STUDIES ON *TRANS*-ACTING FACTORS INVOLVED IN CHLOROPLAST RNA EDITING IN *ARABIDOPSIS THALIANA*

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

Jessica Ann Wagoner

February 2016

STUDIES ON *TRANS*-ACTING FACTORS INVOLVED IN CHLOROPLAST RNA EDITING IN *ARABIDOPSIS THALIANA*

Jessica Ann Wagoner, Ph. D.

Cornell University 2016

RNA editing is a post-transcriptional process that changes specific cytidines to uridines in the organelles of most land plants. Although members of the editosome have begun to be elucidated, the identity of the deaminase is still unknown. Members of the pentatricopeptide repeat (PPR) protein family are known to be the site-recognition factors in RNA editing. Most PPR editing factors also carry extra C-terminal motifs called the E domain and the DYW domain, where the DYW domain features amino acids conserved with the active site in other cytidine deaminases. I examined the DYW motifs of two chloroplast PPR-DYW editing factors: REQUIRED FOR *ACCD* RNA EDITING 1 (RARE1), which is responsible for editing a single site in the *accD* transcript, and QUINTUPLE EDITING FACTOR 1 (QED1), which is responsible for the editing of five plastid sites. Truncation of the DYW motifs greatly decreased the editing efficiency of RARE1 and QED1 when transfected into protoplasts. Using site-directed mutagenesis, the conserved deaminase residues were targeted along with a highly conserved potential second zinc binding site. The mutant proteins could not efficiently complement the RNA editing phenotype when expressed in mutant protoplasts, indicating that these deaminase residues may be important for RNA editing efficiency for certain PPR editing factors. Since QED1 affects five different chloroplast sites including sites that changes the coded protein sequences of *matK*, *ndhB*, and *rpoB*, I further examined the functional consequences of these sites not being edited in *qed1* mutants. *qed1* mutants feature a stunted growth phenotype, the presence of yellow patches that increases with age, a lack of NA(P)H

dehydrogenase (NDH) activity, and an abnormal chloroplast morphology as well as a smaller average cell size which can be observed in mutant protoplasts. These phenotypes were found to be at least partially complemented using *qed1* mutant lines which have a stably integrated QED1 with a 3xFLAG-StrepII tag attached to its N-terminus. The line carrying one stably integrated allele, though not able to complement the mutant phenotypes to wild-type levels, was able to partially complement the mutant phenotypes, likely due to the presence of only one allele rather than two.

BIOGRAPHICAL SKETCH

Jessica Ann Wagoner was born on November 1, 1985 to Steve and Tina Wagoner in Dayton, Ohio. She attended St. Albert the Great grade school and then Archbishop Alter High School before heading to Columbus to study Molecular Genetics at The Ohio State University. She graduated with a B.S. in Molecular Genetics as well as a degree in Italian in 2008 and began graduate school at Cornell University where she joined the Molecular Biology and Genetics department's field of Genetics, Genomics, and Development where she subsequently joined Maureen Hanson's lab.

This thesis is dedicated to my family, friends, and colleagues who supported me in the completion of this endeavor.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Maureen Hanson for the opportunity to work in her lab, for her guidance and suggestions, and especially for her continuing support and patience. I would also like to thank my committee members David Stern and Ailong Ke for their suggestions and help and their support and patience as well. Finally, I would like to thank my family and friends for their continuing support through my graduate experience and especially during the last portion to finish, especially through difficulty. I couldn't have done it without my family, friends, advisor, committee and colleagues.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1:

Introduction

RNA editing is a co- or post-transcriptional process that changes the RNA sequence from what was originally encoded in the genome. This phenomenon was first observed in the mitochondrial transcripts of trypanosomes where the insertion of four uridines was found to restore the reading frame of a frameshifted *coxII* gene (Benne et al., 1986). Since then, RNA editing has been found to occur in the tRNAs, mRNAs, rRNAs, and microRNAs of a wide variety of organisms, including viruses, yeast, plants, and humans (reviewed in Chateigner-Boutin and Small, 2011). Two major forms of RNA editing have evolved: insertion/deletion editing, and substitution editing which most commonly converts adenosine (A) to inosine (I) or cytidine (C) to uridine (U) (reviewed in Chateigner-Boutin and Small, 2011).

RNA EDITING IN PLANTS

In land plants, RNA editing was first discovered in the *coxII* gene in wheat mitochondria when differences were observed between the generated cDNA sequence and what was encoded in the genomic DNA (Covello and Gray, 1989). Soon after, RNA editing was also observed in the chloroplast when it was found that a C to U conversion created a start codon in maize plastid *rpl2* transcripts (Hoch et al., 1991). Since then, RNA editing has been found to occur in the majority of plant lineages (bryophytes, gymnosperms, angiosperms) with the exception of Marchantia liverworts (Steinhauser et al., 1999; Sugita et al., 2006; Takenaka et al., 2008; Wolf et al., 2004). Because RNA editing is not observed in algae or cyanobacteria, the RNA editing process in land plants appears to have evolved independently from the RNA editing processes that are observed in other organisms (Hiesel et al., 1994; Malek et al., 1996).

Many RNA editing events in land plants involve the conversion of specific cytidines to

uridines (C-toU) in the protein-coding regions of organelle mRNAs. However, C-to-U editing is also known to occur in non-coding RNA sequences such as tRNAs, introns, and 5' and 3' UTRs but not in rRNAs (Binder et al., 1994; Börner et al., 1995; Kudla and Bock, 1999; Maréchal-Drouard et al., 1993). Plants are not limited to C-to-U editing however; U-to-C changes have also been observed in mosses and hornworts (Grewe et al., 2009; Kugita et al., 2003; Oldenkott et al., 2014; Yoshinaga et al., 1996, 1997). Adenosine-to-inosine editing can also occur in nuclear and plastid tRNAs and recently the editing enzyme has been identified for this process (Zhou et al., 2013, 2014). Recently evidence of C-to-U editing of nuclear encoded tRNAs has been observed in plants, indicating that C-to-U editing may not be strictly limited to the organelles, although it is unlikely that this editing process proceeds in the same manner as organellar C-to-U editing (Zhou et al., 2014).

Unlike C-to-U RNA editing of apolipoprotein B in mammals, which serves to create functionally diverse proteins from one transcript, C-to-U editing in plant organelles appears to have evolved as a correction mechanism, as editing tends to restore a codon to evolutionary consensus (Covello and Gray, 1989; Koito and Ikeda, 2012; Maier et al., 1992, 1992). RNA editing can also create start codons as well as stop codons (Hoch et al., 1991; Kudla et al., 1992; Neckermann et al., 1994; Wintz and Hanson, 1991). Some experiments have demonstrated that changing a nucleotide in a codon from T to C results in impaired activity of the affected protein. For instance, tobacco, which genomically encodes a T at a particular location in plastid *psbF*, was modified to match the corresponding spinach *psbF* site which is normally edited. No editing of that site occurred in the tobacco mutant and the resulting plants were impaired in growth and photosynthesis (Bock et al., 1994). The *petB* gene, which is edited in maize, was introduced in the algae *Chlamydomonas reinhardtii*, which does not edit its RNA, therefore naturally

containing a genomically encoded T at the homologous position. The resulting transformed cells, containing a C instead of a T, were non-photosynthetic (Zito et al., 1997).

The number of editing sites in plant species can range from zero, such as in Marchantia liverworts, to very few, like in the moss *Physcomitrella patens* containing only two plastid editing sites and eleven mitochondrial editing sites, to 1000-1500 mitochondrial editing sites and over 3400 plastid editing sites in lycophytes, which include firmosses, clubmosses, and spikemosses (Grewe et al., 2009; Ohyama, 1996; Oldenkott et al., 2014; Rüdinger et al., 2009). Arabidopsis has 43 editing sites in the plastid and over 600 editing sites in the mitochondria (Bentolila et al., 2008, 2013a; Ruwe et al., 2013). Though editing in the organelles appears to occur through a similar process, the inability of a chloroplast RNA sequence to be edited in the mitochondria indicates that organelle-specific processes have evolved (Sutton et al., 1995; Zeltz et al., 1996).

Other experiments revealed that land plant RNA editing is a post-transcriptional process, opposed to the co-transcriptional RNA insertion editing found in the slime mold *Physarium polycephalum* (Gott et al., 1993; Visomirski-Robic and Gott, 1995, 1997), because partially edited transcripts have been observed and exogenous synthesized RNAs added to editing competent extracts are able to be edited (Hayes and Hanson, 2007; Lu and Hanson, 1992; Neuwirt et al., 2005). RNA editing was found to be independent of other RNA processing steps as well; for example, it can occur before the splicing of organelle mRNAs (Freyer et al., 1993; Sutton et al., 1991).

Chloroplast RNA editing events are thought to affect the stability of the encoded protein. In the cases of petunia *atp6* and wheat *atp9*, despite observation of partially edited transcripts, only proteins from fully edited transcripts accumulate, even though transcripts in various editing

states have been associated with polysomes (Bégu et al., 1990; Graves et al., 1990; Lu and Hanson, 1994, 1996; Williams et al., 1998). However, using antibodies and mass spectrometry, proteins from unedited transcripts were observed to accumulate in the case of the mitochondrial *rps12* gene (Lu et al., 1996; Phreaner et al., 1996). Editing efficiencies in the plastid tends to be high while there is more variation in editing efficiency in the mitochondria. However, editing efficiency can be tissue dependent, particularly for plastid *ndh* genes, where the high editing efficiency observed in leaves drops drastically in non-green tissues; also, editing efficiency can be developmentally dependent. For example, the extent of RNA editing in the mitochondrial *nad3* gene increases as the development of maize seedlings progressed (Chateigner-Boutin and Hanson, 2003; Grosskopf and Mulligan, 1996; Peeters and Hanson, 2002). Interestingly, editing efficiency can also be affected by growth conditions or environmental stress, though there is currently no evidence that this variation can result in functionally diverse proteins in response to these stresses (Karcher and Bock, 1998, 2002; Nakajima and Mulligan, 2001).

The actual C to U conversion is known to proceed via a deamination reaction rather than a base excision or a nucleotide replacement (Blanc et al., 1995; Rajasekhar and Mulligan, 1993). The alpha-phosphate in cytidine was radioactively labelled and incorporated into mRNA sequences in intact mitochondria. After RNA purification, the RNA was digested to mononucleotides and analyzed by one-dimensional thin-layer chromatography where some of the radioactive signal was observed to co-migrate with uridine monophosphates (Rajasekhar and Mulligan, 1993). This retention of the alpha-phosphate indicated that plant RNA editing does not occur via base excision or replacement as the radioactive signal would not have been observed to co-migrate with uridine monophosphates in these situations. Thus the results of these experiments suggested that the editing enzyme is a cytidine deaminase. In later

experiments, the addition of zinc chelators to editing competent extracts was found to inhibit editing, supporting the role of a zinc-dependent cytidine deaminase in plant RNA editing (Hegeman et al., 2005). However, the known cytidine deaminases encoded in the Arabidopsis genome do not seem to be involved in organellar RNA editing, leaving the exact factors involved in the deamination reaction a mystery (Faivre-Nitschke et al., 1999)

CURRENT MECHANISTIC MODEL: RNA EDITING INVOLVES *CIS*-ELEMENTS AND MULTIPLE *TRANS*-FACTORS

The simple view of plant RNA editing holds that editing occurs through the interaction of sequence elements encoded around the edited C target, known as *cis*-elements, and nuclear-encoded protein *trans*-factors (Figure 1.1). Using petunia lines which exhibited "high-editing extent" or "low-editing extent" of the mitochondrial *nad3* transcript due to an unknown nuclear factor and using petunia lines which contained *nad3* in different transcriptional contexts (transcribed with an abnormal *pcf* gene in cytoplasmic male-sterile (CMS) lines versus wild-type lines where this *pcf* gene is not present), it was observed that selection of C-targets depended on the sequence surrounding the C to be edited and was not dependent on nuclear genotype or transcriptional context (Lu and Hanson, 1992; Wilson and Hanson, 1996). In certain CMS lines missing a 10 nt sequence 5' of a *nad3* editing site, that particular *nad3* site is never edited, although another editing site 5' of the deletion is always edited in both lines, providing further evidence that site selection depends on the sequence immediately around a C-target (Wilson and Hanson, 1996). Subsequent work to define the sequence around the C-target important for editing which was termed the '*cis*-element' was initially performed using transformation of

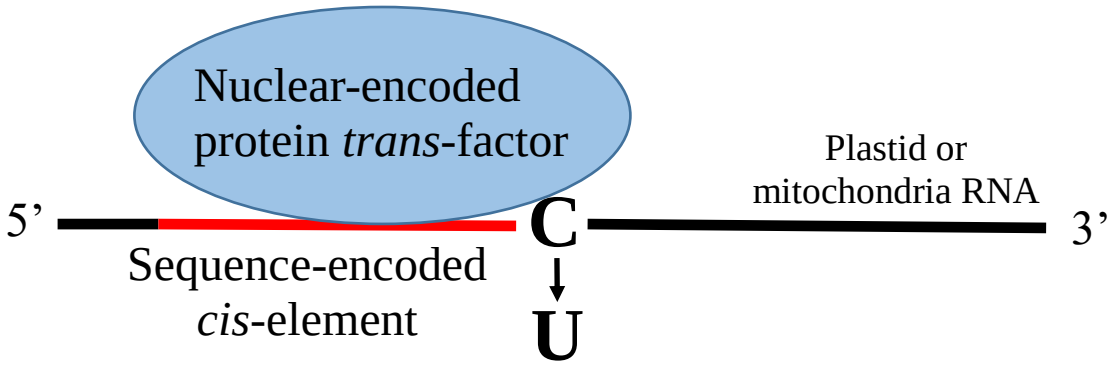


Figure 1.1: Simple model of RNA editing. Nuclear-encoded protein sequence recognition *trans*-factors interact with RNA sequence (termed the *cis*-element) encoded around the targeted C to perform the editing reaction in plastid and mitochondrial RNAs

tobacco plastids with transgenes that expressed different lengths of RNA sequences surrounding the editing site of interest. Later, expression of exogenous RNAs in competent editing extracts allowed further dissection of *cis*-elements, identifying important residues through mutational analysis (Bock et al., 1996, 1997; Chateigner-Boutin and Hanson, 2002; Chaudhuri and Maliga, 1996; Farré et al., 2001; Hayes and Hanson, 2007; Hayes et al., 2006; Reed et al., 2001; Verbitskiy et al., 2008)

In general, around 20-25 nt upstream and 5 nt downstream of the C-target is sufficient for site recognition. For example, only 16 nt upstream and 5 nt downstream of the *psbL* C2 target in tobacco is necessary for editing and about -12/+5 nt around a site in tobacco *ndhB* was sufficient for editing of that site (Bock et al., 1996; Chaudhuri and Maliga, 1996). -20/+6 nt around a C-target in tobacco *rpoB* was also sufficient to specify the target site, although in these cases, a greater amount of sequence was necessary to obtain editing efficiencies near wild-type level (Hayes et al., 2006). Some variation does exist in the length of these *cis*-elements required for editing, though. For example, two tobacco *ndhB* sites were not edited even when transformed

with 42 nt upstream and 42 nt downstream sequences present (Bock et al., 1996). Mutating specific nucleotides or blocks of nucleotides helped to further define *cis*-elements of particular editing sites and indicate specific nucleotides within the *cis*-element that were particularly important for efficient editing function. These studies indicated that most of the sequence important for editing lies 5' of the editing site, though a few nucleotides 3' of the edited C also appeared to have some importance in the efficiency of the editing reaction (Hayes et al., 2006; Heller et al., 2008; Miyamoto et al., 2002; Sasaki et al., 2006). Interestingly, *in vitro* experiments in tobacco assaying the editing extent of chimeric RNAs where one RNA had the 5' region of *psbL* and the 3' region of *ndhB* and the other RNA had the 5' region of *ndhB* and the 3' region of *psbL* resulted in greatly reduced editing of these RNAs compared to wild type *psbL* and *ndhB* RNAs, indicating that editing extent also relies on the proper combination of 5' and 3' sequences around the edited C (Hirose and Sugiura, 2001).

EARLY WORK ON RNA EDITING PROTEIN *TRANS*-FACTORS

Work on the *cis*-elements of editing sites also indirectly provided information about the then unidentified *trans*-factors involved in plant RNA editing. These *trans*-factors were determined to be site-specific when the over-expression of exogenous mRNA carrying a *psbL* editing site caused editing at the corresponding endogenous site to be decreased but did not affect editing frequency of other editing sites (Chaudhuri et al., 1995). To test if RNA was a component in these site-specific *trans*-factors, chloroplast extracts were pre-treated with micrococcal nuclease, and it was observed that the editing activity was not affected even when the nuclease was in high concentrations suggesting that RNA, like guide RNA for instance, was not a component in these site-specific *trans*-factors. AMT (4′-aminomethyl-4,5′,8-trimethyl psoralen) was also used to attempt to cross-link any RNAs potentially interacting with *psbL*

mRNAs, but these attempts resulted only in negative data. These micrococcal nuclease and AMT cross-linking assays indicated that the site-specific *trans*-factors were likely to be proteins and not RNAs (Hirose and Sugiura, 2001). Using UV cross-linking experiments, the site-specificity factors were confirmed to be proteins after the organelle extract was found to contain proteins that were able to bind upstream of editing sites (Hirose and Sugiura, 2001). A 56 kD protein was identified as a site-specific factor for the *psbE* editing site in tobacco when mRNA substrates containing the editing site were incubated in editing competent tobacco extract and subjected to UV cross-linking. This 56 kD protein was not found and editing did not occur when the same mRNA substrate was incubated in pea extract where the corresponding *psbE* site is not normally edited and instead a T is encoded in the genome (Miyamoto et al., 2002). A 70 kD protein was also found to bind to a *petB* editing site in tobacco using the same methods. When this tobacco *petB* mRNA was incubated in pea extract and subjected to UV cross-linking, the mRNA was edited and a 70 kD protein was cross-linked to the RNA. In this case, the corresponding *petB* site is also edited in pea, and these results indicated that *trans*-factors and editing sites likely co-evolve (Miyamoto et al., 2002).

Though the *cis*-elements of C-targets generally do not exhibit much sequence consensus when compared to other editing sites within the same species, *cis*-elements tend to be conserved amongst species that edit a specific site and there are examples of *cis*-elements within a species that are similar in sequence to each other. When a portion of *rpoB* RNA containing a particular editing site was over-expressed in tobacco chloroplasts, not only was editing at the endogenous *rpoB* site decreased, editing at four other sites was reduced; two of these sites were strongly affected and the editing of the other two sites was affected albeit less severely. The *cis*-elements of these sites were found to have sequence similarities to that of the targeted *rpoB* editing site's

cis-element, where the *cis*-elements of the sites that were more drastically reduced had greater consensus to the *rpoB cis*-element than the sites that were less affected (Chateigner-Boutin and Hanson, 2002). Cross-competition was also observed between maize plastid editing sites *rpoB* C467 and *rps14* C80 when exogenous RNAs carrying these editing sites were added to editing competent extracts in excess, causing reduced editing at not only the endogenous site but also the other site. Mutation of these *cis*-elements, which also had sequence similarities, helped identify nucleotides critical for efficient editing (Heller et al., 2008). The *cis*-elements of tobacco *ndhF* C290 and *ndhB* C1481 also have moderate sequence identity and UV cross-linking experiments revealed that the same 95 kD protein bound to both *cis*-elements (Kobayashi et al., 2008). All of this evidence together indicated that *trans*-factors can sometimes recognize multiple *cis*-elements.

THE PPR PROTEIN FAMILY: SEQUENCE RECOGNITION FACTORS IN RNA EDITING

The first evidence that these *trans*-factors were encoded in the nucleus and not the plastid came from the observation that C-targets in maize and barley *rpoB* were still able to be edited in plastid ribosome-deficient plants (Zeltz et al., 1993). Editing *trans*-factors were confirmed to be nuclear-encoded was when it was found that all plastid editing sites were able to be edited efficiently in *iojap* mutant maize that completely lacked plastid ribosomes (Halter et al., 2004). However, the identity of these *trans*-factors was elusive until the discovery of *CHLORORESPIRATORY REDUCTION 4* (CRR4) through an EMS (ethyl methanesulfonate) mutagenesis screen aiming to identify plants deficient in NDH (NAD(P)H dehydrogenase) activity. CRR4 was discovered to be essential in the editing of the second C in plastid *ndhD*, effectively creating a start codon (Kotera et al., 2005).

CRR4 is a member of the pentatricopeptide repeat (PPR) family which is defined by a degenerate 35 amino acid motif that tends to be organized as tandem repeats within a protein (Small and Peeters, 2000). PPR proteins are present in most organisms; the first PPR protein that was identified was the yeast protein Pet309 which participates in translation of mitochondria COX1 (Manthey and McEwen, 1995; Small and Peeters, 2000; Tavares-Carreón et al., 2008). Though most organisms only have a handful of PPR proteins, this family has significantly expanded in land plants, with 458 members in Arabidopsis compared to only 39 in

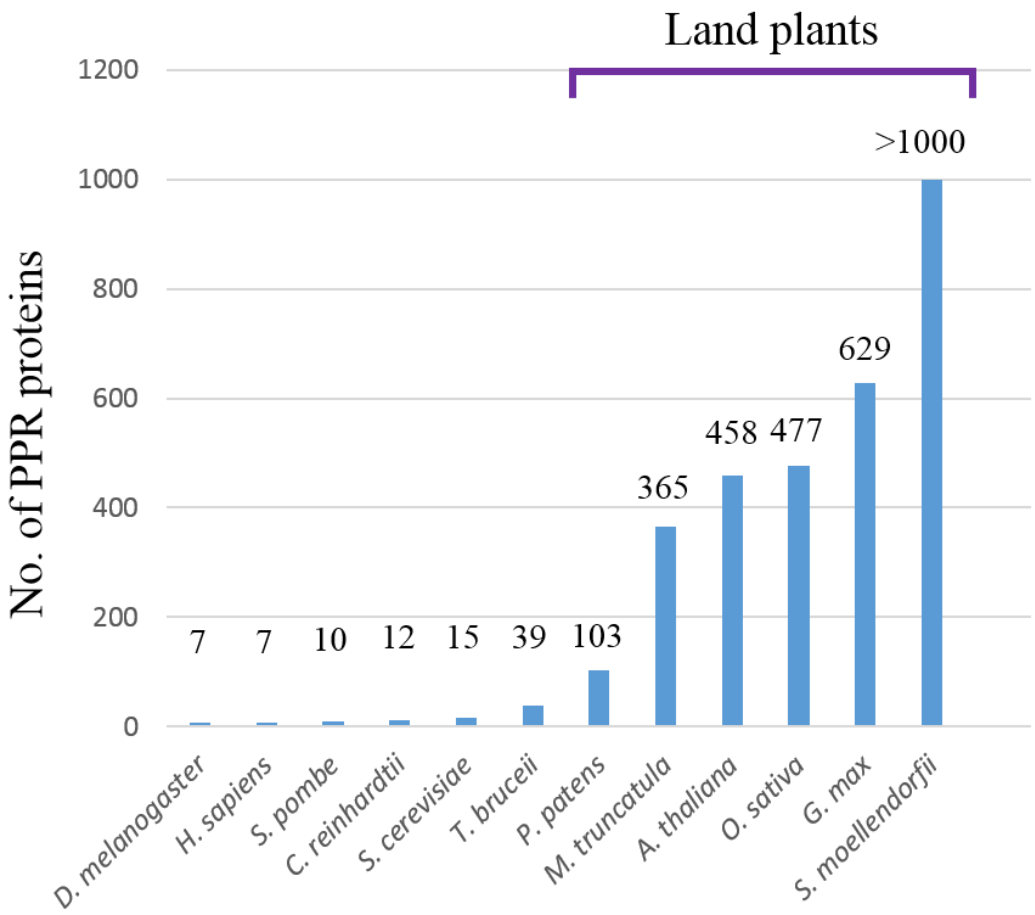


Figure 1.2: Number of PPR proteins found in a variety of species. Land plants are indicated by the purple bar. *D. melanogaster* (fly) , *H. sapiens* (human), *S. pombe* (fission yeast), *C. reinhardtii* (algae), *S. cerevisiae* (budding yeast), *T. brucei* (trypanosomes), *P. patens* (moss), *M. truncatula*, *A. thaliana* (thale cress), *O. sativa* (rice), *G. max* (soybean), *S. moellendorfii* (spikemoss)

trypanosomes, (Figure 1.2) (Aphasizhev and Aphasizheva, 2013; Aubourg et al., 2000; Baggio et al., 2014; Colcombet et al., 2013; Fujii and Small, 2011; Herbert et al., 2013; Lightowlers and Chrzanowska-Lightowlers, 2013; O'Toole et al., 2008; Pusnik et al., 2007). In angiosperms specifically, the PPR gene content appears fairly similar in number across species as orthologous pairs are formed by more than 80% of rice and Arabidopsis PPR proteins (O'Toole et al., 2008). This expansion of the PPR family in plants is most likely due to the role of PPR proteins in nuclear and organellar gene expression (Chateigner-Boutin and Small, 2011; O'Toole et al., 2008). Indeed, the vast majority of PPR proteins in plants are targeted to either organelle and though a few are reported to be found outside of the organelles, the physiological consequence of this localization is not clear (Colcombet et al., 2013; Ding et al., 2006; Hammani et al., 2011).

In plants, the PPR protein family can be divided into two subclasses: the P class which is defined by the presence of the canonical 35 amino acid motif and the PLS class which not only has the 35 aa motif (P) but also up to 38 aa long (L) motifs and 32-34aa short (S) motifs, usually organized in P-L-S blocks within the protein (Lurin et al., 2004). The PLS class is further characterized by the presence of additional C-terminal motifs; most PLS class proteins have what is termed the "E motif" and many of these E motif proteins carry an additional DYW motif, named for the three amino acids found at the end of this motif (Figure 1.3) (Lurin et al., 2004). Members of the PPR protein family serve a variety of functions in many organellar RNA metabolism processes such as transcription, splicing, cleavage, stabilization, translation and editing (Figure 1.3) (reviewed. in Barkan and Small, 2014). P-class PPR proteins participate in many of these RNA metabolic processes but PLS class PPR proteins appear to be primarily involved in RNA editing, as all PPR proteins known to be involved in RNA editing are members of the PLS class (Barkan and Small, 2014). However, there are a few examples of PLS class

PPR proteins that do not function in editing: the PLS class proteins CRR2 and OTP70 are involved in RNA processing and splicing, respectively (Chateigner-Boutin et al., 2011; Hashimoto et al., 2003). In rare cases, P class proteins have been shown to influence editing extent but this could be a secondary effect from changes in RNA turnover or accessibility (Doniwa et al., 2010).

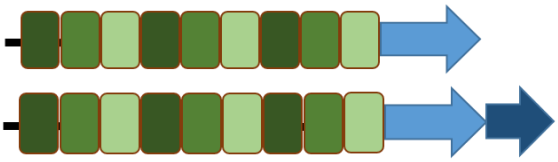
A.

P-class PPR protein

Motifs	Functions	Examples
P ~35 aa	RNA stabilization	CRP1, PPR10, HCF152, MCA1
	RNA splicing	THA8, PPR4, PPR5, OTP51
	RNA cleavage	RPF1, RPF2, RPF3, RPF5
	Transcription	pTAC2, DG1
	Translation	PPR10, PGR3, CRP1

B.

PLS-class PPR protein



Motifs	Functions	Examples
P ~35 aa L ~36-38 aa S ~32-34 aa	RNA editing	CRR4, CRR21, RARE1, QED1, OTP80, OTP82
	RNA splicing	OTP70
E/E+ DYW	RNA cleavage	CRR2, PDM1

Figure 1.3: Architecture and function of P and PLS class PPR proteins. **A.** P-class proteins consist of ~35 aa (P) motifs often present as tandem repeats and function in RNA stabilization, splicing, cleavage, transcription and translation. **B.** PLS-class proteins consist of ~35 aa P motifs, ~36-38 aa L motifs, and ~32-34 aa S motifs often organized as repeating PLS blocks with S motifs occasionally interspersed and often have additional C-terminal E or E-DYW motifs. PLS PPR proteins mostly function in RNA editing though a few proteins function in RNA splicing and cleavage.

PPR PROTEIN STRUCTURE AND FUNCTION: THE PPR CODE

The PPR protein family likely evolved from the closely related tetratricopeptide repeat (TPR) family which is known to mediate protein-protein interactions (Blatch and Lässle, 1999). Early structural predictions of PPR proteins based on known TPR motif structure suggested that, like TPR motifs, PPR motifs consisted of two antiparallel alpha-helices and the tandem organization of the motifs would result in a binding groove (Jínek et al., 2004; Small and Peeters, 2000). However, based on the properties of the residues predicted to project into the groove, PPR proteins were hypothesized to be RNA binding proteins instead of protein-binding proteins like their TPR protein counterparts (Small and Peeters, 2000). Subsequent experiments using gel shift assays, co-immunoprecipitation, and *in vitro* binding assays with recombinant proteins proved this RNA binding hypothesis to be true (reviewed in Nakamura et al., 2012). For example, CRR4 was found to bind the RNA near the site that it edited, which suggested that editing PPR proteins acted as the site-specificity factors. Furthermore, CRR4 lacking its C-terminal motifs was still able to bind RNA, indicating that interaction was mediated through the PPR motifs but not the C-terminal motifs (Okuda et al., 2006a). Since the discovery of CRR4, other PPR proteins involved in RNA editing have been identified, some of which are responsible for more than one editing site (Chateigner-Boutin et al., 2008; Wagoner et al., 2015; Zhou et al., 2009). In many cases, the *cis*-elements of editing sites under the influence of a single PPR factor shared significant sequence similarity, consistent with what was observed in earlier experiments involving *cis*-elements of editing sites, supporting the hypothesis that PPR proteins serve as site-specificity factors in RNA editing (Hammani et al., 2009; Okuda et al., 2009).

Despite strong evidence that PPR proteins function as RNA binding proteins with the PPR motifs mediating the protein-RNA interaction, the mode by which the PPR motifs

recognized specific RNA nucleotides was unknown until a combinatorial amino acid code by Barkan et al (2012) was published (Figure 1.4). The PPR code dictates that the particular combination of amino acids at position 6 of one PPR motif and the first amino acid of the following PPR motif (1') work in conjunction to preferentially recognize certain RNA nucleotides (Figure 1.4A); a tandem arrangement of multiple PPR pairs give rise to a sequence specific recognition of a ssRNA. For example, a threonine (T) at position 6 and an asparagine (N) at position 1' will preferentially recognize an A in the RNA, whereas an asparagine (N) at position 6 and an aspartic acid at position 1' will preferentially recognize a U (Barkan et al., 2012). Using this combinatorial code, P class protein PPR10 was successfully recoded to bind RNA sequences other than its natural target (Barkan et al., 2012). When this code was tested with PLS class proteins, while L motifs could not be incorporated into the code as it was then developed, it was observed that a conserved five nucleotide space between the edited C and the beginning of protein binding was present, probably to accommodate the additional C-terminal motifs (Barkan et al., 2012). Further study on the L motifs and the final S motif in PLS class proteins served to refine the combinatorial code and improve target site recognition for editing factors (Takenaka et al., 2013). Indeed, this refined code enabled a new mitochondrial editing factor, MEF32, to be identified based on the binding sites predicted by the PPR motifs of an uncharacterized protein which was then confirmed through mutant analysis (Takenaka et al., 2013). A three amino acid code was developed that indicated that amino acids at position 1, 4 (position 6 in the two-amino acid code) and ii (position 1' in the two-amino acid code) were responsible for target site selection (Figure 1.4B) (Yagi et al., 2013). Though it appeared that positions 4 and ii (6 and 1') were the more important amino acids in determining binding sites, the addition of the 1 position amino acid again improved predictions for binding sites. This three

amino acid code was also used to find the RNA editing site for which PPR protein abscisic acid hypersensitive germination

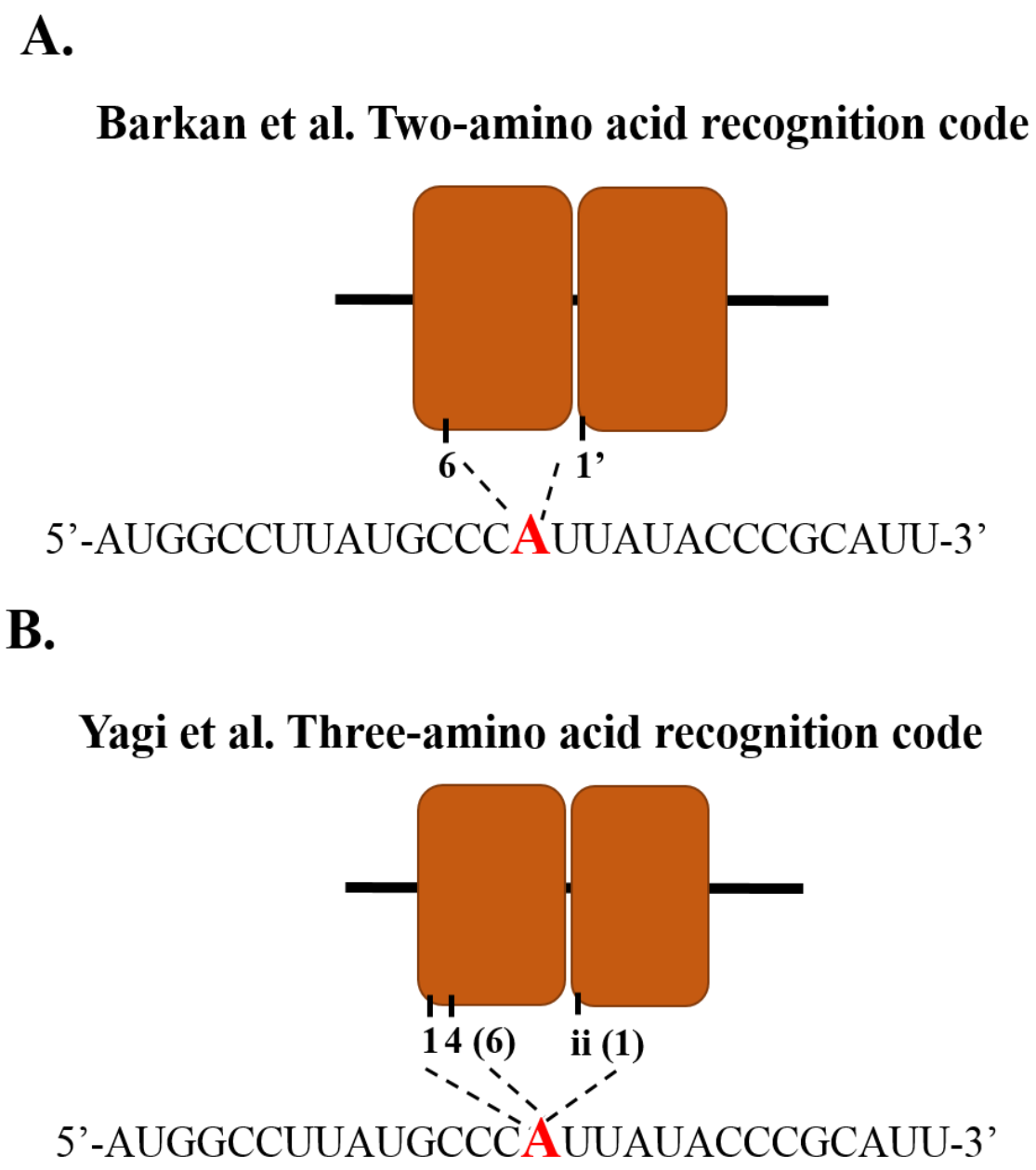


Figure 1.4: Schematics of the two- and three-amino acid PPR codes **A.** The Barkan et al two-amino acid recognition code holds that particular RNA nucleotides are recognized by the sixth amino acid in a PPR motif (6) and the first amino acid of the next PPR motif (1'). **B.** The Yagi et al three-amino acid code consists of particular RNA nucleotides recognized by the first (1) and fourth (4) amino acids of one PPR motif and what is termed the ii amino acid in the next PPR motif. The 4 and ii amino acids of the three-amino acid code correspond to the 6 and 1' amino acids of the two- amino acid code.

11 (AHG11) was responsible (Yagi et al., 2013).

Recently, due to the PPR code, the protein sequence of a PPR protein was able to be adjusted to predictably alter the binding specificity of that protein and a synthetic PPR protein has been created and characterized; both of these accomplishments have implications for designing site-specific RNA binding proteins for practical use via post-transcriptional control of target RNAs in the future (Gully et al., 2015; Kindgren et al., 2015).

The information gained from the elucidation of the combinatorial amino acid PPR code is consistent with the early experiments on the *cis*-elements of RNA editing sites as the code confirms that the PPR proteins bind to the sequence 5' of the C-target. The code also explains the results from the earlier experiments showing that editing sites with similar *cis*-elements were recognized by the same *trans*-factor (Bock et al., 1996; Chateigner-Boutin and Hanson, 2002; Heller et al., 2008). Similarly, the elucidation of the RNA binding targets is congruent with results from previous experiments which showed that the actual editing status of the C-target does not affect the binding affinity of the *trans*-factor to the site it recognizes; for instance, RNA from *ndhF* transgenes which encoded a T at the editing site still competed with endogenous *ndhF* for *trans*-factors at the same rate as unedited transgenes (Reed et al., 2001). However, the code does not explain cases like the two *ndhB* tobacco sites which are not edited in chloroplast minigenes even when 47 nt upstream and 47 nt downstream of the editing sites are included, nor does it explain the requirement for sequence 3' of the editing site for efficient editing (Bock et al., 1996). Possibly other factors with RNA binding capabilities besides the PPR proteins are required to achieve full editing efficiency. The RNA sequence predicted to be recognized by some PPR editing factors like MEF1, MEF14 and CRR28 do not match their target sequence very well, suggesting that further refinement of the code may be necessary (Takenaka et al.,

2013).

Potential refinement of the PPR code could be gleaned from results obtained from crystal structures of PPR proteins bound to its cognate RNA. Despite the tendency of PPR proteins to be insoluble, recently the crystal structure of a few PPR proteins, such as P-class protein PPR10 and PLS-class protein THA8L, have been solved in both the apo and RNA-bound forms (Ban et al., 2013; Gully et al., 2015; Ke et al., 2013; Yin et al., 2013). The resolved crystal structures confirmed much of what had been predicted from related TPR protein structures such as the alpha-solenoid tertiary structure and the modular recognition of RNA residues predicted by the bioinformatic analyses that resulted in the PPR code. Analysis of PPR10 bound to its RNA confirmed that particular amino acids actually contacted the RNA residues they were previously predicted to contact via the PPR code, validating the importance of the sixth residue in a protein repeat in recognizing specific RNA residues with contributions from two other amino acids (Ban et al., 2013; Gully et al., 2015b; Yin et al., 2013). Structural studies of PPR10 also offered some new insights. If the first amino acid of a helix is considered the start of the repeat, the currently established PPR motif boundaries would be shifted either one amino acid upstream or downstream, resulting in a motif that encompasses all residues potentially engaging in RNA residue recognition (Yin et al., 2013).

C-TERMINAL MOTIFS OF EDITING PPR PROTEINS: THE E AND DYW MOTIFS

Though the evidence clearly indicates that the PPR motifs in PPR editing factors function to specifically select the sites to be edited, the role in the editing reaction of the C-terminal motifs is still unclear. All PPR editing factors are members of the PLS class and also contain the E motif, and most of these proteins contain an additional E+ or E+ and DYW motif (Lurin et al., 2004). Truncation experiments determined that the E motif is always essential for editing as

editing was lost at the sites controlled by E-class PPR editing factors CRR4 and CRR21 when CRR4 and CRR21 sans E motifs were transformed into their respective mutant plants to test for complementation (Okuda et al., 2007). The same result was observed when the E and DYW motifs were removed from DYW class editing factors CRR28 and CRR22, although editing still occurred when the DYW motif only was removed (Okuda et al., 2009). E motifs also could be swapped between CRR4 and CRR21 and the E-DYW motifs could be swapped between CRR22 and CRR28 with no adverse effect on editing, indicating that the E motif serves a general role in plastid RNA editing instead of a role dependent on specific editing sites (Okuda et al., 2007, 2009). E motifs between two mitochondrial editing factors could also be exchanged with no effect on editing (Chateigner-Boutin et al., 2013). However, when E motifs were exchanged between mitochondrial editing factors and plastid editing factors, resulting in chimeric proteins that could recognize their specific editing sites, but had the E motif from a PPR editing factor that functioned in the other organelle, editing of their specific target editing sites did not occur (Chateigner-Boutin et al., 2013). Though the E motifs appear to have a general function in RNA editing, some organelle-specific specialization appears to have evolved. Chateigner-Boutin et al (2013) noted that mitochondrial PPR editing factors tend to have longer E motifs than plastid editing factors, but no specific sequence elements were found to be particularly conserved in mitochondrial editing factor E motifs versus plastid editing E motifs (Chateigner-Boutin et al., 2013).

The function of the E motifs which render them critical for RNA editing function is still unknown. Though E motifs between editing factors tend to be more conserved than the PPR motifs, no amino acids are absolutely conserved amongst all of the E motifs and there are no recognizable known domains found within the E motif besides what appears to be two

particularly degenerate PPR motifs (Lurin et al., 2004; Takenaka et al., 2013). However, amongst the five known plastid editing PPR-E proteins there appears to be some conservation around what was designated the PGC box by Hayes et al (2013), and in a block of amino acids 5' of the PGC box (Figure 1.5) (Hayes et al., 2013). Given the conserved spacing between the PPR binding site and the C target, the E motif could serve as a spacer between the PPR motifs and the factor that performs the editing reaction, but the C-target/binding sequence spacing appears to be well-conserved in both organelles and does not seem to reflect the length of the E motif, which may indicate the E motif has a more active role in the RNA editing process than as a spacer (Chateigner-Boutin et al., 2013; Takenaka et al., 2013).

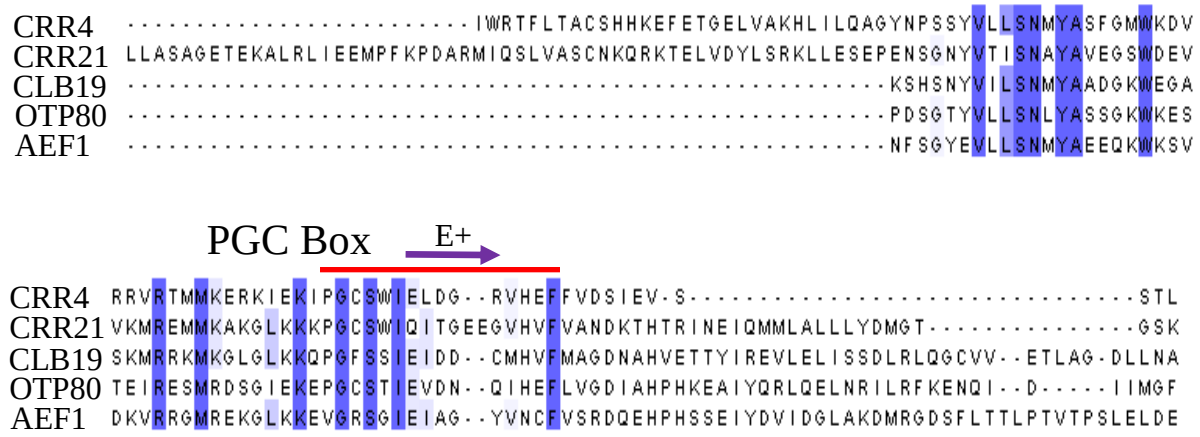


Figure 1.5: C-terminal motif alignment of the five chloroplast PPR-E editing factors. E and partial E+ motif protein sequences from the five plastid PPR-E editing factors (CRR4 NP_182060.2, CRR21 NP_200385.1, CLB19 NP_172066.3, OTP80 NP_200728.2, AEF1 NP_188854.1) were obtained from the NCBI database, aligned using T-Coffee, and presented with Jalview. The degree of conservation is indicated by the intensity of shading. The beginning of the E+ motif is indicated by the purple arrow and the red line indicates the PGC box designated by Hayes et al (2013).

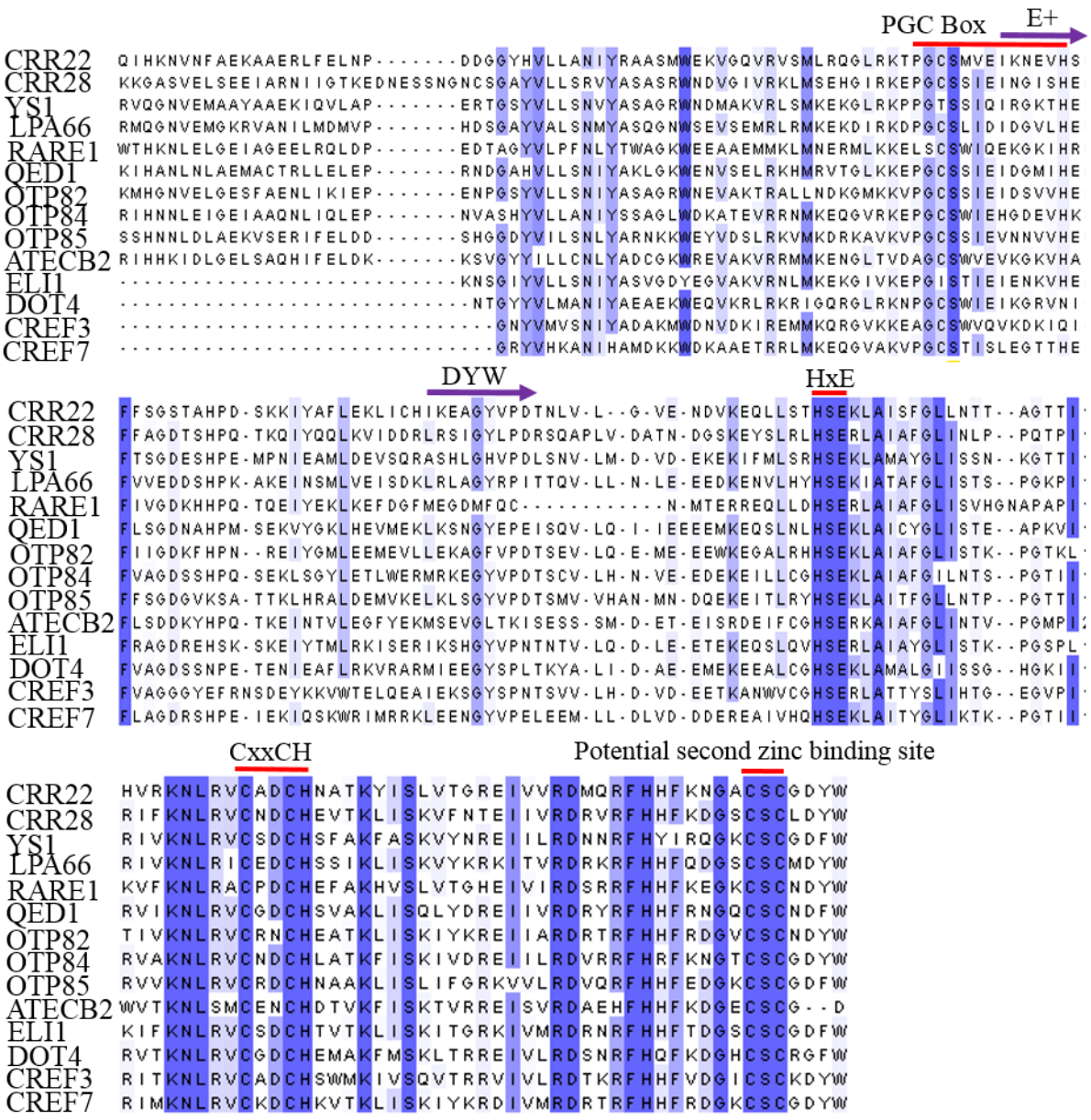


Figure 1.6: C-terminal motif alignment of fourteen PPR-DYW chloroplast editing factors. DYW, E+ and partial E motif protein sequences from fourteen plastid PPR-DYW editing factors (CRR22 NP_172596.1, CRR28 NP_176180.1, YS1 NP_188908.2, LPA66 NP_199702.1, RARE1 NP_196831.1, QED1/OTP81 NP_180537.1, OTP82 NP_172286.1, OTP84 NP_191302.2, OTP85 NP_178398.1, ATECB2 NP_173004.1, ELI1 NP_195454.1, DOT4 NP_193610.1, CREF3 NP_188050.1, CREF7 NP_201453.1) were obtained from the NCBI database, aligned using T-Coffee, and presented with Jalview. Degree of conservation is represented by intensity of shading. The beginnings of the E, E+, and DYW motifs are designated by purple arrows. The conserved CxxCH/HxE deaminase motifs, as well as a second potential zinc binding site and the PGC box designated by Hayes et al (2013) are designated by red lines.

The presence of the two degenerate PPR motifs could signify that the E motif performs a non-site specific RNA or protein binding role in RNA editing. The E motif, however, does not appear to play a role in RNA binding involved in site selection since CRR4 with its E motif removed could still bind RNA near the editing site it affects (Okuda et al., 2006). Besides RNA, PPR motifs are also known to be able to bind protein as well. P-class proteins PPR10 and HCF152 and PLS-class protein THA8 can form homodimers *in vitro* and the P-class PPR protein PNM1, which has been associated with nuclear chromatin, has been shown to interact with chromatin-interacting proteins through yeast two-hybrid (Y2H) assays (Ban et al., 2013; Barkan et al., 2012; Hammani et al., 2011; Nakamura et al., 2003). Also using Y2H assays, interactions between editing PPR proteins like CRR28 and OTP82 and other members of the RNA editing complex have been observed (Sun et al., 2013; Wagoner et al., 2015). The E motif, then, could function to mediate protein-protein interaction between the PPR site-recognition factor and other proteins involved in the RNA editing process, particularly the factor that carries the deaminase function or a factor that can interact with the deaminase.

The other C-terminal motif, the DYW motif, is present in many, but not all of the mitochondrial and plastid PPR editing factors. Of the motifs found in a PPR-DYW protein, the DYW motif tends to display the most conservation between factors (Lurin et al., 2004). Particularly, a set of invariant amino acids (HxE/CxxC) is found in all DYW motifs; these residues are also highly conserved with the active site found in known deaminases (Figure 1.6) (Iyer et al., 2011; Lurin et al., 2004). This observation led to the hypothesis that the DYW motif could function as the deaminase in plant RNA editing (Salone et al., 2007). However, this hypothesis was not supported when experiments showed that editing still occurred even when the DYW motifs of CRR22 and CRR28 were truncated (Okuda et al., 2009). A number of other

PPR-DYW factors are still able to perform specific editing even when the DYW motif is truncated (Hayes et al., 2013; Okuda et al., 2010). Still other editing factors, including CRR4 and CRR21, are able to function despite carrying only an E motif (Okuda et al., 2007). Altogether, these observations appeared to discount the possibility that the DYW motif could function as the deaminase.

THE DYW MOTIF: THE DEAMINASE IN PLANT RNA EDITING?

However, recent studies have shown that DYW motifs can act in *trans*. A protein containing a DYW motif, but no recognizable PPR motifs, was found to interact with E-class editing factor CRR4 to edit the *ndhD* C2 site (Boussardon et al., 2012). Plants mutant for this gene, called DYW1, were unable to edit *ndhD* C2, similarly to *crr4* mutants, and the DYW1 and CRR4 proteins were found to interact *in planta* (Boussardon et al., 2012). A fusion of CRR4 and DYW1 was created and this chimeric protein was able to complement the editing phenotype in *crr4-dyw1* double mutants (Boussardon et al., 2012). Interestingly, this CRR4-DYW1 fusion protein resembled a complete PPR-DYW protein as DYW1 provided the DYW motif and part of an E+ motif that is missing in the CRR4 protein (Boussardon et al., 2012). These results indicate that DYW motifs can act in *trans* and contribute to editing function in conjunction with proteins without a DYW motif, and by extension, possibly with proteins that have a non-functional DYW motif.

Though many PPR-DYW proteins do not appear to require their DYW motif for editing, one mitochondrial editing factor, MEF1, was found to require its DYW motif for editing function, as DYW-truncated MEF1 was not able to complement the editing phenotype in *mef1* mutant plants and protoplasts (Zehrmann et al., 2010). The addition of a 6xHis tag to the end of MEF1 was also found to interfere with editing. This inhibition of function by a C-terminal 6xHis

tag supports the role of its DYW motif in the editing function (Zehrmann et al., 2010). Analysis of the DYW motifs of plastid editing factors ELI1 and DYW1 revealed that these DYW motifs could bind two zinc ions, a feature which is critical for function of other known cytidine deaminases, though the DYW motif of ELI1 was found to be dispensable for editing (Hayes et al., 2013). The observation that these motifs could bind two zinc atoms led to the identification of a possible second zinc binding site formed by a set of residues invariant in all editing factor DYW motifs near the C-terminus of these motifs (Hayes et al., 2013).

Because of its involvement in editing *ndhD* C2, further study on DYW1 was performed to identify amino acids that were important for its function through analysis of TILLING lines containing random EMS mutations and through site-specific mutagenesis targeting its conserved deaminase-like HxE/CxxC motifs as well as the conserved terminal DYW tripeptide (Boussardon et al., 2014). In one DYW1 mutant line, the H and E of the HxE motif were mutated to alanines and in another, the two cysteines of CxxC were mutated to alanines. Lines that carried versions of DYW1 with the mutated deaminase signature were unable to complement the editing defect indicating that these residues are indeed important for the editing process. Unexpectedly, truncation of the terminal DYW tripeptide also resulted in non-complementation of the editing phenotype. Mutation of the individual amino acids of the DYW tripeptide revealed that mutation of D resulted in decreased editing, mutation of the Y did not affect editing extent, and mutation of the W completely abolished editing (Boussardon et al., 2014). This is consistent with what is observed in the DYW motif of DYW proteins where most of the natural variation of this tripeptide occurs in the second position. The importance of the terminal tryptophan can also explain why the addition of a C-terminal tag can severely inhibit editing. Boussardon et al. (2014) hypothesized that the terminal DYW tripeptide could function

as a C-terminal protein-protein interaction signal due to its similarity to other known C-terminal protein-protein interaction signals (charged residue followed by hydrophobic residues); this hypothesis would be consistent with the severe reduction in editing observed when a protein tag is attached to the C-terminus or when the DYW residues are mutated or truncated. The ability of mutagenized versions of DYW1 to bind zinc was also tested in this experiment. One version contained a 32 amino acid deletion of the deaminase-like signature and another version mutated the H and E of the (HxE/CxxC) signature to alanines. These mutagenized versions bound a much lower amount of zinc than the wild type DYW1, indicating that the deaminase-like signature itself is necessary for proper zinc binding (Boussardon et al., 2014). These results indicate that the DYW motif may play an important role in the RNA editing reaction, possibly as the cytidine deaminase, although it is still unclear if the motif has deaminase activity, or if it does not, what other role it may play.

Chapter 2 of this thesis identifies a new PPR-DYW editing factor, *QUINTUPLE EDITING 1* (QED1) that is responsible for the editing of five chloroplast sites and investigates the function of its DYW motif as well as the DYW motif of another PPR-DYW factor, RARE1, in the RNA editing process (Robbins et al., 2009). Using truncation analysis, these DYW motifs were determined to be necessary for editing and that specifically, the conserved deaminase-like residues in both motifs are critical for efficient editing function, consistent with what was observed in mutational analysis of DYW1 (Boussardon et al., 2014). Using site-directed mutagenesis, the importance of other conserved motifs, including the second potential zinc binding site identified in Hayes et al (2013) was investigated and it was found that this second zinc binding site could also play a role in RNA editing (Hayes et al., 2013; Wagoner et al., 2015). After the publication of the investigation of the DYW motifs of QED1 andRARE1, two

other plastid editing PPR-DYW factors, OTP84 and CREF7, which specify editing of three sites and one site, respectively, have been shown to require their DYW domains for efficient editing (Hayes et al., 2015). Interestingly, when the DYW domain of was truncated, the editing of two out of three of the sites specified by OTP84 was severely reduced, while the editing of the third site remained at wild-type efficiency. However, when the glutamate in the conserved HXE deaminase signature was targeted by site-directed mutagenesis, editing at all three sites was severely decreased. Overexpression of mutated OTP84 in wild-type plants did not affect editing efficiency of the three sites, indicating that the decrease in editing efficiency was not due to a dominant negative effect by the mutated gene (Hayes et al., 2015).

Taken together, these results implicate that the DYW motif may indeed be involved in the deamination reaction although currently there is no direct evidence that the DYW motif is capable of deaminase activity. However, there is some evidence that the DYW motif of CRR2, which is responsible for intergenic processing between *rps7* and *ndhB*, could be capable of endonuclease activity (Okuda et al., 2009a). CRR2 lacking its DYW domain could not complement the RNA processing phenotype when transformed into *crr2* mutant plants. Also, despite sequence similarities between the DYW motifs of CRR22, CRR28 and CRR2, CRR2 carrying the DYW motif of CRR22 or CRR28 did not show any endoribonuclease activity (Hashimoto et al., 2003; Okuda et al., 2009a). Conversely, although CRR22 and CRR28 can function without their DYW motif, editing was abolished when their DYW motifs were replaced with the DYW motif of CRR2 (Okuda et al., 2009). Nakamura and Sugita (2008) also characterized another PPR-DYW protein that demonstrated endonuclease activity dependent on the concentration of the fusion protein containing the DYW motif. Mutation of the CxxCH motif to GxxGH greatly reduced the endonuclease activity, suggesting that the activity requires the

deaminase motifs (Nakamura and Sugita, 2008). This protein was later further characterized as OTP85 which is a plastid editing factor responsible for a single site in the *ndhD* transcript (Hammani et al., 2009). Though the DYW motif of OTP85 appears to have endonuclease activity, it is yet unclear if the DYW motif of OTP85 is necessary for its editing function. If it is needed for editing, this observation may support the hypothesis that the DYW motif is the editing deaminase as it would show that the DYW motif of an editing PPR protein is capable of catalytic activity.

Evolutionarily, the presence of the DYW motif correlates very well with the presence of RNA editing, not only in land plants but also in other organisms (Rüdinger et al., 2012) For instance, DYW class proteins are lacking in Marchantia liverworts, which do not have the RNA editing process and C-to-U editing has been observed in protists that appear to have acquired PLS-class PPR-DYW proteins through horizontal gene transfer (Groth-Malonek et al., 2007; Knoop and Rüdinger, 2010; Rüdinger et al., 2011; Schallenberg-Rüdinger et al., 2013). Though DYW motifs can apparently act in *trans*, to date, no other proteins carrying DYW motifs have been associated with any other E-class PPR factors or with PPR-DYW factors with dispensable DYW motifs. 87 DYW-containing proteins are encoded in the Arabidopsis genome and only five of those appear to be good candidates as general editing PPR interacting partners as they have few PPR motifs and a poorly conserved E motif (Boussardon et al., 2012; Colcombet et al., 2013; Lurin et al., 2004). Though these factors could likely be involved in editing sites associated with E-class proteins, or perhaps even PPR-DYW proteins with dispensable DYW motifs like CRR22 and CRR28, given the amount of these E-class or DYW-dispensible PPR proteins involved in both plastid and mitochondrial RNA editing, they would have to participate in associations less specific than that of DYW1, which only is involved with a single editing site

(Boussardon et al., 2012). Though there are currently no sets of plastid PPR editing factors that participate in the same editing site (with the exception of RARE1 and atECB2), other PPR-DYW proteins that are yet unidentified may also act in *trans* to provide a DYW motif that is necessary for editing function (Robbins et al., 2009; Yu et al., 2009). For example, Yagi et al (2013) hypothesize that RARE1 and atECB2 cooperate to edit the *accD* C794 site similarly to how CRR4 and DYW1 cooperate to edit the *ndhD* C2 site, as the *accD* C794 *cis*-element is clearly recognized by RARE1 but not atECB2 via *in silico* analysis and the two PPR proteins appear to be highly related by co-expression profile (Yagi et al., 2013b). RARE1 also affects only one site and shows no mutant phenotype, while atECB2 affects multiple editing sites and atECB2 mutants are albino and seedling lethal (Cao et al., 2011; Robbins et al., 2009; Yu et al., 2009). It is also distinctly possible that there are unidentified proteins that must interact with the E or DYW motif in order to perform the editing reaction.

NON-PPR PROTEIN *TRANS*-FACTORS INVOLVED IN RNA EDITING

Mediation of potential interactions between the PPR editing protein and possible unknown factors involved in catalysis may depend on other proteins that have recently been found in the editosome, which appears to be a complex about 200-400 kD in size (Bentolila et al., 2012). Through co-immunoprecipitation of the PPR-DYW editing factor RARE1, a new non-PPR protein, called RNA EDITING FACTOR INTERACTING PROTEIN 1 (RIP1) was found to be dual-targeted to the chloroplast and the mitochondria and broadly involved in RNA editing in both organelles (Bentolila et al., 2012). RIP1 was found to be part of a small family of proteins named the RIP family. This family is also referred to as the MORF (Multiple Organellar RNA editing Factor) family, as two different labs independently described members of the same family in 2012 (Bentolila et al., 2012, Takenaka et al. 2012). . Like members of the

PPR protein family, members of the RIP/MORF protein family were shown to be targeted to either the plastid or mitochondria and involved in the RNA editing process in that organelle (Bentolila et al., 2013; Takenaka et al., 2012). RIP2/MORF2 and RIP9/MORF9 are targeted to the plastid where, together, they are involved in the editing of nearly all of the plastid sites and can interact with not only each other, but also with PPR editing factors (Bentolila et al., 2013; Takenaka et al., 2012). Though the function of this family of proteins is unclear, they do not appear to be capable of deaminase activity, as they do not have domains that resemble those of deaminases and mutation of these family members often cause a reduction, but not a complete knockout of editing at specific editing sites. This family is also absent in bryophytes and lycophytes that can engage in RNA editing (Bentolila et al., 2013; Takenaka et al., 2012). Using Bimolecular Fluorescence Complementation (BiFC), recent experiments on plastid editing factor CHLOROPLAST BIOGENESIS 19 (CLB19) have shown that its E motif does appear to interact with MORF2/RIP2 but not with MORF9/RIP9, which may imply the RIP/MORF proteins can function as mediators of the interaction between the editing PPR protein and a factor containing deaminase function (Ramos-Vega et al., 2015).

ORRM1, a novel protein that contains twin RIP domains as well as an additional RNA recognition motif (RRM) was discovered to be a major factor in plastid RNA editing, affecting over half of the plastid sites in Arabidopsis (Sun et al., 2013). Surprisingly, in complementation analysis, the RRM of ORRM1, and not the RIP domains, was able to complement the editing defect. Like the RIP family members, ORRM1 was also found to be able to interact directly with editing PPR factors through the PPR motifs (Sun et al., 2013). Two other members in ORRM1's clade, ORRM2 and ORRM3, were found to be mitochondrial editing factors that can interact with each other and, similarly to ORRM1 in plastids, can interact with mitochondrial

members of the RIP/MORF family (Sun et al., 2013). These non-PPR factors that are involved in RNA editing, though their function is unclear, could regulate RNA editing function through mediating protein-protein interactions between the PPR proteins functioning as site-recognition factors and other factors that could be important for catalysis of the RNA editing reaction or have other important regulatory functions.

Given ORRM1's broad involvement in plastid RNA editing, it was identified as a good candidate protein to use as bait to possibly identify other members of the editosome. Using ORRM1 tagged at the N-terminus for co-immunoprecipitation, another non-PPR protein containing two zinc finger domains was found to broadly affect plastid editing efficiencies, affecting 30 chloroplast C-targets (Sun et al., 2015). This protein, named Organelle Zinc finger 1 (OZ1) was also observed to interact with editing PPR proteins as well as ORRM1 via Y2H, although OZ1 did not interact with RIP2/MORF2 or RIP9/MORF9 and only interacted with RIP1 weakly (Sun et al., 2015). OZ1 was found to be a member of a small zinc finger protein family containing 3 other proteins, one predicted to be targeted to the mitochondria, the other two predicted to be targeted to the plastid, all of which are likely to be involved in the RNA editing process (Sun et al., 2015).

Recently, yet another non-PPR member of the editosome has been characterized, as an enzyme functioning in the tetrapyrrole biosynthetic pathway, protoporphyrinogen IX oxidase 1 (PPO1), was found to be involved in the editing of seven plastid sites, including *ndhD* C2, the editing extent of which was completely absent in *ppo1* mutants (Zhang et al., 2014). The RNA editing function of PPO1 was dependent upon its interaction with RIP2 and RIP9, but PPO1 did not interact directly with any of the PPR editing factors that were tested, which supports the role of RIP family members as mediators of protein-protein interactions. Surprisingly, it was also

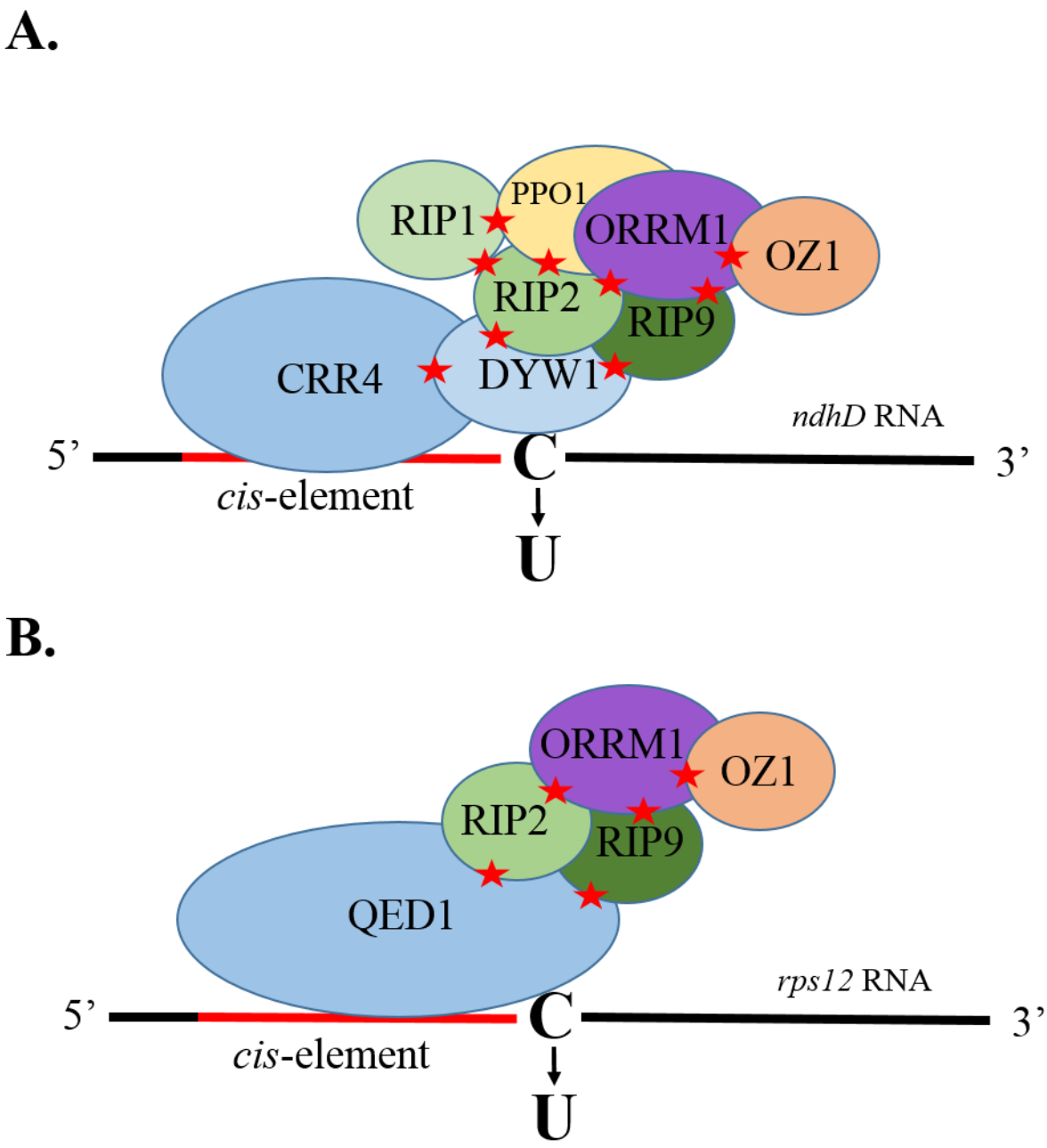


Figure 1.7: Current possible models of RNA editosomes based on mutant analysis and protein-protein interaction data from Y2H and BiFC analyses. **A.** RNA editosome model of CRR4, an E-class PPR editing factor.. **B.** RNA editosome model of QED1, a PPR-DYW editing factor. Red stars indicate protein-protein interaction verified by Y2H and/or BiFC analysis.

found that RIP2 and RIP9 interacted with DYW1, but not with CRR4 which may indicate that the RIP factors may mediate the factors involved in the deaminase reaction if the DYW motif has deaminase activity (Zhang et al., 2014). Despite the recent advances made in identifying members of the editsome, the composition and dynamics of these members in the editosome during the editing reaction is still unclear, although Y2H and BiFC data does give some insight on which proteins may interact in the complex (Figure 1.7). Additionally, there may yet be some components of the editosome that have not been discovered yet, particularly those that may contribute to the deamination reaction. Table 1.1 summarizes which proteins have been associated with specific plastid editing sites based on editing extents that were affected in plants mutant for specific genes, or when particular genes were silenced using Virus Induced Gene Silencing (VIGS) (Bentolila et al., 2013; Boussardon et al., 2012; Sun et al., 2013, 2015; Takenaka et al., 2012; Zhang et al., 2014).

CONSEQUENCES WHEN EDITING FAILS TO OCCUR

Why have plants devoted so many proteins to the RNA editing process? As was discussed previously, this process appears to be a corrective process in order maintain codons that are evolutionarily conserved in other plant species. The consequences of not editing these codons can result in a wide range of phenotypes depending on the transcripts that are affected. In *CHLOROPLAST BIOGENESIS 19* (CLB19) mutants, sites in *rpoA* and *clpP*, which encode a plastid RNA polymerase subunit and a protease respectively, are not edited, resulting in a seedling lethal phenotype (Chateigner-Boutin et al., 2008). In *YELLOW SEEDLING 1* (YS1) mutants, a site in *rpoB*, encoding another subunit of plastid RNA polymerase, is unedited, and seedlings display a yellow phenotype, although these plants eventually green and grow to maturity (Zhou et al., 2009). Mutant plants that have defects in genes such as CRR4, CRR22,

and CRR28 which affect editing in subunits of the NDH complex appear phenotypically normal, but lack the transient increase in chlorophyll fluorescence after shutting off actinic light that can be observed with specialized equipment in wild type plants (Hammani et al., 2009; Kotera et al., 2005; Okuda et al., 2007, 2009). In some cases, plants mutant for certain genes, such as RARE1, OTP80, OTP82, OTP85, and OTP86, do not appear to have any discernible mutant phenotype despite the lack the of editing at these PPR proteins' particular sites (Hammani et al., 2009; Robbins et al., 2009). However, it is possible that problems caused by the alteration of a protein sequence due to lack of RNA editing would not be observed under lab conditions, but instead would cause a detrimental phenotype if the plant was growing in nature and was subjected to particular environmental stresses (Kurihara-Yonemoto and Handa, 2001; Nakajima and Mulligan, 2001). Abnormal phenotypes might also be species-specific: *rare1* mutant Arabidopsis do not edit a site in *accD* which is involved in fatty acid synthesis, but do not display any obvious mutant phenotype (Robbins et al., 2009). However, *in vitro* analysis of pea *accD* revealed that the version of pea *accD* unedited at this same site had no enzymatic activity when expressed in *E. coli* (Sasaki et al., 2001).

Chapter three of this thesis investigates the physiological consequences of the lack of editing at five chloroplast sites in *qed1* mutants, which includes sites located in genes encoding an RNA polymerase subunit, a chloroplast splicing factor, and a subunit of the NDH complex. These mutant plants show a defect in growth and in NDH activity as well as an abnormal chloroplast morphology phenotype. Together, the experiments in chapters two and three aim to gain insight into phenotypical consequences of lack of editing at certain plastid editing sites and also to help address some of the outstanding questions in the plant RNA editing field involving the role of the C-terminal motifs in RNA editing and the identity of the deaminase that drives the

reaction.

Table 1.1: Protein factors associated with specific editing sites

Plastid Editing Site	Associated PPR Protein	Associated RIP	Associated MORF	Other Associated Proteins
ndhB C1255	CREF7 PPR-DYW		2,9	ORRM1, PPO1, OZ1
ndhB C1481	OTP84 PPR-DYW	2, 9	2, 9	PPO1
ndhB C149		2, 9	2, 9	PPO1, OZ1
ndhB C153		2	NT	PPO1
ndhB C467	CRR28 PPR-DYW	1, 2	2, 9	ORRM1, PPO1, OZ1
ndhB C586		1, 9		ORRM1, PPO1, OZ1
ndhB C708			NT	
ndhB C726		9	NT	OZ1
ndhB C746	CRR22 PPR-DYW	9	2, 9	ORRM1, OZ1
ndhB C830	ELI1 PPR-DYW	2	2, 9	ORRM1, PPO1, OZ1
ndhB C836	OTP82 PPR-DYW	2, 9	2, 9	PPO1, OZ1
ndhB C872	QED1 PPR-DYW	2, 9	2, 9	ORRM1, PPO1, OZ1
ndhD C887	CRR22 PPR-DYW	2, 9	2, 9	ORRM1, PPO1, OZ1

Plastid Editing Site	Associated PPR Protein	Associated RIP	Associated MORF	Other Associated Proteins
accD 3'UTR C1568	QED1 PPR-DYW	2, 9	2 (inc) 9 (inc)	ORRM1, OZ1
accD C794	ATECB2 PPR-DYW RARE1 PPR-DYW	1, 9	2, 9	OZ1
atpF C92	AEF1 PPR-E	2	2, 9	
clpP C559	CLB19 PPR-E	9	2, 9	ORRM1, OZ1
matK C640	QED1 PPR-DYW	2	2, 9	ORRM1 , OZ1
rpoA C200	CLB19 PPR-E	9	2 (inc), 9	ORRM1, OZ1
rpoB C2432	QED1 PPR-DYW	2	2, 9	ORRM1, OZ1
rpoB C338	YS1 PPR-DYW		2, 9	OZ1
rpoB C551	CRR22 PPR-DYW		2, 9	ORRM1, OZ1
rpoC1 C488	DOT4 PPR-DYW	1 (inc), 2, 9		PPO1 (inc), OZ1 (inc)
rps12 i1 C58	QED1 PPR-DYW	1 (inc), 2, 9	2, 9	ORRM1, OZ1
rps14 C149		9	2, 9	ORRM1
rps14 C80	OTP86 PPR-DYW	2, 9	2, 9	

Plastid Editing Site	Associated PPR Protein	Associated RIP	Associated MORF	Other Associated Proteins
ndhD C2	CRR4 PPR-E	1, 2, 9	2, 9	ORRM1, PPO1, DYW1, OZ1
ndhD C383	CRR21 PPR-E	2, 9	2, 9	PPO1
ndhD C674	OTP85 PPR-DYW	9	2, 9	ORRM1, PPO1, OZ1
ndhD C878	CRR28 PPR-DYW	2, 9	2, 9	ORRM1, PPO1, OZ1
ndhF C290	OTP84 PPR-DYW	2	2, 9	PPO1, OZ1
ndhG C50	OTP82 PPR-DYW	2, 9	2, 9	ORRM1, PPO1, OZ1
petL C5		1, 9	2, 9	PPO1, OZ1
psbE C214	CREF3 PPR-DYW		2	
psbF C77	LPA66 PPR-DYW		2, 9	
psbZ C50	OTP84 PPR-DYW	2, 9	2, 9	
rpl23 C89	OTP80 PPR-E	2, 9	2, 9	PPO1, OZ1

Table 1.1: Protein factors associated to specific plastid editing sites via mutant or silencing analysis. Associated PPR Protein indicates PPR-E or PPR-DYW editing factor associated with an editing site. Associated RIP indicates which RIP/MORF proteins have been associated to each site via Virus Induced Gene Silencing (VIGS) in Bentolila et al (2013). Associated MORF indicates which RIP/MORF proteins have been associated with each site via mutant analysis in Takenaka et al (2013). RIP1 = MORF8, RIP2 = MORF2, RIP9 = MORF9. Plastid RNA status was not tested in MORF8 mutants. Other Proteins Associated indicates non-PPR, non-RIP/MORF proteins that have been associated with each site via mutant analysis. Inc. indicates increase in editing extent was observed in mutant or silenced plants. NT = Not Tested in that analysis

REFERENCES

Aphasizhev, R., and Aphasizheva, I. (2013). Emerging roles of PPR proteins in trypanosomes: switches, blocks, and triggers. RNA Biol. *10*, 1495–1500.

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

Baggio, F., Bratic, A., Mourier, A., Kauppila, T.E.S., Tain, L.S., Kukat, C., Habermann, B., Partridge, L., and Larsson, N.-G. (2014). *Drosophila melanogaster* LRPPRC2 is involved in coordination of mitochondrial translation. Nucleic Acids Res. *42*, 13920–13938.

Ban, T., Ke, J., Chen, R., Gu, X., Tan, M.H.E., Zhou, X.E., Kang, Y., Melcher, K., Zhu, J.-K., and Xu, H.E. (2013). Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. J. Biol. Chem. *288*, 31540–31548.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. *65*, 415–442.

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

Bégu, D., Graves, P.V., Domec, C., Arselin, G., Litvak, S., and Araya, A. (1990). RNA editing of wheat mitochondrial ATP synthase subunit 9: direct protein and cDNA sequencing. Plant Cell *2*, 1283–1290.

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

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

Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., Wijk, K.J. van, and Hanson, M.R. (2012). *RIP1*, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proc. Natl. Acad. Sci. *109*, E1453–E1461.

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

Binder, S., Marchfelder, A., and Brennicke, A. (1994). RNA editing of tRNA(Phe) and tRNA(Cys) in mitochondria of *Oenothera berteriana* is initiated in precursor molecules. Mol. Gen. Genet. MGG *244*, 67–74.

Blanc, V., Litvak, S., and Araya, A. (1995). RNA editing in wheat mitochondria proceeds by a deamination mechanism. FEBS Lett. *373*, 56–60.

Blatch, G.L., and Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. BioEssays News Rev. Mol. Cell. Dev. Biol. *21*, 932–939.

Bock, R., Kössel, 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.

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

Bock, R., Hermann, M., and Fuchs, M. (1997). Identification of critical nucleotide positions for plastid RNA editing site recognition. RNA N. Y. N *3*, 1194–1200.

Börner, G.V., Mörl, M., Wissinger, B., Brennicke, A., and Schmelzer, C. (1995). RNA editing of a group II intron in *Oenothera* as a prerequisite for splicing. Mol. Gen. Genet. MGG *246*, 739–744.

Boussardon, C., Salone, V., Avon, A., Berthomé, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C. (2012). Two interacting proteins are necessary for the editing of the *NdhD-1* site in Arabidopsis plastids. Plant Cell *24*, 3684–3694.

Boussardon, C., Avon, A., Kindgren, P., Bond, C.S., Challenor, M., Lurin, C., and Small, I. (2014). The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of *ndhD-1*. New Phytol. *203*, 1090–1095.

Cao, Z.-L., Yu, Q.-B., Sun, Y., Lu, Y., Cui, Y.-L., and Yang, Z.-N. (2011). A point mutation in the pentatricopeptide repeat motif of the AtECB2 protein causes delayed chloroplast development. J. Integr. Plant Biol. *53*, 258–269.

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.

Chateigner-Boutin, A.-L., and Hanson, M.R. (2003). Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar *cis*-elements. Nucleic Acids Res. *31*, 2586–2594.

Chateigner-Boutin, A.-L., and Small, I. (2011). Organellar RNA editing. Wiley Interdiscip. Rev. RNA *2*, 493–506.

Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-García, A., Andrés, C., de la Luz Gutiérrez-Nava, M., Cantero, A., Delannoy, E., Jiménez, 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. Cell Mol. Biol. *56*, 590–602.

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

Chateigner-Boutin, A.-L., Colas des Francs-Small, C., Fujii, S., Okuda, K., Tanz, S.K., and Small, I. (2013). The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. Plant J. Cell Mol. Biol. *74*, 935–945.

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.

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.

Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthomé, R., and Lurin, C. (2013). Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA Biol. *10*, 1557–1575.

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

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

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.

Faivre-Nitschke, S.E., Grienenberger, J.M., and Gualberto, J.M. (1999). A prokaryotic-type cytidine deaminase from *Arabidopsis thaliana* gene expression and functional characterization. Eur. J. Biochem. FEBS *263*, 896–903.

Farré, J.C., Leon, G., Jordana, X., and Araya, A. (2001). *cis* Recognition elements in plant mitochondrion RNA editing. Mol. Cell. Biol. *21*, 6731–6737.

Freyer, R., Hoch, B., Neckermann, K., Maier, R.M., and Kössel, H. (1993). RNA editing in maize chloroplasts is a processing step independent of splicing and cleavage to monocistronic mRNAs. Plant J. Cell Mol. Biol. *4*, 621–629.

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

Gott, J.M., Visomirski, L.M., and Hunter, J.L. (1993). Substitutional and insertional RNA editing of the cytochrome c oxidase subunit 1 mRNA of *Physarum polycephalum*. J. Biol. Chem. *268*, 25483–25486.

Graves, P.V., Bégu, D., Velours, J., Neau, E., Belloc, F., Litvak, S., and Araya, A. (1990). Direct protein sequencing of wheat mitochondrial ATP synthase subunit 9 confirms RNA editing in plants. J. Mol. Biol. *214*, 1–6.

Grewe, F., Viehoever, P., Weisshaar, B., and Knoop, V. (2009). A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. Nucleic Acids Res. *37*, 5093–5104.

Grosskopf, D., and Mulligan, R.M. (1996). Developmental- and tissue-specificity of RNA editing in mitochondria of suspension-cultured maize cells and seedlings. Curr. Genet. *29*, 556–563.

Groth-Malonek, M., Wahrmund, U., Polsakiewicz, M., and Knoop, V. (2007). Evolution of a pseudogene: exclusive survival of a functional mitochondrial *nad7* gene supports Haplomitrium as the earliest liverwort lineage and proposes a secondary loss of RNA editing in *Marchantiidae*. Mol. Biol. Evol. *24*, 1068–1074.

Gully, B.S., Shah, K.R., Lee, M., Shearston, K., Smith, N.M., Sadowska, A., Blythe, A.J., Bernath-Levin, K., Stanley, W.A., Small, I.D., et al. (2015a). The design and structural characterization of a synthetic pentatricopeptide repeat protein. Acta Crystallogr. D Biol. Crystallogr. *71*, 196–208.

Gully, B.S., Cowieson, N., Stanley, W.A., Shearston, K., Small, I.D., Barkan, A., and Bond, C.S. (2015b). The solution structure of the pentatricopeptide repeat protein PPR10 upon binding *atpH* RNA. Nucleic Acids Res. *43*, 1918–1926.

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.

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.

Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I., and Giegé, P. (2011). An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. Plant Cell *23*, 730–740.

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. Cell Mol. Biol. *36*, 541–549.

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.

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.

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

Hayes, M.L., Dang, K.N., Diaz, M.F., and Mulligan, R.M. (2015). A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. J. Biol. Chem. *290*, 10136–10142.

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. Cell Mol. Biol. *42*, 124–132.

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.

Herbert, C.J., Golik, P., and Bonnefoy, N. (2013). Yeast PPR proteins, watchdogs of mitochondrial gene expression. RNA Biol. *10*, 1477–1494.

Hiesel, R., Combettes, B., and Brennicke, A. (1994). Evidence for RNA editing in mitochondria of all major groups of land plants except the *Bryophyta*. Proc. Natl. Acad. Sci. U. S. A. *91*, 629–633.

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.

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

Iyer, L.M., Zhang, D., Rogozin, I.B., and Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res. *39*, 9473–9497.

Jínek, M., Rehwinkel, J., Lazarus, B.D., Izaurralde, E., Hanover, J.A., and Conti, E. (2004). The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. Nat. Struct. Mol. Biol. *11*, 1001–1007.

Karcher, D., and Bock, R. (1998). Site-selective inhibition of plastid RNA editing by heat shock and antibiotics: a role for plastid translation in RNA editing. Nucleic Acids Res. *26*, 1185–1190.

Karcher, D., and Bock, R. (2002). The amino acid sequence of a plastid protein is developmentally regulated by RNA editing. J. Biol. Chem. *277*, 5570–5574.

Ke, J., Chen, R.-Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H.E., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.-K., et al. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. Nat. Struct. Mol. Biol. *20*, 1377–1382.

Kindgren, P., Yap, A., Bond, C.S., and Small, I. (2015). Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. Plant Cell *27*, 403–416.

Knoop, V., and Rüdinger, M. (2010). DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? FEBS Lett. *584*, 4287–4291.

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

Koito, A., and Ikeda, T. (2012). Apolipoprotein B mRNA-editing, catalytic polypeptide cytidine deaminases and retroviral restriction. Wiley Interdiscip. Rev. RNA *3*, 529–541.

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

Kudla, J., and Bock, R. (1999). RNA editing in an untranslated region of the Ginkgo chloroplast genome. Gene *234*, 81–86.

Kudla, J., Igloi, G.L., Metzlaff, M., Hagemann, R., and Kössel, 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.

Kugita, M., Yamamoto, Y., Fujikawa, T., Matsumoto, T., and Yoshinaga, K. (2003). RNA editing in hornwort chloroplasts makes more than half the genes functional. Nucleic Acids Res. *31*, 2417–2423.

Kurihara-Yonemoto, S., and Handa, H. (2001). Low temperature affects the processing pattern and RNA editing status of the mitochondrial *cox2* transcripts in wheat. Curr. Genet. *40*, 203–208.

Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2013). Human pentatricopeptide proteins. RNA Biol. *10*, 1433–1438.

Lu, B., and Hanson, M.R. (1992). A single nuclear gene specifies the abundance and extent of RNA editing of a plant mitochondrial transcript. Nucleic Acids Res. *20*, 5699–5703.

Lu, B., and Hanson, M.R. (1994). A single homogeneous form of ATP6 protein accumulates in petunia mitochondria despite the presence of differentially edited *atp6* transcripts. Plant Cell *6*, 1955–1968.

Lu, B., and Hanson, M.R. (1996). Fully edited and partially edited *nad9* transcripts differ in size and both are associated with polysomes in potato mitochondria. Nucleic Acids Res. *24*, 1369–1374.

Lu, B., Wilson, R.K., Phreaner, C.G., Mulligan, R.M., and Hanson, M.R. (1996). Protein polymorphism generated by differential RNA editing of a plant mitochondrial *rps12* gene. Mol. Cell. Biol. *16*, 1543–1549.

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, 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.

Maier, R.M., Hoch, B., Zeltz, P., and Kössel, H. (1992). Internal editing of the maize chloroplast *ndhA* transcript restores codons for conserved amino acids. Plant Cell *4*, 609–616.

Malek, O., Lättig, K., Hiesel, R., Brennicke, A., and Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J. *15*, 1403–1411.

Manthey, G.M., and McEwen, J.E. (1995). The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of Saccharomyces cerevisiae. EMBO J. *14*, 4031–4043.

Maréchal-Drouard, L., Ramamonjisoa, D., Cosset, A., Weil, J.H., and Dietrich, A. (1993). Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. Nucleic Acids Res. *21*, 4909–4914.

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.

Nakajima, Y., and Mulligan, R.M. (2001). Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. Curr. Genet. *40*, 209–213.

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.

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. FEBS *270*, 4070–4081.

Nakamura, T., Yagi, Y., and Kobayashi, K. (2012). Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. Plant Cell Physiol. *53*, 1171–1179.

Neckermann, K., Zeltz, P., Igloi, G.L., Kössel, H., and Maier, R.M. (1994). The role of RNA editing in conservation of start codons in chloroplast genomes. Gene *146*, 177–182.

Neuwirt, J., Takenaka, M., van der Merwe, J.A., and Brennicke, A. (2005). An *in vitro* RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species. RNA N. Y. N *11*, 1563–1570.

Ohyama, K. (1996). Chloroplast and mitochondrial genomes from a liverwort, *Marchantia polymorpha*--gene organization and molecular evolution. Biosci. Biotechnol. Biochem. *60*, 16–24.

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.

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. *104*, 8178–8183.

Okuda, K., Chateigner-Boutin, A.-L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009). Pentatricopeptide repeat

proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. Plant Cell *21*, 146–156.

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

Oldenkott, B., Yamaguchi, K., Tsuji-Tsukinoki, S., Knie, N., and Knoop, V. (2014). Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. RNA N. Y. N *20*, 1499–1506.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. Mol. Biol. Evol. *25*, 1120–1128.

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

Phreaner, C.G., Williams, M.A., and Mulligan, R.M. (1996). Incomplete editing of *rps12* transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. Plant Cell *8*, 107–117.

Pusnik, M., Small, I., Read, L.K., Fabbro, T., and Schneider, A. (2007). Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. Mol. Cell. Biol. *27*, 6876–6888.

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

Ramos-Vega, M., Guevara-García, A., Llamas, E., Sánchez-León, N., Olmedo-Monfil, V., Vielle-Calzada, J.P., and León, P. (2015). Functional analysis of the *Arabidopsis thaliana CHLOROPLAST BIOGENESIS 19* pentatricopeptide repeat editing protein. New Phytol. n/a – n/a.

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.

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 N. Y. N *15*, 1142–1153.

Rüdinger, M., Funk, H.T., Rensing, S.A., Maier, U.G., and Knoop, V. (2009). RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. Mol. Genet. Genomics MGG *281*, 473–481.

Rüdinger, M., Fritz-Laylin, L., Polsakiewicz, M., and Knoop, V. (2011). Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. RNA N. Y. N *17*, 2058–2062.

Rüdinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V. (2012). Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. J. Mol. Evol. *74*, 37–51.

Ruwe, H., Castandet, B., Schmitz-Linneweber, C., and Stern, D.B. (2013). Arabidopsis chloroplast quantitative editotype. FEBS Lett. *587*, 1429–1433.

Salone, V., Rüdinger, 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.

Sasaki, T., Yukawa, Y., Wakasugi, T., Yamada, K., and Sugiura, M. (2006). A simple *in vitro* RNA editing assay for chloroplast transcripts using fluorescent dideoxynucleotides: distinct types of sequence elements required for editing of *ndh* transcripts. Plant J. *47*, 802–810.

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.

Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J.M., and Knoop, V. (2013). A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. RNA Biol. *10*, 1549–1556.

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.

Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing. J. Mol. Evol. *48*, 303–312.

Sugita, M., Miyata, Y., Maruyama, K., Sugiura, C., Arikawa, T., and Higuchi, M. (2006). Extensive RNA editing in transcripts from the *PsbB* operon and *RpoA* gene of plastids from the enigmatic moss *Takakia lepidozioides*. Biosci. Biotechnol. Biochem. *70*, 2268–2274.

Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M.R., and Bentolila, S. (2013). An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc. Natl. Acad. Sci. U. S. A. *110*, E1169–E1178.

Sun, T., Shi, X., Friso, G., Van Wijk, K., Bentolila, S., and Hanson, M.R. (2015). A zinc finger motif-containing protein is essential for chloroplast RNA editing. PLoS Genet *11*, e1005028.

Sutton, C.A., Conklin, P.L., Pruitt, K.D., and Hanson, M.R. (1991). Editing of pre-mRNAs can occur before *cis*- and *trans*-splicing in Petunia mitochondria. Mol. Cell. Biol. *11*, 4274–4277.

Sutton, C.A., Zoubenko, O.V., Hanson, M.R., and Maliga, P. (1995). A plant mitochondrial sequence transcribed in transgenic tobacco chloroplasts is not edited. Mol. Cell. Biol. *15*, 1377–1381.

Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., and Brennicke, A. (2008). The process of RNA editing in plant mitochondria. Mitochondrion *8*, 35–46.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B., and Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc. Natl. Acad. Sci. U. S. A. *109*, 5104–5109.

Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PloS One *8*, e65343.

Tavares-Carreón, F., Camacho-Villasana, Y., Zamudio-Ochoa, A., Shingú-Vázquez, M., Torres-Larios, A., and Pérez-Martínez, X. (2008). The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial COX1 mRNA in yeast. J. Biol. Chem. *283*, 1472–1479.

Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., Brennicke, A., and Takenaka, M. (2008). Multiple specificity recognition motifs enhance plant mitochondrial RNA editing *in vitro*. J. Biol. Chem. *283*, 24374–24381.

Visomirski-Robic, L.M., and Gott, J.M. (1995). Accurate and efficient insertional RNA editing in isolated Physarum mitochondria. RNA N. Y. N *1*, 681–691.

Visomirski-Robic, L.M., and Gott, J.M. (1997). Insertional editing of nascent mitochondrial RNAs in Physarum. Proc. Natl. Acad. Sci. U. S. A. *94*, 4324–4329.

Wagoner, J.A., Sun, T., Lin, L., and Hanson, M.R. (2015). Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. J. Biol. Chem. jbc.M114.622084.

Williams, M.A., Tallakson, W.A., Phreaner, C.G., and Mulligan, R.M. (1998). Editing and translation of ribosomal protein S13 transcripts: unedited translation products are not detectable in maize mitochondria. Curr. Genet. *34*, 221–226.

Wilson, R.K., and Hanson, M.R. (1996). Preferential RNA editing at specific sites within transcripts of two plant mitochondrial genes does not depend on transcriptional context or nuclear genotype. Curr. Genet. *30*, 502–508.

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.

Wolf, P.G., Rowe, C.A., and Hasebe, M. (2004). High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. Gene *339*, 89–97.

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

Yagi, Y., Tachikawa, M., Noguchi, H., Satoh, S., Obokata, J., and Nakamura, T. (2013b). Pentatricopeptide repeat proteins involved in plant organellar RNA editing. RNA Biol. *10*, 1419–1425.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.-W., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature *504*, 168–171.

Yoshinaga, K., Iinuma, H., Masuzawa, T., and Uedal, K. (1996). Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. Nucleic Acids Res. *24*, 1008–1014.

Yoshinaga, K., Kakehi, T., Shima, Y., Iinuma, H., Masuzawa, T., and Ueno, M. (1997). Extensive RNA editing and possible double-stranded structures determining editing sites in the *atpB* transcripts of hornwort chloroplasts. Nucleic Acids Res. *25*, 4830–4834.

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. Cell Mol. Biol. *59*, 1011–1023.

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

Zeltz, P., Hess, W.R., Neckermann, K., Börner, T., and Kössel, 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.

Zeltz, P., Kadowaki, K., Kubo, N., Maier, R.M., Hirai, A., and Kössel, H. (1996). A promiscuous chloroplast DNA fragment is transcribed in plant mitochondria but the encoded RNA is not edited. Plant Mol. Biol. *31*, 647–656.

Zhang, F., Tang, W., Hedtke, B., Zhong, L., Liu, L., Peng, L., Lu, C., Grimm, B., and Lin, R. (2014). Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. Proc. Natl. Acad. Sci. *111*, 2023–2028.

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

Zhou, W., Karcher, D., and Bock, R. (2013). Importance of adenosine-to-inosine editing adjacent to the anticodon in an Arabidopsis alanine tRNA under environmental stress. Nucleic Acids Res. *41*, 3362–3372.

Zhou, W., Karcher, D., and Bock, R. (2014). Identification of enzymes for adenosine-to-inosine editing and discovery of cytidine-to-uridine editing in nucleus-encoded transfer RNAs of Arabidopsis. Plant Physiol. *166*, 1985–1997.

Zito, F., Kuras, R., Choquet, Y., Kössel, 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.

Chapter 2

Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing[1]

[1] J. Wagoner designed and generated the QED1 Y2H constructs and the complementation and mutagenized constructs, genotyped plants, performed protoplast and stable transformation assays, performed PPEs, and performed sequence analysis. T. Sun performed the Y2H assays and created RIP and ORRM1 Y2H constructs. L. Lin performed PPE assays and Western Blots.

Background: Pentatricopeptide repeat (PPR) proteins are site-specific RNA editing factors; many have C-terminal motifs of unknown function.
Results: Site-directed mutagenesis of two PPR-DYW domains significantly reduces editing efficiency.
Conclusion: Residues conserved in cytidine deaminases in the DYW domains are important for editing activity.
Significance: This study strengthens the evidence that the DYW domains of PPR proteins carry the deaminase activity necessary for C-to-U modification.

ABSTRACT

In angiosperm organelles, cytidines are converted to uridines by a deamination reaction in the process termed RNA editing. The C targets of editing are recognized by members of the pentatricopeptide repeat (PPR) protein family. Although other members of the editosome have begun to be identified, the enzyme that catalyzes the C-U conversion is still unknown. The DYW motif at the C-terminus of many PPR editing factors contains residues conserved with known cytidine deaminase active sites; however, some PPR editing factors lack a DYW motif. Furthermore, in many PPR-DYW editing factors, the truncation of the DYW motif does not affect editing efficiency, so the role of the DYW motif in RNA editing is unclear. Here, a chloroplast PPR-DYW editing factor, QUINTUPLE EDITING FACTOR 1 (QED1), was shown to affect five different plastid editing sites, the greatest number of chloroplast C targets known to be affected by a single PPR protein. Loss of editing at the five sites resulted in stunted growth and accumulation of apparent photodamage. Adding a C-terminal protein tag to QED1 was found to severely inhibit editing function. QED1 and RARE1, another plastid PPR-DYW editing factor, were discovered to require their DYW motifs for efficient editing. To identify specific residues critical for editing, conserved deaminase residues in each PPR protein were mutagenized. The mutant PPR proteins, when expressed in *qed1* or *rare1* mutant protoplasts, could not complement the editing defect. Therefore, the DYW motif, and specifically, the deaminase

residues, of QED1 and RARE1 are required for editing efficiency.

Vascular plants modify Cs post-transcriptionally to Us in chloroplasts and mitochondria in a process termed RNA editing. Cytidine can be converted to uridine by deamination; however, the identity of the cytidine deaminase that carries out the reaction has remained elusive. The plant organelle editosome is a small protein complex of 200-400 kD (1) whose entire composition is not yet known, though members of several Arabidopsis protein families have been found to be essential for efficient RNA editing (1-6). One hypothesis is that deaminase activity is provided by the DYW domain that is present C-terminally on many, but not all, pentatricopeptide (PPR) motif-containing editing factors. The DYW domain contains the HxE and CxxCH motifs that are conserved in cytidine deaminases (7). Short *cis*-elements upstream of C targets of editing are known to bind PPR proteins, which evidently serve as the recognition factors that determine which C will undergo editing (8-10). Recently, a recognition code for the interaction of PPR motifs with particular RNA nucleotides has been proposed (11-13).

Many PPR editing factors recognize only one editing site, but there are a number that can recognize multiple editing sites. For example, CLB19 is involved in the editing of two C targets in the chloroplast, *rpoA* C200 and *clpP* C559, while CRR22 and OTP84 can each recognize three different C targets in the chloroplast (14,15). We describe here a PPR-DYW editing factor, *QUINTUPLE EDITING 1* (QED1) that affects five different chloroplast C-targets, currently the largest number of chloroplast sites affected by a known PPR protein. We observed that the DYW domain of QED1 and of RARE1, another PPR-DYW protein we have previously studied (16), are essential for editing of their respective C targets. In order to discern the possible role of the cytidine deaminase signatures HxE and CxxCH found in the DYW domain of PPR proteins, we

performed site-directed mutagenesis and assayed the effect on editing in transfected mutant protoplasts. We also investigated whether the sequences comprising a putative second zinc-binding site in the DYW domain (17) were necessary for RNA editing.

EXPERIMENTAL PROCEDURES

Arabidopsis Lines--The T-DNA insertion line CSHL_GT13864 (Landsberg erecta ecotype) was obtained from the Arabidopsis Biological Resource Center (ABRC) and the presence of the T-DNA insert was verified through PCR with primers 5'-GCTTGCATAGTTGATGTTCTCG-3' and the Ds3-1 primer 5'-ACCCGACCGGATCGTATCGGT-3' (18)using BioRad Red Master Mix (Hercules, CA). The T-DNA insertion line SALK_092402C (Columbia-*0* ecotype) was also obtained from the ABRC and homozygosity of the mutants was confirmed. Arabidopsis seeds were imbibed in water at 4°C for three days and then seeded on MetroMix. Plants were grown in both long day (16h light, 8h dark, 79 µmol/m^2/sec light, 22°C) and short day (10h light, 14h dark 116 µmol/m^2/sec light, 22°C) conditions.

Complementation and Mutational Analysis--RARE1 and the 2217 bp wild-type coding region of QED1 were amplified using Phusion polymerase and TOPO cloned into PCR8/GW/TOPO (Invitrogen Carlsbad, CA**).** These constructs were used in LR Clonase II recombination reactions with pEXSG-EYFP (19) and a modified pBI121 vector (1,20) to generate the full-length, DYW, E+ and E truncated constructs driven by a 35S promoter. Site-directed mutagenesis of conserved deaminase residues was performed by two-step PCR. Both the full-length construct and the mutagenized product were digested with BamHI and SacI and the mutagenized product was ligated into the digested full-length construct. Oligonucleotides used for mutagenesis are listed in Table 2.2.

Generation of transgenic plants--A construct in the modified pBI121 vector carrying the QED1 wild-type coding region was transformed into *Agrobacterium tumefaciens* GV3101. Floral dip transformation of *qed1* CSHL_GT13864 homozygous mutants plants was performed as in (21) .

Protoplast Transfection--The cellulase/macerozyme solution was prepared following the protocol of Yoo et al. (22) and contained 1.5% cellulase R10, 0.4% macerozyme R10, 0.4 M mannitol, 20 mM KCl, 20 mM MES pH 5.7, 10 mM $CaCl_2$, 0.1% BSA. Arabidopsis plants were grown in short days (8 hr light). Fully expanded leaves were sliced into 0.5-1mm strips using razor blades with minimal wounding and immersed in the enzyme solution. The digestion mixture was vacuum-infiltrated for 2 minutes and incubated overnight in the dark without shaking. The protoplasts were harvested in round-bottomed tubes by 2 minutes centrifugation at 300g. The supernatant was discarded and the cells were resuspended in the W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES pH5.7). The protoplasts were left on ice for 30 minutes before 2 minutes centrifugation at 300g. The supernatant was discarded and MMG solution (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES pH 5.7) was used to resuspend the protoplasts. Cell number was counted using a hemocytometer. The final protoplast density was adjusted to 2 x 10^5 /ml. 30µg of plasmid DNA was used to transfect 300µl (6x 10^4) protoplasts as in Jen Sheen's protocol (23). After 30 minutes incubation, a three times volume of W5 solution was added into each tube to stop the transfection. The cells were spun down by 2 minutes centrifugation at 300g and then resuspended in 1 ml W5 solution. Transfected protoplasts were cultured for 3 days in 24-well tissue culture plate in the dark at room temperature to allow accumulation of protein and edited transcripts before RNA extraction. A gene encoding GFP targeted to the ER (24) was also transfected in a sample to monitor transfection efficiency.

Measuring Editing Extent--RNA extraction and RT-PCR was performed as previously described (25). PPE editing assays of VIGS and QED1 mutant lines was carried out as described by Peeters and Hanson (26), except with the use of fluorescently labeled primer (16). Assays of editing efficiency in protoplasts were performed with the amplification primers listed in Robbins et al

Yeast Two-hybrid Assays--Mature coding sequences of QED1, RIP1, RIP2, RIP3, RIP5 and RIP9 were amplified using Phusion polymerase and cloned into the PCR8/TOPO/GW vector. These constructs were used in LR Clonase II recombination reactions with pGADT7GW and pGBKT7GW vectors to fuse the GAL4 activation domain (AD) and DNA binding domain (BD) respectively to the N-termini of the mature coding sequences. The yeast two-hybrid assay was performed as described in (3).

Protein Extraction and Immunoblotting--Samples were stored at -20°C before polyacrylamide gel electrophoresis. 6x SDS loading buffer (300mm Tris-HCl pH 6.8, 60mm EDTA, 12% SDS, 36% glycerol, 6% β-ME, 0.024% bromophenol blue) was added to samples and heated before loading onto a pre-cast gradient SDS-PAGE gel (Bio-Rad Mini-PROTEAN TGX Any kD) After transfer to a nitrocellulose membrane, protein was probed with anti-FLAG antibody (Sigma-Aldrich, St. Louis, MO).

Sequence Analysis--DNA and protein alignments were achieved using Jalview version 2.0 sequence analysis software (27).

RESULTS

Identification of a Candidate Editing Factor--We identified the pentatricopeptide repeat (PPR)-encoding gene At2g29760 as a candidate editing factor by a previously described candidate gene approach (16). Briefly, the Arabidopsis genome was surveyed to identify genes encoding a PPR protein carrying a DYW motif, as a number of known editing factors are PPR-DYW proteins. These genes were then screened to identify ones that were predicted to be chloroplast-targeted according to both Target P and Predotar. The predicted chloroplast targeted proteins were used in reciprocal best-hit analysis with the Arabidopsis and rice nuclear genome to identify ones that had no rice orthologs. Such genes were therefore candidates for editing of C targets that occur in Arabidopsis but not rice. RARE1, which encodes a PPR-DYW editing factor for a site in the Arabidopsis *accD t*ranscript, and that does not exist in rice, was previously identified by this strategy (16).

Expression of AT2G29760 is Required for the Editing of Five Sites on Five Different Chloroplast-encoded Transcripts--To determine whether At2g29760 is required for C-to-U editing events at one or more of the known chloroplast editing sites, Virus-Induced Gene Silencing (VIGS) trials were carried out. Silencing of At2g29760 was achieved by introducing amplified gene-specific DNA (Fig. 2.1), as defined in the Complete Arabidopsis Transcriptome MicroArray (CATMA) database, in a modified co-silencing vector (16). The vector allows direct visual identification of tissue exhibiting GFP silencing. In GFP silenced plants, five of 34 sites assayed by poisoned primer extension (PPE), on five different transcripts, exhibited significantly lower levels of editing: *accD* C1568, *matK* C640, *ndhB* C872, *rpoB* C2432 and *rps12* i1C58. Notably, all five of these sites are either absent or not edited in rice. Editing efficiency of *accD* C794, in *cis* of *accD* C1568, was not impacted by the silencing of At2g29760, nor sites in *cis*

relative to *ndhB* C872 and *rpoB* C2432 (data not shown). Because five editing sites are affected, the gene was named QED1 for Quintuple Editing factor 1. QED1 is a PPR protein containing 14 PPR motifs organized into four PLS blocks as well as an E/E+ and DYW motif (Fig. 2.1).

Homozygous mutant (HM) lines SALK_092402C and CSHL_GT13864 were acquired to assay C-to-U editing at each of the five sites identified by VIGS. Plants of line SALK_092402C (ecotype Columbia) were genotyped and determined to have a T-DNA insertion positioned 15 nucleotides upstream of the QED1 start codon (Fig. 2.1). The 5' UTR is at least 164 nt, as a database search located a QED1 EST that was created by 5' RACE (EST EG446305.1) that extends 164 nt upstream of the QED1 start codon. The CSHL line contains an insertion in the E domain (Fig. 2.1). No QED1 RNA was detected in the two-week old seedling rosette leaves of the CSHL line after 40 cycles of RT-PCR (data not shown). Homozygous mutant plants carrying either mutant allele demonstrated a significant reduction in editing efficiencies of the same five sites found by gene silencing (Fig. 2.2). No editing was detected at any of the five sites in the CSHL plants, which carry the coding region insertion. As this mutant allele exhibits a stronger editing phenotype than the other allele, subsequent experiments utilized the CSHL line. In order to confirm that the *qed1* allele in the CSHL line specifies the defects in RNA editing, we introduced the wild-type coding region into the mutant background. Editing of all five sites was restored, as can be seen in PPE assays of two representative plant transformants (Fig. 2.3).

Editing of three (*matK* C640: H-to-Y; *ndhB* C872: S-to-L; *rpoB* C2432: S-to-L) of the five sites leads to amino acid changes that could affect protein function, while the other two sites reside in non-coding regions. *accD* C1568 lies within the 3' UTR and *rps12* i1C58 is within the first intron of *rps12*, which is trans-spliced (28,29). Nevertheless, plants homozygous for either the weak or strong allele appeared morphologically normal growing in soil in growth chambers.

The CSHL line (ecotype Landsberg erecta) grows more slowly than wild-type and accumulates yellow regions in the leaves after 5 weeks of growth (Fig. 2.4).

Similar Sequences Are Present Upstream of the Five Editing Targets Controlled by QED1--Because *cis*-elements for recognition of RNA by PPR proteins have been found immediately upstream of the C targets of editing, we made an alignment of the 17 nucleotides 5' to the C target in each of the five sites, which revealed considerable sequence similarity (Fig. 2.5). These sequences were considered with respect to the recognition code that has been proposed for interaction of PPR motifs with RNA nucleotides (11-13). The amino acids at the 6 and 1' position were identified and the sequence predicted by these amino acid combinations according to Barkan et al. (11) and Takenaka et al. (12) was generated (Fig. 2.5). When this predicted sequence was aligned with the *cis*-element alignment, we found that there were seven nucleotides predicted by QED1's PPR motifs that were conserved in all five *cis*-elements (nucleotides indicated in red). There were four more nucleotides predicted by QED1's PPR motifs that were found in one or two *cis*-elements, but not all (nucleotides indicated in pink). The *cis*-elements of *rps12* i1C58 and *rpoB* C2432 had two more nucleotides that were predicted with the PPR code outside of the conserved six (Fig. 2.5). There are a number of *cis*-element positions not predicted by the PPR code that exhibit a large degree of sequence identity among the five *cis*-elements. For example, at position 2, a C or U is predicted by the PPR code, but four out of the five *cis*-elements have an A encoded, with the remaining *cis*-element encoding a G.

The QED1 sequences and *cis*-element sequences were also analyzed with respect to the PPR code proposed by Yagi et al. (13) where, in addition to the amino acids at the 6 and 1' position [referred to as 4 and ii in (13)] a third amino acid at position 3 (under the 6 and 1' designation) or position 1 (under the 4 and ii designation) is thought to influence the nucleotide

binding preference by the PPR motifs. The predicted RNA sequence generated by these combinations of three amino acids was very similar to what was predicted by the two amino acid code. However at position 2, where a C or U is predicted by the two amino acid PPR code, an A or G is predicted with the three amino acid code which is what is actually observed at that position. Additionally, at position 11, where only a G is predicted by the two amino acid PPR code, a G or a U is predicted by the three amino acid PPR code, which is what is observed in all five *cis*-elements.

Interaction of QED1 With Other Known Editing Factors--Recently, protein factors other than PPR proteins have been found to be essential for editing in plant organelles. Members of the RIP/MORF protein family have been shown to be either plastid or mitochondrial editing factors and one is active in both organelles (1,4,30). ORRM1, which belongs to a clade of RNA Recognition Motif (RRM) proteins, has been shown to be essential for editing of a subset of chloroplast C targets (3). We were therefore interested in determining whether QED1 interacts with any RIP family members or ORRM1 in Yeast Two-Hybrid (Y2H) assays. QED1 was fused with the Gal4 DNA binding domain (BD) and co-expressed with members of the RIP family (RIP1, RIP2, RIP3, RIP5, and RIP9) as well as ORRM1 fused with the activation domain (AD) in yeast using growth on minimal media (-histidine) as a reporter for protein-protein interactions. QED1 was also fused with the AD and co-expressed with BD-QED to test for possible homodimerization. Each gene was also co-expressed with the empty vector of the opposite vector type to ensure any yeast growth that was observed was not caused by auto-activation. Growth of yeast when BD-QED1 was co-expressed with each of the RIP family members indicated significant protein-protein interaction (Fig. 2.6). Unexpectedly, no protein-protein interaction was detected when BD-QED1 was co-expressed with AD-ORRM1 despite ORRM1

affecting the same sites as QED1 (Fig. 2.6A). Assays with BD-ORRM1 are uninformative because of autoactivation (data not shown). When BD-QED1 was co-expressed with AD-QED1, some growth was observed indicating a weak protein-protein interaction (Fig. 2.6B).

A C-terminal Tag Interferes with QED1 Editing while an N-terminal Tag Does Not--In order to produce a tagged QED1 for use in biochemical and complementation studies, we produced a construct with a 3xFLAG-StrepII tag fused to the C-terminus of QED1, separated by a GGSGGSGS linker. When we transfected the C-tagged QED1 into *qed1* protoplasts, we observed no complementation. We then moved the 3xFLAG-StrepII tag to the N terminus following a RecA chloroplast transit sequence, and placed the GGSGGSGS linker between the tag and the PPR motifs. This construct was able to restore editing (Fig. 2.7). These results are consistent with reports that several mitochondrial PPR-DYW editing factors cannot tolerate C-terminal fusions to GFP nor to a His-tag (31,32). It is likely that C-terminal tags interfere with the function of the C-terminal DYW motif in some PPR-DYW proteins.

Deletion of the E and/or DYW motifs of QED1 and RARE1 Results in Greatly Reduced Editing Efficiency--We examined the effect of deletions of the C-terminal region of QED1 on editing efficiency by comparing the restoration of editing efficiency in mutant protoplasts when the entire coding region of QED1 was transfected vs. constructs deleted beginning with the E motif, the E+ motif, or the DYW motif. Each of these QED1 constructs were cloned into an expression vector under a 35S promoter and expressed in *qed1* protoplasts. A GFP marker that labels the ER was also expressed in the *qed1* negative control in order to visualize transfection efficiency (>80%). The editing efficiencies of the five editing sites controlled by QED1 in each set of transfected protoplasts were measured using PPE (Fig. 2.8A). Wild-type Landsberg erecta protoplasts and *qed1* protoplasts were transfected with GFP to control for transfection effects on

editing efficiency. The full-length QED1 construct was able to complement the *qed1* mutant's editing phenotype at all five sites, restoring editing to about 40%, the lower efficiency than in wild-type likely due to incomplete transfection or inadequate expression levels. Editing was almost undetectable when constructs with truncations encompassing either the E or E+ domain were transfected, and only a very low level of editing is seen when the DYW deletion construct was transfected (Fig. 2.8A).

Several PPR-DYW editing factors had been reported to complement their respective mutants' editing defects when their DYW domains were deleted (17,33,34). We therefore investigated whether RARE1, another PPR-DYW protein we had previously studied (16), could also tolerate a DYW deletion and restore editing to *rare1* mutant protoplasts. Very little editing was detected when the RARE1 DYW truncation was transfected into mutant protoplasts (Fig. 2.8B).

Further Analysis of the E and DYW Domains--The lack of complementation by truncated proteins allowed us to carry out mutagenesis of the E and DYW domains of QED1 and RARE1. The DYW motif in editing PPR-DYW factors shares some sequence and structural similarities to deaminases, including highly conserved residues that resemble the canonical HxE/CxxC deaminase active sites (Fig. 2.9) (35,36). To investigate whether these specific residues are important to the editing of C-targets of QED1 and RARE1, we performed site-directed mutagenesis on the highly conserved HSE and CxxCH residues in the DYW motif in QED1 and in RARE1 (Fig. 2.10A, 2.11A).

For QED1, we created various constructs that targeted these conserved deaminase residues (Fig. 2.10A). In one construct, the HSE residues were all mutated to alanines (HSE) while in another construct, the five residues including the conserved cysteines (CDGCH) were

all mutated to alanines. Because mutating five sequential residues could have an effect on the stability or folding of QED1, further constructs were created that targeted only the two cysteines.

Constructs were made in which the two cysteines were changed to two alanines or to two serines, the latter amino acid having greater structural similarity to cysteine than does alanine. Each of these mutated constructs were cloned into an expression vector under a 35S promoter and expressed in *qed1* protoplasts along with full length QED1 and ΔDYW for comparison. When we assessed editing at the five editing sites controlled by QED1, we observed the protoplasts transfected by the mutated constructs exhibited little to no complementation of the editing defect. The QED1ΔDYW and QED1 with mutated HSE were able to support a greater level of editing than proteins in which the CxxCH signature was altered (Fig. 2.10B).

To investigate whether the cytidine deaminase signature of RARE1 was also essential, we performed site-directed mutagenesis on the DYW motif of RARE1. The HSE residues were again all mutated to alanines (H), while the targeted cysteines were mutated to serines (S) (Fig 2.11A). These mutated versions of RARE1 expressed under a 35S promoter in *rare1* mutant protoplasts along with full length RARE1 and RARE1ΔDYW for comparison. When editing at *accD* C794 was measured in each of these sets of protoplasts, the full-length RARE1 was able to complement the editing defect while neither the protoplasts transfected with the ΔDYW construct nor any of the protoplasts transfected with the mutated constructs showed any complementation of the editing defect (Fig. 2.11B).

Two Terminal Cysteines in QED1's DYW Motif are Important for Editing Function--We also investigated additional C-terminal regions of QED1 by site-directed mutagenesis. Hayes et al (17) postulated that the deaminase-like residues CxxCH as well as CSC and a nearby H, four highly conserved residues near the C-terminus of the DYW motif bind the two zinc ions they

detected in PPR-DYW factors ELI1 and DYW1. Hayes et al. (17) also observed that a truncation of ELI1 at the "PG box," a 15 amino acid sequence spanning the boundary of the E and E+ motif that is highly conserved in editing PPR-DYWs, was unable to complement an *eli1* mutant. The three terminal residues containing the two cysteines were mutated to alanines along with three well-conserved residues of the PG box, PGC (Fig. 2.11A).

When these mutated versions of QED1 were expressed under a 35S promoter in *qed1* protoplasts along with full length QED1 and QED1ΔDYW for comparison, we found that editing was severely reduced when the terminal cysteines were mutated (Fig. 2.11B). Editing was only slightly reduced when PGC was altered to AAA (Fig. 2.11B).

QED1 Mutated through Site-Directed Mutagenesis is Stably Expressed in Protoplasts--A lack of complementation by a construct carrying a mutated coding region could occur if the mutant protein is highly unstable. While it seems unlikely that multiple constructs targeting different residues could all fail to accumulate, we examined expression of two constructs by tagging them with a 3xFLAG-StrepII tag at the N terminus (Fig. 2.12). The mature wild-type QED1 and QED1 (CxxCH mutated to SxxSH) were tagged at their N-termini with a 3xFLAG-StrepII tag as well as a RecA chloroplast targeting sequence and a GGSGGSGS linker. These N-tagged constructs were expressed from a 35S promoter in *qed1* mutant protoplasts. RNA was extracted from a portion of the protoplasts and the remainder of the cells was lysed to obtain protein for immunoblots. The protein extract from each set of transfected protoplasts, as well as protein extract from untransfected wild-type and *qed1* protoplasts and along with bacteria alkaline phosphatase (BAP) tagged with FLAG (Sigma), were then probed with α-FLAG antibody to probe for expression of the N-tagged constructs (Fig. 2.12). As expected, no signal was observed in the wild-type and *qed1* protein extracts from protoplasts that were not

transfected by the FLAG-tagged constructs. In each of the samples of protoplasts that were transfected by the FLAG-tagged versions of QED1, we observed a band of about 80 kD which is the expected size of mature QED1 with the added FLAG-StrepII tag (Fig. 2.12).

DISCUSSION

Until this report, no chloroplast PPR protein had been found to control more than 3 editing sites. We have found that QED1 is responsible for the editing of five different C-targets in five different chloroplast genes, *accD, matK, ndhB, rpoB*, and *rps12.* Three of these sites (*matK, ndhB, rpoB*) are found in the coding region of the gene, while the two other sites affect non-coding regions. A PPR protein had been assigned as a site-specific editing factor only to the site within the first intron of *rps12*. While this work was in progress, another group published RT-PCR bulk sequencing assays of mutants containing only the weak *qed1* allele from the Salk collection in the gene they named OTP81, and they detected a deficiency only in *rps12* intron editing (15). They reported that failure to edit this site did not affect splicing of *rps12* (15).

Our truncations of the E or E+ motif resulted in a complete loss of editing activity, a result consistent in other studies in which an editing PPR protein was truncated before the E or E+ motif (17,33,37). The E/E+ motif is an essential part of the RNA editing reaction, although its role is unclear. Although there is higher conservation between the DYW motifs of editing factors, the E/E+ motif still shows a higher degree of conservation compared to the PPR motifs. The E/E+ motif also appears to contain two particularly degenerate PPR motifs which could serve in an RNA or protein binding role. If the DYW motif is the deaminase, perhaps the E/E+ motif plays a role in recruiting or mediating protein-protein interactions between the factors that provide the RNA editing activity, which may include DYW motifs that act in *trans.* E motifs

swapped between two mitochondrial and two chloroplast editing factors could not restore editing at those specific sites, but E domains exchanged with an E domain from a protein targeted to the same organelle did give a functional chimeric protein (38), implying there may be some organelle specialization in the function of these motifs.

PPR-DYW proteins undoubtedly interact with other components of the protein complexes that carry out plant organelle RNA editing. Recent studies of non-PPR protein editing factors have revealed that the PPR-DYW site recognition factors are not the only proteins that act site-specifically to mediate editing. For example, RIP1, RIP2, RIP9, and ORRM1 are each essential for editing of particular chloroplast C targets, but dispensable for others (1,3,30). However, mutation in *RIP2, RIP9,* or *ORRM1* affects the efficiency of editing of the five sites controlled by QED1, suggesting the RIP and ORRM1 proteins may act together with QED1 within a subset of editosomes. As would be predicted from such a model, RIP1, RIP2, and RIP9 all interacted with QED1 in our Y2H assays. Unexpectedly, no interaction was detected between QED1 and ORRM1. One possible explanation is that QED1, plastid RIP proteins and ORRM1 are in the same complex, but QED1 directly associates with RIP proteins but not with ORRM1, which may instead interact only with one or more RIP proteins. It is also possible that QED1 does interact with ORRM1 in chloroplast, but not in yeast. If QED1/ORRM1 interaction requires the presence of the chloroplast RNA transcript carrying the *cis*-element to which QED1 binds, the interaction will not occur in yeast where the transcript does not exist. However, ORRM1 has been shown to directly interact with other editing PPR proteins in yeast, so this latter explanation is less likely.

The chloroplast factor QED1 also interacts with two mitochondrial RIP proteins—RIP3 and RIP5. A similar promiscuous interaction between PPR proteins and RIP protein was also

reported by Takenaka et al. (4) in which mitochondrial PPR editing factors MEF1, MEF9, and MEF21 interacted with the plastid RIP proteins RIP2 and RIP9 (also named MORF2 and MORF9). Such interactions are unlikely to occur *in planta* because the PPR proteins are compartmentalized in different organelles.

QED1 exhibited weak interaction with itself in the yeast two-hybrid assay. A PPR protein named PPR10, was reported to form an antiparallel homodimer in RNA-free crystals (39). However, the authors also reported that binding of RNA weakened the homodimerization. Another PPR protein, THA8, was also reported to form an asymmetric homodimer, but unlike PPR10, RNA binding induced dimerization (40). As yeast does not contain the target chloroplast RNA with which QED1 normally interacts, it is possible that no dimerization of QED1 occurs *in vivo*. However, the presence of the target chloroplast RNA could also strengthen the interaction.

Salone et al. (7) hypothesized that the DYW motif served as the deaminase in the RNA editing reaction because it contains a set of residues that are not only highly conserved between DYW motifs, but are also highly conserved with residues that comprise the active site in known deaminases. The predicted DYW protein structure also correlates with the actual or predicted structure of other known deaminases (35). Furthermore, evolutionary analysis indicates that the presence of the DYW motif correlates with the presence of C to U editing in plants (36,41-43). An issue regarding this hypothesis is the observation that the DYW motif is not needed for the editing function of four PPR-DYW editing factors (CRR22, CRR28, OTP82, and ELI1) (17,33,34) and there are editing factors with an E motif that lack a DYW motif (2,14,37,44). A possible explanation that is consistent with a role of the DYW motif in the C-to-U modification is the presence of additional DYW-containing proteins in the editosome, in those cases where a required PPR protein lacks one or carries a dispensable motif. This hypothesis was strengthened

by the finding that the protein DYW1, which lacks any PPR motifs, can act with CRR4, a PPR-E editing factor, to edit *ndhD* C2 (45). Furthermore, when DYW1 was fused with CRR4, the fusion protein resembled a complete PPR-DYW protein with DYW1 supplying the missing DYW motif plus a partial E+ motif that was also missing in CRR4 (45). The mitochondrial PPR protein MEF1 also requires its DYW motif in order to restore editing to *mef1* protoplasts (32).

We have discovered two more editing factors, QED1 and RARE1, which not only require the DYW motif for efficient function, but specifically, the conserved deaminase residues. In some cases, QED1 lacking its DYW motif conferred a greater amount of editing than samples with site-specific mutations within the motif. This finding might be due to another DYW-motif-containing protein being able to act *in trans* at low efficiency when the DYW motif of QED1 is absent. We have verified that some of the mutant proteins accumulate in chloroplasts, so the lack of editing activity cannot be ascribed to absence of the protein. When we mutated the cysteines in the putative second zinc binding site described by Hayes et al (17), we observed an effect on RNA editing similar to what was observed when we mutated the cysteines in the deaminase motif. We also observed that the placement of a 3xFLAG-StrepII tag at the C-terminus of QED1 greatly interferes with RNA editing function, while a 3XFLAG-StrepII tag at the N-terminus does not, which also indicates that the C-terminal DYW motif's function is critical for editing activity.

Though we have identified PPR-DYW proteins that require their DYW domain for editing activity, it is still unclear why other PPR-DYW proteins, such as CRR22 and CRR28, can function without the DYW domain. The composition of editosomes is diverse, with different members of other genes families present in editosomes with certain PPR proteins but not with others. Until more is known about the complete composition of the editosome and its molecular

structure, it will not be possible to understand why the DYW domain is required in some PPR proteins but not in others.

While this manuscript was in preparation, an article by Boussardon et al. (46) appeared in which zinc binding and the deaminase signature of DYW1 was analyzed by mutagenesis. In contrast to the work by Hayes et al (17) that described binding of two zinc atoms by PPR-DYW editing factors, ELI1 and DYW1, binding of only one zinc by a GST fusion with DYW1 was reported by Boussardon et al.(46), which is at odds with our finding that the putative second zinc binding site is important for RNA editing in QED1 and the previously published results of Hayes et al (17). Unlike the proteins we have analyzed, DYW1 lacks PPR motifs, but apparently works in conjunction with the PPR protein CRR4 to mediate editing to specify an AUG start codon in the chloroplast *ndhD* transcript. The HxE and CxxC residues conserved in DYW1 were mutated, resulting in a loss of editing (46), which is consistent with our finding that these residues characteristic of cytidine deaminases also present in PPR-DYW factors are essential for chloroplast RNA editing.

REFERENCES

1. Bentolila, S., Heller, W. P., Sun, T., Babina, A. M., Friso, G., van Wijk, K. J., and Hanson, M. R. (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1453-1461
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. Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M. R., and Bentolila, S. (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1169-E1178
4. Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Hartel, B., and Brennicke, A. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5104-5109
5. 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
6. Zhang, F., Tang, W., Hedtke, B., Zhong, L., Liu, L., Peng, L., Lu, C., Grimm, B., and Lin, R. (2014) Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2023-2028
7. Salone, V., Rüdinger, 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
8. Bock, R., Hermann, M., and Kossel, H. (1996) *In vivo* dissection of *cis*-acting determinants for plastid RNA editing. *EMBO J.* **15**, 5052-5059
9. 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
10. 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
11. Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8**
12. Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* **8**, e65343
13. Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* **8**
14. Chateigner-Boutin, A. L., Ramos-Vega, M., Guevara-Garcia, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., Jimenez, L. F., Lurin, C., Small, I., and Leon, P. (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J.* **56**, 590-602

15. 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
16. 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
17. Hayes, M. L., Giang, K., Berhane, B., and Mulligan, R. M. (2013) Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J. Biol. Chem.* **288**, 36519-36529
18. Tsugeki, R., Kochieva, E. Z., and Fedoroff, N. V. (1996) A transposon insertion in the Arabidopsis SSR16 gene causes an embryo-defective lethal mutation. *Plant J.* **10**, 479-489
19. Jakoby, M. J., Weinl, C., Pusch, S., Kuijt, S. J. H., Merkle, T., Dissmeyer, N., and Schnittger, A. (2006) Analysis of the subcellular localization, function, and proteolytic control of the Arabidopsis cyclin-dependent kinase inhibitor ICK1/KRP1. *Plant Physiol.* **141**, 1293-1305
20. 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
21. 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
22. Yoo, S. D., Cho, Y. H., and Sheen, J. (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565-1572
23. Sheen, J. (2001) Signal transduction in maize and Arabidopsis mesophyll protoplasts. *Plant Physiol.* **127**, 1466-1475
24. 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
25. Bentolila, S., Elliott, L. E., and Hanson, M. R. (2008) Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* **178**, 1693-1708
26. 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
27. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191
28. Hildebrand, M., Hallick, R. B., Passavant, C. W., and Bourque, D. P. (1988) Trans-splicing in chloroplasts: the *rps12* loci of *Nicotiana tabacum*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 372-376
29. Koller, B., Fromm, H., Galun, E., and Edelman, M. (1987) Evidence for i*n vivo* trans-splicing of pre-mRNAs in tobacco chloroplasts. *Cell* **48**, 111-119
30. Bentolila, S., Oh, J., Hanson, M. R., and Bukowski, R. (2013) Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet.* **9**
31. Brehme, N., Zehrmann, A., Verbitskiy, D., Hartel, B., and Takenaka, M. (2014)

Mitochondrial RNA editing PPR proteins can tolerate protein tags at E as well as at DYW domain termini. *Front. Pl. Sci.* **5**, 127
32. Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., and Takenaka, M. (2010) RNA editing competence of trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain. *FEBS Lett.* **584**, 4181-4186
33. 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
34. Okuda, K., Hammani, K., Tanz, S. K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. *Plant J.* **61**, 339-349
35. Iyer, L. M., Zhang, D., Rogozin, I. B., and Aravind, L. (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.* **39**, 9473-9497
36. 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
37. 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
38. Chateigner-Boutin, A.-L., Colas des Francs-Small, C., Fujii, S., Okuda, K., Tanz, S. K., and Small, I. (2013) The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J.* **74**, 935-945
39. Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H. W., Wang, J., Zhu, J. K., Shi, Y., and Yan, N. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168-171
40. Ke, J., Chen, R. Z., Ban, T., Zhou, X. E., Gu, X., Tan, M. H., Chen, C., Kang, Y., Brunzelle, J. S., Zhu, J. K., Melcher, K., and Xu, H. E. (2013) Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nature Struct. Mol. Biol.* **20**, 1377-1382
41. Rudinger, M., Polsakiewicz, M., and Knoop, V. (2008) Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol. Biol. Evol.* **25**, 1405-1414
42. Knoop, V., and Rudinger, M. (2010) DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett.* **584**, 4287-4291
43. Rudinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V. (2012) Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J. Mol. Evol.* **74**, 37-51
44. Takenaka, M., Verbitskiy, D., Zehrmann, A., and Brennicke, A. (2010) Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J. Biol. Chem.* **285**, 27122-27129
45. Boussardon, C., Salone, V., Avon, A., Berthomé, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C. (2012) Two interacting proteins are necessary for the editing of the *ndhD-1* site in Arabidopsis plastids. *Plant Cell* **24**, 3684-3694

46. Boussardon, C., Avon, A., Kindgren, P., Bond, C. S., Challenor, M., Lurin, C., and Small, I. (2014) The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of *ndhD-1*. *New Phytol.* **203**, 1090-1095

FOOTNOTES

*****This work was supported by NSF grants MCB-0929423 and MCB-1330294 to M.R.H.
[1]To whom correspondence may be addressed: Department of Molecular Biology and Genetics, Cornell University, Biotechnology Building, Ithaca, NY14853, USA, Tel: (607) 254-4833; Fax: (607) 255-6249; E-mail: mrh5@cornell.edu

The abbreviations used are AD (Activation Domain), BAP (Bacterial Alkaline Phosphatase), BD (Binding Domain), CATMA (Complete Arabidopsis Transcriptome MicroArray), GFP (Green Fluorescent Protein), QED1 (Quintuple Editing Factor 1), PPE (Poisoned Primer Extension), PPR (pentatricopeptide repeat), RARE1 (Required for *accD* RNA editing), RIP (RNA editing factor Interacting Protein), RRM (RNA Recognition Motif), Y2H (yeast two-hybrid assay).

FIGURE LEGENDS

FIGURE 2.1. **Structure of the At2g29760 locus and the predicted protein**. The locations of two insertion alleles and their corresponding insertion lines are indicated. The 168 nt region of transcripts of At2g29760 targeted by VIGS is delimited by facing arrows. Depicted below the gene model is the modular organization of the predicted QED1 protein. Starting at residue 61, four PLS blocks (I-IV) and two S repeats (comprising two S-PLS-PLS blocks) are shown followed by the E, E+ and DYW domains. The P, L, and S forms of the PPR motifs are indicated left to right by three different shadings of green boxes. Figure is drawn to scale. Two T-DNA lines from the ABRC were obtained: SALK 092402C (*qed1-1*) located -15nt from the start codon and CSHL GT13864 (*qed1-2*) located in the E+ motif 1852 nt into the coding region and homozygous mutants were obtained. Nucleotides 2-169 of the RNA were chosen to target for VIGS.

FIGURE 2.2. **QED1 affects five editing sites in the chloroplast**. The effect of QED1 on chloroplast editing was assayed through Virus Induced Gene Silencing (VIGS), and through analysis of homozygous mutant insertion lines *qed1-1* and *qed1-2*. Editing was assessed using Poisoned Primer Extension (PPE) assay. Edited product (e), unedited product (u), oligonucleotide (o), wild-type (+/+), mutant (-/-). PPE assays of ***A,*** *accD* C1568. ***B,*** *matK* C640. ***C,*** *ndhB* C872. ***D,*** *rpoB* C2432. ***E,*** *rps12* intron1 C58. *F,* Quantifications of edited versus unedited product for *qed1-1* and *qed1-2* compared to wild-type plants.

FIGURE 2.3: **Complementation of the *qed1* editing phenotype by stable transformation.** Using PPE, editing was assessed at all five sites in two independently generated plant transformants carrying the QED1 wild-type coding region in the *qed1-2* background. Edited product (e), unedited product (u). Wild-type (+/+), *qed1* mutant (-/-), transformant (T). PPE assays of *ndhB* C872, *rpoB* C2432, *matK* C640, *rps12* i1C58, *accD* C1568.

FIGURE 2.4. **Growth phenotype of the *qed1* mutant.** An 8.5-week-old homozygous mutant *qed1-2* plant (right) shows a growth defect compared to an 8.5-week-old Landsberg erecta wild-type plant (left) in growth room conditions. Yellow areas that might be due to photodamage (indicated by white arrows) accumulate in leaves as they age in *qed1* plants.

FIGURE 2.5. **Alignment of PPR motifs of QED1 with the *cis*-elements of the editing sites it targets**. The *cis*-elements (-16 to -1 nt) of the five editing sites affected by QED1 were aligned using Tcoffee. Residues shaded in blue indicate degree of conservation and the edited C is indicated by the arrow. The bolded amino acids are at the 6th position in the PPR motif while non-bolded amino acids are at the 1' position. The RNA sequence predicted to be recognized by QED1 was generated using both the Barkan et al. (11) PPR code (top sequence) and Yagi et al's (13) modified PPR code (bottom sequence) and aligned with the *cis*-elements of the five editing sites that QED1 targets. Nucleotides in red in the predicted sequence are residues that are conserved in all five *cis*-elements. Nucleotides in pink are only found in one or two of the five *cis*-elements which are boxed in red.

FIGURE 2.6. **QED1 interacts with members of the RIP family in a yeast two-hybrid assay.** ***A,*** Growth of yeast colonies on selective media when QED1 fused with the BD is combined with RIP family members (RIP1, RIP2, RIP3, RIP5 and RIP9) fused with the AD. When BD-QED1 was co-expressed with AD-ORRM1 in yeast, no growth was observed. *B,* Co-expression of AD-QED1 with BD-QED1 along with auto-activation controls.

FIGURE 2.7. **Ability of QED1 variants with N-terminal vs. C-terminal tags to complement *qed1* mutants.** *A,* Chimeric versions of QED1 with a 3xFLAG-StrepII (3FS) were cloned where the 3FS tag was either attached to the C-terminus of QED1 or at the N-terminus of the mature QED1 protein with the RecA target peptide (TP) and controlled by the 35S promoter (p35S). *B,* PPE assay of editing of the *ndhB* C872 site following protoplast complementation in *qed1* protoplasts. edited (e) and unedited (u) products (o) oligonucleotide. (+) wild type protoplasts, (-) *qed1* protoplasts transfected with GFP, (Full) transfection with full-length QED1, N3FS (N) results from transfection with QED1 tagged at the N-terminus and C3FS (C) results from transfection with QED1 tagged at the C-terminus. *C,* Quantification of editing efficiency from PPE assay.

FIGURE 2.8. **Truncation analysis of QED1 and RARE1** ***A,*** Full length QED1 (Full) and various truncations of QED1 were cloned into an expression vector under a 35S promoter and expressed in *qed1* mutant protoplasts each in two separate transfections. The arrows indicate the position of three truncations that progressively remove each of the C-terminal motifs of QED1. Data for the *matK* C640 and *rps12* i1C58 sites are shown. ***B,*** Full-length RARE1 and DYW-truncated RARE1 were cloned into an expression vector under a 35S promoter and expressed in *rare1* mutant protoplasts. Arrow indicates where the protein was truncated. WT, *qed1,* and *rare1* indicate editing efficiencies at the *accD* C794 site of wild type and mutant protoplasts, each transfected with a GFP gene. Full indicates editing efficiencies of full length proteins while ΔDYW indicate editing efficiencies of DYW-truncated QED1 or RARE1 each expressed in two separate transfections into the respective mutant protoplasts.

FIGURE 2.9. **C-terminal motif alignment of nine chloroplast PPR-DYW editing factors**. DYW, E+ and partial E motif protein sequences from nine chloroplast PPR-DYW editing factors (CRR22 NP_172596.1, CRR28 NP_176180.1, YS1 NP_188908.2, LPA66 NP_199702.1, RARE1 NP_196831.1, QED1 NP_180537.1, OTP82 NP_172286.1, OTP84 NP_191302.2, OTP85 NP_178398.1) were obtained from the NCBI database, aligned using Tcoffee, and

presented by Jalview. The intensity of the shading indicates the degree of conservation between the motifs. The beginnings of the E+ and DYW motifs are marked with green arrows. The conserved deaminase-like motifs (HxE, CxxCH) are denoted by the red lines along with the second putative zinc binding site and the PGC box designated by Hayes et al. (17).

FIGURE 2.10. **Mutagenesis of the cytidine deaminase signature residues in QED1 and RARE1 significantly reduces editing. *A,*** The deaminase residues in the DYW motif (designated in red) of QED1 were targeted using site-directed mutagenesis. Changes that were made are indicated above the targeted residues. *B,* PPE assay of editing in *qed1* protoplasts following two separate transfections of the same batch of protoplasts with full-length (full), GFP (*qed1*-), DYW-truncated QED1 (ΔDYW) and mutagenized versions of QED1 (CxxCH, HSE) compared to wild-type protoplasts (WT+). *C,* PPE assay of editing in *qed1* protoplasts using a different batch of protoplasts than in *B* and additional mutagenized versions of QED1 (AxxAH, SxxSH). PPE gel for *matK* C640 is shown; graphs of editing quantifications for *matK* C640 and *rps12*i1C58 PPE gels are provided. *D,* PPE assay of editing in the same batch of *rare1* protoplasts following two separate transfections with full-length *RARE1* (Full), *GFP* (rare1-), DYW-truncated *RARE1* (ΔDYW), and mutated versions SxxSH and HSE. Edited product (e), unedited product (u), oligonucleotide (o).

FIGURE 2.11. **Mutagenesis of a second potential zinc binding site in the DYW motif of QED1 significantly reduces editing** *A,* Residues in the E and DYW motif of QED1 were targeted for site-directed mutagenesis. Three residues (in red) in a conserved 15aa sequence called the PGC box as defined by Hayes et al (17) (in bold) and three residues (in red) near the end of the DYW motif of QED1 were mutated to alanines. ***B,*** Results from the PPE assay in wild-type protoplasts (WT) and following transfection of GFP (*qed1*- or *rare1*-), the full-length un-mutagenized construct (Full), and mutagenized constructs targeting the PGC box (PGC or P) and the terminal cysteines (CC) into *qed1* protoplasts. *matK* C640 PPE gel is shown; graphs of PPE data from *rpoB* C2432 and *matK* C640 sites are provided. Edited product (e), unedited product (u), oligonucleotide (o).

FIGURE 2.12. **Expression of full-length QED1 and mutagenized QED1 (SxxSH) in mutant protoplasts and effect on editing of three C targets.** *A,* Immunoblot probed with α-FLAG antibody. *qed1* protoplasts were transfected with a construct expressing either the entire wild-type QED1 coding region (full) or mutant QED1 (SxxSH), tagged on their N-termini with a 3xFLAG-StrepII tag. Protein was extracted from the transfected protoplasts, from un-transfected wild-type (WT) protoplasts, and *qed1* protoplasts transfected with a *GFP* gene (*qed1*). 4 μg protein from each sample was run on an SDS-PAGE. 10 μg of FLAG-BAP (Bacterial Alkaline Phosphatase) was included as a positive control. Mature QED1 plus the 3xFLAG-StrepII tag is predicted to be approximately 80 kD in size. *B,* Editing quantifications for *ndhB* C872, *rpoB* C2432, and *matK* C640 from RNA extracted from the same protoplast samples are shown underneath.

Table 2.1. Oligonucleotides

Primer Name	Sequence 5'-3'	Purpose
QED1 start	ATGGCTATCTTCTCCACAGCAC	amplification
QED1 stop	GAAATATCACCAGAAATCGTTAC	amplification
QED1 5'2531	GCTTGCATAGTTGATGTTCTCG	genotyping
Ds3.1	ACCCGACCGGATCGTATCGGT	genotyping
QED1 DYWstop	CTACAGTTTCTCCATAACCTCGTG	amplification
QED1 Eplusstop	CTACTCATGAATCATGCCATCGA	amplification
QED1 Edomainstop	CTAAGTGCTTGGAGGAATCGGCAT	amplification
QED1 BamHI	TAGGTAGATGGATCCATAGCTAC	cloning
pexyfp saci R	GGACTCTAGAGCTCTTACTTGTT	cloning
QED1 SalI-mat	TATATAGTCGACCATGGTCACATGATCCGTAC	cloning
QED1 stop-SpeI	GACTTCCACTAGTTCACCAGAAATCGTTACAGG	cloning
QED1 CxxCHmutF	ATTAAGAATCTAAGGGTGGCTGCAGCTGCTGCTTCAGTGGCTAAA	mutagenesis
QED1 CxxCHmutR	TTTAGCCACTGAAGCAGCAGCTGCAGCCACCCTTAGATTCTTAAT	mutagenesis
QED1 HSEmutF	GAACAATCTCTAAATCTCGCAGCAGCAAAGCTGGCTATTTGTTAT	mutagenesis
QED1 HSEmutR	ATAACAAATAGCCAGCTTTGCTGCTGCGAGATTTAGAGATTGTTC	mutagenesis
qed1 CC sermutF	ATTAAGAATCTAAGGGTGTCTGGAGATTCTCATTCAGTGGCTAA	mutagenesis
qed1 CC2 sermutR	TTAGCCACTGAATGAGAATCTCCAGACACCCTTAGATTCTTAAT	mutagenesis
QED1 PGC F	GAGTCACTGGACTTAAGAAAGAAGCAGCAGCCAGCTCAATTGAAATCGATG	mutagenesis
QED1 PGC R	CATCGATTTCAATTGAGCTGGCTGCTGCTTCTTTCTTAAGTCCAGTGACTC	mutagenesis
QED1 CCend mutF	CATCATTTCAGAAATGGGCAGGCTTCCGCTAACGATTTCTGGTGATATTTC	mutagenesis
QED1 CCend mutR	GAAATATCACCAGAAATCGTTAGCGGAAGCCTGCCCATTTCTGAAATGATG	mutagenesis
RARE1 start	CTATGACGATTCTCACTGTAC	amplification
RARE1 fullstop	GATCATCAGTCACCAGTAATCG	amplification
RARE1 dywstop	CTACATGAAACCATCAAACTCCT	amplification
RARE1 5'1566	GCAATCATCTCCGGATACTGTCA	sequencing
RARE1 HSEmutF	GAGAACAGCTTCTTGATGCTGCTGCGAGACTTGCCATTGCAT	mutagenesis
RARE1 HSEmutR	ATGCAATGGCAAGTCTCGCAGCAGCATCAAGAAGCTGTTCTC	mutagenesis
RARE1 CCsermutF	TTCAAGAATCTTCGGGCAAGCCCAGATAGTCATGAGTTTGCAAAGCAT	mutagenesis
RARE1 CCsermutR	ATGCTTTGCAAACTCATGACTATCTGGGCTTGCCCGAAGATTCTTGAA	mutagenesis

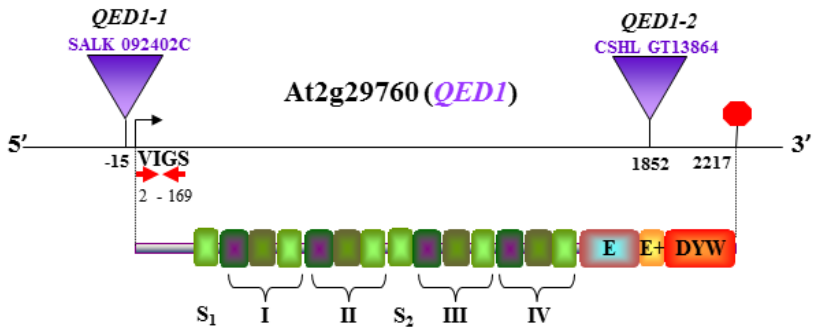


FIGURE 2.1. **Structure of the At2g29760 locus and the predicted protein**. The locations of two insertion alleles and their corresponding insertion lines are indicated. The 168 nt region of transcripts of At2g29760 targeted by VIGS is delimited by facing arrows. Depicted below the gene model is the modular organization of the predicted QED1 protein. Starting at residue 61, four PLS blocks (I-IV) and two S repeats (comprising two S-PLS-PLS blocks) are shown followed by the E, E+ and DYW domains. The P, L, and S forms of the PPR motifs are indicated left to right by three different shadings of green boxes. Figure is drawn to scale. Two T-DNA lines from the ABRC were obtained: SALK 092402C (*qed1-1*) located -15nt from the start codon and CSHL GT13864 (*qed1-2*) located in the E+ motif 1852 nt into the coding region and homozygous mutants were obtained. Nucleotides 2-169 of the RNA were chosen to target for VIGS.

FIGURE 2.2. **QED1 affects five editing sites in the chloroplast**. The effect of QED1 on chloroplast editing was assayed through Virus Induced Gene Silencing (VIGS), and through analysis of homozygous mutant insertion lines *qed1-1* and *qed1-2*. Editing was assessed using Poisoned Primer Extension (PPE) assay. Edited product (e), unedited product (u), oligonucleotide (o), wild-type (+/+), mutant (-/-). PPE assays of ***A,*** *accD* C1568. ***B,*** *matK* C640. ***C,*** *ndhB* C872. ***D,*** *rpoB* C2432. ***E,*** *rps12* intron1 C58. *F,* Quantifications of edited versus unedited product for *qed1-1* and *qed1-2* compared to wild-type plants.

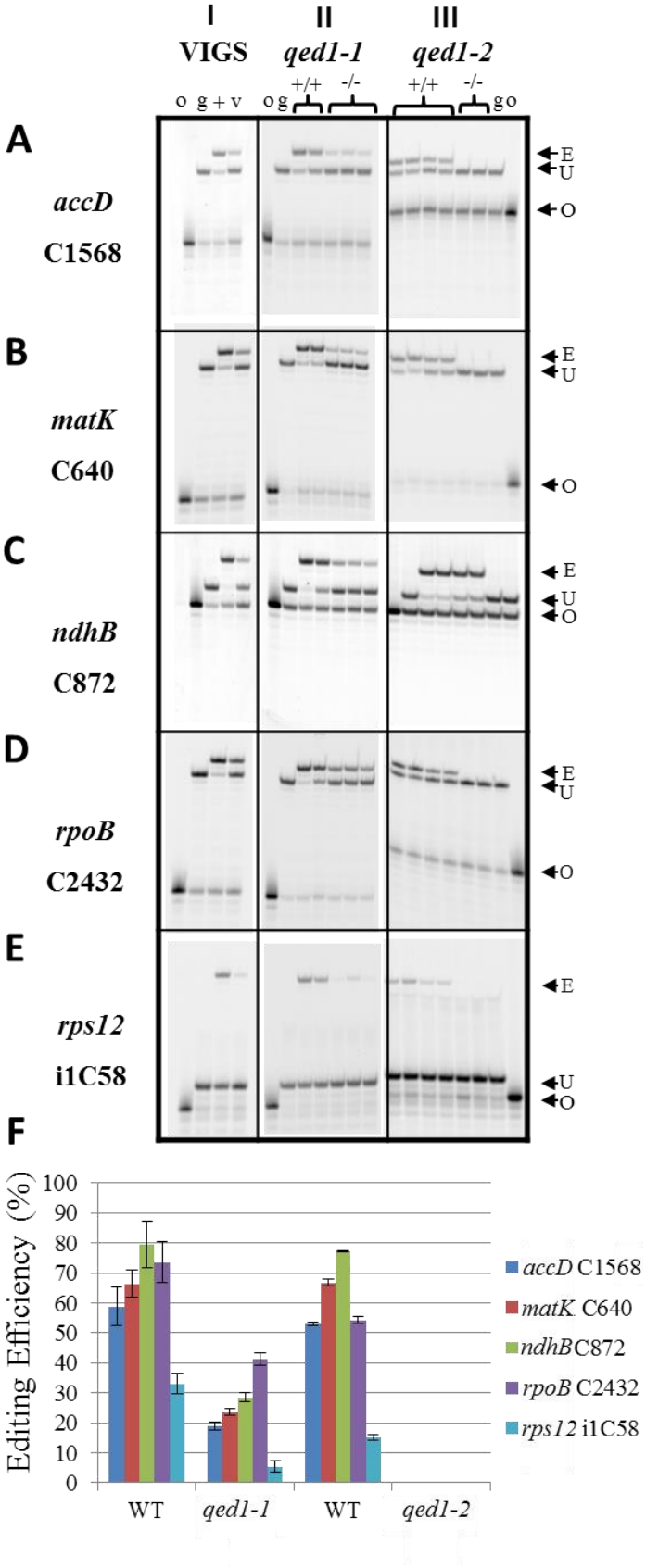


FIGURE 2.3: **Complementation of the *qed1* editing phenotype by stable transformation.** Using PPE, editing was assessed at all five sites in two independently generated plant transformants carrying the QED1 wild-type coding region in the *qed1-2* background. Edited product (e), unedited product (u). Wild-type (+/+), *qed1* mutant (-/-), transformant (T). PPE assays of *ndhB* C872, *rpoB* C2432, *matK* C640, *rps12* i1C58, *accD* C1568.

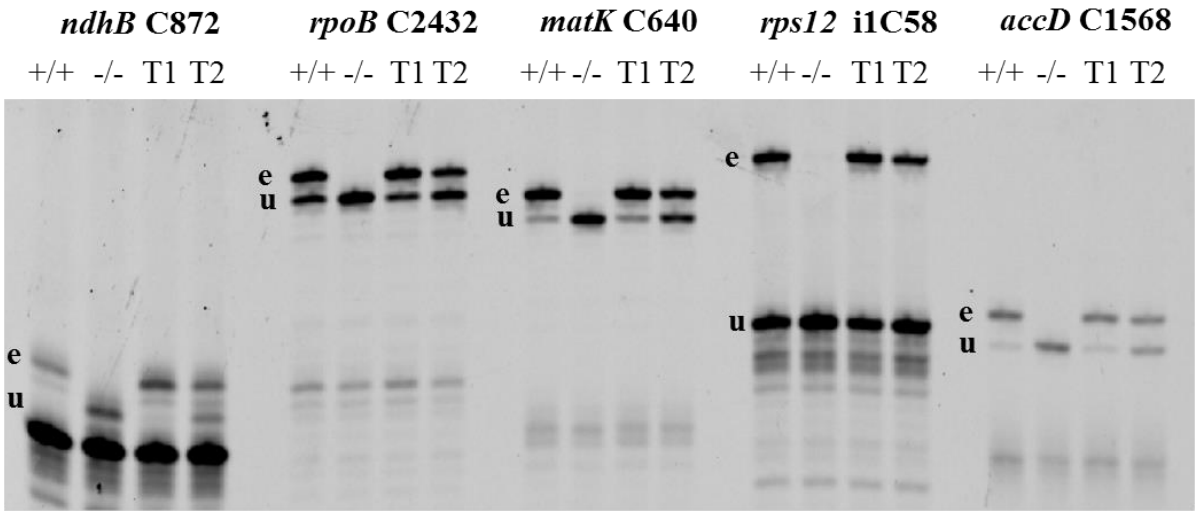


FIGURE 2.4. **Growth phenotype of the *qed1* mutant.** An 8.5-week-old homozygous mutant *qed1-2* plant (right) shows a growth defect compared to an 8.5-week-old Landsberg erecta wild-type plant (left) in growth room conditions. Yellow areas that might be due to photodamage (indicated by white arrows) accumulate in leaves as they age in *qed1* plants.

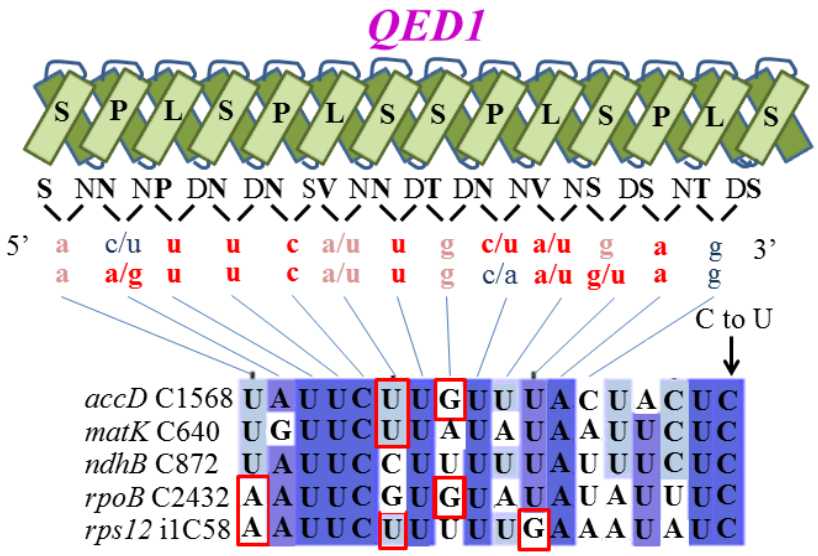


FIGURE 2.5. **Alignment of PPR motifs of QED1 with the *cis*-elements of the editing sites it targets**. The *cis*-elements (-16 to -1 nt) of the five editing sites affected by QED1 were aligned using Tcoffee. Residues shaded in blue indicate degree of conservation and the edited C is indicated by the arrow. The bolded amino acids are at the 6th position in the PPR motif while non-bolded amino acids are at the 1' position. The RNA sequence predicted to be recognized by QED1 was generated using both the Barkan et al. (11) PPR code (top sequence) and Yagi et al's (13) modified PPR code (bottom sequence) and aligned with the *cis*-elements of the five editing sites that QED1 targets. Nucleotides in red in the predicted sequence are residues that are conserved in all five *cis*-elements. Nucleotides in pink are only found in one or two of the five *cis*-elements which are boxed in red.

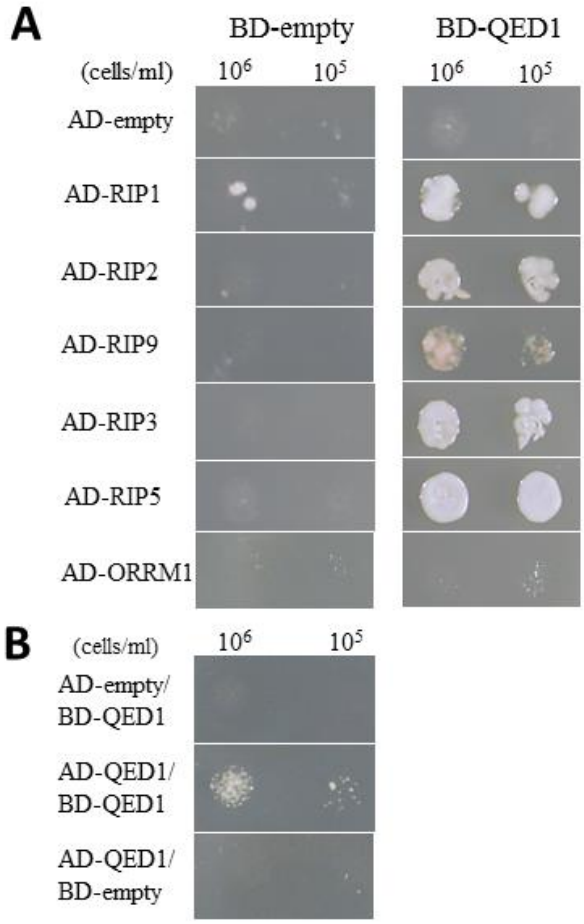


FIGURE 2.6. **QED1 interacts with members of the RIP family in a yeast two-hybrid assay.** ***A,*** Growth of yeast colonies on selective media when QED1 fused with the BD is combined with RIP family members (RIP1, RIP2, RIP3, RIP5 and RIP9) fused with the AD. When BD-QED1 was co-expressed with AD-ORRM1 in yeast, no growth was observed. *B,* Co-expression of AD-QED1 with BD-QED1 along with auto-activation controls.

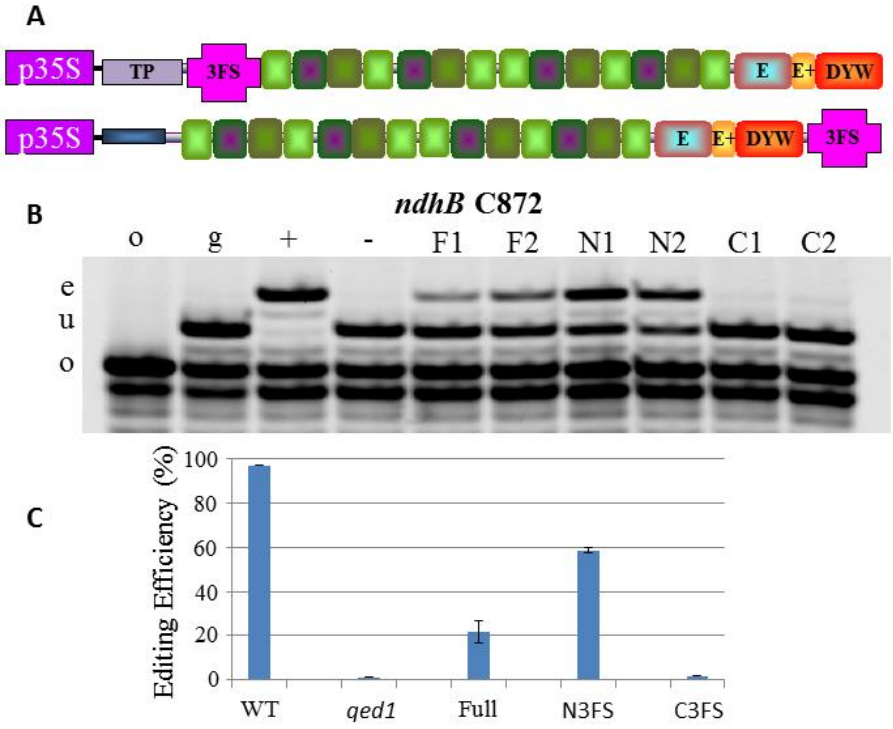


FIGURE 2.7. **Ability of QED1 variants with N-terminal vs. C-terminal tags to complement *qed1* mutants.** *A,* Chimeric versions of QED1 with a 3xFLAG-StrepII (3FS) were cloned where the 3FS tag was either attached to the C-terminus of QED1 or at the N-terminus of the mature QED1 protein with the RecA target peptide (TP) and controlled by the 35S promoter (p35S). *B,* PPE assay of editing of the *ndhB* C872 site following protoplast complementation in *qed1* protoplasts. edited (e) and unedited (u) products (o) oligonucleotide. (+) wild type protoplasts, (-) *qed1* protoplasts transfected with GFP, (Full) transfection with full-length QED1, N3FS (N) results from transfection with QED1 tagged at the N-terminus and C3FS (C) results from transfection with QED1 tagged at the C-terminus. *C,* Quantification of editing efficiency from PPE assay.

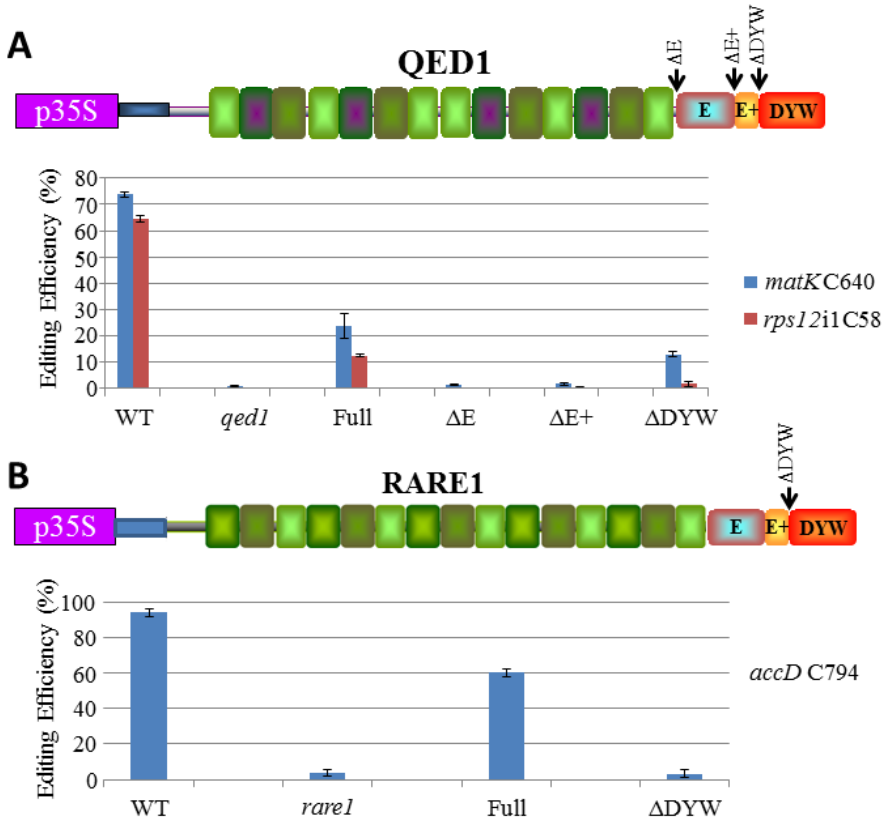


FIGURE 2.8. **Truncation analysis of QED1 and RARE1** ***A,*** Full length QED1 (Full) and various truncations of QED1 were cloned into an expression vector under a 35S promoter and expressed in *qed1* mutant protoplasts each in two separate transfections. The arrows indicate the position of three truncations that progressively remove each of the C-terminal motifs of QED1. Data for the *matK* C640 and *rps12* i1C58 sites are shown. ***B,*** Full-length RARE1 and DYW-truncated RARE1 were cloned into an expression vector under a 35S promoter and expressed in *rare1* mutant protoplasts. Arrow indicates where the protein was truncated. WT, *qed1*, and *rare1* indicate editing efficiencies at the *accD* C794 site of wild type and mutant protoplasts, each transfected with a GFP gene. Full indicates editing efficiencies of full length proteins while ΔDYW indicate editing efficiencies of DYW-truncated QED1 or RARE1 each expressed in two separate transfections into the respective mutant protoplasts.

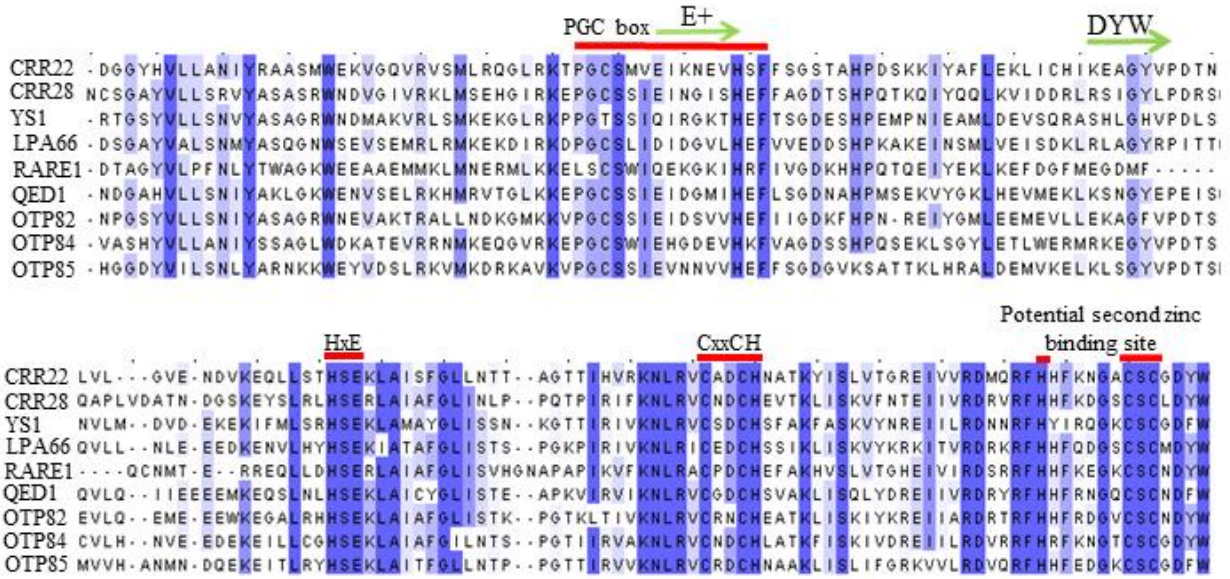


FIGURE 2.9. **C-terminal motif alignment of nine chloroplast PPR-DYW editing factors**. DYW, E+ and partial E motif protein sequences from nine chloroplast PPR-DYW editing factors (CRR22 NP_172596.1, CRR28 NP_176180.1, YS1 NP_188908.2, LPA66 NP_199702.1, RARE1 NP_196831.1, QED1 NP_180537.1, OTP82 NP_172286.1, OTP84 NP_191302.2, OTP85 NP_178398.1) were obtained from the NCBI database, aligned using Tcoffee, and presented by Jalview. The intensity of the shading indicates the degree of conservation between the motifs. The beginnings of the E+ and DYW motifs are marked with green arrows. The conserved deaminase-like motifs (HxE, CxxCH) are denoted by the red lines along with the second putative zinc binding site and the PGC box designated by Hayes et al. (17).

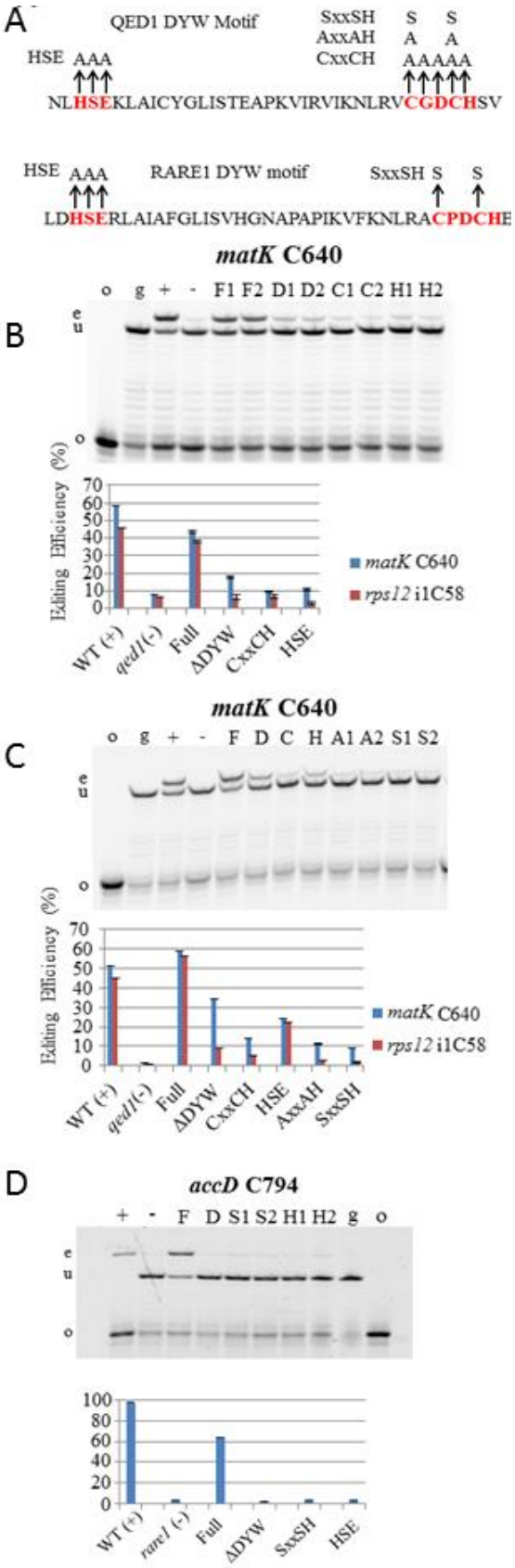


FIGURE 2.10. **Mutagenesis of the cytidine deaminase signature residues in QED1 and RARE1 significantly reduces editing. *A,*** The deaminase residues in the DYW motif (designated in red) of QED1 were targeted using site-directed mutagenesis. Changes that were made are indicated above the targeted residues. *B,* PPE assay of editing in *qed1* protoplasts following two separate transfections of the same batch of protoplasts with full-length (full), GFP (*qed1*-), DYW-truncated QED1 (ΔDYW) and mutagenized versions of QED1 (CxxCH, HSE) compared to wild-type protoplasts (WT+). *C,* PPE assay of editing in *qed1* protoplasts using a different batch of protoplasts than in *B* and additional mutagenized versions of QED1 (AxxAH, SxxSH). PPE gel for *matK* C640 is shown; graphs of editing quantifications for *matK* C640 and *rps12*i1C58 PPE gels are provided. *D,* PPE assay of editing in the same batch of *rare1* protoplasts following two separate transfections with full-length *RARE1* (Full), *GFP* (rare1-), DYW-truncated *RARE1* (ΔDYW), and mutated versions SxxSH and HSE. Edited product (e), unedited product (u), oligonucleotide (o).

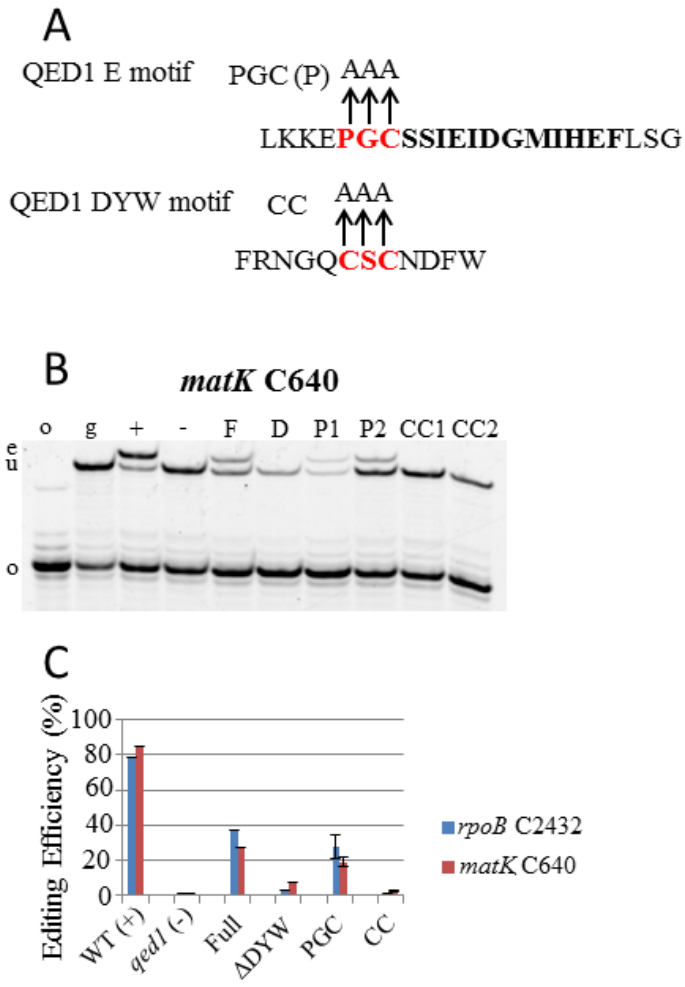


FIGURE 2.11. **Mutagenesis of a second potential zinc binding site in the DYW motif of QED1 significantly reduces editing** *A*, Residues in the E and DYW motif of QED1 were targeted for site-directed mutagenesis. Three residues (in red) in a conserved 15aa sequence called the PGC box as defined by Hayes et al (17) (in bold) and three residues (in red) near the end of the DYW motif of QED1 were mutated to alanines. ***B,*** Results from the PPE assay in wild-type protoplasts (WT) and following transfection of GFP (*qed1*- or *rare1*-), the full-length un-mutagenized construct (Full), and mutagenized constructs targeting the PGC box (PGC or P) and the terminal cysteines (CC) into *qed1* protoplasts. *matK* C640 PPE gel is shown; graphs of PPE data from *rpoB* C2432 and *matK* C640 sites are provided. Edited product (e), unedited product (u), oligonucleotide (o).

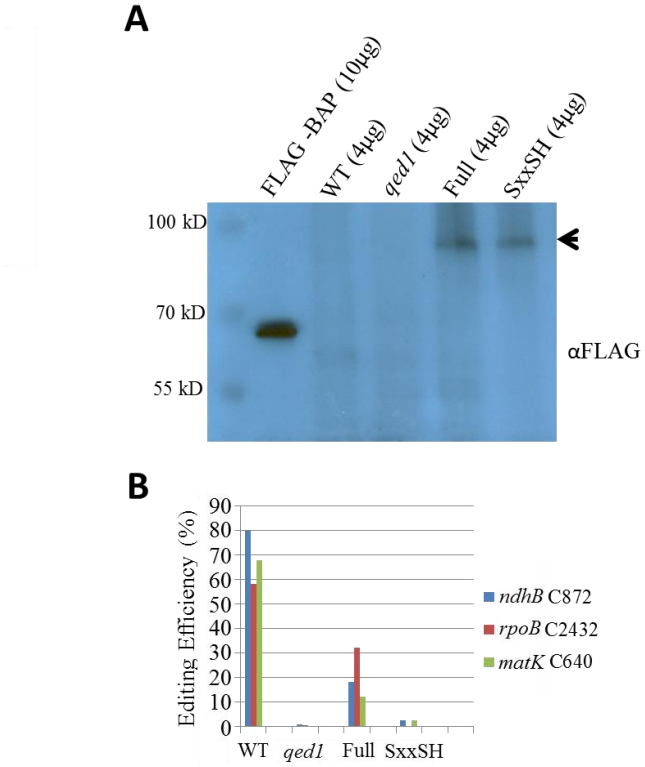


FIGURE 2.12. **Expression of full-length QED1 and mutagenized QED1 (SxxSH) in mutant protoplasts and effect on editing of three C targets.** *A*, Immunoblot probed with α-FLAG antibody. *qed1* protoplasts were transfected with a construct expressing either the entire wild-type QED1 coding region (full) or mutant QED1 (SxxSH), tagged on their N-termini with a 3xFLAG-StrepII tag. Protein was extracted from the transfected protoplasts, from un-transfected wild-type (WT) protoplasts, and *qed1* protoplasts transfected with a *GFP* gene (*qed1*). 4 μg protein from each sample was run on an SDS-PAGE. 10 μg of FLAG-BAP (Bacterial Alkaline Phosphatase) was included as a positive control. Mature QED1 plus the 3xFLAG-StrepII tag is predicted to be approximately 80 kD in size. *B*, Editing quantifications for *ndhB* C872, *rpoB* C2432, and *matK* C640 from RNA extracted from the same protoplast samples are shown underneath.

Chapter 3

Mutation of a PPR protein editing factor affects chloroplast morphology and function[2]

ABSTRACT

Cytidine (C) to uridine (U) editing occurs in many transcripts in the mitochondria and chloroplasts of higher plants. Plant RNA editing is thought to be a corrective process, as editing usually restores a codon to what is conserved in other species, and failure to edit particular sites can result in a range of phenotypes, from no discernable effect to seedling lethality depending on the transcripts that are affected. These C-targets are recognized by members of the pentatricopeptide repeat (PPR) protein family, which are known to bind to *cis*-elements encoded around the edited C and are thought to recruit the rest of the editing complex. While many PPR editing factors are known to be responsible for specifying a single editing site, some factors affect multiple sites. We investigate a chloroplast-targeted PPR protein, QED1, which is responsible for the editing of five different chloroplast C-targets, including three sites in the coding regions of *rpoB*, *matK*, and *ndhB*. In these mutants, we observe a range of physiological consequences which include stunted growth, severe reduction of NDH activity, the accumulation of yellow patches in aging leaves, and abnormal chloroplast morphology. We also investigate the phenotypes of *qed1* mutant plants transformed with the wild-type QED1 coding sequence or QED1 tagged at its N-terminus with a 3XFLAG-StrepII tag.

INTRODUCTION

RNA editing is a modification that changes the nucleotide sequence in RNA from what

[2] J. Wagoner grew the plants used in this study, created complementation constructs and created stable transformants, performed RNA isolation and PPE assays, generated protoplasts, performed protoplast size quantification, and performed sequence analysis. A. Sattarzadeh performed confocal microscopy and protoplast size

was originally encoded in the genome and is a process that is exhibited by a wide range of organisms (Chateigner-Boutin and Small, 2011). In vascular plants, cytidine to uridine RNA editing occurs in the organelles. In *Arabidopsis thaliana*, over 600 C targets are edited in the mitochondria and 43 C targets are edited in the chloroplast (Bentolila et al., 2008, 2013; Ruwe et al., 2013). The site specificity factors in plant RNA editing are members of the pentatricopeptide repeat (PPR) family that function to recognize the sequence elements encoded around the C-target (Okuda et al., 2006). Recently, the mode by which the PPR motifs recognize specific RNA nucleotides has been elucidated (Barkan et al., 2012; Takenaka et al., 2013; Yagi et al., 2013). This process is thought to be a correctional mechanism rather than one resulting in functional diversity, as the editing event usually results in changing the codon that was encoded in the genome to a codon that is conserved in other plant species (Gray and Covello, 1993). Failure to edit particular C-targets can result in a wide variety of physiological consequences, ranging from no obvious phenotype to seedling lethality depending on the specific transcripts that are affected. For example, PPR editing factor CHLOROPLAST BIOGENESIS 19 (CLB19) is responsible for editing a site in *rpoA* and a site in *clpP* and failure to edit these sites result in a seedling lethal phenotype (Chateigner-Boutin et al., 2008). YELLOW SEEDLING1 (YS1) is responsible for editing a site in *rpoB* and failure to edit this site results in seedlings that are yellow but eventually green and are able to grow to maturity (Zhou et al., 2009). Some mutants of CHLORORESPIRATION REDUCTION (CRR) genes affect editing sites in genes encoding subunits of the NDH complex which causes a lack of transient increase in chlorophyll fluorescence after terminating exposure to actinic light (Kotera et al., 2005; Okuda et al., 2007, 2009). RARE1 is responsible for editing a site in *accD*, but failure to edit this site in *rare1*

---

quantification. K. Hine performed confocal microscopy. T. Owens performed the NDH activity assays. L. Lin performed PPE assays

mutants does not appear to cause any discernable mutant phenotype (Robbins et al., 2009). PPR-DYW editing factor QUINTUPLE EDITING FACTOR 1 (QED1) was recently reported to affect five chloroplast editing sites, the most plastid sites known to be associated with a single PPR editing factor (Wagoner et al., 2015). Plants homozygous mutant for *qed1* were reported to grow more slowly than wild-type and appeared to accumulate yellow patches in the leaves, perhaps due to altered amino acid sequences in *rpoB*, *matK*, and *ndhB* due to lack of editing. Here we investigate the physiological consequences of loss of editing at these sites which includes stunted growth, impaired function of the NDH complex and an abnormal chloroplast morphology phenotype.

MATERIALS AND METHODS

Arabidopsis line and growth assays

Both wild-type and mutant Arabidopsis plants were of the *Landsberg erecta* ecotype. The T-DNA insertion line CSHL_GT13864 was obtained from the Arabidopsis Biological Resource Center (ABRC) and the presence of the T-DNA insert was verified through PCR with primers 5'-GCTTGCATAGTTGATGTTCTCG-3' and the Ds3-1 primer 5'-ACCCGACCGGATCGTATCGGT-3' (Martienssen and Springer, 1997) using BioRad Red Master Mix (Hercules, CA).

Arabidopsis seeds were imbibed in water for three days at 4C and then sown on MetroMix. Plants were grown under both long day (16h light, 8h dark, 80 µmol/m2/sec light, 22°C) and short day (10h light, 14h dark, 115 µmol/m2/sec light, 22°C) conditions. Fresh weight was measured by removing plants from soil, washing off leftover soil, and removing excess water with a paper towel. Weight was then measured on a Mettler balance.

Complementation analysis

The 2217 bp wild-type coding region of At2g29760 (*QED1*) was amplified using Phusion polymerase and gene-specific primers: 5'-ATGGCTATCTTCTCCACAGCAC-3' containing the start codon and 5'-TCACCAGAAATCGTTACAGG-3' containing the stop codon (Table 1). The amplified fragment was cloned into PCR8/GW/TOPO cloning vector (Invitrogen, Carlsbad, CA) and verified by sequencing. Using LR Clonase II, the gene was then cloned into a modified pBI121 (Jefferson et al, 1987). This plasmid, carrying the wild-type QED1 cloning region, was then transformed into *Agrobacterium tumefaciens* GV3101. Transformation of *qed1* CSHL_GT13864 mutant plants was performed as in Zhang et al (2006).

Transgene integration site determination and genotyping

Determination of the transgene integration site was performed through ligation of adaptors to genomic DNA digested by HindIII and EcoRI restriction enzymes (Invitrogen) using the protocol and adaptor sequences as described in O'Malley et al (2007). PCR was performed on adaptor-ligated genomic DNA fragments using adaptor-specific primers as designed by O'Malley et al (2007) and a primer specific to the RecA targeting sequence 5'-GCTTCAGAGACAAGACTAGCTG-3'. PCR products were sequenced and the resulting sequence was analyzed in BLAST to determine the location of the transgene. Determination of the homozygous or hemizygous state of the transgene was determined using PCR with primers 5'-GTCATCGTCATCCTTGTAATCG-3', 5'-AACATGGCTCTCAGCCTTTCAA-3', and 5'-CATCGGTAGTGATGAAGTCT-3' (Table 3.1)

Protoplast preparation

The cellulase/macerozyme solution was prepared following the protocol of Yoo et al (2007). and contained 1.5% cellulase R10, 0.4% macerozyme R10 (Yakult Pharmaceutical Ind. Co., Ltd., Japan), 0.4 M mannitol, 20 mM KCl, 20 mM MES pH 5.7, 10 mM $CaCl_2$, 0.1% BSA (Yoo

et al., 2007). Arabidopsis plants were grown in short days (8 h light). Fully expanded leaves were sliced into 0.5-1mm strips using razor blades with minimal wounding and immersed in the enzyme solution. The digestion mixture was vacuum-infiltrated for 2 minutes and incubated overnight in the dark without shaking. The protoplasts were harvested in round-bottomed tubes by 2 minutes centrifugation at 300g. The supernatant was discarded and the cells were resuspended in the W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES pH5.7) (Yoo et al., 2007).

Measurement of Editing Extent

RNA extraction and RT-PCR was performed as described by Bentolila et al. (2008). The primers used to amplify chloroplast transcripts are listed in Robbins et al. (2009). The Poisoned Primer Extension (PPE) assay to analyze RNA editing extent was described by Peeters and Hanson (2002).

Chlorophyll Fluorescence Measurement

Chlorophyll fluorescence transients indicative of an active NDH complex were measured using a Hansatech FMS-2 fluorometer (Norfolk, UK) following the general method of Shikanai et al (1998). Attached leaves from 5 week old wild-type, *qed1* mutant, and T_1 3FS-QED1 transformant plants were dark-adapted for 20 minutes prior to measurement of minimum fluorescence yield (F_o) and maximum fluorescence yield (F_m). The leaves were then exposed to actinic light at 150 µmol photons m^{-2} s^{-1} for 8 minutes. Fluorescence transients were measured for 2 minutes following the termination of actinic illumination. All measurements were normalized to the dark adapted F_o value. Curves were created from data averaged from three physiologically equivalent leaves each from two different plants of each type.

Microscopy

Confocal laser scanning microscopy was performed on a Zeiss 710 confocal microscope (Zeiss, Jena, Germany). The 633 nm line of argon laser was used to excite chlorophyll. Emission of chlorophyll fluorescence was detected in the range of 647–721 nm. Protoplast area was measured using the Image J software (http://rsbweb.nih.gov/ij/) and also Zeiss Zen Blue software. The original picture was imported into Image J and protoplasts area was measured via ROI manager in Image J (Ounis et al., 2005).

Sequence Analysis

DNA and protein alignments were achieved using Jalview version 2.0 sequence analysis software (Waterhouse et al., 2009)

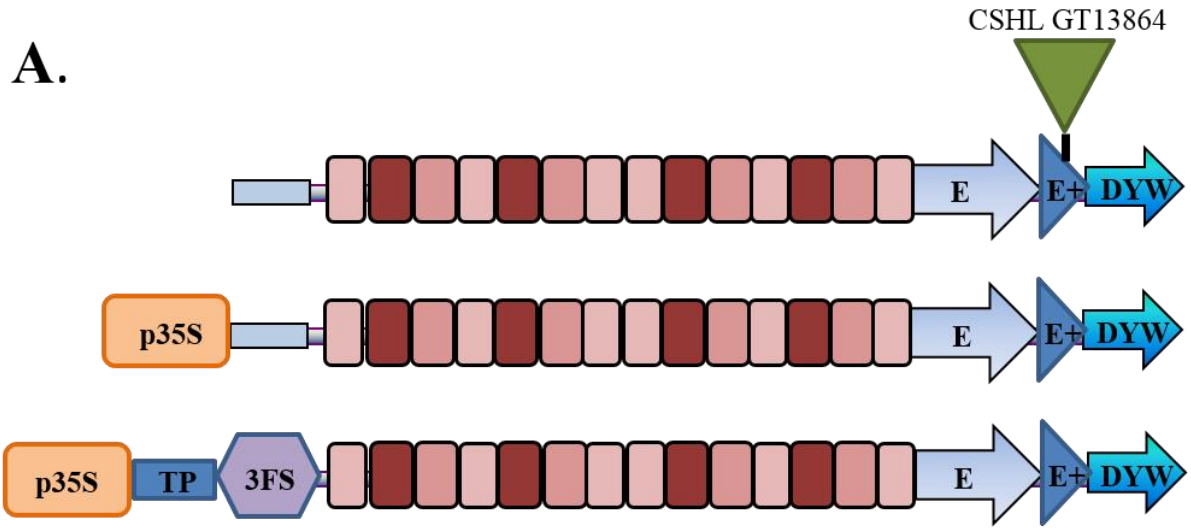
RESULTS

Complementation of editing defects at five chloroplast sites with full-length and N-tagged QED1

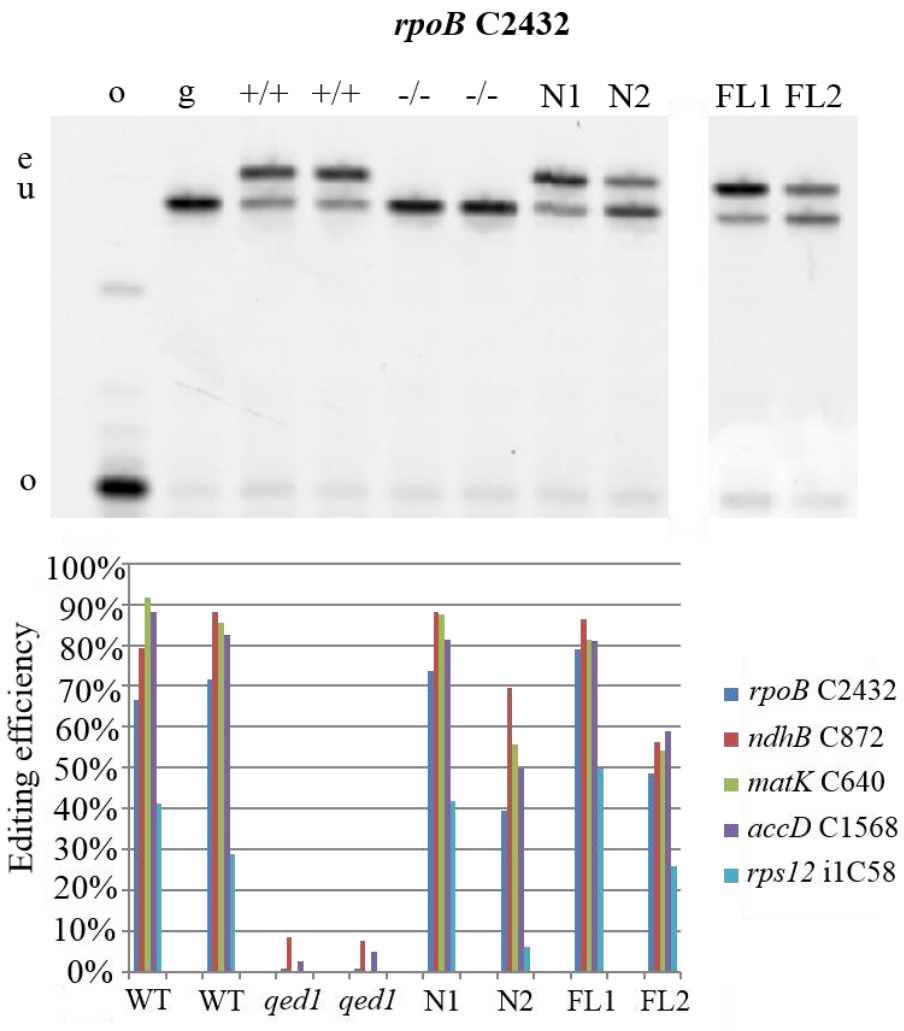
The identification and study of the functional motifs of a new chloroplast PPR-DYW editing factor was previously described in Wagoner et al. (2015) Briefly, *QUINTUPLE EDITING FACTOR 1* (QED1) is responsible for the editing of five C-targets in the chloroplast: *rpoB* C2432, *ndhB* C872, *matK* C640, *rps12* i1C58, and *accD* C1568. The editing of the sites in *rpoB*, *ndhB*, and *matK* all result in non-silent codon changes to the sequence of these proteins. The *rps12* and *accD* sites are found in non-coding regions: the *rps12* site is located in the intron and the *accD* site is located in the 3'UTR. No editing was detected in these five sites in homozygous mutant plants of the CSHL_GT13864 line, which has a T-DNA insertion located in the E+ motif (Fig. 3.1A). These homozygous mutants were observed to have a stunted growth phenotype and accumulation of yellow patches in leaves as they aged. The mutants were still

Figure 3.1: Structure of endogenous *QED1* and of *QED1* transgenes constructed for complementation analysis. **A.** The location of the insertion allele is indicated in the *qed1* line used for analysis of the mutant phenotype. CSHL GT13864 is located in the E motif and homozygous mutants for this insertion allele were previously found to be defective in the editing of five chloroplast sites. The full-length wild-type coding region of QED1 and mature QED1 tagged at its C-terminus with a 3xFLAG-StrepII tag were cloned under a 35S promoter for complementation analysis through stable transformation. **B.** Using the Poisoned Primer Extension (PPE) assay, editing extent was quantified in two independent T_0 transformants carrying N-tagged QED1 (N) and two independent T_0 transformants carrying untagged full-length QED1 (FL) for *rpoB* C2432, *ndhB* C872, *matK* C640, *accD* C1568 and *rps12* i1C58. Editing product (e), unedited product (u), oligo (o). Wild-type (+/+), *qed1* mutant (-/-), N-tagged QED1 transformant (N), full length QED1 transformant (FL) PPE gel for the *rpoB* C2432 site is shown.

A.


CSHL GT13864
E
E+
DYW
p35S
E
E+
DYW
p35S
TP
3FS
E
E+
DYW

B.


rpoB C2432
o
g
+/+
+/+
-/-
-/-
N1
N2
FL1
FL2
e
u
o
Editing efficiency
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
WT
WT
qed1
qed1
N1
N2
FL1
FL2
rpoB C2432
ndhB C872
matK C640
accD C1568
rps12 i1C58

able to flower and set seed. Transgenes containing the QED1 coding region were constructed to create stable transformants: one construct contains the wild-type coding region placed under a 35S promoter and another construct contains the mature coding sequence tagged at its N-terminus with a 3xFLAG-StrepII tag joined by a GGSGGSGS protein linker (Fig. 3.1A). These constructs were transformed into *qed1* mutant plants via floral dip and editing was assessed in the stable transformants (Fig. 3.1B). To determine the insertion point of the transgene, adaptors were ligated to digested genomic fragments and the fragment containing a portion of the transgene and the flanking sequence was PCR amplified and sequenced using adaptor and transgene specific primer pairs (O'Malley et al., 2007). In the T_1 progeny of N1, the N-tagged QED1 transgene was found to be inserted into the second exon of At1g79600, an ABC1-like kinase gene, ABC1K3. Under normal growth conditions, *abc1k3* mutants are physiologically and photosynthetically wild-type, aside from slightly scattered thylakoid grana (Huang et al., 2015; Lundquist et al., 2013; Martinis et al., 2013). Three T_1 progeny tested by PCR indicated that the transgene was in a hemizygous state in each of these plants (Fig. 3.2). The N-tagged construct was able to restore editing to wild-type or near wild-type levels at all five sites, thus the N1 progeny were used in subsequent phenotype studies. A version of QED1 tagged at its C-terminus with the 3xFLAG-StrepII tag was previously constructed and tested for ability to complement the editing defect, but the C-terminal tag was found to interfere with editing function (Wagoner et al., 2015).

Expression of QED1 at different developmental stages of Arabidopsis

To determine the expression pattern of QED1 transcripts at different developmental time points in Arabidopsis, we utilized the GENEVESTIGATOR Arabidopsis microarray gene expression database (Hruz et al., 2008). We also assessed the expression pattern of plastid

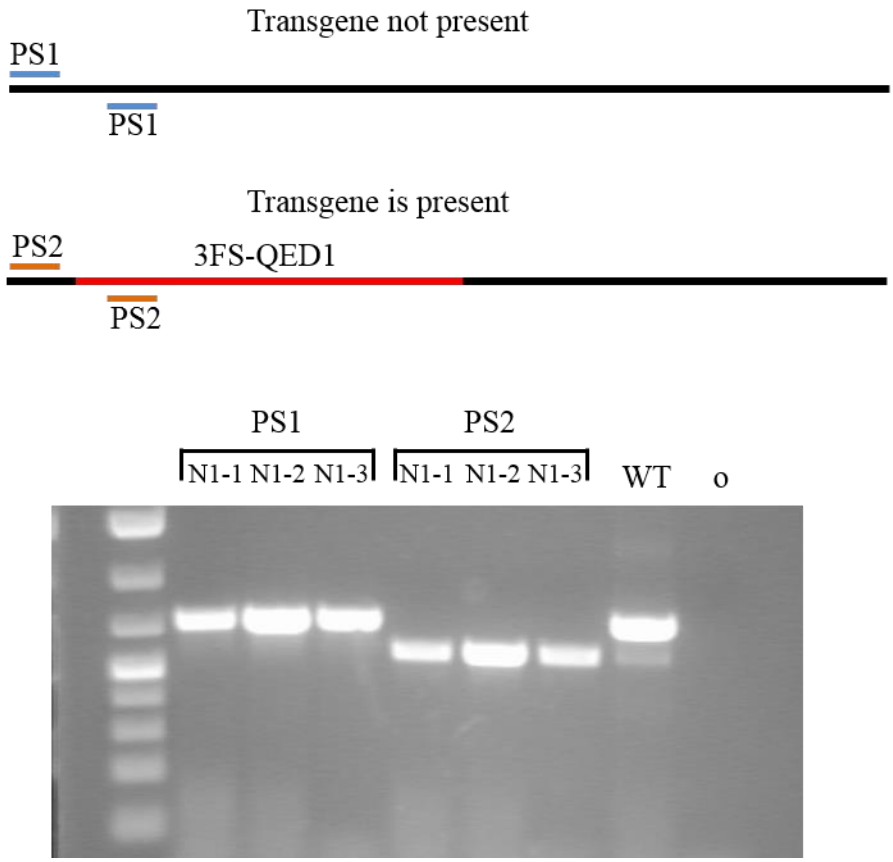


Figure 3.2: Genotyping of three T_1 3FS-QED1 progeny of the N1 transformant for the transgene. Genomic DNA from three T_1 plants (N1-1, N1-2, N1-3) carrying the 3FS-QED1 transgene obtained from N1 selfing was extracted and analyzed by PCR using two primer sets (PS). PS1 yields a band when the transgene is not present, PS2 yields a band when the transgene is present. WT reaction template contained wild type Landsberg DNA, o contained no DNA template. The extra band in the wild type lane is a result of spillover from loading the adjacent N1-3 lane.

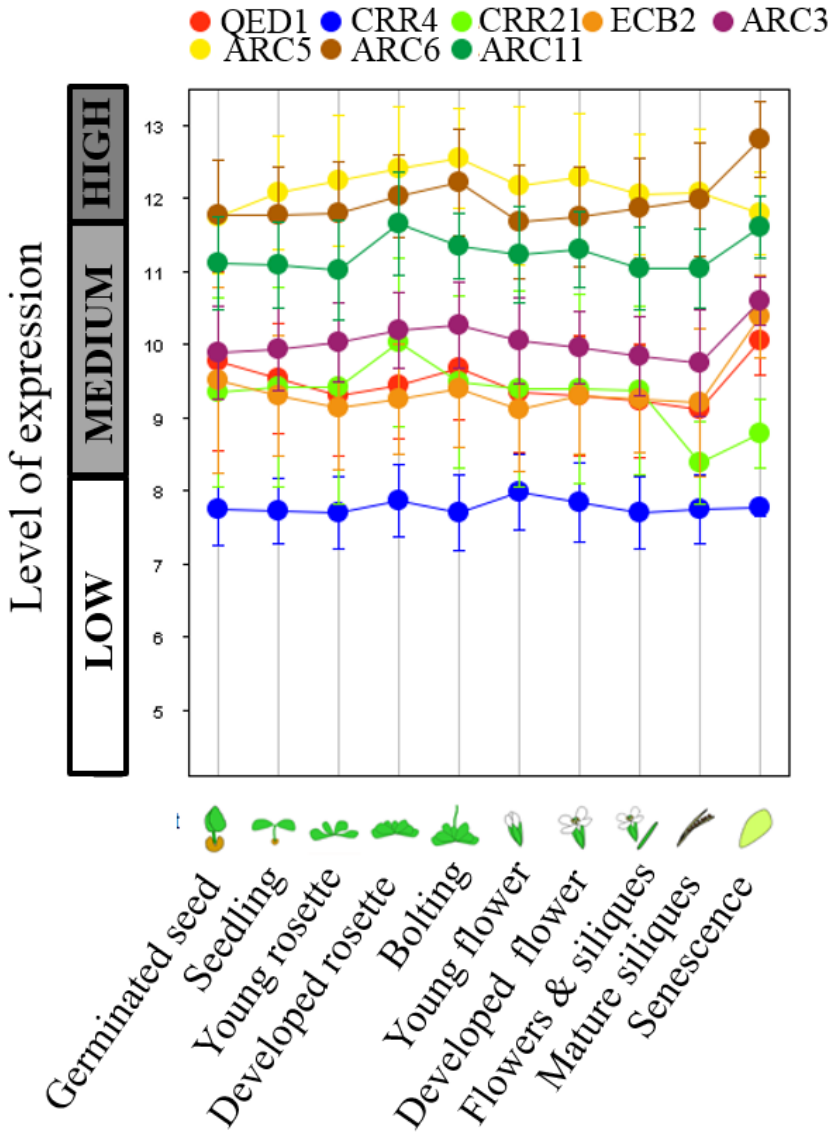


Figure 3.3: Expression levels of four plastid PPR editing factors, QED1, ECB2, CRR4, and CRR21, and four ACCUMULATION AND REPLICATION OF CHLOROPLAST (ARC) genes, ARC3, ARC5, ARC6, and ARC11 at different stages of Arabidopsis development according to online microarray data available from GENEVESTIGATOR.

editing factor genes *CRR4*, *CRR21*, and *ECB2* as well as genes encoding members of the nuclear-encoded and plastid-targeted ACCUMULATION AND REPLICATION OF CHLOROPLASTS (ARC) family, ARC3, ARC5, ARC6, and ARC11 (Fig. 3.3). These genes were chosen because mutation of ARC family genes affect chloroplast morphology, like QED1 (discussed further below). Expression of transcripts encoding each factor did not differ greatly throughout each developmental stage. Transcripts of *QED1* were the highest at germination, during bolting, and in the seeds. Generally, *ARC5, ARC6* and *ARC11* were more highly expressed than the other factors. *QED1, ECB2, CRR21* and *ARC3* were expressed at a similar level and *CRR4* was expressed at a lower level than the other factors. The expression levels of *QED1* and *ECB2* are closely mirrored throughout development.

QED1 proteins and editing status of the target sites of Arabidopsis QED1 in other species

Using BLAST, protein sequences with significant sequence identity to QED1 were found in *A. lyrata*, poplar (*P. trichocarpa)*, grape (*V. vinifera)*, soybean (*G. max)*, flowering tobacco (*N. sylvestris)*, potato (*S. tuberosum*), and tomato (*S. lycopersicum)* (Fig. 3.4). The E and DYW motifs exhibited the greatest conservation, along with the residues near the beginning of each repeat, where the amino acids that are involved with RNA recognition are located. The DYW motif contains a HxE/CxxC signature that is conserved with the active site in known deaminases and is known to be capable of binding zinc (Hayes et al., 2013; Iyer et al., 2011) These residues have been shown to be important for the function of QED1 in Arabidopsis (Wagoner et al., 2015). These deaminase residues are invariant among these protein sequences, along with the residues that encompass a putative second zinc binding site as defined by Hayes et al. (2013).

Three of the editing sites affected by QED1 are found in the coding regions of three genes: *matK*, which encodes a splicing factor, *ndhB*, which is a subunit of the NDH complex,

F**igure 3.4:** Amino acid sequence alignment of QED1. Homologous QED1 protein sequences from other species were found by using BLAST with the Arabidopsis QED1 protein sequence. Protein sequences were aligned using Jalview software and residues were shaded according to degree of conservation. Black lines above the sequence indicate PPR repeats while the red and orange arrows indicate the E motif and DYW motif respectively. Green lines indicate the conserved deaminase residues and the purple line denotes the second potential zinc binding site. Amino acid sequences: *A. thaliana* NP_180537.1, *A. lyrata* XP_002879234.1, *P. trichocarpa* XP_002314110.1, *V. vinifera* XP_002279360.1, *G. max* XP_003520267.1, *N. sylvestris* XP_009770229.1, *S. tuberosum* XP_006350917.1, *S. lycopersicum* XP_004241167.1

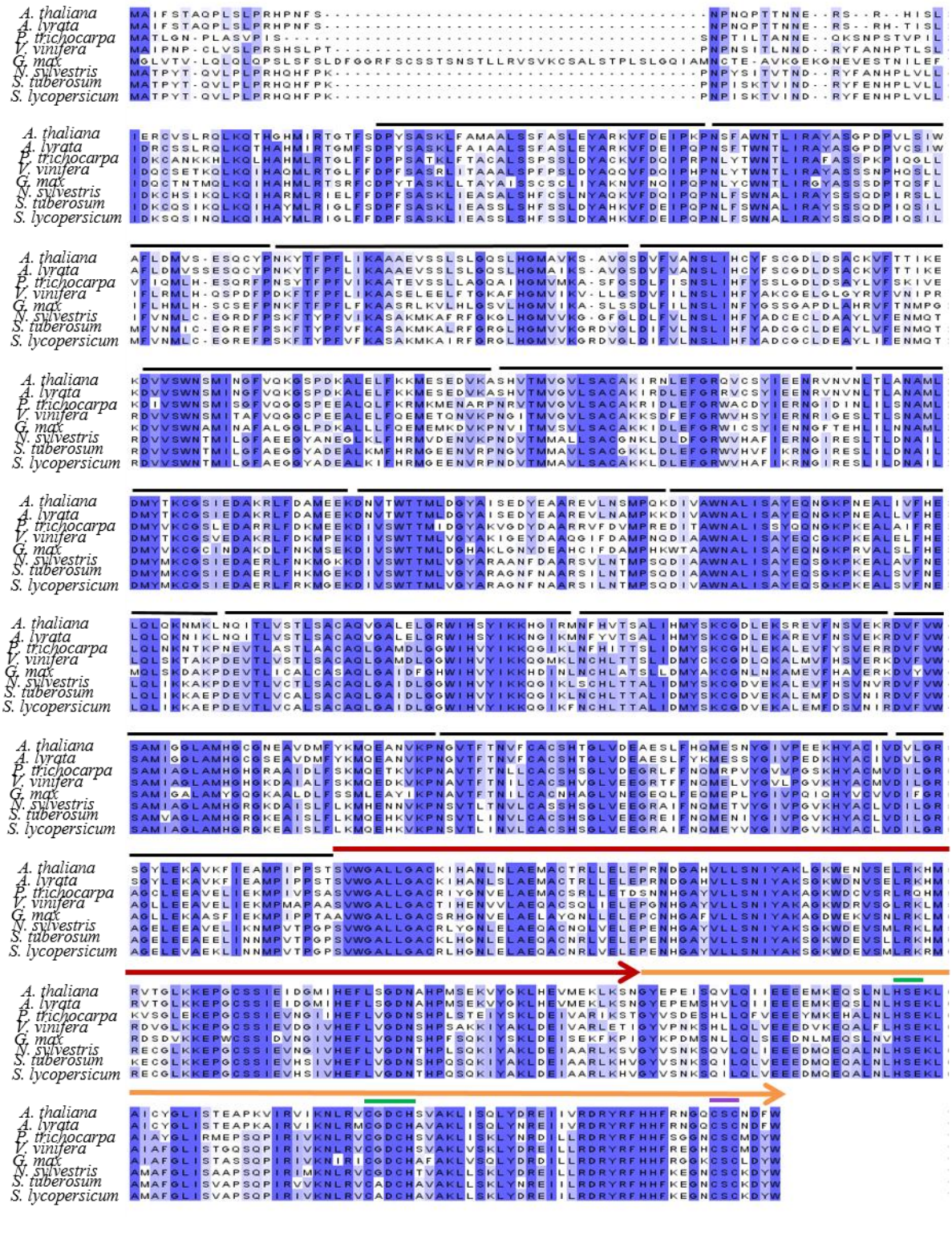
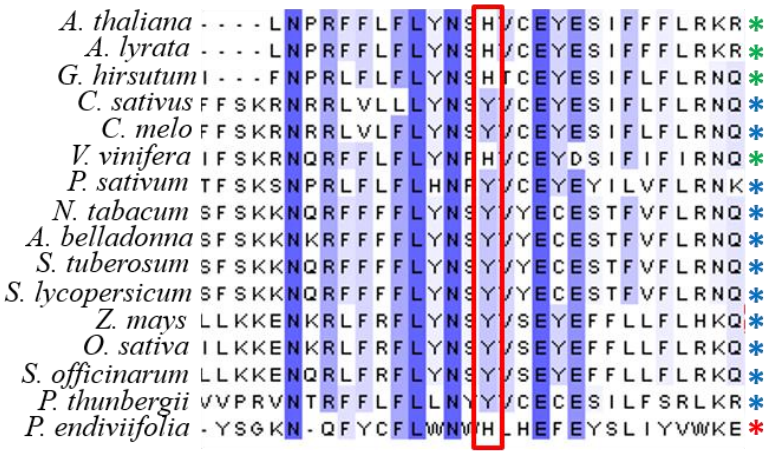
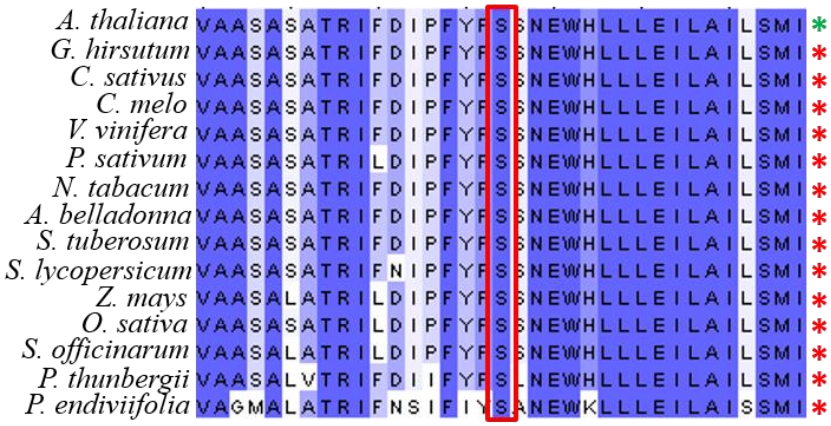

A. thaliana
A. lyrata
P. trichocarpa
V. vinifera
G. max
N. sylvestris
S. tuberosum
S. lycopersicum

Figure 3.5: Comparison of protein sequences containing the codons affected by QED1 editing in different plant species. Protein sequences were retrieved from NCBI, aligned with T-coffee, and presented with Jalview. Green asterisks denote sites that have a C encoded in the genome and that are edited, red asterisks denote sites that have a C encoded in the genome but are not edited, blue asterisks denote sites with a genomically encoded T. *A. thaliana*: rpoB NP_051051, matK NP_051051, ndhB NP_051103.2. *A. lyrata*: rpoB XP_002886863, matK AF144342.1. V. vinifera: rpoB YP_567068.1, matK YP_567057.1, ndhB YP_567138.1. *P. trichocarpa*: rpoB YP_001109491, matK YP_001109481.1, ndhB YP_001109571.1. *S. lycopersicum*: rpoB YP_008563080, matK YP_008563069.1, ndhB YP_008563149.1. *Z. mays*: rpoB NP_043015.1, matK NP_043005.2, ndhB NP_043102.1. *O. sativa*: rpoB AY522330.1, matK P0C381, ndhB NP_039459.1. *V. vinifera* editing (Jansen et al. 2006), *S. lycopersicum* editing (Kahlau et al. 2006) *Z. mays* editing (Maier et al. 1995)

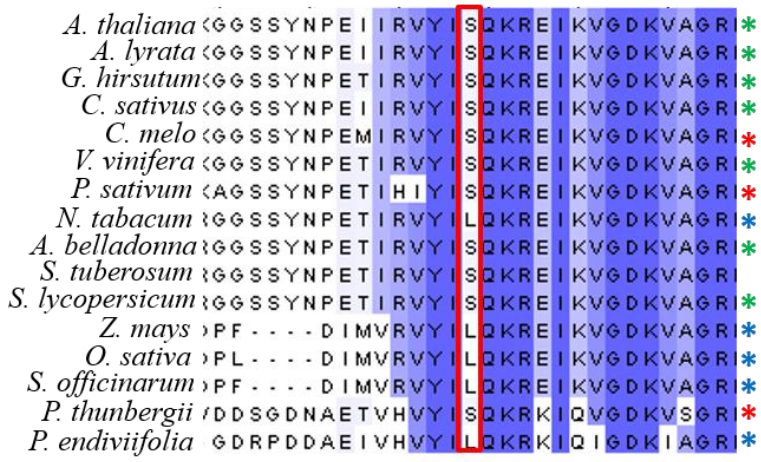
matK C640
His -> Tyr



ndhB C872
Ser -> Leu



rpoB C2432
Ser -> Leu



and *rpoB*, which is a subunit of plastid encoded RNA polymerase. To investigate the editing status of these sites in diverse plant species, protein sequences were collected from the NCBI database and aligned, and editing status of these sites was determined from a literature search of published chloroplast transcriptomes (Fig. 3.5). Most species contain a genomically encoded T at the *matK* site and all those that encode a C edit that site to a U, except for liverwort. Most species either contain a genomically encoded T or edit the C at the *rpoB* site except for melon, pea, and black pine. Interestingly, all species encode a C at the *ndhB* site, but of the species investigated, only Arabidopsis edits that site. These observations indicate that maintenance of the conserved codon is critical at the *matK* and *rpoB* sites; thus, these particular codons are likely to be important for protein function. The few plants that do not edit these codons may have compensating mutations in other codons. In contrast, of the species that were analyzed, the identity of amino acid encoded by editing of the *ndhB* site evidently is important only in Arabidopsis.

qed1 Mutants Exhibit Stunted Growth

When homozygous mutant lines were generated for the CSHL_GT13864 insertion line, the *qed1* mutants were observed to grow more slowly than wild-type plants, so growth rate was compared between wild-type and *qed1* mutant plants. Plants (ecotype Landsberg *erecta*) homozygous mutant for *qed1* were grown alongside Landsberg *erecta* wild-type plants under short day conditions and then assessed for fresh weight after 5 weeks (Fig. 3.6A) and eight weeks (Fig. 3.6B). At five weeks, the average fresh weight of *qed1* plants was approximately 0.5 g, while wild type plants averaged approximately 1.25g. The growth defect became more severe as the plants aged. At eight weeks, *qed1* mutant plants remained at approximately 0.5 g, while wild-type plants were now at 3.0 g. *qed1* mutants and wild-type plants were then grown

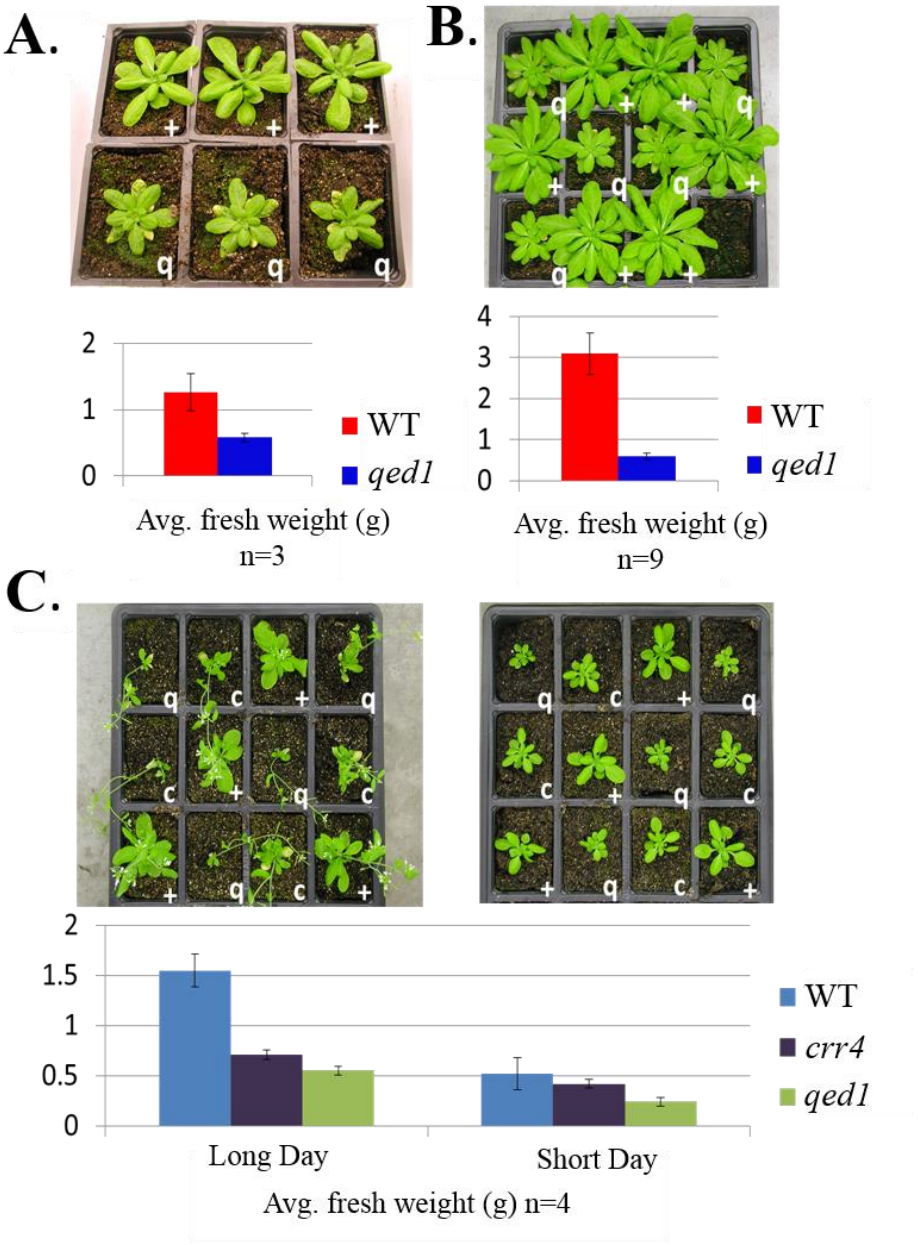


Figure 3.6: *qed1* plants display a stunted growth phenotype. Wild type and *qed1* plants homozygous mutant for CSHL GT13864, ecotype Landsberg erecta, were grown under short day conditions. Fresh weight was assessed at **A.** 5 weeks (n=3) and **B.** 8 weeks (n=9) and averaged. The bar graphs indicate fresh weight averages at five and eight weeks with wild type fresh weight average indicated in red and qed1 fresh weight average indicated in blue. Error bars indicate standard deviation. **C.** Wild type (+), *crr4* homozygous mutant (c), and *qed1* (q) mutant plants were assessed for growth under both long and short day conditions at 5 weeks. Fresh weight was measured and averaged (n=4 each). The bar graph indicates these averages under both long and short day conditions. Wild type is indicated by blue, *crr4* is indicated by dark blue and *qed1* is indicated by green.

WT *qed1* FL2

WT *qed1* N1

Figure 3.7: Phenotype of T_0 plants transformed with the wild type QED1 coding region (FL2) or QED1 tagged at its N-terminus with a 3xFLAG-StrepII (3FS) (N1) tag at 6 weeks old, grown under long day conditions.

under both long and short day conditions along with *crr4* (*chlororespiratory reduction 4*) mutants and assessed for fresh weight after five weeks. Relative to wild-type, the impairment of growth of the both the *crr4* and the *qed1* mutants were greater under long day conditions (Fig. 3.6C). Like QED1, CRR4 encodes a chloroplast editing factor that specifically affects an *ndh* transcript (Kotera et al., 2005). Yellow patches accumulated in both *qed1* and *crr4* as the leaves aged. The growth rate of T_0 plants transformed with full length QED1 (*qed1* + QED1) or with N1 plants, which contains QED1 tagged at the N-terminus with a 3xFLAG-StrepII tag, (*qed1* + 3FS-QED1) was greater than the untransformed mutant (Fig. 3.7)

NDH Activity is Lower in *qed1* Mutants Relative to Wild-type

Since QED1 affects editing site *ndhB* C872, which is in a subunit of the NDH dehydrogenase complex, *qed1* mutants were assessed for NDH activity. The NDH complex is involved in cyclic electron flow around photosystem I (Munekage et al., 2004; Shikanai et al., 1998). Using fluorescence imaging, activity of this complex may be assayed by measuring a transient increase of chlorophyll fluorescence due to dark reduction of the plastoquinone pool after exposure to actinic light (Hashimoto et al., 2003; Shikanai et al., 1999). When NDH activity was measured in five-week old wild type, *qed1* mutants and T_1 *qed1* plants carrying the 3FS-QED1 transgene (Ntag), a transient increase of chlorophyll fluorescence was observed in wild-type plants, characteristic of normal NDH activity. However, this transient increase was severely reduced in *qed1* mutants, indicating reduced NDH activity. In the 3FS-QED1 plants, the 3FS-QED1 transgene was able to partially restore the NDH activity (Fig. 3.8). Therefore, the failure to edit *ndhB* C872 affects the ability of the NDH complex to carry out its normal function.

***qed1* Mutants Exhibit Abnormal Chloroplast Morphology**

Because the *qed1* and *crr4* mutants accumulated yellow patches in the leaves as the

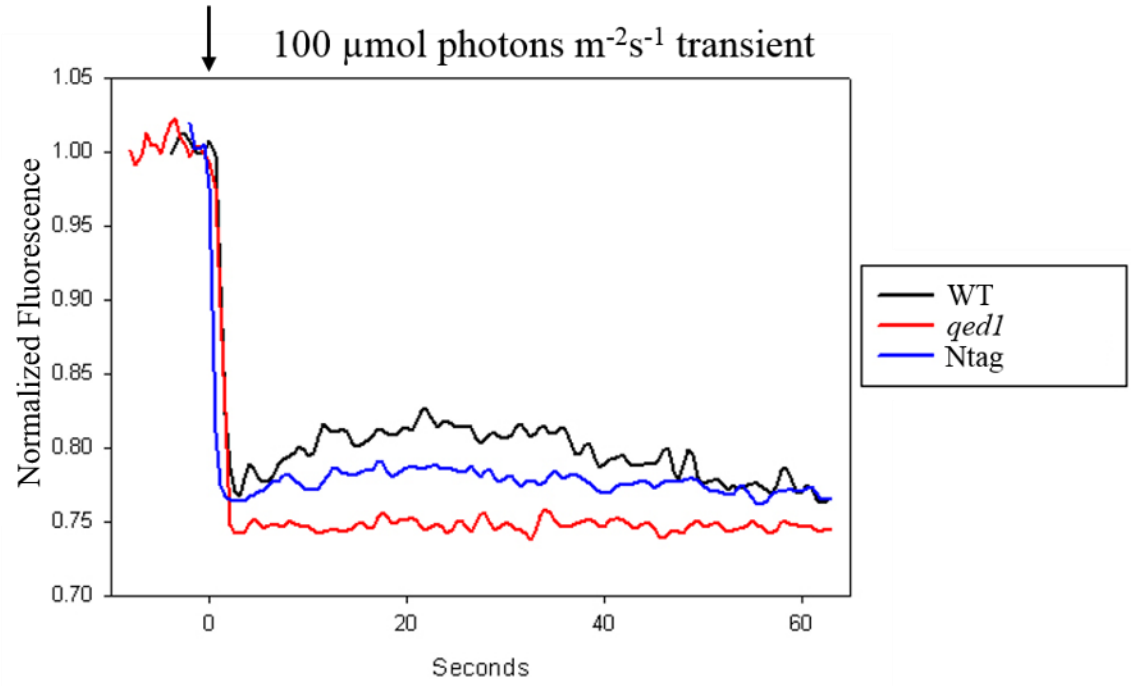


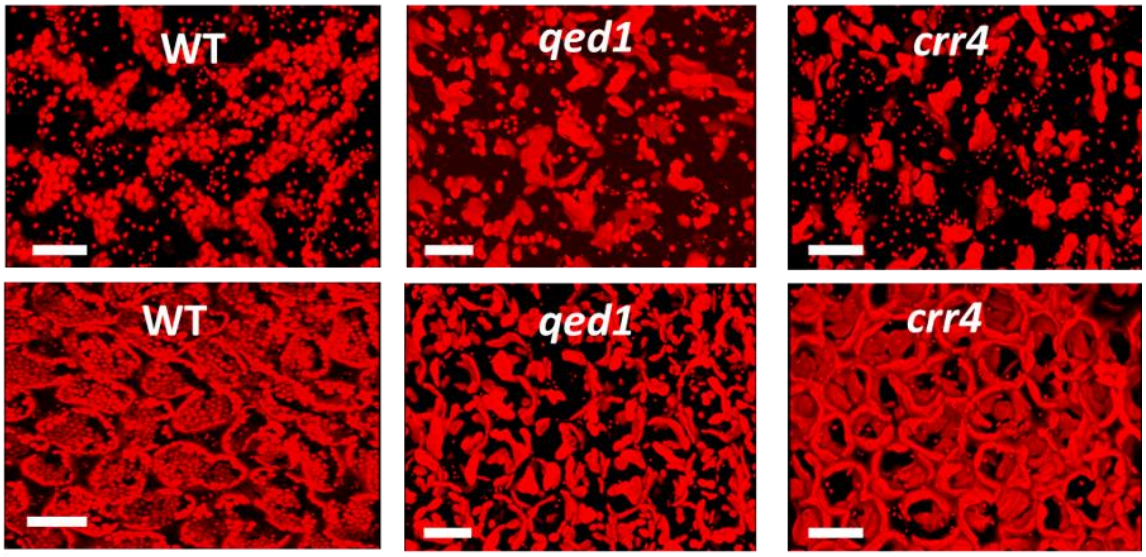
Figure 3.8: Evaluation of chlorophyll fluorescence in wild type, *qed1* mutant, and T_1 3FS-QED1 complemented plants obtained from selfing the N1 plant (Ntag). Leaves were exposed to actinic light at 150 µmol photons m-2 s-1 for 8 minutes followed by 2 minutes of darkness. AL was turned off at the point indicated by the downward arrow. The curves are averages of data from three physiologically equivalent leaves each from two plants of each type. The black curve represents the transient increase of chlorophyll fluorescence typical in wild type plants indicative of NDH activity. The red curve represents chlorophyll fluorescence in *qed1* mutant plants. The blue curve represents chlorophyll fluorescence in Ntag complemented plants. Fluorescence was normalized to the dark-adapted F_o value (=1.0)

leaves aged, morphology of chloroplasts was examined in these mutants by confocal microscopy. Using chlorophyll autofluorescence to visualize the chloroplasts, images were taken of the top layer and in the bottom layer of the leaves of six-week old plants. In the wild-type leaf, the chloroplasts exhibited the characteristic round, regular shape expected. However, in both *qed1* and *crr4* mutant leaves, the chloroplasts exhibited an abnormal morphology where many of the chloroplasts were much larger than the wild-type chloroplasts and irregularly shaped (Fig. 3.9). Both mutants also appeared to have fewer chloroplasts per cell on average compared to wild type.

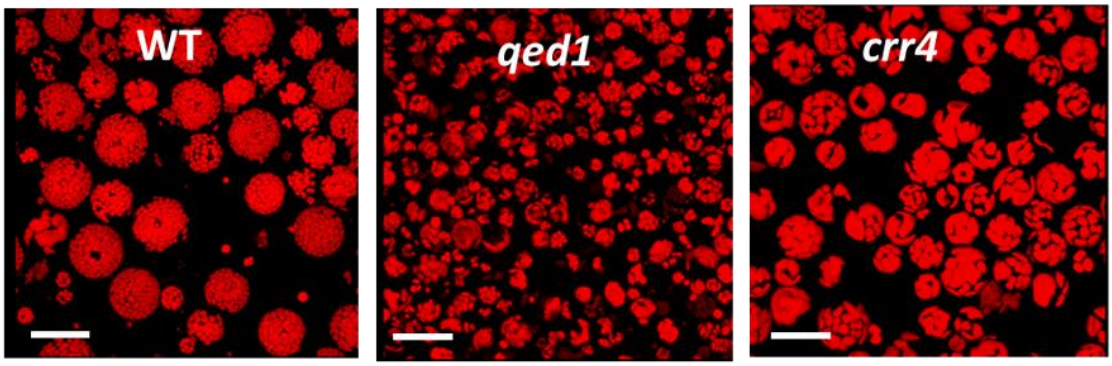
Protoplasts were isolated from wild type, *qed1* and *crr4* plants and observed using confocal microscopy in order to further characterize the chloroplast morphology phenotype. The *qed1* protoplast cells were smaller on average than wild-type protoplasts (Fig. 3.9). To quantify this size difference, two hundred protoplasts from each plant type were measured. On average, *qed1* protoplasts were half the size of wild type protoplasts while the total area calculated for *crr4* protoplasts was similar to that of the wild-type protoplasts (Fig. 3.9). The abnormal chloroplasts observed in total leaf in *qed1* and *crr4* plants were also seen in the protoplasts. Chloroplasts were visualized in a hemizygous T_0 transformant carrying a full-length QED1 transgene (*qed1*+QED1) and in one of the heterozygous T1 progeny carrying the 3FS-QED1 transgene genotyped in Figure 2 (*qed1*+3FS-QED1) to determine whether the presence of the QED1 transgene altered the chloroplast morphology (Fig. 3.10). Chloroplasts in a plant transformed by the full-length QED1 or 3FS-QED1 gene were more similar to wild-type chloroplasts than those in the mutant, but many large and abnormally shaped chloroplasts were still observed. Protoplasts from wild-type plants, *qed1* mutant plants, and T_0 plants transformed with full length QED1 hemizygous for the transgene were prepared and observed using confocal

Figure 3.9: *qed1* and *crr4* mutants exhibit a small cell and abnormal chloroplast morphology phenotype. **A.** Chlorophyll auto-fluorescence (red) was visualized in wild type, *crr4*, and *qed1* mutant leaves. **B.** Wild type, *crr4*, and *qed1* protoplasts were prepared and observed using a Zeiss microscope. **C.** Two hundred protoplasts of each type were measured and analyzed using ImageJ software. The y-axis indicates total protoplast area in μm^2. White bars = 50μm

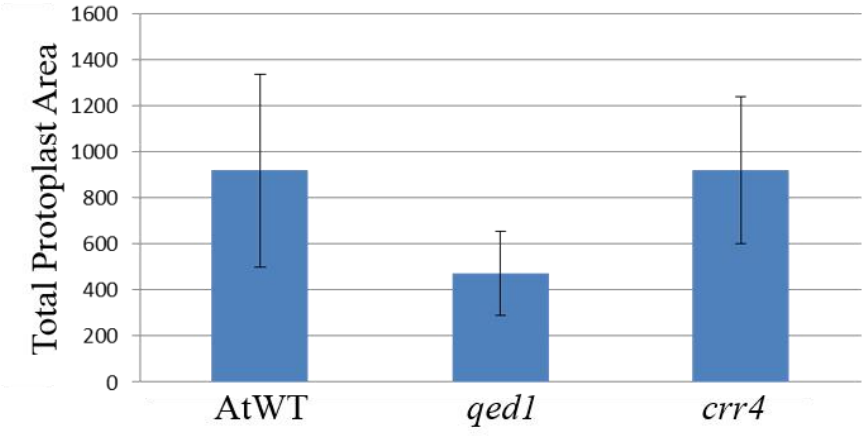
A.


WT
qed1
crr4
WT
qed1
crr4

B.


WT
qed1
crr4

C.


Total Protoplast Area
1600
1400
1200
1000
800
600
400
200
0
AtWT
qed1
crr4

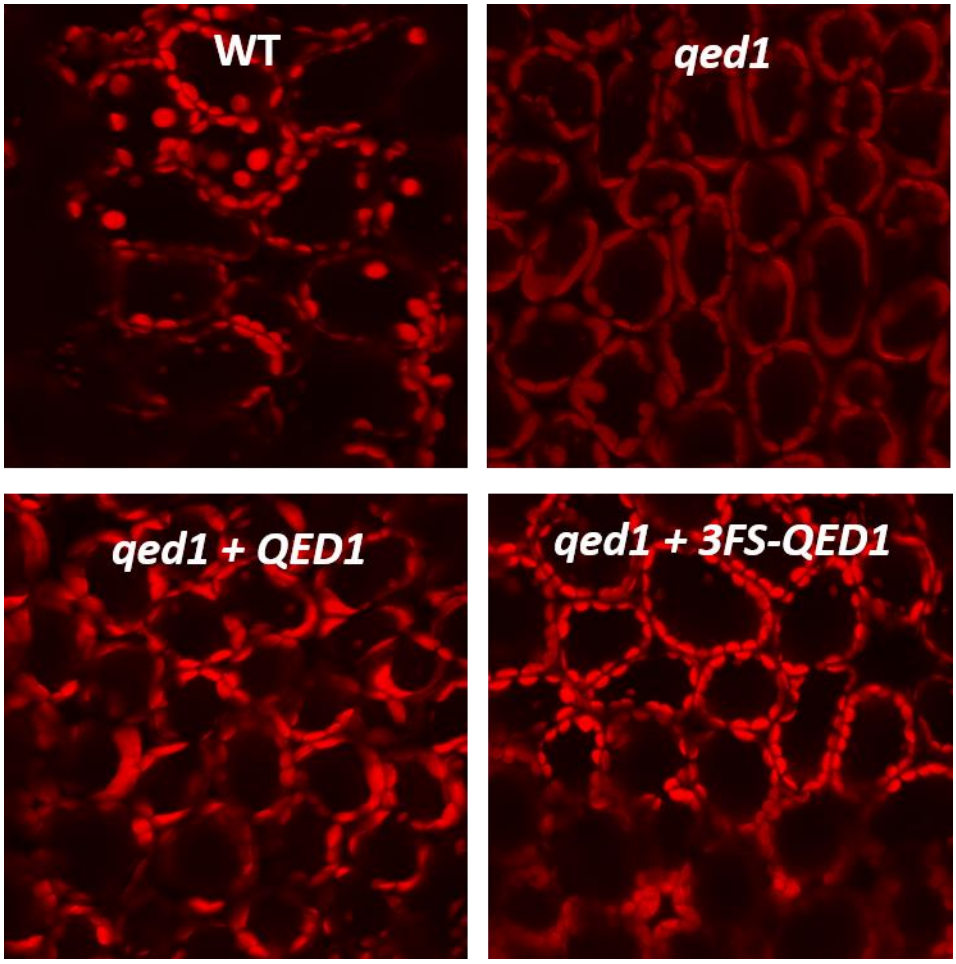
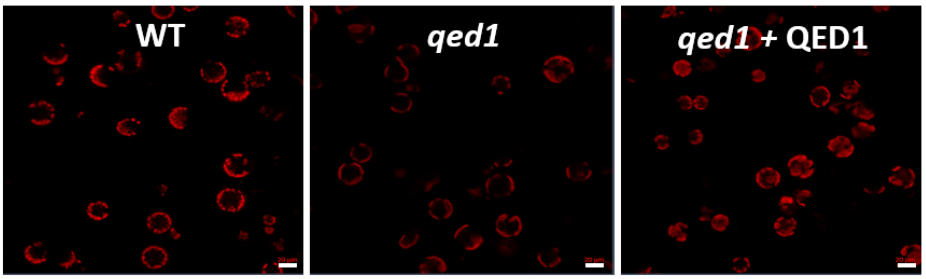
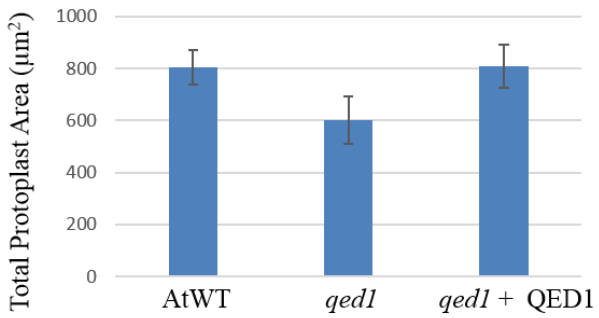


Figure 3.10: Chloroplast morphology in plants transformed with a QED1 transgene. Chlorophyll auto-fluorescence was visualized in the leaves of wild type, a *qed1* mutant, a T_0 *qed1* mutant plant transformed with the QED1 coding sequence, and a T_1 plant containing QED1 carrying a 3FLAG-StrepII tag at its N-terminus obtained from selfing the N1 transformant. Leaves were observed using a Leica microscope.

A.



B.



C.

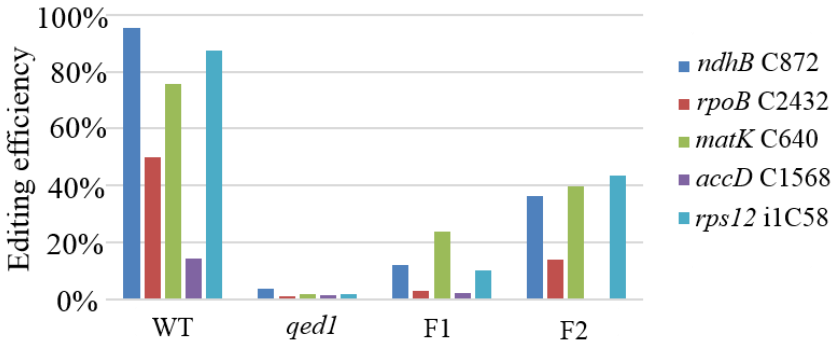


Figure 3.11: *qed1* complemented with QED1 exhibits partial complementation in cell size and chloroplast morphology. **A.** Chlorophyll auto-fluorescence (red) was visualized in wild type and *qed1* protoplasts and protoplasts prepared from two T_0 transformant progeny which were prepared and observed using a Zeiss microscope. **B.** Twenty protoplasts of each type were measured and analyzed using Zen Blue software (Zeiss). The y-axis indicates total protoplast area in μm^2. White bars = 20µm **C.** Editing status of the wild type plants (WT), *qed1* mutant plants (*qed1*), and two T_0 transformants carrying the QED1 transgene (F1, F2) used to prepare protoplasts.

microscopy to determine whether the small cell size phenotype was complemented by the presence of the full length QED1 transgene (Fig. 3.11). Protoplasts from the transformants are on average closer in size to wild-type, though some small cells are still present. The chloroplasts within the protoplasts of the transformed lines are less abnormal than the mutant, but are not fully normal in appearance.

DISCUSSION

Inactivation of the *QED1* gene, which is required for editing of 5 Cs to Us, results in slow growth, defective NDH activity, and abnormal chloroplast morphology. Three of the affected editing sites are in the coding regions of *matK*, a splicing factor, *rpoB,* a subunit of RNA polymerase, and *ndhB,* which is a subunit of the NDH complex. Though sites are also affected in *accD* and *rps12*, these sites are in non-coding regions: the *accD* site is in the 3'UTR and the *rps12* site is in the intron. Plants carrying a weak mutant allele of *QED1,* carrying a T-DNA insert 5' of the coding region, were previously observed to be morphologically normal; furthermore, lack of editing at this site did not appear to affect the splicing of the *rps12* intron (Hammani et al., 2009). Therefore, the lack of editing at this site is not likely to contribute to the phenotype seen in the *qed1* mutants.

The *accD* editing site controlled by QED1 is also affected in mutants of a different PPR-DYW factor, atECB2. Plants mutant in *atECB2* were not able to edit this site, displayed a seedling lethal albino phenotype and had defects in early chloroplast biogenesis (Yu et al., 2009). Further analysis of lines mutant for this gene, including a line described as VANILLA CREAM 1 (VAC1) and a line containing a point mutation in ECB2 that did not cause a seedling lethal phenotype, revealed that both the *ndhF* C290 and *accD* C1568 sites were affected in these mutants as well as the *accD* C794 site (Cao et al., 2011; Tseng et al., 2010; Yu et al., 2009). The

severe mutant alleles of this gene result in a lethal albino seedling phenotype, an abnormal chloroplast phenotype in which chloroplasts were smaller than wild-type chloroplasts, and had almost no organized thylakoid membranes and no starch grains. Plastid-encoded polymerase (PEP) transcripts were also decreased in these mutants (Tseng et al., 2010). The weak mutant allele of this gene caused a delayed greening phenotype, in which the seedlings were observed to have a variegated phenotype and delayed chloroplast development, but these seedlings could green and develop normally as they aged (Cao et al., 2011). In plants carrying the weak mutant allele, RNA editing of the *ndhF* site and the two *accD* sites was very low in the young plants, but increased to near wild-type levels as the plants aged (Cao et al., 2011).

Variegation and delayed chloroplast development in the weak *ecb2* mutants may be associated with lack of editing of sites in *accD* and a subunit of the NDH complex. Unlike *qed1* mutants, *ecb2* mutants exhibit decreased editing in *accD* C794. Mutants in the PPR protein-coding gene *RARE1* that lack editing of *accD* C794 appeared to be physiologically wild-type. Given that *rare1* mutants were phenotypically normal, including chloroplast phenotype, these observations indicate that if the abnormal phenotypes in *qed1* mutants are due to defects in *accD* editing, then it is due to defects in editing the *accD* C1568 site, but not the C794 site (Cao et al., 2011; Robbins et al., 2009). In tobacco, *accD* function is evidently essential, as homoplasmic plants could not be recovered when the wild-type plastid *accD* gene was replaced with a mutant allele interrupted with *aadA* (Kode et al., 2005). Leaves of transformed, non-homoplasmic tobacco had a variegated phenotype and analysis of the pale sectors revealed the presence of aberrant structures in the chloroplasts that disrupted the outer membranes, as well as the presence of vesicle-like structures, neither of which were observed in wild-type chloroplasts (Kode et al., 2005).

Like ECB2, QED1 controls editing of sites in *accD* and in subunits of the NDH complex. Leaf variegation and abnormal chloroplast morphology is also observed in *qed1* mutants, although the mutant phenotype differs from that observed in *ecb2* mutants. As the accD1568 site is in the 3'UTR, lack of editing at this site could result in decreased stability of the *accD* transcript. However, upon assaying chloroplast gene expression in weak allele *ecb2* mutants, PEP dependent plastid gene expression was similar to expression observed in wild type plants while expression of NEP (Nuclear Encoded Polymerase) dependent plastid genes, which includes *accD*, was actually observed to be upregulated which means the phenotypes are not likely to be associated with a lack of AccD protein.

A decrease in activity of the plastid NDH is most strongly implicated in the abnormal chloroplast morphology seen in the *qed1* mutant. The *crr4* mutant, which exhibits loss of editing in the *ndhD* transcript, preventing formation of the stop codon and thus eliminating enzyme activity, also exhibits larger and aberrantly shaped chloroplasts (Kotera et al., 2005). Possibly some of the abnormalities in chloroplast phenotype in *ecb2* mutants could also be due to loss of editing in the gene encoding NdhF subunit (Cao et al., 2011).

PEP activity appeared to be linked to the severity of phenotype in ECB2/VAC1 mutants. Although the editing of the *rpo* sites appeared to be at wild type levels in these mutants, defects in growth and development have been associated with lack of editing in the subunits of plastid RNA polymerase. Plants mutant for CHLOROPLAST BIOGENESIS 19 (CLB19) do not edit sites in *rpoA* and in *clpP* which results in a seedling lethal phenotype (Chateigner-Boutin et al., 2008). YELLOW SEEDLING 1 (YS1) is responsible for editing a different site in *rpoB* and seedlings mutant for *YS1* are yellow, although they eventually green and grow to maturity (Zhou et al., 2009). The *rpoB* C2432 site is not edited in *qed1* mutants and this lack of editing may also

contribute to the growth defect and chloroplast morphology phenotypes.

However, the *qed1* mutant plants that were transformed with the 3FS-QED1 sequence do not seem to display a strong complementation of the physical phenotypes observed in the *qed1* mutant lines although the editing phenotype is complemented to near wild-type level. If these phenotypes are indeed due to lack of editing at these five sites, the lack of phenotype complementation in the stable transformants may be due to the location the transgene was inserted. QED1 appears to be expressed at a medium level (Fig. 3.3) throughout the plant's development. If the transgene was under the control of a promoter that was specific to a particular development stage later in the plant's life, the lack of expression in early development could also cause non-complementation of the mutant phenotype. For the T_0 plants transformed with full length QED1, the editing phenotype was complemented, but not well; this could also be due to the location of the full-length QED1 transgene. The selection process for the transgene also relies on the use of herbicide resistance. Though plants carrying the transgene can survive contact with the herbicide, they do tend to be sickly which in turn may also affect editing efficiency.

The T_1 transformants that were genotyped were found to have only one copy of the transgene while the T_0 transformants are known to only have one copy of the transgene. A T_2 plant carrying two copies of the transgene must be examined to determine whether greater QED1 expression might result in complete restoration of wild-type phenotype. The physical phenotypes could also be due to a second site mutation, but the observation of the same chloroplast morphology phenotype in mutant *crr4* plants, in which lack of editing affects the same complex, makes this scenario unlikely.

The analysis of the codons affected by QED1 editing in other species indicates that the

conservation of the codons at the particular positions in *rpoB* and *matK* is important, as the conservation of the *rpoB* site is observed in most of the species investigated, except for in melon and in pea, and the codon affected in *matK* is conserved in all species except for liverwort. Interestingly, the site edited by QED1 in *ndhB* only appears to be edited in Arabidopsis; the other species that were investigated have a genomically encoded C that does not appear to be edited. If editing at this *ndhB* site is important for its function in Arabidopsis, it may not be important in the other species, possibly due to compensatory mutations in other sites of the *ndhB* gene or in other NDH subunits in these species. Although these species do not edit this particular site, this appears that this transcript is generally highly edited in these species, as this gene has been reported to have 9-11 editing sites in the species that were investigated (Freyer et al., 1995). Generally, the lack of editing at particular chloroplast sites can result in an array of phenotypes. Because multiple phenotypes have been associated with lack of editing in the genes affected by QED1 in mutants of other PPR editing factors, more work will have to be performed in order to dissect which phenotypes are associated with the lack of editing at the sites controlled by QED1.

REFERENCES

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

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

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

Cao, Z.-L., Yu, Q.-B., Sun, Y., Lu, Y., Cui, Y.-L., and Yang, Z.-N. (2011). A point mutation in the pentatricopeptide repeat motif of the AtECB2 protein causes delayed chloroplast development. J. Integr. Plant Biol. *53*, 258–269.

Chateigner-Boutin, A.-L., and Small, I. (2011). Organellar RNA editing. Wiley Interdiscip. Rev. RNA *2*, 493–506.

Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-García, A., Andrés, C., de la Luz Gutiérrez-Nava, M., Cantero, A., Delannoy, E., Jiménez, 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. Cell Mol. Biol. *56*, 590–602.

Freyer, R., López, C., Maier, R.M., Martín, M., Sabater, B., and Kössel, H. (1995). Editing of the chloroplast *ndhB* encoded transcript shows divergence between closely related members of the grass family (Poaceae). Plant Mol. Biol. *29*, 679–684.

Gray, M.W., and Covello, P.S. (1993). RNA editing in plant mitochondria and chloroplasts. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *7*, 64–71.

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.

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. Cell Mol. Biol. *36*, 541–549.

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

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv. Bioinforma. *2008*, 420747.

Huang, H., Yang, M., Su, Y., Qu, L., and Deng, X.W. (2015). Arabidopsis atypical kinases ABC1K1 and ABC1K3 act oppositely to cope with photodamage under red light. Mol. Plant *8*, 1122–1124.

Iyer, L.M., Zhang, D., Rogozin, I.B., and Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res. *39*, 9473–9497.

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
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. Cell Mol. Biol. *44*, 237–244.

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

Lundquist, P.K., Poliakov, A., Giacomelli, L., Friso, G., Appel, M., McQuinn, R.P., Krasnoff, S.B., Rowland, E., Ponnala, L., Sun, Q., et al. (2013). Loss of plastoglobule kinases ABC1K1 and ABC1K3 causes conditional degreening, modified prenyl-lipids, and recruitment of the jasmonic acid pathway. Plant Cell *25*, 1818–1839.

Martienssen, R.A., and Springer, P. (1997). Enhancer and gene trap transposons in Arabidopsis. Inser. Mutagen. Pract. Approach.

Martinis, J., Glauser, G., Valimareanu, S., and Kessler, F. (2013). A chloroplast ABC1-like kinase regulates vitamin E metabolism in Arabidopsis. Plant Physiol. *162*, 652–662.

Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K., Endo, T., Tasaka, M., and Shikanai, T. (2004). Cyclic electron flow around photosystem I is essential for photosynthesis. Nature *429*, 579–582.

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.

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. *104*, 8178–8183.

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

O'Malley, R.C., Alonso, J.M., Kim, C.J., Leisse, T.J., and Ecker, J.R. (2007). An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome. Nat. Protoc. *2*, 2910–2917.

Ounis, A., Cerovic, Z.G., Briantais, J.M., and Moya, I. (2005). Rasband WS, Image J. US National Institutes of Health, Bethesda, Maryland, USA.

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

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 N. Y. N *15*, 1142–1153.

Ruwe, H., Castandet, B., Schmitz-Linneweber, C., and Stern, D.B. (2013). Arabidopsis chloroplast quantitative editotype. FEBS Lett. *587*, 1429–1433.

Shikanai, T., Endo, T., Hashimoto, T., Yamada, Y., Asada, K., and Yokota, A. (1998). Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. Proc. Natl. Acad. Sci. U. S. A. *95*, 9705–9709.

Shikanai, T., Munekage, Y., Shimizu, K., Endo, T., and Hashimoto, T. (1999). Identification and characterization of Arabidopsis mutants with reduced quenching of chlorophyll fluorescence. Plant Cell Physiol. *40*, 1134–1142.

Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PloS One *8*, e65343.

Tseng, C.-C., Sung, T.-Y., Li, Y.-C., Hsu, S.-J., Lin, C.-L., and Hsieh, M.-H. (2010). Editing of *accD* and *ndhF* chloroplast transcripts is partially affected in the Arabidopsis *vanilla cream1* mutant. Plant Mol. Biol. *73*, 309–323.

Wagoner, J.A., Sun, T., Lin, L., and Hanson, M.R. (2015). Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. J. Biol. Chem. jbc.M114.622084.

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. Bioinforma. Oxf. Engl. *25*, 1189–1191.

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

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

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. Cell Mol. Biol. *59*, 1011–1023.

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.

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

Table 3.1: Oligonucleotides

Primer Name	Sequence 5'-3'	Purpose
QED1 start	ATGGCTATCTTCTCCACAGCAC	amplification
QED1 stop	GAAATATCACCAGAAATCGTTAC	amplification
QED1 5'2531	GCTTGCATAGTTGATGTTCTCG	genotyping
Ds3.1	ACCCGACCGGATCGTATCGGT	genotyping
RecA Reverse	GCTTCAGAGACAAGACTAGCTG	amplification
Ntag geno 3fs	GTCATCGTCATCCTTGTAATCG	genotyping
Ntag geno pkin F	CATCGGTAGTGATGAAGTCT	genotyping
Ntag geno pkin R	AACATGGCTCTCAGCCTTTCAA	genotyping

Chapter 4:
Conclusions and Future Studies

In chapter 2, we report a new PPR-DYW editing factor, QUINTUPLE EDITING FACTOR 1 (QED1) which is responsible for the editing of five chloroplast sites, the most sites known to be affected by a single plastid PPR editing factor. The DYW motif is known to have sequence and structural similarities to known deaminases, including highly conserved residues that resemble the deaminase active site (HxE/CxxC), but this motif for a number of factors has been shown to be dispensable for editing (Hayes et al., 2013b; Iyer et al., 2011; Okuda et al., 2009c). A number of PPR editing factors also function with only the E motif which has been shown to be essential for editing function for all editing factors tested. However, recent experiments have shown that the DYW motif can act in *trans* with the discovery of DYW1 which interacts with PPR-E editing factor CRR4 to edit the *ndhD* C2 site (Boussardon et al., 2012; Kotera et al., 2005). Site-directed mutagenesis was performed on DYW1 to identify residues critical for editing function and it was discovered that the deaminase-like signature was critical for editing along with the terminal DYW residues for which the motif is named (Boussardon et al., 2014). The DYW motif was also discovered to be able to bind zinc in DYW1 and another PPR-DYW editing factor, ELI1, an ability which is an important contributor in deaminase catalysis. The presence of zinc was found to be important for the plant RNA editing reaction itself, as editing was discovered to be sensitive to zinc chelators added to editing competent extracts (Boussardon et al., 2014; Hayes et al., 2013; Hegeman et al., 2005). Removing the deaminase signature of DYW1 or mutagenizing its deaminase signature significantly inhibited its ability to bind zinc ions, specifically associating the motif's ability to bind zinc with the deaminase signature (Boussardon et al., 2014)

We have found that the deaminase signature in the DYW motifs of two plastid PPR editing factors is critical for efficient editing function, an observation which is consistent with the hypothesis that the DYW motif carries out the deaminase function. The importance of the DYW motif was also observed with mitochondrial PPR-DYW editing factor MITOCHONDRIAL EDITING FACTOR 1 (MEF1), as well as plastid factors OTP84 and CREF7, which also required their DYW motifs for efficient editing function, although the deaminase signature has not been targeted in MEF1 and only the E in the HxE portion of the deaminase signature was targeted in OTP84 and CREF7's DYW motifs (Hayes et al., 2015; Zehrmann et al., 2010). The DYW motif can act in *trans* and can contribute to editing function for at least one E-class editing factor; therefore, a *trans*-acting DYW motif can possibly perform the same function for other E-class editing factors or even other PPR-DYW factors that do not need its own DYW motif for editing (Boussardon et al., 2012; Okuda et al., 2009). Additionally, the DYW motif has been shown to interact with the RNA around the C-target (Okuda et al., 2014). Despite the compelling evidence that the DYW motif may carry the deaminase activity in the RNA process, this motif has not yet been shown directly to exhibit deaminase activity in vitro.

The editing PPR proteins were discovered to be site-specific factors when CRR4 was demonstrated to bind the RNA near the site it edited when the protein and RNA was expressed in *E. coli*, but editing did not occur with these two components only (Okuda et al., 2006). In the case of CRR4, it is now known that it requires DYW1 as well as additional protein factors, such as RIP2 and RIP9, OZ1, ORRM1, and PPO1, which are known to interact with CRR4 and DYW1 and affect the editing of *ndhD* C2 (Bentolila et al., 2013a; Boussardon et al., 2012; Sun et al., 2013, 2015; Takenaka et al., 2012; Zhang et al., 2014). In the future, CRR4 and DYW1

could be expressed in *E. coli* with RNA containing the *ndhD* C2 site to assay whether these components together can edit the protein. Alternatively, since QED1's own DYW motif is important for editing function, it may be more straightforward to express a single protein and RNA rather than attempt to express multiple factors at once in an *E. coli* system. Either editing will occur, which would provide strong evidence that the DYW motif itself is capable of deaminase activity, or the lack of editing would indicate that other factors are needed to drive the reaction (members of the RIP family, ORRM1, etc).

If the currently known factors cannot reconstitute the editing reaction, then other currently unknown protein factors likely also function in the editosome to perform the editing reaction. The RIP family was identified through co-immunoprecipitation of RARE1 (Bentolila et al., 2012). However, this co-immunoprecipitation was done with RARE1 tagged at its C-terminus with a 3xFLAG tag and this construct could complement the editing phenotype but not to wild-type levels (Bentolila et al., 2012). We observed that a 3xFLAG-StrepII tag appended to the C-terminus of QED1 could not complement the editing phenotype well which is consistent with what was observed with MEF1 when it was appended with a C-terminal 6xHis tag, implying that a C-terminal tag greatly interferes with editing function (Wagoner et al., 2015; Zehrmann et al., 2010). When we tested QED1 tagged at its N-terminus with the 3xFLAG-StrepII tag in *qed1* protoplasts, this construct was able to more efficiently complement the editing phenotype than the C-tagged construct (Wagoner et al., 2015). Boussardon et al (2014) found that the mutagenesis of the terminal tryptophan in the DYW motif completely eliminated editing of the *ndhD* C2 site, indicating that this residue is important for editing function (Boussardon et al., 2014). Boussardon et al (2014) hypothesized that the terminal DYW tripeptide may function as a docking signal as it shares some characteristics with other known C-

terminal protein-protein interaction signals (Boussardon et al., 2014). Therefore, the presence of a tag at that location may prevent or weaken interactions with factors that may contain deaminase function or interact with a DYW motif to help it perform the deamination if the DYW motif itself is the deaminase. The inability of certain editing factors to efficiently specify editing when they carry a C-terminal tag is consistent with this hypothesis.

In order to identify other protein factors that function in the editosome and may possibly function to drive the editing reaction, we have constructed a version of QED1 with a 3xFLAG-StrepII tag at its N-terminus which can efficiently complement the editing phenotype (Wagoner et al., 2015). This construct has been stably integrated into plants and can be used for co-immunoprecipitation. The identity of the components in the isolated complex can then be determined using mass spectrometry, similar to the experiments performed with RARE1 and ORRM1 (Bentolila et al., 2012; Sun et al., 2015). It would be interesting to compare the identity of components immunoprecipitated with QED1 to those found when RARE1 was immunoprecipitated. If co-immunoprecipitation was performed with an E-class editing PPR protein, or with a PPR-DYW protein that does not need its DYW motif, the proteins identified in these complexes could be particularly informative, especially when compared to proteins identified in co-immunoprecipitations with editing proteins that do need their DYW motif, like QED or RARE1. A cursory examination of a protein alignment comparing the C-terminal motifs of plastid editing factors that have been shown to require their DYW motifs through truncation

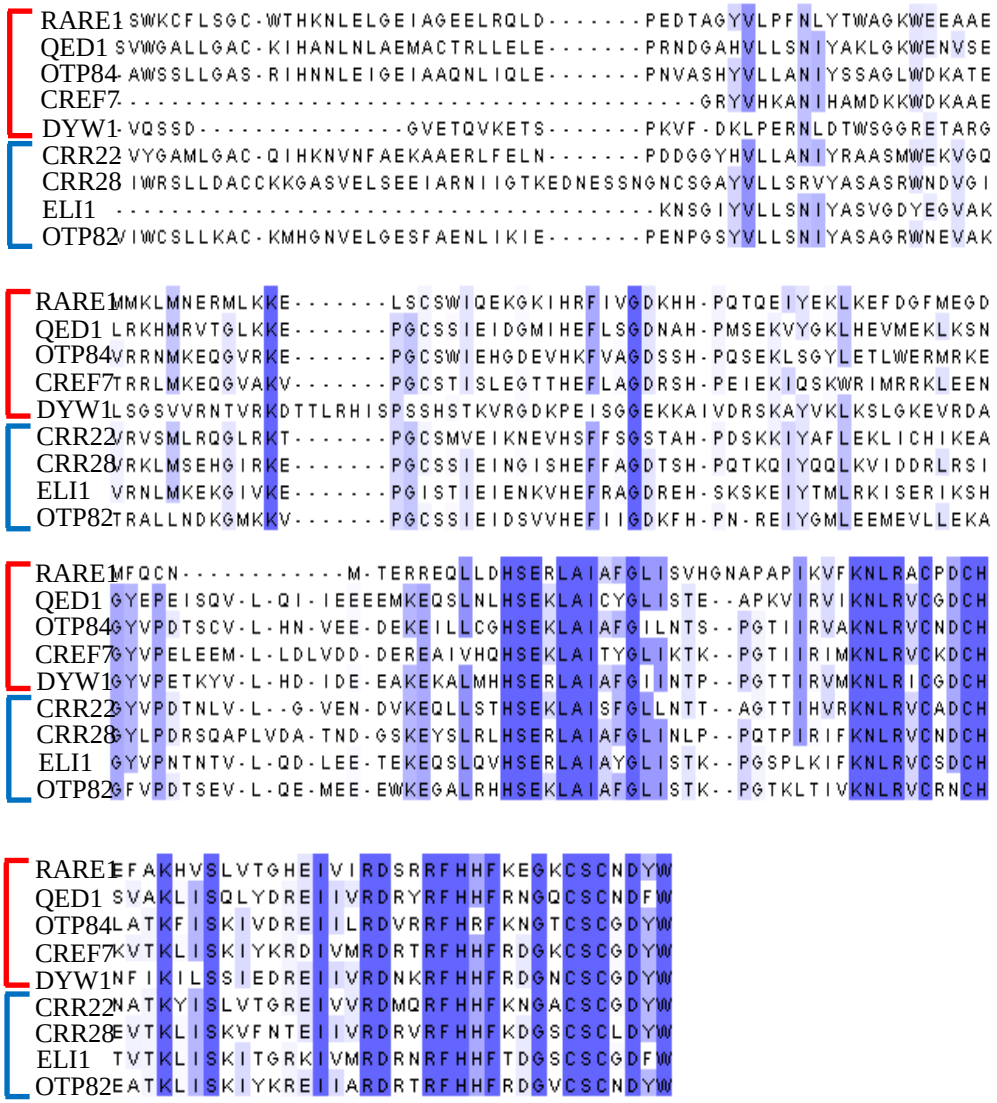


Figure 4.1: Comparison of C-terminal motifs between plastid editing PPR-DYW proteins that require the DYW motif for editing and those that do not require the DYW motif for editing. DYW, E+ and partial E motif protein sequences were obtained from the NCBI database for nine plastid PPR-DYW editing factors whose DYW motifs were shown to be required or not required through truncation analysis (RARE1 NP_196831.1, QED1 NP_180537.1, OTP84 NP_191302.2, CREF7 NP_201453.1, DYW1 NP_175189.2, CRR22 NP_172596.1, CRR28 NP_176180.1, ELI1 NP_195454.1, OTP82 NP_172286.1). Red bracket denotes editing PPR-DYW proteins that require their DYW motif for editing, blue bracket denotes editing PPR-DYW proteins that do not require their DYW motif.

analysis (RARE1, QED1, OTP84, CREF7) as well as DYW1, and plastid editing factors for which the DYW motif has been shown to be dispensable also through truncation analysis (CRR22, CRR28, ELI1, OTP82) does not reveal any obvious protein sequence elements shared within one group but absent in other which may indicate why the DYW motif is needed for some proteins but not for others (Figure 4.1) (Boussardon et al., 2012; Hayes et al., 2013, 2015; Okuda et al., 2009, 2010; Wagoner et al., 2015). This observation may imply that other protein interactors could render the DYW motif dispensable or indispensable rather than the presence or absence of particular protein motifs. This information could also be gleaned from a comparison of editosomes involving editing PPR-DYW proteins that need their DYW motif and editing PPR-DYW proteins that do not need their DYW motifs. It was also observed that the E motifs for PPR editing factors seem to have evolved some organelle specificity, as PPR-E class editing factors with their E motif exchanged with an editing factor targeted to the other organelle could not complement the editing phenotype, although complementation is observed when E motifs are exchanged between editing factors both targeted to the same organelle (Chateigner-Boutin et al., 2013; Okuda et al., 2009). To probe organelle specificity, the DYW motif could be exchanged between QED1, which requires its DYW motif and is targeted to the plastid, and MEF1, which also requires its DYW motif and is targeted to the mitochondria. This motif swap experiment can indicate if the DYW motif has also undergone organelle specialization, similarly to what was observed with the E motifs and may confirm that specific organellar proteins are necessary to drive the editing reaction and not some general factor that can be dual targeted. If complementation does happen when the DYW motif is swapped, this result could help support the hypothesis that the DYW motif is capable of deaminase activity.

In chapter 3, we investigated the physiological consequences of the loss of editing at the

five chloroplast sites affected by QED1. The *qed1* mutant plants were observed to grow more slowly than wild-type plants, appeared to accumulate photodamage in the leaves, and had an abnormal chloroplast morphology phenotype. We investigated the phenotype of *qed1* mutant plants transformed with the wild-type QED1 coding region but did not observe a strong complementation of the observed physical phenotypes in the few transformants that were observed, although the editing phenotype was complemented to near wild-type levels. Progeny of these transformants, particularly those carrying two copies of the transgene, will have to be identified and assessed for complementation of the mutant phenotypes.

The analysis of the QED1 protein sequence in a variety of species and the investigation of the editing status of the codons that QED1 affects in the *matK, rpoB,* and *ndhB* coding regions in a number of plant species does have some interesting evolutionary implications. When the editing status of these particular codons was investigated, it was found for most of the species that the *matK* and *rpoB* codons were particularly conserved such that if a T was not already encoded in the genome, the site was edited to a U in the RNA to maintain the conserved codon. For *matK*, many of the species already had a genomically encoded T at that position, and the species that encoded a C edited that site. While more species had a C encoded at the *rpoB* site, most of them did edit that site to a U to change the codon to what is observed in *rpoB* of the monocots. Interestingly, all of the species that were investigated encoded a C in the genome for the *ndhB* site, but only Arabidopsis appears to edit that C to a U, changing the codon to something not conserved in other species. Unexpectedly, the sequence upstream of the *ndhB* site is perfectly conserved in the few species that were investigated that also appear to have a version of QED1 encoded in the genome (Figure 4.2). Therefore, the protein sequences of genes orthologous to QED1 in other species were compared to ascertain whether differences in protein

sequences could explain the differences in editing status at the *ndhB* site. The protein sequences from other plant species found to have significant sequence similarity with QED1 were particularly conserved in the C-terminal domains, where the deaminase signature was found to be invariant between the species, and in the regions in specific PPR motifs that encode the amino acids that specify RNA recognition (Barkan et al., 2012; Takenaka et al., 2013b; Yagi et al., 2013). Therefore, the versions of QED1 found in different species were analyzed with respect to the PPR code. When the RNA sequence predicted by each version of QED1 was compared, the predicted RNA sequences were very well conserved which makes the *ndhB* result even more surprising (Figure 4.3). If QED1 exists in these other species and should recognize that *ndhB* *cis*-element, why does it not edit it? It would be an interesting experiment to express Arabidopsis QED1 in some of these other species to determine if Arabidopsis QED1 can edit the *ndhB* site. If it can edit the site, this may indicate that there is a feature of QED1 in other species that prevents it from editing that specific site. If it cannot edit that site, it may indicate that there may be other factors other than the PPR proteins that might be involved in site selection.

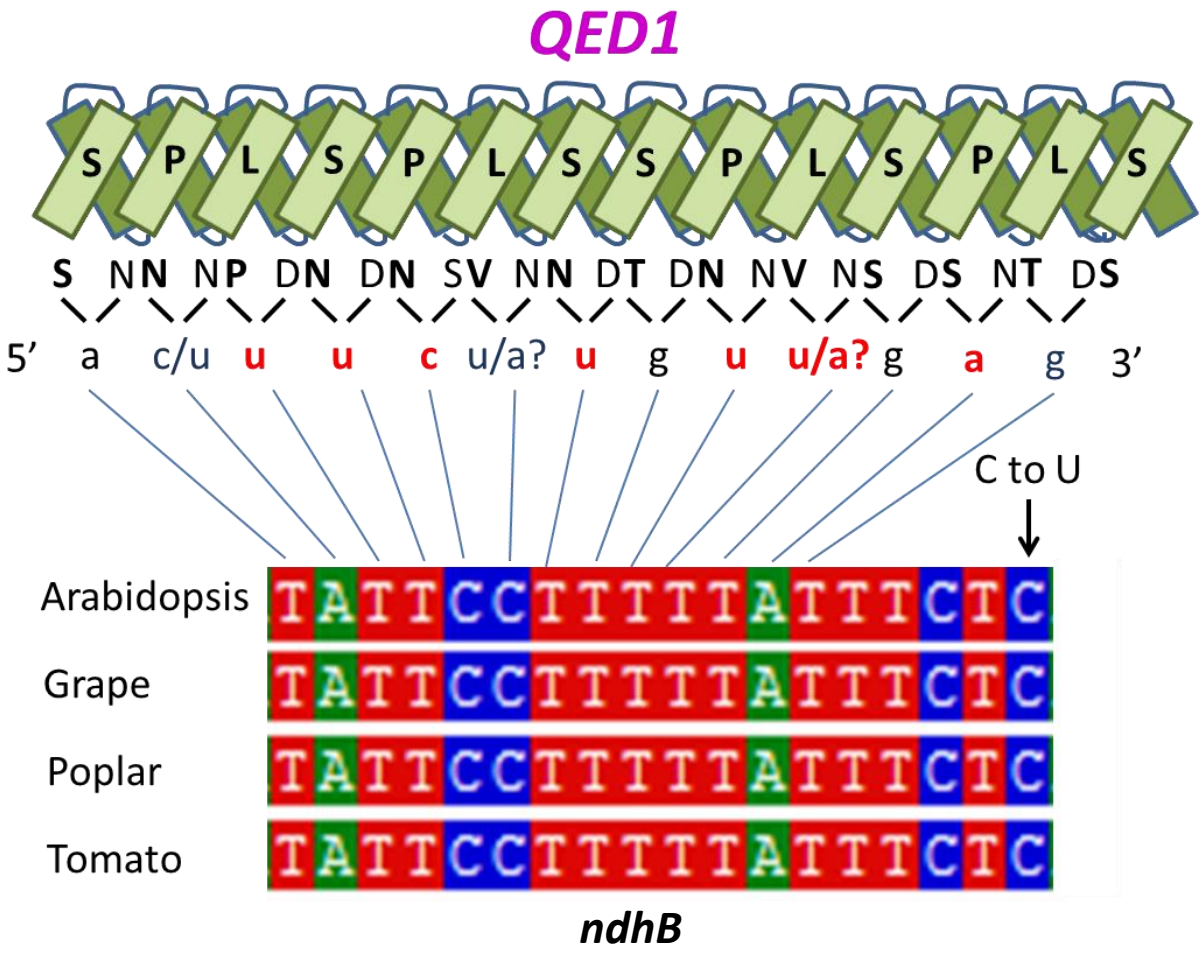


Figure 4.2: Comparison of the consensus RNA sequence predicted to be recognized by QED1 and a DNA alignment of the *cis*-element of the *ndhB* C-target edited by QED1 in Arabidopsis. The predicted RNA sequence was generated through analysis of QED1's PPR motifs using the Barkan et al PPR code. The partial ndhB DNA sequence for *Arabidopsis thaliana* (Arabidopsis) NC_000932.1, *Vitis vinefera* (grape) NC_007957.1, *Populus trichocarpa* (poplar) NC_009143.1, and *Solanum lycopersicum* (tomato) NC_007898.3 were obtained from the NCBI database and aligned using Sequencher.

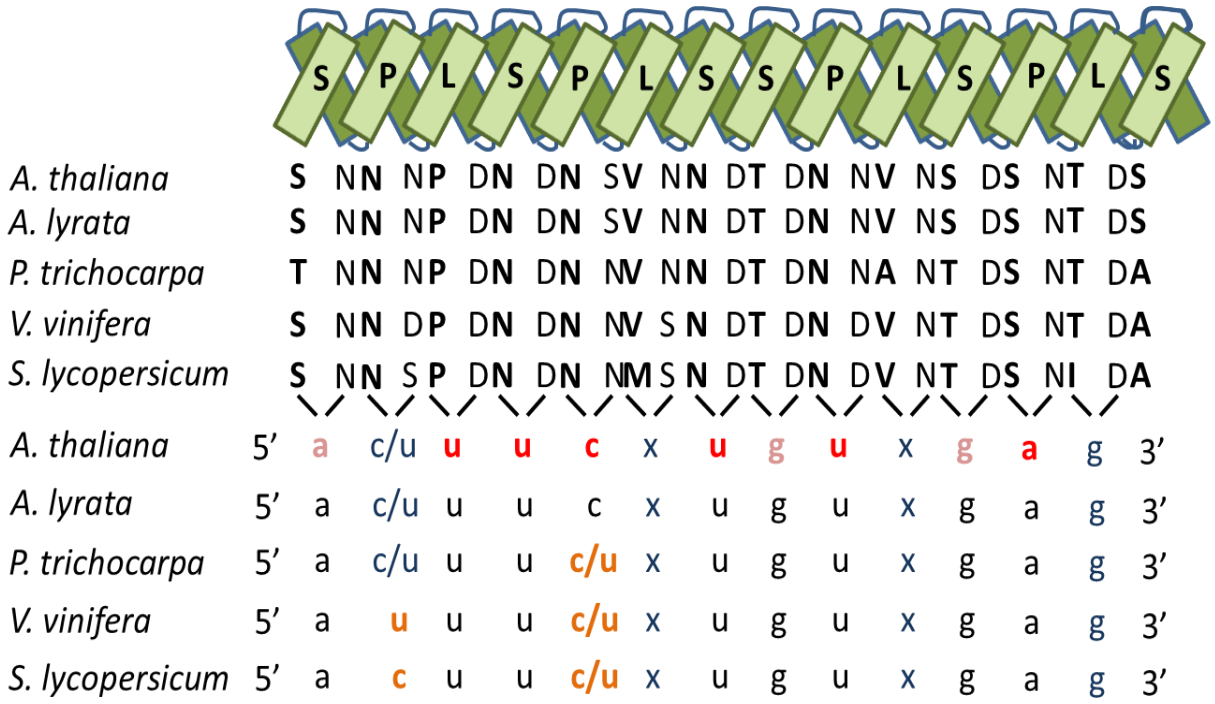


Figure 4.3: Comparison of the RNA sequence predicted to be recognized by QED1's orthologs in a variety of species. The predicted RNA sequences were generated through analysis of QED1's PPR motifs using the Barkan et al PPR code. Orthologs of Arabidopsis QED1 were identified through NCBI's BLAST using the Arabidopsis QED1 protein sequence as the query. Red letters indicate RNA nucleotides predicted by Arabidopsis QED1 that are conserved in all five of QED1's editing sites' *cis*-elements in Arabidopsis. Pink letters are RNA nucleotides predicted by Arabidopsis QED1 that are conserved in some of QED1's editing sites' *cis*-elements in Arabidopsis. Orange letters indicate preferred RNA nucleotides that differ from what is preferred by Arabidopsis QED1.

REFERENCES

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

Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., Wijk, K.J. van, and Hanson, M.R. (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proc. Natl. Acad. Sci. *109*, E1453–E1461.

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

Boussardon, C., Salone, V., Avon, A., Berthomé, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C. (2012). Two interacting proteins are necessary for the editing of the *NdhD-1* site in Arabidopsis plastids. Plant Cell *24*, 3684–3694.

Boussardon, C., Avon, A., Kindgren, P., Bond, C.S., Challenor, M., Lurin, C., and Small, I. (2014). The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of *ndhD-1*. New Phytol. *203*, 1090–1095.

Chateigner-Boutin, A.-L., Colas des Francs-Small, C., Fujii, S., Okuda, K., Tanz, S.K., and Small, I. (2013). The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. Plant J. Cell Mol. Biol. *74*, 935–945.

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

Hayes, M.L., Dang, K.N., Diaz, M.F., and Mulligan, R.M. (2015). A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. J. Biol. Chem. *290*, 10136–10142.

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.

Iyer, L.M., Zhang, D., Rogozin, I.B., and Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res. *39*, 9473–9497.

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

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.

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.

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

Okuda, K., Shoki, H., Arai, M., Shikanai, T., Small, I., and Nakamura, T. (2014). Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. Plant J. Cell Mol. Biol. *80*, 870–882.

Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M.R., and Bentolila, S. (2013). An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc. Natl. Acad. Sci. U. S. A. *110*, E1169–E1178.

Sun, T., Shi, X., Friso, G., Van Wijk, K., Bentolila, S., and Hanson, M.R. (2015). A zinc finger motif-containing protein is essential for chloroplast RNA editing. PLoS Genet *11*, e1005028.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B., and Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc. Natl. Acad. Sci. *109*, 5104–5109.

Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PloS One *8*, e65343.

Wagoner, J.A., Sun, T., Lin, L., and Hanson, M.R. (2015). Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. J. Biol. Chem.

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

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

Zhang, F., Tang, W., Hedtke, B., Zhong, L., Liu, L., Peng, L., Lu, C., Grimm, B., and Lin, R. (2014). Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. Proc. Natl. Acad. Sci. *111*, 2023–2028.