to 3 spots per chloroplast (Figure AIV.1). As mentioned in (3) the first PPR repeat of RARE1 begins at the 47th amino acid. If transit peptide cleavage occurs N-terminal to the first PPR motif, the resulting processed fusion protein in the infiltrated plants would contain about one and a half PPR motifs. Since PPR motifs bind nucleic acid, one possible explanation for the punctuate pattern observed in the infiltrated plants is the formation of aggregates by virtue of RARE1's N-terminal PPR motifs binding to cellular RNA or DNA; another possibility is that overexpression of the fusion protein leads to the formation of insoluble aggregates caused by 'stickiness' of the PPR motifs.

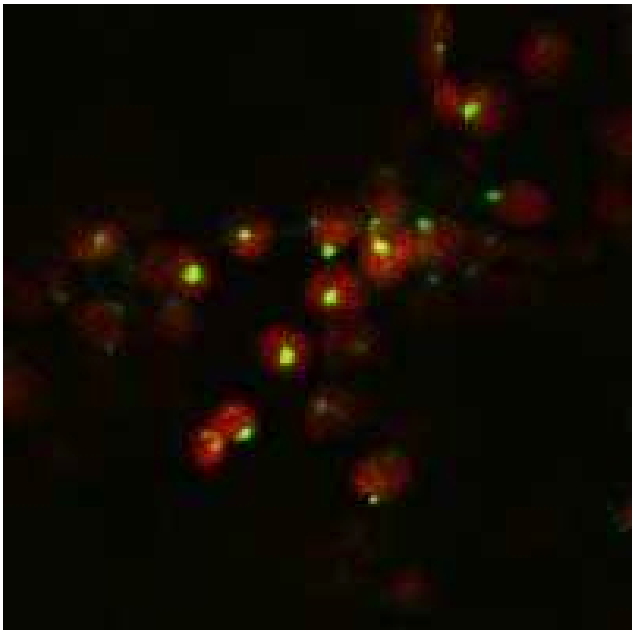


FIGURE AIV.1. RARE1 transit peptide fused to GFP localizes to punctuate spots within *N. benthamiana* chloroplasts. Agro-infiltrated leaves were observed by confocal microscopy three days post-infiltration. Red, chlorophyll autofluorescence. Green, GFP from RARE1-GFP fusion protein.

REFERENCES

1. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

2. Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*, **45**, 616-629..

3. Robbins, J.C., Heller, W.P. and Hanson, M.R. (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.