SEQUENCE ELEMENTS ESSENTIAL FOR RNA EDITING IN CHLOROPLASTS

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

Michael Lloyd Hayes

January 2008

SEQUENCE ELEMENTS ESSENTIAL FOR RNA EDITING IN CHLOROPLASTS

Michael Lloyd Hayes, Ph. D.

Cornell University 2008

Organellar transcripts of angiosperms contain cytidines that are specifically edited to uridines. Almost all editing events lead to the restoration of a codon for a conserved amino acid instead of one predicted from the organelle genome. The sequences immediately surrounding an edited C are required for proper selection of the C target. There is no single consensus sequence near all known editing sites in a plant species, suggesting that different sequences are required to edit different C targets in plant organelles. However, chloroplast sites can be grouped into clusters of two to five sites that share similar sequences within 30 nt 5' and occasionally 30 nt 3' to the C target of editing. The importance of nucleotides proximal to two tobacco chloroplast editing sites was investigated *in vivo* and *in vitro*.

A site in *rpoB* was chosen for analysis because sequences around the C target are shared with two other sites. In order to determine the nucleotides critical for editing, an *in vitro* chloroplast editing assay was developed with chloroplast extracts from multiple plant species. Using this *in vitro* assay, the essential sequence elements for editing were delimited to a region containing sequences of *rpoB* that are shared with two other editing sites. Wild-type substrates were edited to a greater extent *in vitro* compared to substrates with nucleotide differences in the common sequences.

Because substrates containing a site in the *psbE* gene were efficiently edited *in vitro*, the nucleotides surrounding this C target were also thoroughly analyzed. A 6 nucleotide motif was discovered to be necessary for editing. The precise distance between a proximal cis-element and the C target was found to affect selection of the

proper C. Unlike the common sequences around the site examined in *rpoB*, the 6 nt essential cis-element was not present around other known editing sites. The motif is commonly present in species that have a homologous editing site in *psbE*, but is less frequently present in plants that have a genomically encoded T at the same position. Therefore, it is likely that the crucial motif has co-evolved with the editing site.

BIOGRAPHICAL SKETCH

Michael Lloyd Hayes was born in Fort Worth, Texas on April 18, 1980. He grew up in Lexington, KY and graduated from Henry Clay High school in 1998. He attended the University of Kentucky and graduated in 2002 with a B.S. degree, *magna cum laude*, in Agricultural Biotechnology. During his four years at the University of Kentucky he diligently worked in the laboratory of Professor Glenn B. Collins. He entered the graduate program in field of Biochemistry, Molecular and Cellular Biology at Cornell University in 2002. In March 2003 he became a member of Dr. Maureen Hanson's laboratory until the completion of his degree.

This dissertation is dedicated to my girlfriend Gudrun Reiterer who has supported me throughout my studies. Also this dissertation is dedicated my mother Pamela Hancock and my grandmother Myrene Hancock for giving me the skills necessary for my pursuit of higher education. Finally this dissertation is dedicated to my grandfather Lloyd Hancock who instilled in me a love for plants and the natural world.

ACKNOWLEDGMENTS

I would like to especially thank my major advisor Professor Maureen R. Hanson for her support and guidance throughout my graduate studies. I also thank my committee members Professor David B. Stern and Professor Klaas van Wijk for critical feedback and advice concerning my project.

I thank all the members of the Hanson lab past and present that have helped me throughout the years. The expertise of the skilled Dr. Carla Hegeman was instrumental to the development of my graduate research project. Insightful discussions with Jason D. Gillman and Dr. John R. Robbins in particular further aided my work.

I thank the federal grants that have made this accomplishment possible. This includes the NIH pre-doctoral fellowship and NSF grant MCB 0344007 to Professor Maureen R. Hanson.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

SIGNALS NECESSARY FOR SELECTING TARGETS OF PLANT ORGANELLE RNA EDITING-A REVIEW

ABSTRACT

Chloroplast and mitochondrial genomes encode a number of genes required for each organelle's distinct function. C-to-U RNA editing alters the sequence of many organellar transcripts in all but one group of embryophytes, land plants. Another form of RNA editing, conversion of U to C, modifies nucleotides in transcripts from the organelles of plants, but is absent in spermatophytes. No specific motif is frequently observed around all editing sites, raising questions about how the editing mechanism recognizes the correct nucleotide for editing. Sequences surrounding edited nucleotides are diverse, suggesting the action of varied trans-factors, though some general factors might be shared among editing targets. Several tobacco chloroplast editing sites can be placed into groups of 2-5 sites exhibiting limited upstream sequence similarity. For all organelle sites that have been examined in detail, an essential core cis-element is located about 11 nucleotides 5′ of the C target. Although the location of the element is similar, the critical region exhibits considerable sequence variability between sites. Also important for the efficient editing of some C targets are enhancer-like sequences in regions outside of the typical core element. Here I review the current knowledge of RNA editing cis-elements in mRNAs and tRNAs within the organelles of plants.

Prevalence of mRNA editing in plants

Many genes encoded by the chloroplast and mitochondrial genomes have transition mutations in codons that encode amino acids that are conserved in other plants. Organellar mRNAs and tRNAs of some speices do not contain the genomic

mutations present in others, so that the genomic DNA encodes the conserved amino acid (Covello & Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989; Hoch et al., 1991). Thus plants have evolved a post-transcriptional RNA editing mechanism capable of compensating for certain genomic mutations by modifying the identity of specific pyrimidine nucleotides of transcripts. Although many RNA processing events are known to modify organellar transcripts, in plants only two different forms of editing have been described, C-to-U and U-to-C (Hiesel et al., 1994; Yoshinaga et al., 1996). The number of events and the type of editing varies across plant lineages (Table 1.1).

All major groups of land plants with the exception of Marchantiidae, complex thalloid liverworts, modify RNAs by C-to-U editing (Steinhauser et al., 1999). None of the chlorophyte algae or Marchantiidae studied to date utilize either C-to-U or U-to-C RNA editing. The absence of C or U targets in Marchantiidae is likely due to a secondary loss of the editing mechanism, lending support for a single emergence of plant organellar editing in a common ancestor of all land plants (Groth-Malonek et al., 2007).

Editing sites vary in number and location in chloroplasts and mitochondria from different plant species. The extent to which some C targets are edited differs between tissue types, environmental conditions, and ecotypes (Karcher & Bock, 2002; Peeters & Hanson, 2002; Bentolila et al., 2005). Chloroplasts of seed plants typically contain 30-40 C targets, with many editing sites specific to a single species (Maier et al., 1995; Wakasugi et al., 1996; Corneille et al., 2000; Tillich, 2001; Peeters & Hanson, 2002; Schmitz-Linneweber et al., 2002; Sasaki et al., 2003; Inada et al., 2004; Tillich et al., 2005; Kahlau et al., 2006; Sasaki et al., 2006; Zeng et al., 2007). Alignment of genes that contain editing sites in one plant with chloroplast genes of other plants indicates that genomically encoded Ts are frequently found at the same

Table 1.1. Number, location, and type of editing sites identified in plant mRNAs

Species	Organelle	C→U	U→C	References
Anthoceros formosae	Chloroplast	509	433	(Kugita et al., 2003)
Adiantum capillus-veneris	Chloroplast	315	35	(Wolf et al., 2004)
Pinus thunbergii	Chloroplast	26	0	(Wakasugi et al., 1996)
Nicotiana tabacum	Chloroplast	38	0	(Kahlau et al., 2006; Sasaki et al., 2006)
Nicotiana sylvesetris	Chloroplast	33	0	(Sasaki et al., 2003)
Nicotiana tomentosiformis	Chloroplast	32	0	(Sasaki et al., 2003)
Arabidopsis thaliana	Chloroplast	28	0	(Tillich et al., 2005)
Zea mays	Chloroplast	28	0	(Maier et al., 1995; Corneille et al., 2000; Tillich, 2001; Peeters & Hanson, 2002)
Pisum sativum	Chloroplast	27	0	(Inada et al., 2004)
Phalaenopsis Aphrodite	Chloroplast	44	0	(Zeng et al., 2007)
Atropa belladonna	Chloroplast	31	0	(Schmitz-Linneweber et al., 2002)
Oryza sativa	Chloroplast	26	0	(Corneille et al., 2000; Inada et al., 2004)
Brassica napus	Mitochondrion	427	0	(Handa, 2003)
Arabidopsis thaliana	Mitochondrion	456	0	(Giege & Brennicke, 1999)
Oryza sativa	Mitochondrion	491	0	(Notsu et al., 2002)
Beta vulagris	Mitochondrion	357	0	(Mower & Palmer, 2006)

position as edited Cs (Wintz & Hanson, 1991). In fact, there is no known editing site that has been found to be conserved in the chloroplasts from all examined seed plants.

Chloroplasts and mitochondria of seedless plants have an order of magnitude more editing targets relative to chloroplasts from seed plants. 350 and 942 nucleotides are known to be edited in the chloroplasts of the fern *Adiantum capillus-veneris* and the hornwort *Anthoceros formosae*, respectively (Kugita et al., 2003; Wolf et al.,

2004). The average number of editing targets in chloroplasts of mosses and lycopods is not known. Sequencing of cDNA from transcripts of mosses and lycopods suggests they contain many editing sites (Freyer et al., 1997; Sugita et al., 2006; Tsuji et al., 2007). The mitochondria of all higher plants examined have over 350 C targets of editing (Giege & Brennicke, 1999; Notsu et al., 2002; Handa, 2003; Mower & Palmer, 2006). The number of angiosperm mitochondrial editing sites may be underestimated when RNAs from only one tissue type is analyzed, as 67 editing events were detected in Arabidopsis leaf RNA that were not found in suspension culture mitochondria RNA (Bentolila et al., 2007). A typical number of editing targets in the mitochondria of a seedless plant has not been determined although there is evidence that editing pervades the transcriptome (Malek et al., 1996).

Seed plant chloroplasts maintain around 30 C targets of editing

Within chloroplasts of seed plants, an unknown mechanism may limit the number of editing sites to approximately 30 to 40. Some of the variability in the set of chloroplast sites in different species can be explained by ancestral transfers of genes to the nucleus. For instance, the chloroplast genome of *Phalaenopsis aphrodite* encodes some *ndh* genes with insertion/deletions and others with large truncations, but all have frameshifts compared with tobacco (Zeng et al., 2007). All of the *ndh* genes encoded in the chloroplast of this species are likely to be pseudogenes, implicating the transfer of functional genes for the complex to the nucleus. Although around 40% of the editing sites in most angiosperms are found in *ndh* genes, only one editing site was identified in the *ndhB* pseudogene in *Phalaenopsis aphrodite* chloroplasts (Zeng et al., 2007). The relative absence of editing sites in the *ndh* genes of *P. aphrodite* is probably due to the lack of selection for coding regions in the pseudogenes. In pines, *ndh* genes are not encoded by the chloroplast genome (Wakasugi et al., 1996). Despite the absence of the functional *ndh* genes, which exhibit a majority of the editing sites in

most angiosperms, *Pinus thunbergii* and *Phalaenopsis aphrodite* still have at least 26 and 44 distinct Cs that are edited, respectively.

It is not currently clear why plastids maintain C targets when Ts are present at the same positions in other plants, suggesting that most sites could be lost through a C to T genomic mutation with little consequence (Wintz & Hanson, 1991). Editing site emergence must be dynamic because of the large number of C targets that are species specific. Possibly there is an energetic cost for editing that limits the accumulation of new editing targets in seed plants, preventing the total number of sites from greatly exceeding 40.

C-to-U modification editing in tRNAs

In addition to editing of mRNAs, the organelles of several species modify Cs to Us in immature tRNAs. Changes in tRNAs have often been referred to as modification events instead of editing, but a number of different alterations in the nucleotides of tRNAs are now known from a variety of organisms (Gott and Emeson, 2000). C-to-U changes in plant tRNAs are likely to be mechanistically similar to C-to-U mRNA editing and therefore there is little reason to segregate these changes from editing sites in mRNAs.

tRNA editing has been observed in the chloroplasts of seedless plants and the mitochondria of particular gymnosperms and dicots (Binder et al., 1994; Marechal-Drouard et al., 1996b; Fey et al., 2000; Kugita et al., 2003; Wolf et al., 2004) (Table 1.2). Of all identified C targets, the one in *trnC* is edited in the largest group of plants including cycads and many dicots (Fey et al., 2000). Like editing sites in mRNAs, the number and location vary between species; however, editing sites in tRNAs are by far rarer, with between 0-3 sites known per organelle of a plant species (Fey et al., 2002). Arabidopsis mitochondria are unlikely to have any tRNA editing sites (Giege & Brennicke, 1999). Several other plants probably also do not require tRNA editing in

the chloroplasts and/or mitochondria. Since all known tRNA editing events are in species with extensive C-to-U editing of mRNAs, a similar or related editing mechanism for sites in both types of RNA can be suspected.

The function of tRNA editing, like mRNA editing, appears to be a compensation mechanism for genomic mutations. Unedited tRNAs that contain the genomic encoded sequence are not fully processed (Marchfelder et al., 1996; Marechal-Drouard et al., 1996a; Marechal-Drouard et al., 1996b). Currently all tRNA editing sites discovered in plant mitochondria (Table 1.2) restore base pairing of loops critical for the secondary structure and function of tRNAs. On the other hand, the two sites identified in chloroplasts modify the anticodon of a tRNA (Kugita et al., 2003; Wolf et al., 2004). The modified anticodon better matches the codon usage of the plastid genome.

Table 1.2. RNA editing sites identified in tRNAs

Species	Organelle	tRNA	C→U	References
Anthoceros formosae	Chloroplast	*tnrK*	1	Kugita et al., 2003
Adiantum capillus-veneris	Chloroplast	*trnL*	1	Wolf et al., 2004
Larix leptoeuropaea	Mitochondrion	*trnH*	3	Marechal-Drouard et al., 1996b
Oenothera berteriana *Phaseolus vulgaris* *Solanum tuberosum*	Mitochondrion	*trnF*	1	Binder et al., 1994, Marechal-Drouard et al., 1993
Oenothera berteriana *Solanum tuberosum* *Magnolia grandiflora* *Ceratozamia mexicana*	Mitochondrion	*trnC*	1	Binder et al., 1994, Fey et al., 2002

U-to-C editing occurs in seedless plants

U-to-C editing of mRNA is currently believed to be restricted to the mitochondria and chloroplasts of seedless plants. Plants found to perform U-to-C

editing in mitochondria exhibit the same activity in chloroplasts (Tillich et al., 2006). One seed plant, *Oenothera berteriana*, was reported to contain one U target in the mitochondria, but this report awaits further confirmation (Schuster et al., 1990). Using our nomenclature, initiated in Hayes et al. (2006), the site is called OBcob U311, where OB represents the species *Oenothera berteriana*, cob symbolizes the gene name, U indicates the nucleotide modified, and 311 is the number of nucleotides from the A of the initiation codon. Most dicots have a T at the corresponding location in the mitochondrial genome and there is no evidence that modification to C occurs in other species. Also, no U-to-C editing has been described in any other seed plants. The U-to-C event is likely either a sequencing artifact or an unusual case restricted to *Oenothera berteriana*. The shared time of emergence (ancestor of land plants) and loss (ancestor of seed plants) of U-to-C editing in both chloroplasts and mitochondria suggests a similar or shared biochemical mechanism between the two organelles.

Most C targets in plants are in the second position of the codon and are surround by particular nucleotides

A single common sequence around all C targets with enough complexity to direct the highly specific editing mechanism has not been discovered. Nevertheless, some nucleotides are more frequently found next to editing targets. Computer programs have been developed that use the nucleotide frequencies to predict with reasonable accuracy whether a C will be modified in plant mitochondria, based on the nucleotides surrounding a C (Cummings & Myers, 2004; Mower, 2005; Du et al., 2007). The most common sequence feature around editing sites is the presence of a 5′ pyrimidine and a 3′ purine adjacent to the edited C in both mitochondria and chloroplasts. One difference between organelles is that in chloroplasts an A is frequently present 3′ of a C target whereas in mitochondria a G is usually found 3′ (Mulligan et al., 2007).

Editing sites are more frequent in certain codons than others in both chloroplasts and mitochondria (Maier et al., 1995; Giege & Brennicke, 1999). The bias for nucleotides that flank editing sites can be explained by restrictions on emergence of new sites due to the genetic code and nucleotide preferences for the editing machinery. Most editing sites are in coding regions and at the second position of a codon (Mulligan et al., 2007). The relative absence of editing sites at third codon positions and UTRs has been theorized to be due to the lack of evolutionary pressure and/or lack of maintenance of editing machinery for superfluous nucleotide changes. Although rare, a few Cs have been discovered to be edited in codons that either do not encode an evolutionarily conserved amino acid or would not be altered by editing. Seemingly unnecessary targets may be edited due to the presence of signals similar to those recognized by a trans-factor for a C target where editing is essential. Even at first and second positions, the codons in which Cs are edited demonstrate some partiality. Editing sites rarely occur in codons that would encode a similar amino acid after modification (Giege & Brennicke, 1999). Also, only a few editing sites have been discovered in infrequent codons, such as stop codons (Giege & Brennicke, 1999).

Local sequence restrictions influenced by the genetic code are not sufficient to explain the codon bias of editing sites. In Arabidopsis chloroplasts, modification of UCN codons to UUN would have similar effects on the encoded polypeptide based on the genetic code. However, UCA is by far the most common codon containing a C targeted by editing in chloroplasts (Giege & Brennicke, 1999). Likewise CCN modification to CUN leads to similar alterations in the amino acid encoded. Yet four distinct Cs are edited at the second position of CCA codons and only one C is edited at the second positions of a CCU and a CCC codon from Arabidopsis chloroplasts (Giege & Brennicke, 1999). These codon preferences must represent a neighboring nucleotide bias for the acquisition of a C that can be edited.

Editing is most efficient when a 5' pyrimidine and a 3' purine flanks the C target

Initial methods used to determine chloroplast editing cis-elements utilized the expression of exogenous editing templates in transplastomic plants (Bock et al., 1996; Chaudhuri & Maliga, 1996; Bock et al., 1997; Reed et al., 2001b, Figure 1.1). Expression of editing substrates by chloroplast transformation requires multiple cloning steps as well as time consuming transformation and regeneration steps. The laborious process limits the amount of editing substrates that can be surveyed. A method for mitochondrial transformation has not been reported, preventing analysis of exogenous editing substrates in transformed mitochondria. The development of *in vitro* editing assays for C targets using both mitochondrial (Takenaka & Brennicke, 2003) and chloroplast (Hirose & Sugiura, 2001) extracts and an *in organello* editing assays for mitochondrial targets (Farre & Araya, 2001; Staudinger & Kempken, 2003) allowed for the survey of critical cis-elements in greater detail (Figure 1.2). The location and sequences of important cis-elements can now be compared between mitochondria and chloroplasts.

Several editing substrates constructed with a different nucleotide 5′ of the editable C compared to the native sequence are reduced in the extent they are edited (Bock et al., 1996; Chaudhuri & Maliga, 1996; Farre et al., 2001). Preferences for certain adjacent nucleotides in editing substrates *in vitro* are similar to those suggested from genome analysis and transplastomic plants but there is some flexibility depending on the C target. In chloroplasts editing of NTrpoB C473 requires a 5' purine (ML Reed and MR Hanson, unpublished); NTpsbE C214 prefers a 5′ adjacent U or A (Miyamoto et al., 2004); a 5′ neighboring G inhibits editing of NTpetB C611 and NTndhB C746 (Bock et al., 1996; Miyamoto et al., 2004); and editing of NTpsbL C2 is not tolerant of a 5′ C (Chaudhuri & Maliga, 1996). Nucleotide preferences for editing of mitochondrial sites are similar. PSatp9 C20 prefers a 3′ adjacent A (Neuwirt

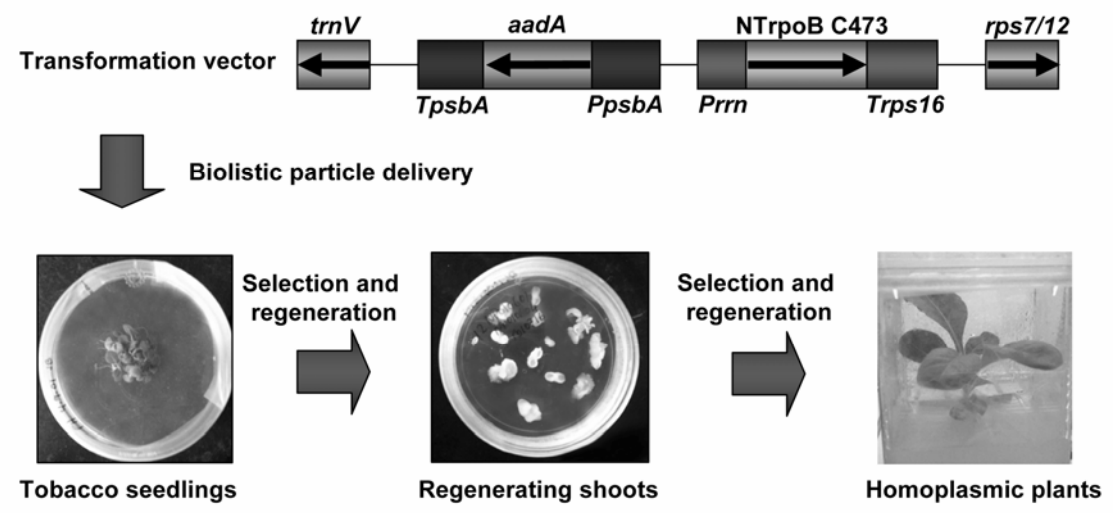


Figure 1. 1 The method for expression of editing sites *in vivo* is represented in the above diagram. At the top of the figure a series bars, arrows, and lines signify a typical transformation vector that can be constructed; light grey bars denote coding regions that are labeled above the appropriate bar; arrows indicate the direction of transcription of the coding regions; dark shaded bars illustrate promoter and terminator regions and are labeled below the corresponding region; lines represent the regions of the vector that are neither promoters nor coding regions. In the lower left panel, tobacco seedlings at the stage of particle bombardment are shown. In the center, regeneration shoots on antibiotic containing medium are illustrated. In the lower right a regenerated homoplasmic plant is shown.

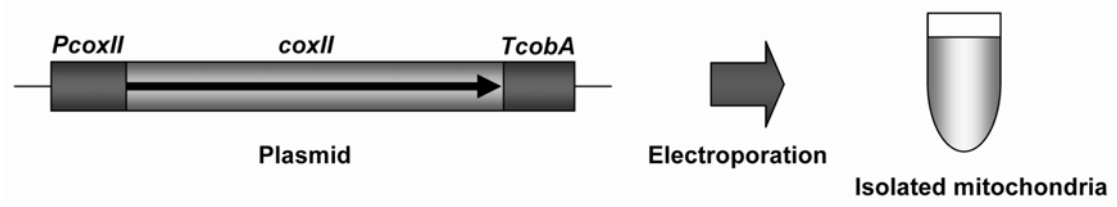
et al., 2005); TTcoxII C77 (where TT represents *Triticum timopheevi*) prefers a 5′ and 3′ U although a 5′ A and C are tolerated (Choury et al., 2004); and TTcoxII C259 is inhibited by a 5′ G and requires a 3′G (Farre et al., 2001). Overall templates with different nucleotides around the C target compared to wild-type are generally less edited *in vitro*. A 5′ neighboring G renders a C unable to be edited. Therefore the neighboring nucleotides are known to form part of the signal necessary for the targeting of most Cs.

A common location for core cis-elements

The location of essential sequence elements has been delimited from *in vivo* studies to a small region around the C target (Chaudhuri & Maliga, 1996). In all known cases, a greater length of sequence 5′ to the targeted nucleotide is necessary than 3′ sequence for editing of the substrate (Table 1.3). In most cases, at least 20 nucleotides of 5′ sequence and 5 nucleotides of 3′ sequence around the C are required for editing. The sequence of cis-elements in the -20/+5 regions vary between sites but the location and length of salient sequence was similar.

The location of cis-elements identified for mitochondrial sites are similar to ones identified around chloroplast targets (Table 1.4.). Regions from -10/-6, -15/-1, and -20/-5 are necessary for chloroplast editing *in vitro* of NTndhB C467, NTndhF C290, and NTrpoB C473 respectively (Hayes et al., 2006; Sasaki et al., 2006). For the mitochondrial site, TTcoxII C77, the sequence GCANNU from -11/- 6 and the dinucleotides AU at -14/-13 are crucial for *in organello* editing (Choury et al., 2004). Alternatively, for TTcoxII C259 the nucleotides AU at -12/-11 as well as GNAU at +1/+4 are necessary for editing in the mitochondria (Choury et al., 2004). Also in mitochondria, PSatp9 C20 sequences at positions -40/-36 and -25/-1 are crucial, but the essential core cis-element is contained in the -15/-6 region (Takenaka et al., 2004). Therefore, critical cis-elements for both organelles reside around 11 nucleotides

A



B

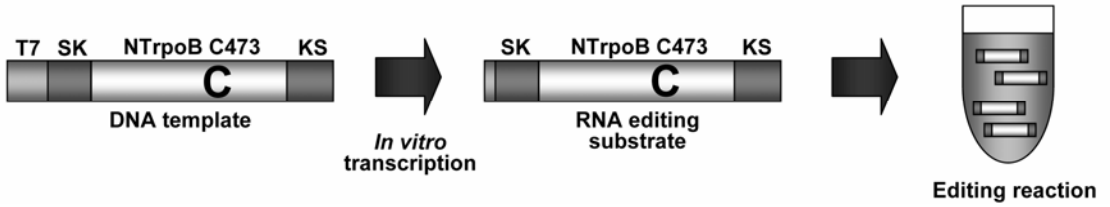


Figure 1.2. RNAs can be edited *in vitro* using either isolated mitochondria or organellar extracts. (A) A method for expression of exogenous editing sites in isolated mitochondria is summarized. Dark bars represent promoter and terminator regions of the expression plasmid and are identified by the above text. The light grey bar indicates the gene fragment containing editing sites. An arrow connotes the direction of transcription. The DNA plasmid is electroporated directly into isolated mitochondria in an aqueous suspension. (B) A summary of the methods used to study editing *in vitro*. Bars are labeled below as either representing a DNA template or RNA substrate. White bars indicate an organellar gene fragment with the editing site denoted above. Dark grey bars signify bacterial derived universal amplification sequences with the sequence classified above each region. Light grey bars symbolize viral derived promoters, T7 or T3, critical for *in vitro* transcription. RNA substrates are purified after transcription and added to editing reactions composed of crude chloroplast extracts under defined conditions. SK and KS sequences are dervived from pBluescript.

Table 1.3. Cis-acting regions defined for chloroplast editing sites

Plastid[a]	Gene[b]	Position[c]	5′ sequence[d]	3′ sequence[e]	References
Tobacco	NTrpoB	473	20	0	Hayes et al., 2006
Tobacco	NTpsbE	214	11	8	Hayes & Hanson, 2007b
Tobacco	NTndhB	746	21	2	Bock et al., 1996
Tobacco	NTndhB	737	12	11	Bock et al., 1996
Tobacco	NTndhG	50	10	0	Sasaki et al., 2006
Tobacco	NTndhG	347	15	0	Sasaki et al., 2006
Tobacco	NTpetB	611	20	11	Miyamoto et al., 2002
Tobacco	NTpsbL	2	16	5	Chaudhuri & Maliga, 1996
Tobacco	ATpsbE	214	13	15	Hayes & Hanson, 2007b
Arabidopsis	ATpsbE	214	31	15	Hegeman et al., 2005
Maize	NTrpoB	473	20	0	Hayes et al., 2006

[a]The species that was the source of the plastids or plastid extracts used for editing

[b]Gene in which the editing site was examined

[c]Position in nucleotides of the C target of editing from the A of the initiation codon

[d]Number of nucleotides 5′ of the editing site containing delimited cis-elements

[e]Number of nucleotides 3′ of the editing site containing delimited cis-elements

upstream of the C target. However, the identity and precise location of the crucial nucleotides varies between sites.

Cis-elements necessary for editing NTpsbE C214 have been extensively studied using chloroplast extracts and synthetic editing templates. The sequence GCCGUU from -11/-6 is critical for editing (Hayes & Hanson, 2007b). Any nucleotide change at any position within the cis-element reduces editing compared to the wild-type substrate. Nucleotide requirements for editing at each position in the important -5/-1 region are more relaxed. In this region only particular nucleotide changes at each of the positions reduces the ability of substrates to be edited. The

Table 1.4. Cis-acting regions defined for mitochondrial editing sites

Species	Gene	Position[a]	5′ sequence[b]	3′sequence[c]	References
Cauliflower	BOatp4	248	20	3	Verbitskiy et al., 2006
Cauliflower	BOatp4	251	23	2	Verbitskiy et al., 2006
Cauliflower	BOatp9	20	40	49	Neuwirt et al., 2005
Cauliflower	BOatp9	50	70	19	van der Merwe et al., 2006
Cauliflower	PSatp9	20	20	0	Notsu et al., 2002
Cauliflower	PSatp9	50	30	19	van der Merwe et al., 2006
Pea	PSatp9	20	20	0	Takenaka et al., 2004
Wheat	TTcoxII	259	15	4	Farre et al., 2001
Wheat	TTcoxII	77	14	6	Choury et al., 2004

[a]Position in nucleotides of the C target of editing from the A of the initiation codon

[b]Number of nucleotides 5′ of the editing site containing delimited cis-elements

[c]Number of nucleotides 3′ of the editing site containing delimited cis-elements

addition of competitor RNA molecules with nucleotide differences throughout the same region can reduce editing of the wild-type substrate *in vitro* (Miyamoto et al., 2002). Lack of competition can be explained if the -5/-1 region is not necessary for binding of a critical trans-factor. An alternative explanation for the lack of competition is that the region serves as a signal for a critical trans-factor that is not in limiting quantities, making it insensitive to the same level of competition as other cis-elements. Nucleotides at positions +2, +3 and +8 from NTpsbE C214 are also sensitive to some nucleotide changes in editing templates and probably form part of the editing signal.

Cis-elements necessary but not essential for editing are located 5′ of the core element for some C targets

Transgenic mRNAs containing longer native 5′ and 3′ sequences were edited in transplastomic plants to a greater extent than minimal substrates composed predominantly of the essential core element (Bock et al., 1996; Reed et al., 2001b).

Substrates containing the chloroplast site NTndhF C290 that have nucleotides differing from wild-type at positions -40/-36 are poorly edited *in vitro* (Sasaki et al., 2006). The -40/-36 region does not contain editing cis-elements in other editing sites, as substrates with alterations in this region are edited to the same level as substrates with wild-type sequences for NTpsbE C214, NTpetB611, and NTndhB C467 (Miyamoto et al., 2002; Sasaki et al., 2006). The simplest explanation for this observation is that cis-elements outside of the core region influence editing efficiency. Alternatively, differences in RNA secondary structure might also cause some substrates to be less edited.

In mitochondria, a short sequence necessary for editing *in vitro* was discovered and deemed an enhancer element upstream of PSatp9 C20 (Takenaka et al., 2004). Substrates with nucleotide differences compared to wild-type in the -40/-36 region were not edited *in vitro*. While wild-type RNA used as a competitor in the reaction *in vitro* resulted in decreased editing, competitor RNA molecules with nucleotide transversions at positions -40/-36 had little effect on editing of the wild-type template (Takenaka et al., 2004) (Takenaka et al., 2004).

As in chloroplasts, mitochondrial editing enhancer-like elements do not always reside in the -40/-36 region relative to the C target. The mitochondrial site TTcoxII C259 does not appear to have cis-elements in this region (Farre et al., 2001). Perhaps enhancer-like sequences vary in proximity to the editing site. A variable location of enhancer-like elements could explain why some templates that contain nucleotides -20/+5 around the editing site cannot be edited *in vitro* (Hayes & Hanson, 2007a). Editing substrates with more native sequence may sometimes be required in order to express all necessary cis-elements or to allow substrate RNAs to assume proper secondary structures. Perhaps essential editing signals are at a great distance from some C targets.

Editing sites with common sequences share critical editing trans-factors

Overexpression of an editing site in transplastomic plants reduces editing of the corresponding endogenous site (Chaudhuri et al., 1995). Presumably, critical editing factors are limiting and sequestered by the over-expressed transgenes, preventing interactions with native transcripts. Assay of the editing extent in two such transgenic plants for every known C target revealed small clusters of 3-5 sites, including the endogenous site, that are less edited compared to wild type plants (Chateigner-Boutin & Hanson, 2002). The abundance of transgenic transcript correlates with the level of editing reduction (Hayes et al., 2006). Plants that highly express an edited transgene have endogenous C targets that are less edited. Clustered editing sites are generally edited to similar levels in different tissues (Chateigner-Boutin & Hanson, 2003). Alignment of sequences around the edited C of all cluster members exposes small stretches of common consecutive nucleotides (Chateigner-Boutin & Hanson, 2002, 2003). Some of the common sequences are irregularly spaced relative to the C target, requiring introduction of gaps for identification. NTrpoB C473 substrates containing nucleotide differences in the common sequences are poorly edited, revealing their importance for editing (Hayes et al., 2006). Critical editing trans-factors are likely shared among cluster members and presumably recognize short sequences common around allied C targets.

Not every site is likely to share a specificity factor with another editing site. An example is NTpsbE C214, which though similar to sequences around two other C targets, appears to have a unique specificity factor. Editing of a substrate containing NTpsbE C214 is reduced due to self competition in reactions with addition of excess RNA templates with *psbE* sequences. Addition of the same amount of RNA templates containing the editing sites that have shared sequences does not reduce editing of a substrate with NTpsbE C214 (Hayes & Hanson, 2007b). Therefore the sequence

similarities are probably serendipitous. Furthermore, plants with mutations in *crr4* and *crr21* are each defective in editing a single C target (Kotera et al., 2005; Okuda et al., 2007) and editing of no other known C targets is affected. Therefore some sites are members of a genuine cluster while others are not, despite occasional inconsequential 3' sequence similarities.

A molecular ruler determines the cytosine that is modified

Altering the spacing relative of a cis-element to a C target can shift the editing site to a previously unedited cytosine (Hermann & Bock, 1999; Choury et al., 2004; Hayes & Hanson, 2007b). Therefore which C is selected for editing is influenced by a molecular ruler. In chloroplasts, the role of precise spacing of cis-elements relative to targets for editing sites NTpsbE C214 and NTndhB C746 have been examined (Hermann & Bock, 1999, Hayes & Hanson, 2007b). In both cases the editing target is located a certain number of nucleotides from a 5′ cis-element. Precise spacing of a cis-element proximal to a C target is also important for mitochondrial editing of templates containing PSatp9 C20, TTcoxII C259 and TTcoxII C77 (Choury et al., 2004; Neuwirt et al., 2005). Like the chloroplast sites studied, the distance from a C to the 5′ cis-element determines the editing target for substrates with PSatp9 C20 or TTcoxII C77. Conversely, for editing of TTcoxII C259, the relative distance of 3′ cis-element and the C target is crucial (Choury et al., 2004; Neuwirt et al., 2005). Due to the small numbers of sites studied, it is yet unclear whether the distance from a 3′ cis-element is critical for other C targets or specific to mitochondrial editing targets.

For NTpsbE C214 in chloroplasts and TTcoxII C259 and TTcoxII C77 in mitochondria, the nucleotide at -11 is important for efficient editing (Choury et al., 2004). The -1l nucleotide is within all known chloroplast and mitochondria editing cis-elements that have been identified to date. A-form or A'-form RNA forms a helical structure where one complete turn is composed of 11 and 12 nucleotides respectively

(Tanaka et al., 1999). Although the structure of the RNA editing template is not known, if the RNA is in A-form or A'-form the -11 nucleotide would be opposite the major groove and physically close to the C target of editing. RNA binding proteins Tat and Rev are known to interact with their target RNAs through the major groove (Puglisi et al., 1995; Battiste et al., 1996). Though the RNA structure of editable substrates has yet to be determined and the local sequence around editing targets is not highly structured, the native sequence could be double-stranded either due to base pairing with itself or other RNAs. Possibly the folding of the RNA functions as a "molecular ruler" by the precise placement of the catalytic domain and C target.

A single protein could be specificity factor and editing enzyme

Models constructed to describe the editing mechanism remain extremely crude due to the lack of identified proteins that are part of the editing complex (Figure 1.3). The protein(s) responsible for the catalysis of the editing reaction have not been elucidated. UV-crosslinking studies indicate a close interaction of putative editing factors with the C target and a 5′ cis element for NTpsbE C214 and NTpetB C611 (Miyamoto et al., 2004). The relative amount of putative trans-factor crosslinked to isotope-labeled NTpsbE C214 varies based on the identity of the 5′ neighboring nucleotide. The amount of trans-factor labeled correlates with the level of editing observed *in vitro* for templates with different 5′ adjacent nucleotides. Templates that are better edited *in vitro* also are better substrates for UV crosslinking studies. In contrast, UV crosslinking using templates with a labeled -10 nucleotide are unaffected by the identity of the -1 nucleotide. Therefore binding of the trans-factor is independent of downstream sequences, though the trans-factor and C target are in very close physical distance. This is consistent with other observations in which templates expressed *in vivo* that contain a U in place of the C target can act as effective competitors for editing of endogenous mRNAs (Reed et al., 2001a). Perhaps there

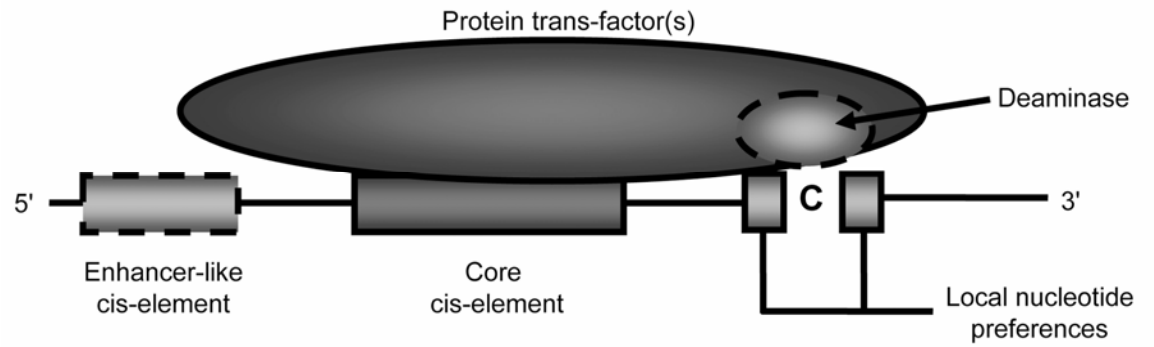


Figure 1.3. Signals necessary for organellar RNA editing are illustrated above. The dark bar framed by solid lines represents the core cis-element always found around the nucleotide 11 nt 5′ of the C target. The light bar with dashed lines signifies an enhancer-like element that is either only present in some sites or at variable locations between C targets. The light grey boxes with solid lines symbolize preferences of the editing mechanism for certain nucleotides adjacent to the C targeted. A dark oval represents the critical protein trans-factor(s) that can recognize the cis-elements. A smaller light grey oval denotes a domain of a protein trans-factor or individual protein responsible for the catalytic activity of the editing mechanism.

may be specificity factors that are capable not only of recognizing RNA substrates but also catalyzing the editing reaction when downstream nucleotides are favorable.

Genes *crr4* and *crr21* are necessary for the editing of two different C targets in chloroplasts (Kotera et al., 2005; Okuda et al., 2007). One of the corresponding gene products, CRR4, exhibits specific RNA binding activities to sequences around NTndhD C2 (Okuda et al., 2006). Unfortunately neither CRR4 nor CRR21 have demonstrated editing activity *in vitro*. They also lack a canonical deaminase domain. Either the *in vitro* conditions and/or the templates are not optimal for editing or other proteins may be involved. Because other RNA protein processing mechanisms involve large complexes, perhaps the editing activity also requires many components. Sharing of general editing components between sites would be economical, given the large number of editing sites in each organelle. For NTpsbL C2 and NTndhB C737 the depletion of cp31, a general RNA binding protein, renders the chloroplast extract incapable of *in vitro* editing (Hirose & Sugiura, 2001). Adding back cp31 restores the ability of the extract to edit NTpsbL C2 and NTndhB C737, but the effect is dosage dependent. Until editing can be reconstructed *in vitro* with defined components and conditions any models generated are somewhat speculative.

Conclusions

A wealth of knowledge is now known about the signals essential for RNA editing. Although the location of cis-elements share many generalities across all sites examined, the individual nucleotides necessary for editing must be diverse. Elements that are outside of the core region may not only vary in sequence but also in distance from the C target. Therefore the critical cis-elements cannot merely be estimated but must be experimentally examined for each C target of interest.

In addition to the diverse sequence signals for editing, individual C targets also have other site-specific features that impact editing. Editing *in vitro* has different

energy optima for different sites. For instance, 10 mM of ATP is optimal for NTpsbE C214 while 1 mM concentrations are preferred for NTrpoB C473 (Hayes & Hanson, 2007a). High temperature can also influence editing of individual sites (Karcher & Bock, 2002). The different features observed for distinct sites are probably due to variation in the site-specific trans-factors. The proteins responsible for editing are almost certainly as numerous and diverse as the editing signals for individual sites.

ACKNOWLEDGMENTS

This work was supported by NSF grant MCB 0344007 to M.R.H.

REFERENCES

Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. 1996. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science 273*:1547-1551.

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

Bentolila S, Elliott LE, Hanson MR 2007. Genetic Architecture of Mitochondrial Editing in *Arabidopsis thaliana*. *Genetics* in press

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

Bock R, Hermann M, Fuchs M. 1997. Identification of critical nucleotide positions for plastid RNA editing site recognition. *RNA 3*:1194-1200.

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

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

Chateigner-Boutin AL, Hanson MR. 2003. Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res 31*:2586-2594.

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

Chaudhuri S, Maliga P. 1996. Sequences directing C to U editing of the plastid *psbL* mRNA are located within a 22 nucleotide segment spanning the editing site.

EMBO J 15:5958-5964.

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

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

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

Cummings MP, Myers DS. 2004. Simple statistical models predict C-to-U edited sites in plant mitochondrial RNA. *BMC Bioinformatics 5*:132.

Du P, He T, Li Y. 2007. Prediction of C-to-U RNA editing sites in higher plant mitochondria using only nucleotide sequence features. *Biochem Biophys Res Commun 358*:336-341.

Farre JC, Araya A. 2001. Gene expression in isolated plant mitochondria: high fidelity of transcription, splicing and editing of a transgene product in electroporated organelles. *Nucleic Acids Res 29*:2484-2491.

Farre JC, Leon G, Jordana X, Araya A. 2001. cis Recognition elements in plant mitochondrion RNA editing. *Mol Cell Biol 21*:6731-6737.

Fey J, Tomita K, Bergdoll M, Marechal-Drouard L. 2000. Evolutionary and functional aspects of C-to-U editing at position 28 of tRNA(Cys)(GCA) in plant mitochondria. *RNA 6*:470-474.

Fey J, Weil JH, Tomita K, Cosset A, Dietrich A, Small I, Marechal-Drouard L. 2002. Role of editing in plant mitochondrial transfer RNAs. *Gene 286*:21-24.

Freyer R, Kiefer-Meyer MC, Kossel H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A 94*:6285-6290.

Giege P, Brennicke A. 1999. RNA editing in Arabidopsis mitochondria effects 441 C

to U changes in ORFs. *Proc Natl Acad Sci U S A 96*:15324-15329.

Groth-Malonek M, Wahrmund U, Polsakiewicz M, 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.

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

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

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

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

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

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

Hermann M, Bock R. 1999. Transfer of plastid RNA-editing activity to novel sites suggests a critical role for spacing in editing-site recognition. *Proc Natl Acad Sci U S A 96*:4856-4861.

Hiesel R, Combettes B, 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.

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

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

Hoch B, Maier RM, Appel K, Igloi GL, Kossel H. 1991. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature 353*:178-180.

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

Kahlau S, Aspinall S, Gray JC, Bock R. 2006. Sequence of the tomato chloroplast DNA and evolutionary comparison of Solanaceous plastid genomes. *J Mol Evol 63*:194-207.

Karcher D, Bock R. 2002. Temperature sensitivity of RNA editing and intron splicing reactions in the plastid *ndhB* transcript. *Curr Genet 41*:48-52.

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

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

Maier RM, Neckermann K, Igloi GL, Kossel H. 1995. Complete sequence of the

maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol 251*:614-628.

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

Marchfelder A, Brennicke A, Binder S. 1996. RNA editing is required for efficient excision of tRNA(Phe) from precursors in plant mitochondria. *J Biol Chem 271*:1898-1903.

Marechal-Drouard L, Cosset A, Remacle C, Ramamonjisoa D, Dietrich A. 1996a. A single editing event is a prerequisite for efficient processing of potato mitochondrial phenylalanine tRNA. *Mol Cell Biol 16*:3504-3510.

Marechal-Drouard L, Kumar R, Remacle C, Small I. 1996b. RNA editing of larch mitochondrial tRNA(His) precursors is a prerequisite for processing. *Nucleic Acids Res 24*:3229-3234.

Marechal-Drouard L, Ramamonjisoa D, Cosset A, Weil JH, 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, 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.

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

Mower JP. 2005. PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics 6*:96.

Mower JP, Palmer JD. 2006. Patterns of partial RNA editing in mitochondrial genes of

Beta vulgaris. *Mol Genet Genomics 276*:285-293.

Mulligan RM, Chang KL, Chou CC. 2007. Computational analysis of RNA editing sites in plant mitochondrial genomes reveals similar information content and a sporadic distribution of editing sites. *Mol Biol Evol*., in press.

Nakajima Y, Mulligan RM. 2005. Nucleotide specificity of the RNA editing reaction in pea chloroplasts. *J Plant Physiol 162*:1347-1354.

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

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

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

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

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

Puglisi JD, Chen L, Blanchard S, Frankel AD. 1995. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science 270*:1200-1203.

Reed ML, Hanson MR. 1997. A heterologous maize *rpoB* editing site is recognized by

transgenic tobacco chloroplasts. *Mol Cell Biol 17*:6948-6952.

Reed ML, Lyi SM, Hanson MR. 2001a. Edited transcripts compete with unedited mRNAs for trans-acting editing factors in higher plant chloroplasts. *Gene 272*:165-171.

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

Sasaki T, Yukawa Y, Miyamoto T, Obokata J, Sugiura M. 2003. Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. *Mol Biol Evol 20*:1028-1035.

Sasaki T, Yukawa Y, Wakasugi T, Yamada K, 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.

Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM. 2002. The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol 19*:1602-1612.

Schuster W, Hiesel R, Wissinger B, Brennicke A. 1990. RNA editing in the cytochrome b locus of the higher plant *Oenothera berteriana* includes a U-to-C transition. *Mol Cell Biol 10*:2428-2431.

Staudinger M, Kempken F. 2003. Electroporation of isolated higher-plant mitochondria: transcripts of an introduced *cox2* gene, but not an *atp6* gene, are edited *in organello*. *Mol Genet Genomics 269*:553-561.

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. 1999. Plant mitochondrial RNA editing. *J Mol Evol 48*:303-312.

Sugita M, Miyata Y, Maruyama K, Sugiura C, Arikawa T, 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.

Takenaka M, Brennicke A. 2003. *In vitro* RNA editing in pea mitochondria requires NTP or dNTP, suggesting involvement of an RNA helicase. *J Biol Chem 278*:47526-47533.

Takenaka M, Neuwirt J, Brennicke A. 2004. Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res 32*:4137-4144.

Tanaka Y, Fujii S, Hiroaki H, Sakata T, Tanaka T, Uesugi S, Tomita K, Kyogoku Y. 1999. A'-form RNA double helix in the single crystal structure of r(UGAGCUUCGGCUC). *Nucleic Acids Res 27*:949-955.

Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM. 2005. Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J 43*:708-715.

Tillich M, Lehwark P, Morton BR, Maier UG. 2006. The evolution of chloroplast RNA editing. *Mol Biol Evol 23*:1912-1921.

Tillich M, Schmitz-Linneweber, C, Herrmann, RG, Maier, RM. 2001. The plastid chromosome of maize (*Zea mays*): Update of the complete sequence and transcript editing sites. *Maize Genetics Cooperation Newsletter 75*:42-44.

Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T, Yamaguchi K. 2007. The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. *J Plant Res 120*:281-290.

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

Verbitskiy D, Takenaka M, Neuwirt J, van der Merwe JA, Brennicke A. 2006. Partially edited RNAs are intermediates of RNA editing in plant mitochondria. *Plant J 47*:408-416.

Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kossel H, Sugiura M. 1996. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc Natl Acad Sci U S A 93*:8766-8770.

Wintz H, Hanson MR. 1991. A termination codon is created by RNA editing in the petunia *atp9* transcript. *Curr Genet 19*:61-64.

Wolf PG, Rowe CA, 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.

Yoshinaga K, Iinuma H, Masuzawa T, 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.

Zeng WH, Liao SC, Chang CC. 2007. Identification of RNA editing sites in chloroplast transcripts of *Phalaenopsis aphrodite* and comparative analysis with those of other seed plants. *Plant Cell Physiol 48*:362-368.

CHAPTER 2

∗SEQUENCE ELEMENTS CRITICAL FOR EFFICIENT RNA EDITING OF A TOBACCO CHLOROPLAST TRANSCRIPT *IN VIVO* AND *IN VITRO*

ABSTRACT

34 nucleotides in tobacco chloroplast transcripts are known to be efficiently edited to U. No common consensus region is present around all editing sites; however, sites can be grouped in clusters that share common sequence elements. Transgene transcripts carrying either the wild-type -31/+22 or -31/+60 sequence near NTrpoB C473, an editing site within tobacco *rpoB* transcripts, or three different mutated sequences, were all highly edited *in vivo.* Endogenous transcripts of *rpoB, psbL*, and *rps14*, all of which contain common sequences S1, S2, and S3 5' to NTrpoB C473, NTpsbL C2, and NTrps14 C80, were less edited in transgenic plants that over-express transcripts from NTrpoB C473 transgenes. The extent of reduction of endogenous editing differed between transgenic lines expressing mutated -31/+22 regions, depending on the abundance of the transgene transcripts. The -20/-5 sequence was discovered to contain critical 5' sequence elements. Synthetic RNA templates with alterations in S1, S2, and S3 within the 5' sequence element were less efficiently edited *in vitro* than wild-type templates, when assayed with either tobacco or maize chloroplast extracts. The tobacco chloroplast extract supports both RNA editing and processing of 3' transcript termini. We conclude that within the 5' sequence element, sequences common to editing sites in the transcripts of *rpoB*, *psbL* and *rps14* are critical for efficient NTrpoB C473 editing.

∗ Hayes ML, Reed ML, Hegeman CE, Hanson MR (2006) Nucleic Acids Res 34: 3742-54, ©2006 by Oxford University Press. Reed and Hegeman produced the transformation vectors and the transgenic plants used here.

INTRODUCTION

RNA editing, a form of RNA processing, occurs in both nuclear and organelle transcripts of diverse organisms. In vascular plants, around 30 C targets of editing typically exist in chloroplast transcripts, while over 400 such targets have been observed in plant mitochondria (Giege & Brennicke, 1999; Tsudzuki et al., 2001; Notsu et al., 2002; Handa, 2003). Tobacco chloroplasts are a particularly good model system for editing because deliberate alteration in editing substrates can be assayed i*n vivo* in chloroplast transgenic plants or *in vitro* with chloroplast extracts (Chaudhuri et al., 1995; Reed & Hanson, 1997; Hirose & Sugiura, 2001). Tobacco chloroplasts have 34 editing sites within the 155,939 bp organelle genome, and all are modifications from cytidine to uridine. Of the 34 identified editing sites, 32 sites are known to be efficiently edited, with 70-100% of transcripts modified from C-to-U at each edited position (Chateigner-Boutin & Hanson, 2003). In vascular plant chloroplasts, identified editing sites are almost exclusively within coding regions and occur most frequently at the second position of a codon (Tsudzuki et al., 2001). Editing of all but one of the C targets identified in tobacco results in a change of encoded amino acid (Tsudzuki et al., 2001; Chateigner-Boutin & Hanson, 2003). Edited codons commonly encode amino acids that are conserved among orthologous proteins of other plants (Maier et al., 1996; Peeters & Hanson, 2002; Schmitz-Linneweber et al., 2002). Defects in editing of some transcripts result in plants with severe phenotypes, producing dysfunctional proteins (Zito et al., 1997; Sasaki et al., 2001; Schmitz-Linneweber et al., 2005). Editing in plant organelles is likely a mechanism for the correction of genomic T-to-C mutations rather than for creation of protein diversity (Covello & Gray, 1993; Smith et al., 1997). The low number of C-to-U modifications, accompanied by the high extent of editing of C targets, suggests the existence of a highly efficient and specific editing mechanism within chloroplasts.

In vivo and *in vitro* editing studies have focused on the sequence elements responsible for directing editing in chloroplasts as well as in mitochondria. Both organelles edit C-to-U and may share similar mechanisms for editing (Takenaka & Brennicke, 2003). Regions critical for editing are primarily located in nearby regions 5′ of the editing sites, and *in vivo* studies have identified a number of editable substrates that carry only 20-40 nt 5′ and 10-20 nt 3′ to the C editing target.

The editing site NTrpoB C473 is within the tobacco *rpoB* transcript; the C at position 473 from the A of the initiation codon is edited. This editing site has been previously referred to as rpoB-2, but because of the number of additional species in which chloroplast editing has been characterized, a previous nomenclature system (Tsudzuki et al., 2001) for chloroplast C editing targets has become unwieldy. We propose here to name editing sites by initials of genus and species, gene name, then nucleotides from the A of the closest gene's initiation codon. C473 in tobacco *rpoB* is at the second position of the codon, altering the encoded amino acid from serine to leucine.

The cis-requirements for editing for NTrpoB C473 were previously examined *in vivo* by expressing transgenes carrying a small portion of the *rpoB* gene surrounding the C editing target (Reed & Hanson, 1997; Reed et al., 2001b; Chateigner-Boutin & Hanson, 2002; Hegeman et al., 2005a). A 27 nt sequence flanking the editing site NTrpoB C473 is sufficient for editing; however, a more highly edited template contained 92 nt around the editing site. The lower amount of editing observed in the smaller 27 nt template is most likely due to the loss of important nucleotides in the reduced 3′ and 5′ regions around the edited C, compared to the larger substrate. Therefore, we have created transgenes and transcripts with 54 nt around the edited C to better define the important cis-acting region *in vivo* and *in vitro*.

Although no consensus sequence is common to all editing sites, groups of sites

with common sequences can be gathered into clusters of sites that may also share sequence-dependent specificity factors. Over-expression of sequences flanking NTrpoB C473 or NTndhF C290 in tobacco chloroplasts results in a reduction in editing of a small group of endogenous editing sites which contain some short common sequence elements (Chateigner-Boutin & Hanson, 2002). All of the known editing sites within tobacco can be grouped into clusters based on short common sequences. Some of these clusters exhibit similar changes in efficiency of editing depending on tissue type (Chateigner-Boutin & Hanson, 2003). Upon over-expression of a template containing a region surrounding NTrpoB C473 in tobacco chloroplast transgenic plants, endogenous *rpoB, psbL*, and *rps14* transcripts exhibit less editing at NTrpoB C473, NTpsbL C2, and Ntrps14 C80 than in wild-type plants. These three editing sites carry three common sequence elements of 2-3 nt that we have termed S1, S2, and S3. We have constructed *in vitro* templates to examine the importance of S1, S2, and S3 in editing of NtrpoB C473 in both tobacco and maize chloroplast extracts. These studies indicate that all three elements are important for efficient editing of C473 in *rpoB* transcripts. We have also explored the effect of 5' and 3' flanking sequences on editing efficiency *in vivo* and *in vitro*. Furthermore, we report the production of a maize chloroplast extract that is capable of editing the tobacco C473 editing site.

MATERIALS AND METHODS

In vitro processing assay

0.1 fmol of 5′ or randomly labeled RNA was incubated in tobacco chloroplast extract for 2h under the *in vitro* editing assay conditions previously described (Hegeman et al., 2005b). RNA was purified by phenol:chloroform extraction and precipitated. Resuspended RNAs were separated on 6% poly-acrylamide gels.

Construction of plastid transformation vectors

The editing site and adjacent bases were amplified by PCR from tobacco leaf genomic DNA and specific sequence alterations were generated by mutagenic PCR (Table 2.1). Five different editing templates were constructed. Editing templates were flanked by *Nco*I and *Xba*I restriction sites. Transformation constructs were then created by the integration of the editing templates into the vector pLAA24A (Zoubenko et al., 1994). Restriction enzyme digestion at sites NcoI and XbaI were used to remove the *uidA* coding sequence from pLAA24A and insert the NTrpoB C473 gene fragment, creating constructs for bombardment.

Transformation and tissue culture

Standard methods were used to create chloroplast transgenic plants (Chaudhuri et al., 1995; Reed & Hanson, 1997; Reed et al., 2001a; Reed et al., 2001b; Hegeman et al., 2005a). Young tobacco seedlings were bombarded with plasmid-coated tungsten particles. After bombardment plants were selected on regeneration media containing 500 mg/L spectinomycin (Svab & Maliga, 1993). A number of initial transformants were created for each construct, and one line for each construct was maintained for further analysis. Plants were assayed for homoplasmicity after selection through Southern blotting. All plants with integrated constructs remained homoplasmic throughout this investigation, except for R54. Despite continued rounds of selection, line R54 never achieved homoplasmicity and was analyzed as a heteroplasmic plant.

Expressed transcripts from all constructs include an ATG codon because of the use of the *Nco*I restriction site. Translation beginning at this AUG would be out of frame of *rpoB* and would proceed just 16 amino acids before reaching a stop codon.

Table 2.1. Oligonucleotides (Integrated DNA Technologies, Coralville, IA) used in experiments reported here.

Name	Sequence 5′-3′	Purpose
500f	GATCCCCATGGGGCACCATAATA TCAGATTGGGGAGG	Transgene construction
500r	CCGTCTAGATTTTCTATCAATTTC TAATTCTGATCTTC	Transgene construction
501f	GATCCCCATGGGGCACCATAATTT CAGATTGGGGAGG	Transgene construction
502f	GATCCCCATGGGGCACCATAATTT CAGATTGGGGAGG	Transgene construction
502r	CCGTCTAGATTTTCTTTCAATTTCT AATTCTGATCTTC	Transgene construction
505f	GATCCCCATGGGGCACCATAATA TCAGATTGGGGA	Transgene construction
505r	GATCCCCATGGGGCACCCTGATAT CAGATTGGGGA	Transgene construction
506f	CCGTCTAGAATTTTTTGTTTCCTA CTTACACGAGCCCA	Transgene construction
500sreverse_Lg	TGTCCATTTTTCGGGGTCTCAAAG GGGCGTGGAAA	S1 nuclease mapping
T7_5'_500s	TAATACGACTCACTATAGGGGCG AACTCCGGGCGAATA	Transcript production
PC1.1	TCTTGAACAACTTGGAGCCGGGC C	Southern probe
PCα1.2	GAGGATAGCAAGTTCCAAATTCT GTCTCGG	Southern probe
PPrrn2	AATACGAAGCGCTTGGATACAGT TGTAGGGA	PCR transgenic transcript
Trps16sh	TCCTTAATTTATTTCCTTAATTGA ATTTCTCTAGA	PCR 3′ end II
Trps16lg	AATTCAATGGAAGCAATGATAAA AAAATACAAATA	PCR 3′ end I
FRpoB2	ACTCCAGGTTCCTCGGGGTAAA	PCR endogenous *rpoB*
RRpoB2	TTGCGGAGTAAATGGGCTTCTAA	PCR endogenous *rpoB*
RpoB-2(C)	GGCACCATAATATCAGATTGGGG AGGAAG	PPE NTrpoB C473
FpsbL	TACCGTCTTTTTTTTGGGATC	PCR endogenous *psbL*
RpsbL	ATTTTGTTCGTTCGGGTTTGA	PCR endogenous *psbL*
PsbL(G)	AACATTTTGTTCGTTCGGGTTTGA TTGTGT	PPE NTpsbL C2

Table 2.1. (Continued)

FRps14	CAGAGGGAGAAGAAGAGGC	PCR endogenous *rps14*
RRps14	GCTCCTGGCAACAAACAT	PCR endogenous *rps14*
Rps14-1(C)	GGAACAGAAATATCATTCGATTCGTCG ATCC	PPE NTrps14 C80
Rps14-2(A)	CGATGAAGGCGTGTAGGTGCACTATTCC	PPE NTrps14 C149
RpoB54Rev KS	TATCTTTTCTATCAATTTCTAATTCTGAT CTTCCTCCCCA	R54KS *in vitro* substrate
R54R2For	CTAGAACTAGTGGATCGGCACCATAATT AGAGATTGGGGA	R2 and R14 *in vitro* substrate
R54R7For	CTAGAACTAGTGGATCGGCACCATAAT ATCAGTAAGGGGA	R7 *in vitro* substrate
R54R5Rev	TATCTTTTCTATCAATTTCTAATTCTGAA GTTCCTCCCCA	R9 *in vitro* substrate
R54R7Rev	TATCTTTTCTATCAATTTCTAATTCTGAT CTAGCTCCCCA	R7 and R15 *in vitro* substrate
R54R6For	ATCGGCACCATAATATCAGTAAGGGGA GCTAGATCAGAAT	R12 *in vitro* substrate
R54R7For	ACTAGTGGATCGGCACCATAATTAGAG TAAGGGGAGGAAG	R13 *in vitro* substrate
R54R9For	ACTAGTGGATCGGCACCATAATTAGAG TAAGGGGAGCTAG	R15 *in vitro* substrate
R54R1For	CTAGAACTAGTGGATCGGCACCATAATT TCAGATTGGGGA	R1 *in vitro* substrate
T7SK	TAATACGACTCACTATAGGGCGCTCTAG AACTAGTGGATC	*in vitro* substrate construction
R54R3For	CGCTCTAGAACTAGTGGATCGGCACCAT AATATCTCTTTGGGGAGGAAGATCAGA ATTAG	R3 *in vitro* substrate
R54R3Rev	TCGAGGTCGACGGTATCTTTTCTATCAA TTTCTAATTCTGATCTTCCTCCC	R3, R16, R17, R18 *in vitro* substrate
R54R5For	CGCTCTAGAACTAGTGGATCGGCACCAT AATATCAGATTCCCGAGGAAGATCAGA	R5 *in vitro* substrate
R54R5Rev	TCGAGGTCGACGGTATCTTTTCTATCAA TTTCTAATTCTGATCTTCCTCGGGAAT	R5 *in vitro* substrate
R54R6For	CGCTCTAGAACTAGTGGATCGGCACCAT AATATCAGATTGGGCTCGAAGATCAGA	R6 *in vitro* substrate
R54R6Rev	TCGAGGTCGACGGTATCTTTTCTATCAA TTTCTAATTCTGATCTTCGAGCCCAAT	R6 *in vitro* substrate

Table 2.1. (Continued)

R54R8For	CGCTCTAGAACTAGTGGATCGGCACCATAAT ATCAGATTGGGGAGGATCTTCAGA	R8 *in vitro* substrate
R54R8Rev	TCGAGGTCGACGGTATCTTTTCTATCAATTTC TAATTCTGAAGATCCTCCCCAAT	R8 *in vitro* substrate
R54R10For	CGCTCTAGAACTAGTGGATCGGCACCATAAT ATCAGATTGGGGAGGAAGATCTCT	R10 *in vitro* substrate
R54R10Rev	TCGAGGTCGACGGTATCTTTTCTATCAATTTC TAATAGAGATCTTCCTCCCCAAT	R10 *in vitro* substrate
R54R11For	CGCTCTAGAACTAGTGGATCGGCACCATAAT ATCAGATTGGGGAGGAAGATCAGAT	R11 *in vitro* substrate
R54R11Rev	TCGAGGTCGACGGTATCTTTTCTATCAATTTC TTTATCTGATCTTCCTCCCCAAT	R11 *in vitro* substrate
R54R16For	TAGAACTAGTGGATCGGCACCATAATGCTAG ATTGGGGAGGAAGATCAGA	R16 *in vitro* substrate
R54R17For	TAGAACTAGTGGATCGGCACCATAATATCGA GTTGGGGAGGAAGATCAGA	R17 *in vitro* substrate
R54R18For	TAGAACTAGTGGATCGGCACCATAATATCAG GCCGGGGAGGAAGATCAGA	R18 *in vitro* substrate
R54R19For	TAGAACTAGTGGATCGGCACCATAATATCAG ATTAAAGAGGAAGATCAGAAT	R19 *in vitro* substrate
R54R19Rev	GAGGTCGACGGTATCTTTTCTATCAATTTCTA ATTCTGATCTTCCTC	R19 *in vitro* substrate
R54R20For	TAGAACTAGTGGATCGGCACCATAATATCAG ATTGGGAGAGAAGATCAGAATTAG	R20 *in vitro* substrate
R54R20Rev	TCGAGGTCGACGGTATCTTTTCTATCAATTTC TAATTCTGATCTTC	R20 *in vitro* substrate
R54R21For	TAGAACTAGTGGATCGGCACCATAATATCAG ATTGGGGAGAGAGATCAGAATTAGAA	R21 *in vitro* substrate
R54R21Rev	TCGAGGTCGACGGTATCTTTTCTATCAATTTC TAATTCTGATCTCTCTCC	R21 *in vitro* substrate

DNA blot analysis

Total DNA was isolated from transgenic and wild-type leaves from shoots grown on RMOP media. DNA (1 μg) was digested using BamHI, electrphoresed on 1% agarose, and blotted onto positively charged nylon (Amersham) using a turboblotter (Schleicher and Schuell). Oligonucleotides (PC1.1 and PCα1.2) were used

to amplify a 350 nt genomic probe from wild-type genomic DNA overlapping the insertion site. The probe was random labeled using the DECAprime II kit and α^{32}P-(dCTP) and hybridized to the DNA blot for 24h at 65ºC.

S1 nuclease assay

A DNA probe was constructed by PCR using oligonucleotides T7_5'_500s and 500sreverse_Lg. The PCR product was designed to hybrizidize with the transgenic transcript and overlap the 3′ end terminator sequence from tobacco *rps*16. The DNA probe was restricted with *Nco*I (Invitrogen) and the 3′ end of the antisense strand was labeled using a Klenow fill reaction. Labeled probe and 1, 10, or 25 µg of RNA were hybridized overnight. 500U/ml of S1 nuclease (Promega) was added to the nuclease reaction for 1hr at 37 ^{o}C and the products were electrophesed on a 5% polyacrylamide gel.

Immunobloting

Immunoblotting was performed as in Hegeman and Hanson (Hegeman et al., 2005a). Total leaf protein was obtained from shoots grown on RMOP medium using homogenization Buffer containing 50mM Tris-HCl, 1mM EDTA, 1X Protease Inhibitor cocktail (Complete, Roche), and 0.1% (v/v) Triton X-100. Protein was quantified using Bio-Rad Protein Assay kit and a BSA standard curve. 20 µg of total proteins was boiled in SDS-PAGE, electrophoretically separated onto 10% acrylamide gels, and transferred to nitrocellulose membranes (Pierce). Membranes were blocked overnight in blocking buffer (5% dried milk powder, 1%TBS-T) after which the primary antibody was added to 1:500 dilution from crude serum. The washed blots were incubated in secondary antibody (horseradish peroxidase conjugated goat anti-rabbit; Amersham) diluted to 1:50,000 and proteins were visualized using SuperSignal West Dura Extended Duration Substrate to manufacturer specifications (Pierce Biotechnology).

Editing analysis

Total RNA was isolated using Trizol (Invitrogen) for transgenic plants and wild-type leaves. Contaminating DNA was removed using Turbo DNAse (Ambion) and cDNA was synthesized by reverse transcription (Omniscript, Qiagen) using degenerate hexamers. Transgenic transcripts were amplified using PPrrn2 and either Trps16lg for transcripts with 3′ end I or Trps16sh for amplification of transcripts with both 3′ ends. Amplified transcripts were then assayed for editing extent using the poisoned primer extension assay as previously described (Peeters & Hanson, 2002; Hegeman et al., 2005a; Hegeman et al., 2005b).

S*ubstrates for analysis *in vitro

For substrates equivalent to the transgenic transcripts, DNA substrates were produced from PCR amplifications using primers T7_5'_500s and either Trps16sh or Trps16lg. For substrates with sequence alterations in the -31 to +22 region around NTrpoB C473, the respective mutagenic PCR primers were used. The bacterial sequences SK and KS were added to flank the region of *rpoB* to prevent amplification from endogenous nucleic acids. A T7 sequence was added to the 5' end of the substrate also by PCR amplification. RNA substrates were then produced using the PCR products as template by *in vitro* transcription using the T7 MEGAshortscript kit (Ambion). RNAs were then purified using the RNA clean-up kit-5 (Zymo Research).

Editing reactions *in vitro*

The editing reactions were performed as previously described (Hegeman et al., 2005b). 0.1fmol of RNA was added to 80 µg of tobacco, competent, chloroplast extract (Hegeman et al., 2005b) in assay conditions. Maize extracts were prepared from 7 to 10 day old maize plants grown in the same conditions as tobacco (Hegeman et al., 2005b). Leaves were homogenized and plastids isolated using a Percoll (Amersham Biosciences) gradient. Intact chloroplasts were lyzed using Triton X-100,

and dialyzed in Dialysis Buffer (Hegeman et al., 2005b). Conditions for the maize *in vitro* assay were identical to the conditions used for tobacco except only 20 μg of chloroplast extract was used. Editing of RNA substrates was analyzed using the poisoned primer extension assay (Peeters & Hanson, 2002; Hegeman et al., 2005a; Hegeman et al., 2005b).

Semi-quantitative RT-PCR

cDNA was synthesized for both short and long transcripts by reverse transcription (Sensiscript, Qiagen) from 50ng of total RNA isolated from transplastomic plants and the Trps16sh primer. PCR amplification of the cDNA templates utilized the primers Trps16sh and PPrrn2. RT-PCR products from different rounds of PCR were then separated on 3% agarose gels. After 22 cycles differing quantities of RT-PCR products could be distinguished corresponding to the varying amounts of initial transcript. After 40 cycles all bands were of similar intensity and reactions without reverse transcriptase showed no specific amplification.

RESULTS

Production of transgenic plants for further analysis of an *rpoB* editing site *in vivo*

Previously, transgenic plants were analyzed that contained 27 nt surrounding the NTrpoB C473 editing site. Approximately 25% of the transcripts carrying the wild-type tobacco sequence or a sequence altered at either -7 or +2 from the C target were edited, on average (Reed et al., 2001b). However, transcripts carrying a T rather than the wild-type A at -20 were poorly edited; only 3 of 221 individually analyzed transcripts exhibited editing. Furthermore, a homologous sequence from black pine (which contains T rather than A at -20) was also not edited. Because a homologous 92 nt region from maize *rpoB* had previously been observed to be highly edited (~ 50%), we produced transgenic plants carrying a sequence larger than 27 nt, but smaller than

92 nt to further define sequence requirements for editing. A 54 nt region (-31 to +22) from tobacco was inserted into vector pLAA24A, as well as a 92 nt tobacco sequence (-31/+60) for use as a control analogous to the maize region previously expressed in tobacco (Reed et al., 2001b). Several mutated versions were also produced. R92m1 carries two nucleotides found in the black pine sequence at -23 and -25. R54m1 contains the -20 T change found inimical to editing in the 27 nt transgene. Because the region -16 to -21 was observed to be complementary to the +12 to +17 sequence, the +12 T was changed to A as a potential compensatory mutation for the -20A to T change (Figure 2.1).

The transformation vector was introduced to young tobacco leaves by biolistic delivery, and plastid genome composition was determined in regenerating plants by Southern blotting (Figure 2.2). A single R54-containing shoot did not reach homoplasmy despite lengthy efforts and this line was therefore examined as a heteroplasmic plant. Single transformation events of the 4 other lines achieved homoplasmy after repeated selection on antibiotic medium (Figure 2.2).

Regenerating shoots were treated with auxin and grown on rooting medium. Only R54m2 could not be induced to root, all other plants were taken to soil (Figure 2.3). R54 as a heteroplasmic plant exhibited sensitivity to light and leaves became bleached in exposure to levels of light equal to the normal growing conditions for the other plants. R54 plants did not survive in soil even in low light conditions. R54m1 plants thrived on soil but had severe leaf abnormalities. Leaves appeared malformed, with defects in venation, long trichomes, and short internodes. Plants were severely stunted and did not flower. R92 plants were stunted and flowers were male sterile. R92m1 plants grew normally with no other defects observed from the transgene. Progeny resulting from an outcross using male pollen from wild-type plants and chloroplast transformant flowers with R92 were male sterile, but did not display any

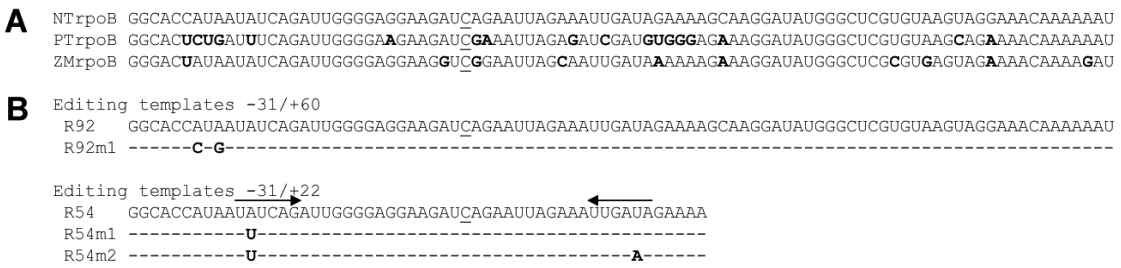


Figure 2.1. RNA substrates were created with regions of sequence around the editing site NTrpoB C473 to study template requirements of editing *in vivo*. (**A**) The sequence 31 nt 5′ to 60 nt 3′ around the editing site at position 473 from the initiation codon of *Nicotiana tobacum,* NTrpoB, was aligned with *Pinus thunbergii* PTrpoB and *Zea mays* ZMrpoB sequences. Bold characters indicate nucleotides different from NTrpoB sequence. The position of the edited nucleotide is indicated by an underlined C. (**B**) The sequences of created templates are represented with differences at nucleotides that are divergent between NTrpoB and PTrpoB. Bold characters indicate nucleotides that differ from the wild-type NTrpoB sequence. Dashes indicate positions where wild-type sequence is present in the template. Complementary sequences present around the editing site are indicated by arrows.

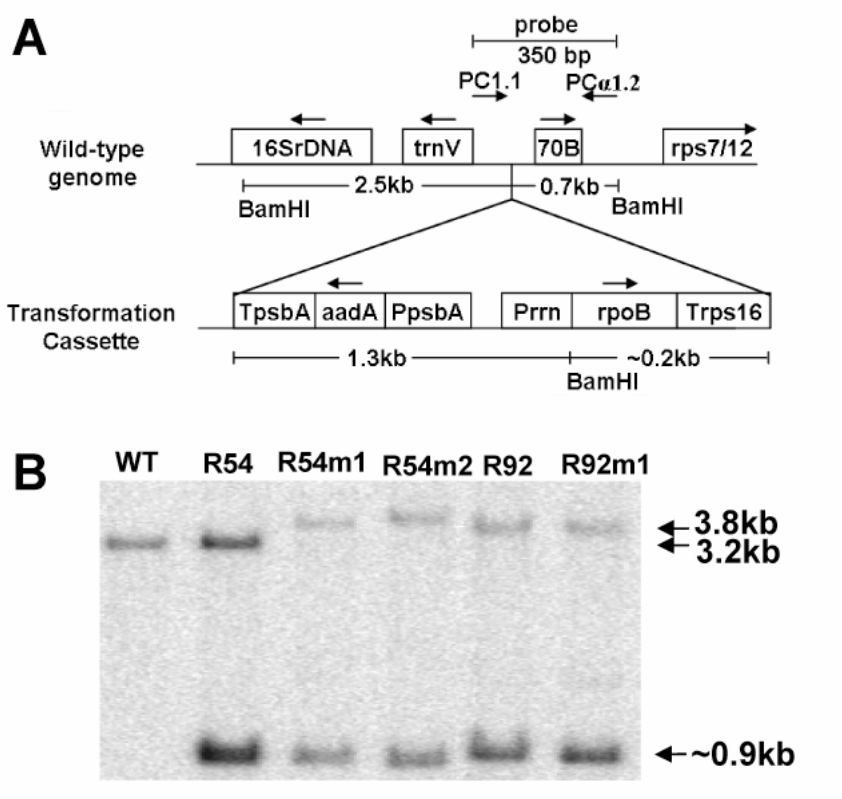


Figure 2.2. Creation of transplastomic plants. (**A**) Diagram representing the wild-type tobacco chloroplast genome and insertion of the transformation cassette. The probe used in the Southern blot is indicated at the top right of the diagram, and spans the wild-type insertion site. (**B**) Southern blots containing BamHI digested DNA from transplastomic plant leaves were probed with a labeled 350 bp PCR product. Bands at 3.2 kb are due to the untransformed genome. 3.8 kb and 0.9 kb bands indicate an integrated transgene.

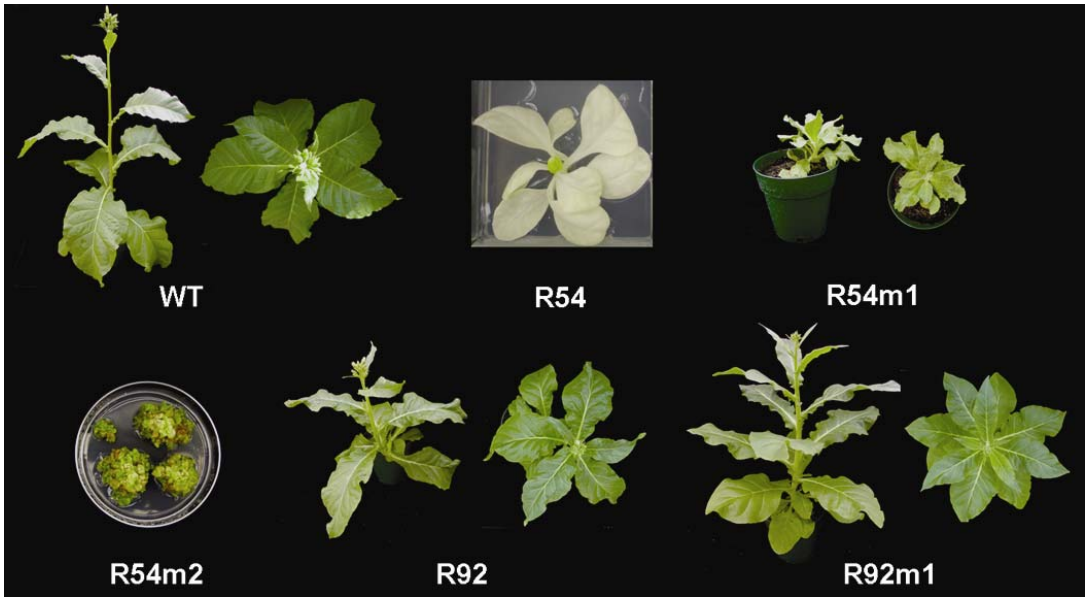


Figure 2.3. Phenotypes of transplastomic plant lines. R54 and R54m2 could not grow on soil and were photographed at their last developmental stage in which they would thrive. R54 displayed a bleached phenotype when grown on sucrose containing media under low light intensity. Transplastomic R54m2 shoots grown on regeneration media are shown. All other lines were grown in soil and are shown from two different perspectives. Viewed from the side, left, it is clear many lines display stunted phenotypes characterized by short internodes. From the top view, right, abnormal leaf morphology is evident in plant line R54m1.

other significant phenotypes compared to wild-type tobacco plants (data not shown), suggesting that mutant phenotypes have resulted from the bombardment and tissue culture for plant regeneration.

Total protein was isolated from plants and RpoB protein levels determined through immunoblotting. There was no correlation between RpoB protein levels and severity of phenotype, and only R92m1 appeared to have reduced amounts of RpoB protein compared to total protein (Figure 2.4). The transplastomic progeny of R92 and R92m1 plants did not display reduced RpoB protein levels (Figure 2.4).

Assay of editing extent of transgene transcripts

Editing of transgene transcripts in leaves of the 5 transgenic lines were assessed by poisoned primer extension. R92m1 transcripts, which carry altered -23 and -25 nucleotides were 65% edited compared to 67% for the R92 construct, the template with the same region of wild-type sequence. The -23 and -25 nucleotides appear to be of little importance for the editing process. Surprisingly, even the lines carrying the -20A to T change that prevented editing in transgene transcripts of shorter length (-20 to +6) (Reed et al., 2001b), exhibited editing over 60% (Figure 2.5). In fact, editing of the two lines containing -20A to T changes exhibited somewhat higher editing than the R54 line containing wild-type sequence. Evidently either the addition of 11 nucleotides at the 5' end or 15 nt 3' to the edited C has affected the editing efficiency of the transgene transcripts in comparison to the transgenes carrying only 27 nt of chloroplast sequence.

Further characterization of transgene transcripts

In order to understand how the increased size of the transcripts carrying 54 nt rather than 27 nt of *rpoB* sequence has enhanced editing efficiency, it was necessary to determine the exact RNA species produced *in vivo*. The 5′ end of the transcript is determined by the *rrn16* promoter, which initiates transcription downstream of

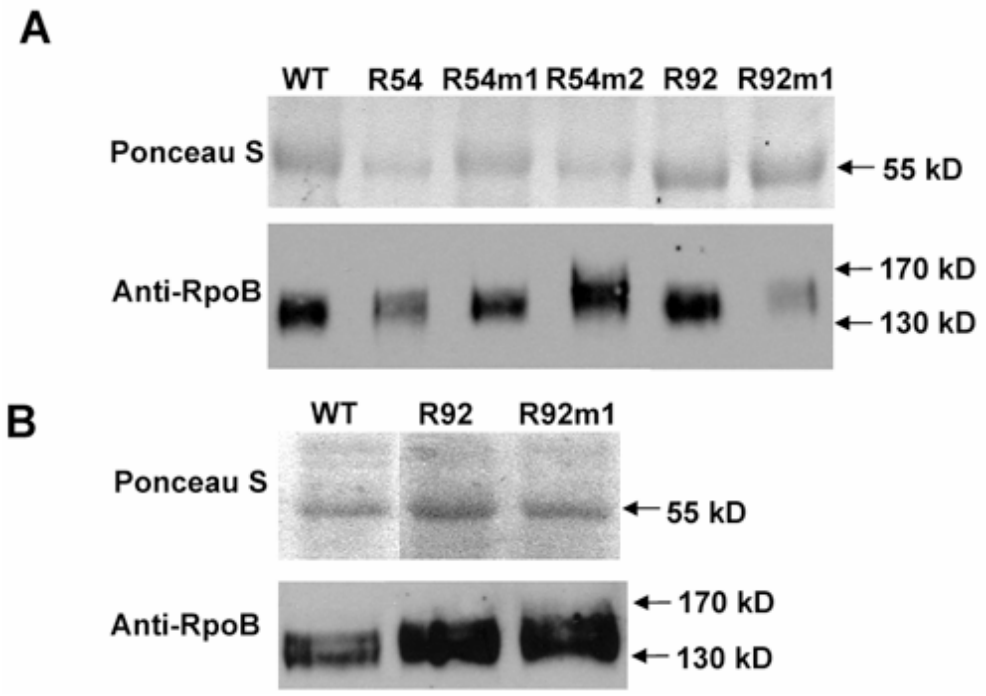


Figure 2.4. RpoB protein levels in transplastomic plants. To verify equal SDS-PAGE sample loading, blots were stained with Ponceau S to compare levels of the abundant Rubisco large subunit. The 120 kD RpoB protein was detected using primary antisera raised against a RpoB peptide. (A) Total and RpoB protein levels were determined in protein preparations isolated from leaves of transplastomic plants growing on regeneration media. (B) Determination of total and RpoB protein levels isolated in preparations from transplastomic plants resulting from a cross between wild-type pollen onto stigmas from transfomed plants grown in soil.

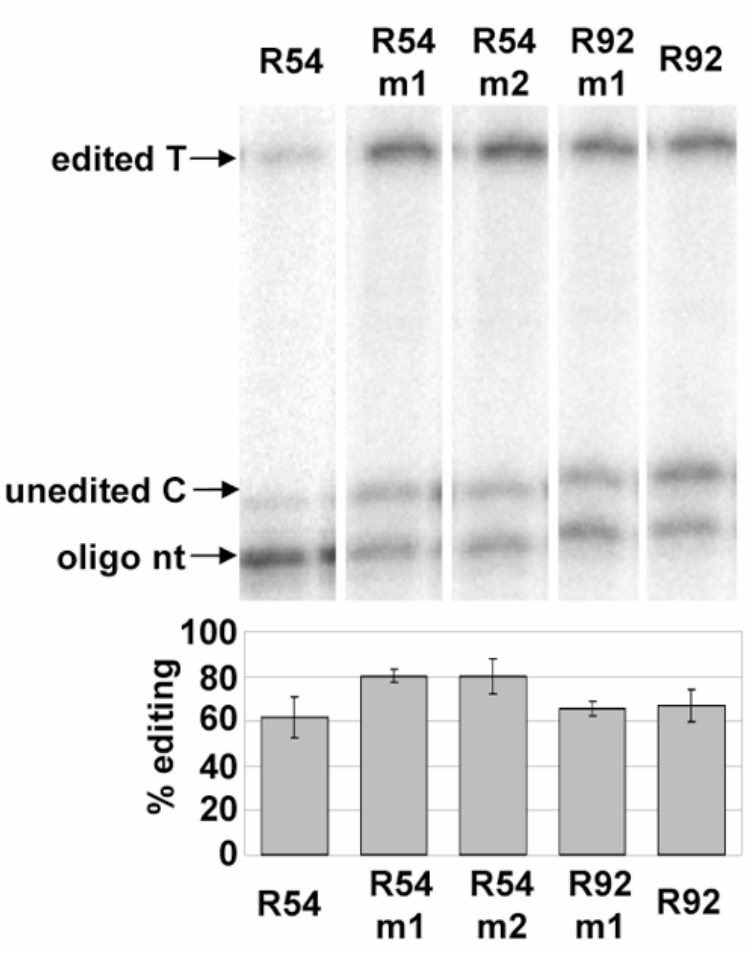


Figure 2.5. Poisoned primer extension reactions comparing editing in transgene transcripts between transplastomic plants. Primers PPrrn2 and Trps16sh, which amplify the region from -73 to +57 around the editing site in the transgene, yielded a 131bp fragment by RT-PCR. Percent editing was calculated from band intensity for duplicate reactions. Error bars represent one standard deviation from the mean.

eubacterial-like -35/ -10 promoter elements (Svab & Maliga, 1993). The 3′ end of the transcript is determined by the terminator sequence, Trps16, from the 3′ UTR of tobacco *rps16* (Zoubenko et al., 1994). The 3′ end of the *rps16* gene had not been previously mapped in tobacco; however, it had been determined in white mustard (Neuhaus et al., 1989). In white mustard, a nuclease has been purified and implicated in cleavage at a recognition sequence creating the 3′ end (Nickelsen & Link, 1991, 1993). We aligned the Trps16 sequence from tobacco with white mustard and found the tobacco sequence carries a sequence similar to the white mustard nuclease recognition element (Figure 2.6). We performed 3' end mapping in the tobacco transgenic plants and found that two RNA species are the product of the transcribed transgene in tobacco. One of the two, 3′ end I, might be a precursor to the more abundant 3′ end II. 3′ end II also matches a region of similar sequence where *rps16* in white mustard is cleaved to form the mature 3′ end.

After determining that two RNA species were the product of *in vivo* transcription of the transgene, we wanted to assess whether transcripts of both sizes were edited to the same extent. Through selective RT-PCR, it was possible to amplify the transcripts with 3′ end I only, unlike the data shown in Figure 2.5, in which primers were used so that transcripts of both sizes were assayed simultaneously. Of the two RNA species within transgenic plants, the smaller RNAs accumulated more edits in every construct than the longer precursor (Figure 2. 6). This could either result from an inhibition of editing by the extra 3' sequence, or because the longer transcripts are processed quickly to the smaller size, before significant editing has occurred.

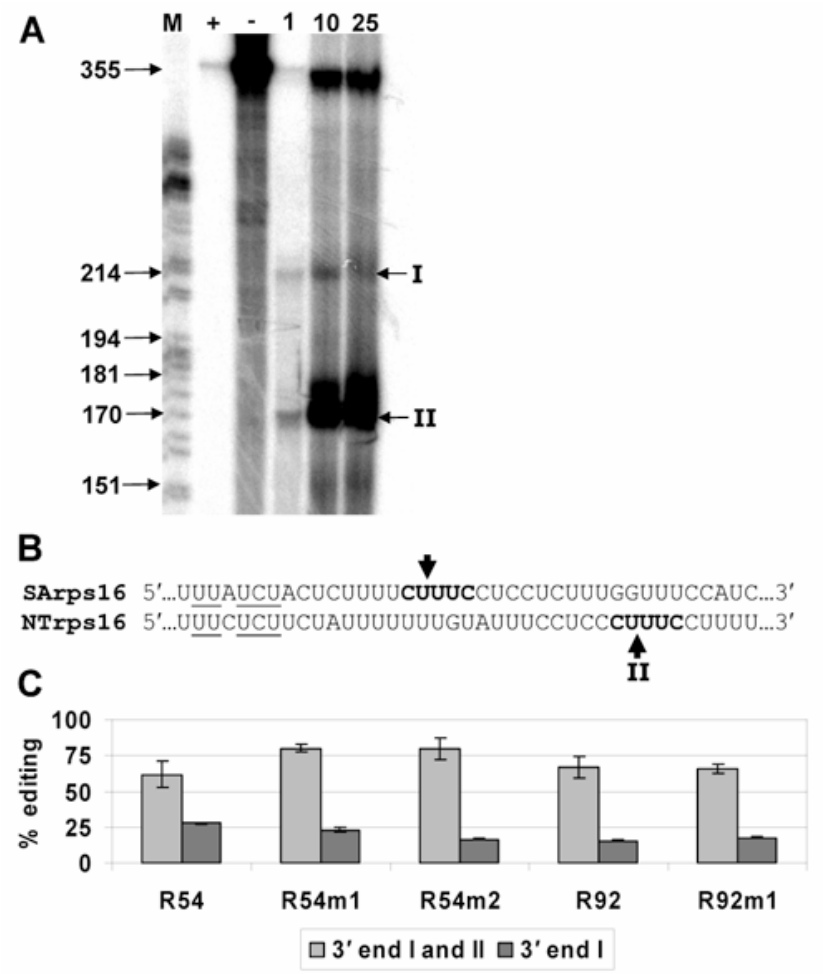


Figure 2.6. The 3′ ends of transgenic transcripts were determined using S1 nuclease protection mapping. (**A**) The 355 nt antisense probe was mixed with S1 nuclease (lane+), without S1 nuclease (lane -), and 1, 10, and 25 µg (lanes 1,10, 25) of total RNA from leaves of transplastomic shoots. Two 3′ ends were observed (I and II). Lane M contains a DNA sequencing reaction, which served as a molecular weight standard. (**B**) The 3′ UTR sequences from *rps16* of *Sinapsis alba* (SArps16) and *Nicotiana tobaccum* (NTrps16) were aligned. Arrows indicate mapped cleavage sites and underlined characters represent a 7-mer protein binding region from Nickelsen *et al.* (27). (**C**) The percentage of edited transgenic transcripts calculated by poisoned primer extension reactions with either 3′ end I or both 3′ ends. Error bars represent one standard deviation from the mean.

Editing of endogenous transcripts in transgenic plants carrying NTrpoB C473 transgenes

Previously we described editing of chloroplast transcripts in tobacco chloroplast transgenics that were overexpressing a 92 nt maize *rpoB* gene fragment encompassing the maize chloroplast sequence homologous to NTrpoB C473. Editing of the endogenous site NTrpoB C473 as well as sites NTpsbL C2, and NTrps14 C80, is reduced in these transgenic plants (Chateigner-Boutin & Hanson, 2002). All three sequences were observed to share common elements (Figure 2.7) and were therefore described as a "cluster" of editing sites. We analyzed the editing extent of endogenous transcripts of all three cluster members in the 5 new transgenic lines. All transgenic plants exhibit some reduction, varying from 10-30%, of endogenous NTrpoB C473 editing (Figure 2.7). In every homoplasmic plant, a reduction of endogenous NTrpoB C473 editing was accompanied by a 10-32% reduction of endogenous NTpsbL C2 editing. A reduction in editing extent of cluster member NTpsbL C2 was not observed in R54. Endogenous NTrps14 C80 editing is reduced 12% and 32% in the wild-type sequence-containing R54 and R92 plants, respectively. Reduced Rps14 C80 endogenous editing correlates with plants with large, over 20%, reductions in endogenous NTrpoB C473 and NTpsbL C2. Constructs with differences at -20 had little effect on NTrps14 C80 editing. NTrps14 C80 is less sensitive to over-expression of NTrpoB C473 substrates than is NTpsbL C2 editing. Editing extent of another editing cluster of endogenous sites that share common sequences with each other, NTrps14 C149 and NTndhB C737, but that do not carry S1, S2, nor S3, are unaffected by expression of transgenes carrying NTrpoB C473 (Figure 2.7).

A possible reason for differences in extent of inhibition of editing of related endogenous sequences between different *rpoB* transgenic lines could be different levels of transgene transcripts. The reduction in endogenous transcript editing is

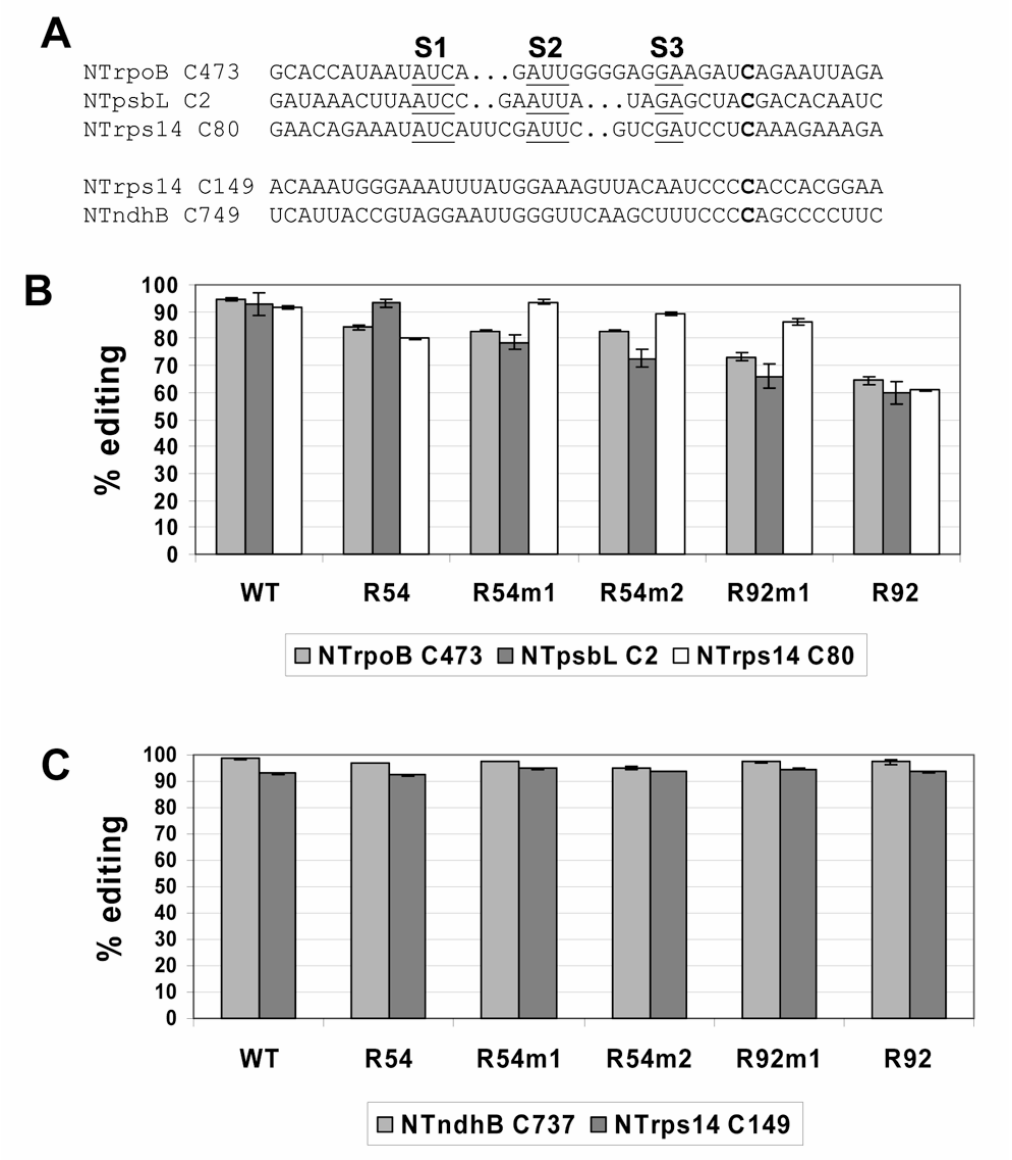


Figure 2.7. Percentage of sites edited in endogenous transcripts within transgenic plants. (**A**) Sequence alignment with induced spacing, of a cluster of three editing sites. Cluster members: NTrpoB C473, NTpsbL C2, and NTrps14 C80. Underlined characters represent common sequence elements: (S1), (S2), and (S3). (**B**) Editing in sites that contain S1, S2, and S3 sequences. (**C**) Editing in two sites that do not contain S1, S2, and S3 sequences. (**B, C**) Error bars represent one standard deviation from the mean.

thought to be due to competition among transcripts for a limited quantity of editing factors. Transgene transcripts presumably must be expressed at a sufficient level to engender this competition effect. We assayed the abundance of transgene transcripts in the 5 lines by semi-quantitative PCR. R92, the transgenic line which exhibited most reduction of editing of cluster members, also exhibited the highest transcript levels (Figure 2.8). R54, which exhibited no reduction in endogenous editing, had very low transcript levels, presumably due its heteroplasmic state. The other three lines exhibited intermediate amounts of transcripts and less reduction in editing of endogenous transcripts of cluster members than in R92 (Figure 2.8).

Analysis of chimeric transcripts *in vitro*

Analyzing the requirements for editing of transcripts incubated in chloroplast extracts *in vitro* allows more rapid examination of a large number of mutated RNA substrates than is possible with chloroplast transgenic plants, which require considerable labor and months of time for regeneration. We carried out further analysis of the cis-requirements for editing of *rpoB* using an *in vitro* editing system. Previously, the RNAs that have been assayed for editing efficiency *in vitro* have either been synthetic RNAs comprised only of chloroplast sequence and a 3' KS amplification sequence (Hirose & Sugiura, 2001; Miyamoto et al., 2002, 2004) or have been surrounded by SK and KS amplification sequences following transcription *in vitro* by T7 polymerase (Hegeman et al., 2005b). In order to mimic the transgene transcripts for *in vitro* analysis, a T7 promoter was placed 5' of the 92 nt *rpoB* sequence, which was followed by the sequence corresponding to the longer 3' end (3' end I). A second substrate was created that carried the short 3' end (3' end II). Both were incubated with tobacco chloroplast extract, and editing efficiency was assessed. Transcripts with the shorter 3' ends were more highly edited than those with the longer 3' ends (Figure 2.9), but the difference in editing extent between the long and

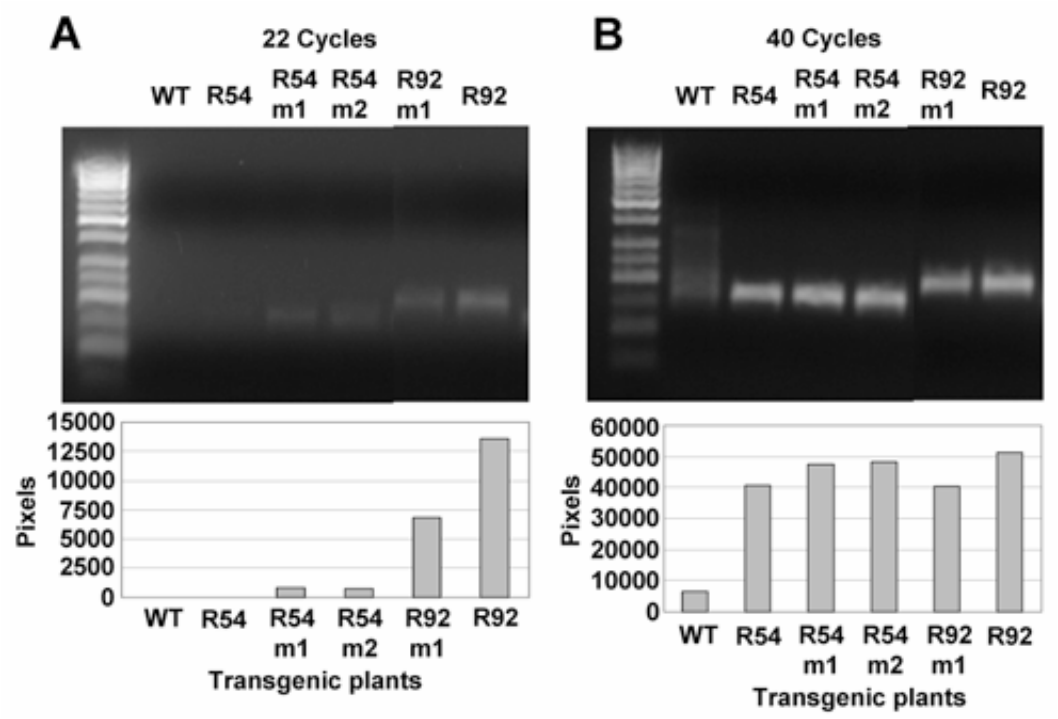
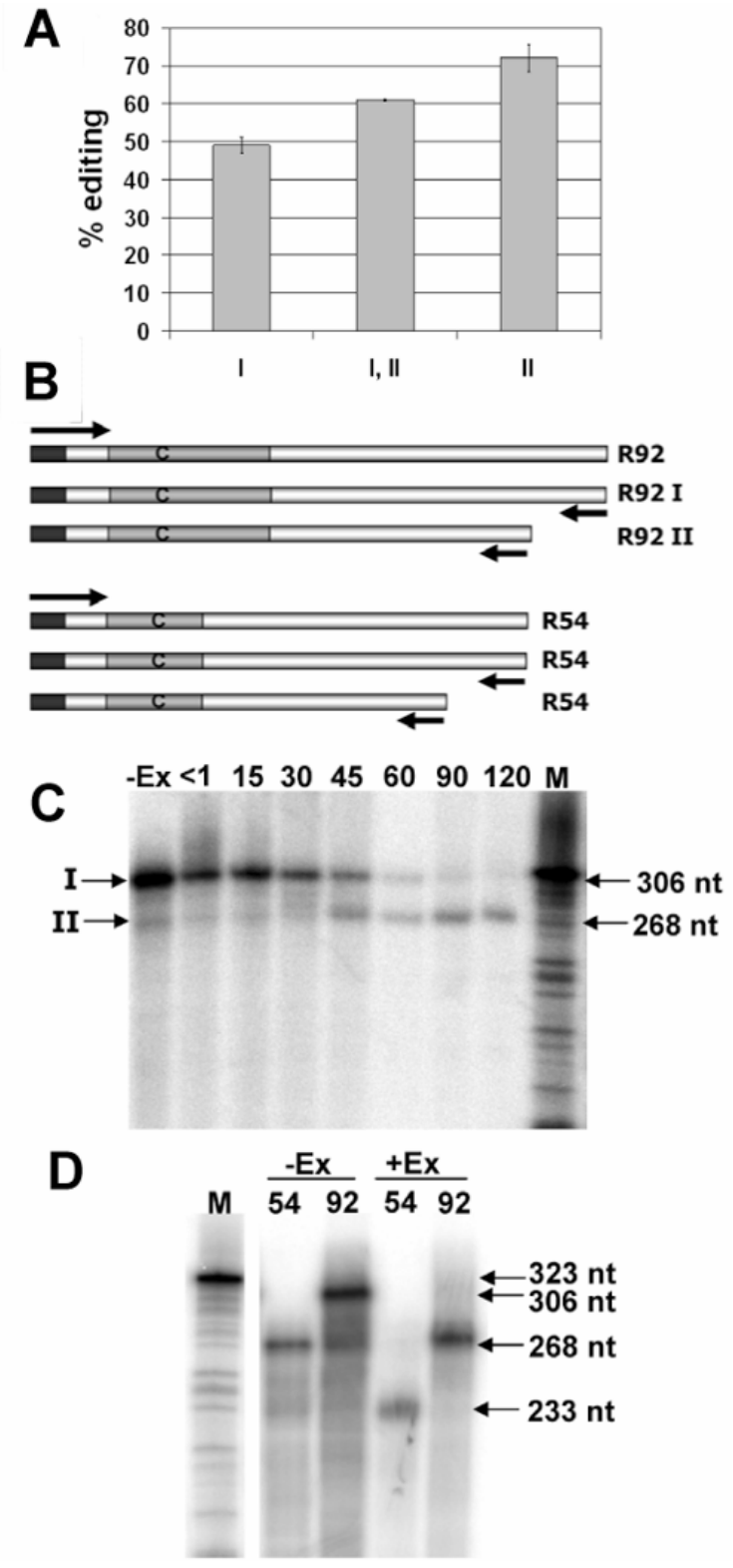


Figure 2.8. Semi-quantitative RT-PCR of transgene transcripts from transplastomic plants. (**A**) At 22 cycles of RT-PCR, bands of different intensities correspond to varying transcript abundances. (**B**) After 40 cycles of RT-PCR all bands are at equivalent intensities.

Figure 2.9. *In vitro* editing of substrates corresponding to transgenic transcripts were incubated in chloroplast extract under *in vitro* editing conditions. Editing percentages were calculated by comparing poisoned primer extension reaction intensity and error bars represent one standard deviation from the mean. (**A**) Lanes (I) and (I, II): Substrates with either 3' end I or both 3' ends were amplified through selective RT-PCR from an initial RNA template equivalent to 3' end I, respectively. Lane (II): Substrate was amplified with 3' end II from an RNA template with 3' end II. (**B**) Diagram of DNA substrates created to express RNA templates corresponding to the transgenic transcripts. Arrows indicate primers used for PCR amplification. Bars represent DNA substrates, black bars symbolize T7 sequence used for transcription *in vitro*, and gray bars indicate the region of *rpoB*. (**C**) Incubating 5′ end-labeled RNA substrates under *in vitro* editing conditions. Lane –Ex: without chloroplast extract for 120 minutes. Lanes <1 through 120: with tobacco chloroplast extract for points indicated up to 120 minutes. Bands that correspond to S1 nuclease mapped ends I and II are indicated to the left of the figure. (**D**) Internally labeled RNA substrates for R54 and R92 were incubated with tobacco chloroplast extract, +Ex, and without extract, -Ex, for 120 minutes. (**C, D**) Lane M: sequencing reactions serving as a molecular weight standard with molecular weights in nucleotides are indicated to the right.


A
% editing
80
70
60
50
40
30
20
10
0
I
I, II
II
B
R92
R92 I
R92 II
R54
R54
R54
C
-Ex <1 15 30 45 60 90 120 M
I
II
306 nt
268 nt
D
-Ex
+Ex
M 54 92 54 92
323 nt
306 nt
268 nt
233 nt

short transcripts was less than *in vivo*. When a primer was used that would amplify both long and short transcripts from extract initially incubated with the longer transcript, a higher editing efficiency was obtained. This suggested that the chloroplast extract used for editing *in vitro* might be capable of 3' end processing.

To assess the processing capacity of the chloroplast extracts used for editing analysis, RNA substrates equivalent to R92 with 3' end I were amplified by PCR and transcribed *in vitro*. Substrates were then radiolabeled at their 5' ends and incubated in chloroplast extracts to determine whether the longer RNA can be processed *in vitro*. The RNA was cleaved over time and was nearly fully digested after 2 hours (Figure 2.9). RNA substrates equivalent to R54 and R92 were randomly labeled to show if they were equally cleaved and that no other cleavage products were formed (Figure 2.9). Only substrates with 3′ ends equivalent to *in vivo* 3′ ends I and II were observed. Thus, chloroplast extracts used for assaying editing also exhibit 3' RNA processing activity. Since most transcripts have been processed by 2 hours of incubation (Figure 2.9), most of the transcripts assayed with the primer that amplifies both short and long transcripts are actually the shorter transcripts.

RNA substrates were constructed with the same 5' and 3' ends as *in vivo* transgenic transcripts to compare editing i*n vitro* vs. *in vivo*. The extent of editing within transcripts with sequence differences versus substrates carrying wild-type sequence correlates with the differences observed *in vivo* (Figure 2.10). To determine whether the presence of KS and SK sequence surrounding a chloroplast sequence affects editing efficiency of wild-type sequences *in vitro*, we compared a substrate with a KS sequence 3' to the 54 nt *rpoB* sequence vs. substrates carrying the rps16 3' end II (Figure 2.10). The insignificant difference between substrates carrying the 54 nt sequence and either the 5' rrn16/MCS 5' end and the rps16 3' end vs. the SK and KS 5' and 3' ends led us to utilize the convenient SK/KS sequences for further

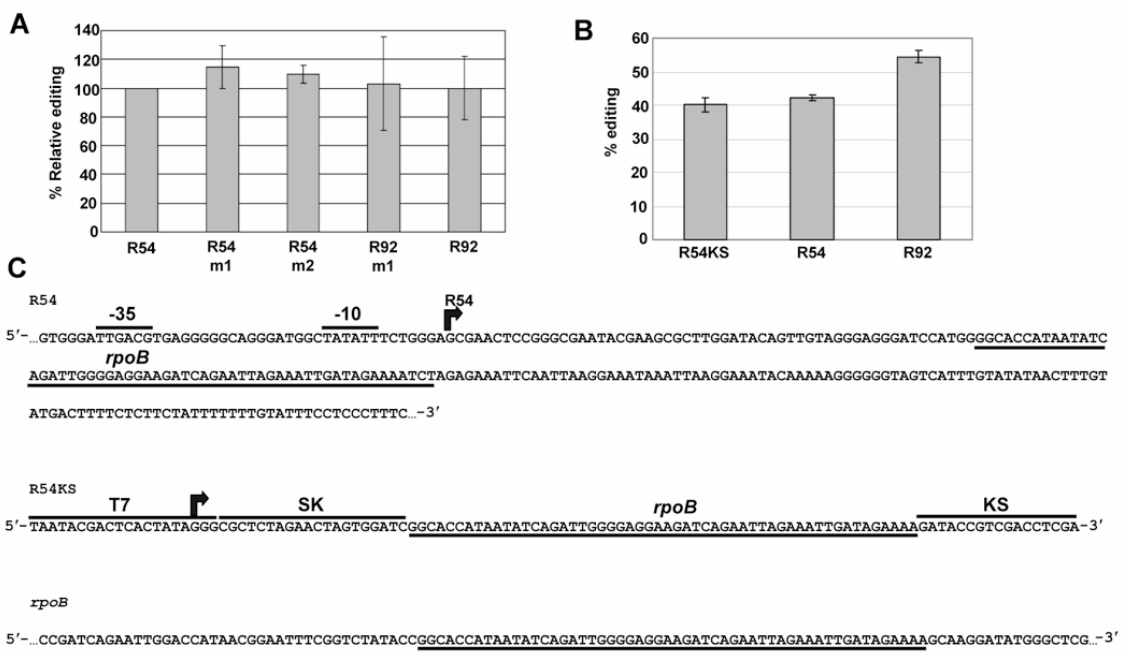


Figure 2.10. Editing *in vitro* of substrates analogous to transgenic transcripts. (A) Relative *in vitro* editing of substrates analogous to transgenic transcripts created with 3' end II. Editing is expressed as % of the wild-type R54 editing because substrates were assayed by two different extracts, one editing the constructs from 50-80% and the other from 22-36%. (**B**) *In vitro* editing of the R54KS substrate with bacterial sequences SK and KS around the editing site and substrates equivalent to the transgene transcript expressed in transplastomic plants. (**C**) Diagram of editing templates. Arrows represent the +1 position of transcription initiation. Underlined nucleotides signify the common *rpoB* sequence around the editing site.

assessment of editing efficiency of substrates in chloroplast extracts. The 5' and 3' environment of the NTrpoB C473 region is necessarily different in the transgene and *in vitro* substrates than in actual *rpoB* transcripts (Figure 2.10).

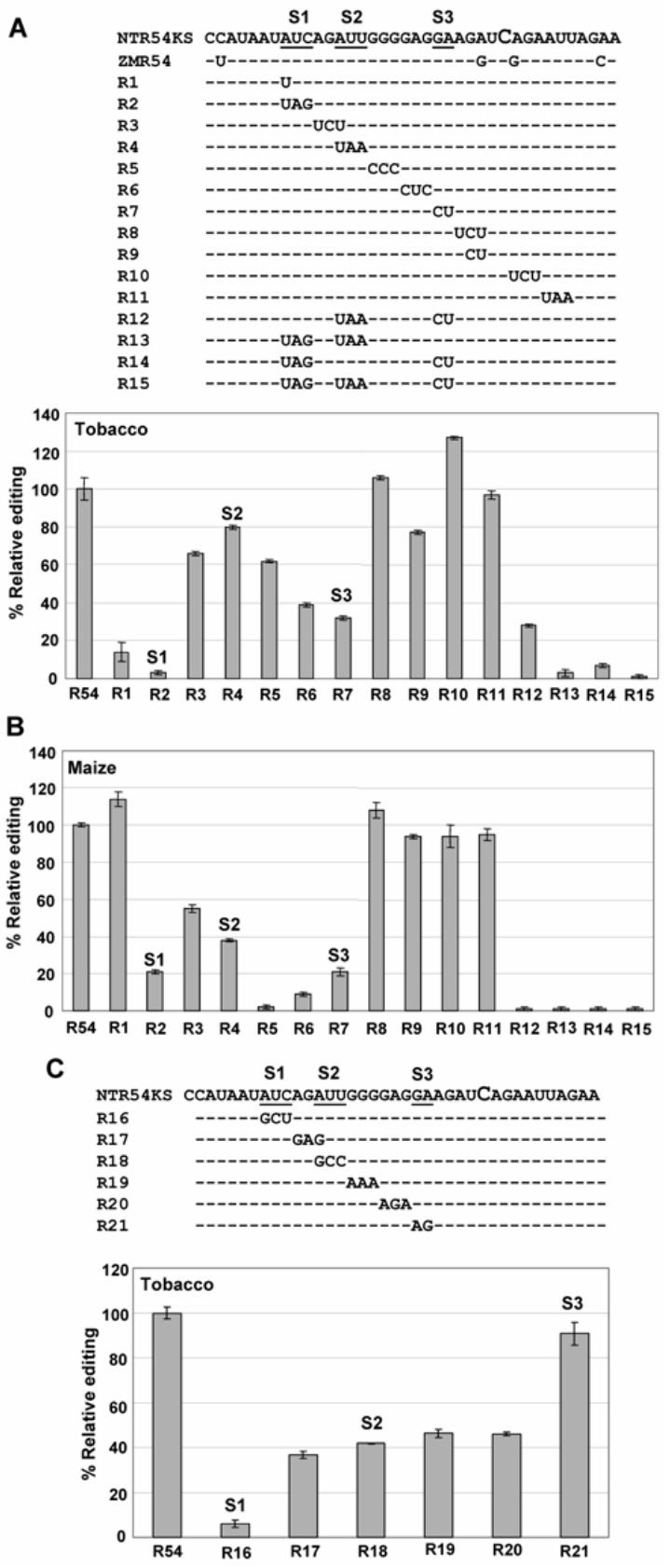
Sequence changes in S1, S2, and S3 affect editing efficiency

We produced a number of RNA substrates to assess the effect of alterations within the conserved elements on editing efficiency and to identify any other important sequences. DNA templates were created by PCR with bacterial T7 promoter and SK and KS sequences flanking the *rpoB* region and then transcribed *in vitro* to produce the substrate RNA. All substrates carried the -31/+22 chloroplast sequence present in the transgenes tested *in vivo*.

Substrates R1-R11 were created to determine the critical elements within -20/+6 region (Figure 2.11). Together these substrates cover the -20/+6 region found to be important *in vivo*. The editing efficiency of substrates R1-R7 and R9 in tobacco extracts was significantly reduced compared to the wild-type substrates. Of the substrates with reduced editing efficiencies, R4 and R9 had the smallest reduction (20%) in relative editing, and alterations were either not in an important region or not a sufficient sequence change to alter processing by the editing apparatus. Substrate R8 contains the same alterations as R9 but was not reduced in its editing efficiency, suggesting that the -3 and -2 nucleotides contained in R9 are not part of the critical sequence element. Curiously, R10 was a better substrate than R54KS perhaps indicating that the sequence restraints within coding regions result in endogenous sequences that do not always represent the optimal sequences for editing. The increase in editing is only around 30% and may be due to enhancement of RNA structure rather than representing a positive alteration in part of an important sequence element.

A single -20 A→U change nearly completely abolished editing, in contrast to the results obtained *in vivo* with the -31/+22 transgene (Figure 2.5), but in agreement

Figure 2.11. Effect of substrate sequence alterations of S1, S2, and S3. (**A**) Top: Alignment of RNA templates differing from wildtype sequence by purine to pyrmidine changes or visa versa. Bottom: RNA templates were incubated with tobacco or maize (**B**) chloroplast extracts. (**A, B**) The percentage of edited substrates, as calculated by poisoned primer extension. Percentage relative editing equals the percent editing of the substrate divided by the percent editing of the wildtype substrate. Data in (**A**) derived from three experiments in which wildtype R54KS was edited at either 30, 34, or 72%. Data in (**B**) derived from three experiments in which R54KS editing was either 30, 60, or 64%. **(C)** Top: Alignment if RNA templates differing from wildtype sequence by purine to purine or pyrimidine to pyrimidine changes. Bottom: RNA templates were incubated with tobacco chloroplast extracts and relative editing calculated after one experiment with R54KS editing of 73%. (**A,B, C**) Error bars represent one standard deviation from the mean from two replicates.


A
S1 S2 S3
NTR54KS CCAUAAUAUCAGAUUGGGGAGGAAGAUCAGAAUUAGAA
ZMR54 -U----------G--G-------C-
R1 -------U----------
R2 -------UAG----------
R3 ----------UCU----------
R4 ----------UAA----------
R5 ----------CCC----------
R6 ----------CUC----------
R7 ----------CU----------
R8 ----------UCU----------
R9 ----------CU----------
R10 ----------UCU-------
R11 ----------UAA----
R12 ----------UAA------CU----------
R13 -------UAG--UAA----------
R14 -------UAG----------CU----------
R15 -------UAG--UAA------CU----------
Tobacco
% Relative editing
S1
S2
S3
R54 R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14 R15
B
Maize
% Relative editing
S1
S2
S3
R54 R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14 R15
C
S1 S2 S3
NTR54KS CCAUAAUAUCAGAUUGGGGAGGAAGAUCAGAAUUAGAA
R16 ------GCU----------
R17 ---------GAG----------
R18 ----------GCC----------
R19 ----------AAA----------
R20 ----------AGA----------
R21 ----------AG----------
Tobacco
% Relative editing
S1
S2
S3
R54 R16 R17 R18 R19 R20 R21

with the low editing efficiency when only -20/+7 chloroplast sequence is included in the transgene (Reed et al., 2001b). Evidently the presence of the SK sequence 5' to the chloroplast sequence results in greater sensitivity of the transcripts to editing perturbation by alteration of the -20 nucleotide than does the presence of a few nucleotides derived from the *rrn16* promoter.

Substrates R2, R4, R7, and R12-R15 were created with differences in S1, S2, S3 to test their importance for editing NTrpoB C473 (Figure 2.11). R4 showed small, around 20%, reductions in editing relative to wild-type. This can be compared to the 65% reduction in R7 and abolition of editing in R2. Substrates carrying a combination of altered common elements, R12-R15, were created to study the effect of multiple changes. The effects of changes in S1, S2 and S3 are evident in cumulative reductions of editing in substrates with combinations of altered elements compared to wild-type substrates. All substrates with differences in S1 could not be edited by tobacco chloroplast extracts.

Substrates R1-R15 all had sequence changes from purine to pyrimidine and visa versa. Substrates R3-R5 only exhibited minor 20-40% reductions in editing compared to wild-type (Figure 2.11). To test whether purine to purine and pyrimidine to pyrimidine mutations might disrupt editing more significantly, such mutations were created within the previously defined critical region in substrates R16-R21 (Figure 2.11). Again nucleotide changes in S1, as in R16, had the greatest effect on editing. R17-R19 had more severe effects than substrates R3-R5. R18, carrying a change in S2 is a much poorer editing substrate than R4 or R54KS and confirms that S2 is a critical sequence. Overall, purine to purine and pyrimidine to pyrimidine base changes had a larger impact on editing than purine to pyrimidine and pyrimidine to purine. In spite of this, R21 had no effect compared to R7 although they both alter S3. S3 is flanked by GA nucleotides and the mutations in R21 happen to create two S3 sequences shifted

one nucleotide 5' and 3'of the endogenous position from the editing site. The sequence of R21 is evidently not sufficiently altered to reduce the editability of the substrate.

When maize *rpoB* is aligned with tobacco *rpoB*, ZMrpoB C467 is at the same position as NTrpoB C473, and the surrounding sequence is very similar to that in tobacco. Tobacco chloroplasts can edit a template expressing a -31/+61 region around ZMrpoB C467 (data not shown). Therefore, it is likely that tobacco and maize contain factors that can recognize similar cis-acting elements around NTrpoB C473. The same substrates tested *in vitro* using tobacco extracts were tested *in vitro* using maize extracts. Substrates R2-R7, which have sequence alterations in the -20/-5 region, exhibited major reductions in editing efficiency. The same region of sequence is therefore critical for editing in maize as in tobacco. All substrates with differences in S1, S2, and S3 were less efficient editing substrates in maize extracts compared to wild-type (Figure 2.11). As in the tobacco extracts, changes in S1 and S3 affected editing of the substrate more than S2. In contrast, the single -20 T to A change that severely affected editing in tobacco had little effect on editing in maize extracts. The editing efficiencies of the substrates with mutations within the -20/-5 region were reduced to a greater extent in maize compared to tobacco. Also substrates R5 and R6 had the largest reductions in editing compared to R2, which had the largest reduction in tobacco. These results indicate that, while the RpoB C467/C473 editing factors in the two species may be similar, some evolutionary divergence has probably occurred that has resulted in differential preference for sequences surrounding the editing site.

DISCUSSION

With the addition of maize extracts described here, assays for chloroplast editing *in vitro* are available in four species including tobacco, pea, and Arabidopsis as

well (Hirose & Sugiura, 2001; Hegeman et al., 2005b; Nakajima & Mulligan, 2005). *In vitro* systems have been particularly useful in studying the cis-acting elements for editing sites due to their relative speed, flexibility of species, and cost advantages. Studies *in vivo* have presently been limited to tobacco because of the technical difficulty of chloroplast transformation. The maize *in vitro* assay described here should facilitate identification of editing factors in an organism with better genetic resources. This is also the first monocot *in vitro* editing system and should allow comparisons of editing between dicot and monocot species.

The same substrates have not been previously assayed both *in vivo* and *in vitro* to test the biological relevance of the *in vitro* system. Here we describe the assay of comparable templates for NTrpoB C473 *in vivo* and *in vitro*, which allows a direct comparison between systems. Five substrates with sequence differences but the same approximate flanking sequences have been compared to wild-type substrates *in vivo* and *in vitro*. All 5 substrates showed similar *in vivo* and *in vitro* editing efficiencies, indicating that data from the *in vitro* system are applicable to editing in intact plant chloroplasts. Chloroplast extracts are therefore likely to be complete in their complement of critical editing factors.

Transcripts containing -31 to +22 nt surrounding NTrpoB C473 were more efficiently edited *in vivo* than transcripts with only -20 to +6 nt. Thus, substrates reduced to the immediate flanking and adequate region for editing do not represent the full complement of cis-acting elements that influence editing. Sequence elements that enhance editing are probably present outside of the -20/+6 region around the editing site. Nevertheless, the -20 to +6 region is sufficient to specify editing the proper C target *in vivo*.

Substrates have been constructed and tested *in vivo* or *in vitro* with changes in the nucleotides within the -20/+6 minimal region (Figure 2.12). Changes in the

Figure 2.12. Ssequences common around tobacco editing sites. (**A**) Nucleotides that have been altered within *in vivo* and *in vitro* substrates the editing site NTrpoB C473. Underlined characters represent positions where substrates with sequence alterations have been assayed *in vivo* or *in vitro*. The (*) indicates nucleotides that were changed in a substrate containing a -20 to +6 region of *rpoB* expressed in transplastomic plants by Reed *et al.* (14). Bold characters represent nucleotides that when are altered within substrates are edited less efficiently than wild-type. (**B**) Alignment of editing sites that contain S1, S2, and S3 sites. (**C**) Alignment of sequence of substrates containing 54 and 27 nt regions around NTrpoB C473 expressed *in vivo* and sequence from the substrates 3′ end II. Underlined characters represent regions of complementarity. (**D**) An alignment of sequence from the 3′ end I and wild-type sequence around the NTrpoB C473 site. Underlined sequences represent S1, S2, and S3 and similar sequences.

sequence from -20/-5 had large reductions in the editing efficiency of the corresponding substrate and the critical sequence element is contained within that region. Differences in S1, S2, and S3 consistently affected editing of the templates when assayed in extracts from tobacco and maize. Consistent with our hypothesis that elements conserved between cluster members are important site recognition features, S1, S2, and S3 are critical for RNA substrate editing in both tobacco and maize extracts. A substrate that contains a -20 A to U change is very poorly edited in tobacco but well edited in maize. This suggests that in maize the -20 nucleotide is not critical. The editing factor in maize apparently sufficiently differs from the tobacco factor that it can tolerate a change in the S1 sequence whereas the tobacco factor cannot. In addition, substrates with changes within the -20/-5 region have different reductions in editing efficiency between tobacco and maize. Although editing factors in maize recognize the same region of sequence they have diverged to the extent that different nucleotides are critical for editing.

11 of the 34 editing sites within tobacco chloroplasts share elements similar to S1, S2, and S3, although not all are reduced by over-expression of NTrpoB C473 (Figure 2.12). One explanation for the lack of a competition effect is the existence of different trans-factors for some or all of these sites. Another possible reason that sites with common cis-acting elements are differently affected by over-expression of NTrpoB C473 is that the same sequestered factor may be important for editing of multiple sites, but may not always be limiting. This seems to be the case even among sites whose editing is reduced by over-expression, as there are significant differences in the strength of competition between different members of the RpoB C467 cluster. Some sites may have stronger affinities for a particular trans-factor than other sites. Therefore, all 11 sites may share important cis-acting elements, but the possible relationship of such sites will not be obvious by over-expression of a single RNA

editing site in a transgenic plant.

The presence of S1, S2, and S3 in the three genes whose editing is altered in transgenic plants could only be identified by allowing significant gaps, suggesting critical sequence elements could be irregularly spaced from the edited site. This complicates understanding how editing sites are specifically targeted, since any common distance from a cis-acting element to the editing site might differ among sites. Possibly, either a processive mechanism specifies a particular nucleotide for editing that is different between sites from a cis-acting element, or there are elements that require particular spacing that may serve roles in locating the editing components to the particular C to be edited. Any processive mechanism within the NTrpoB C473 cluster would have to differ significantly between sites due to different sequences around the editing site, and the position of 5′ and 3′ Cs around editing sites. Sensitivity of editing site substrates to altered spacing has been observed in mitochondria, suggesting a potential for a "molecular ruler" determining the target for editing by its distance from a cis-acting element (Choury et al., 2004; Neuwirt et al., 2005). In the NTrpoB C473 cluster, S3 is 4 to 5 nucleotides away from the edited nucleotide and could be important for editing site targeting. Substrate R21 is not reduced in its editing efficiency and suggests that small perturbations in the precise position of this sequence do not affect editing, at least in *rpoB* transcripts (Figure 2.11). In mitochondria, editing sites separated by only a short length of sequence were found to share cis-acting elements (Hermann & Bock, 1999; van der Merwe et al., 2006). Interestingly, sites NTatpA C791 and NTatpA C795 contain two S3 elements 5 nt and 6 nt upstream of the edited sites respectively, and share the same upstream elements S1 and S2 (Figure 2.12), but their editing efficiency is unaffected by overexpression of NTrpoB C473.

Some editing sites in plant mitochondria and chloroplasts have been shown to be sensitive to changes in nucleotides immediately adjacent to the editing site

(Miyamoto et al., 2004; Neuwirt et al., 2005). However, an inserted ZMrpoB C467 editing site from maize sequence, flanked by 92 bases, differs from the tobacco sequence by a 3′ adjacent G and is edited as well as the tobacco sequence (Reed & Hanson, 1997; Hegeman et al., 2005a). Endogenous editing of NTrpoB C473 and cluster member NTpsbL C2 are reduced when the NTrpoB C473 gene fragment is over expressed, although the NTpsbL C2 editing site differs in its immediate 5′ and 3′ nucleotides (Chateigner-Boutin & Hanson, 2002). If NTpsbL C2 shares factors involved in the editing mechanism, then the editing complex for NTrpoB C473 may be insensitive to differences in nucleotides flanking the editing site.

The -20A nucleotide has been determined to be critical *in vivo* using a 27 nt (-20, +6) template and *in vitro* using a 54 nt (-31, +22) template surrounded by SK and KS sequences. *In vivo*, however, a substrate containing a 54 nt region around NTrpoB C473 with the -20U alteration can be more highly edited than a wild-type substrate. A substrate analogous to the *in vivo* transgenic transcript is also well edited *in vitro* compared to wild-type. Therefore, the reduction in editing caused by the -20U alteration depends upon the particular sequences 5' to the -20 nt. Since two independent observations show a critical role for the -20A, it is likely it is important for editing. Possibly the importance of the region 5' to -20 relates to the existence of complementary sequence in the mature 3′ end, which could be involved in a secondary structure near the editing site (Fig. 12). The extent of the complementarity would differ between the 54 nt region around NTrpoB C473 and the 27 nt region, and would not be present in the *in vitro* substrate that carries SK and KS flanking sequences. Possibly this region of complementarity can compensate for the -20U mutation in the *in vivo* R54m1 and R54m2 transcripts. Other unknown RNA secondary structures or cis-elements might also result in the compensation for the -20U alteration.

The presence or absence of critical nucleotides near the C editing target does

not solely determine the extent to which an RNA substrate is edited *in vivo* or *in vitro*. RNA editing sites are sensitive to not only to the amount of local sequence around the edited site, but can also be affected by sequence features significantly distal to the editing site. *In vivo* and *in vitro* RNA substrates with 3′ ends containing 44 nt more *rps16* sequence, 111 nt away from the edited nucleotide, are edited less efficiently. The endogenous NTrpoB C473 editing site is located far from the 3′ end of its transcript, and possibly the edited machinery adapted to edit this site may have done so without the bulky stem loops and other elements that normally characterize plastid transcript 3′ ends. Another possibility is that there is sequence in the longer 3′ end that has the ability to sequester editing factors. Indeed, near the end of the longer *rps16* 3' end present in the transgene transcripts, there is some sequence similar to the common elements known to be critical for NTrpoB C473 editing (Figure 2.12). Substrate editing is therefore sensitive to the presence of nearby cis-acting elements, more distant enhancing elements, and flanking sequence. These sensitivities could be responsible for the early difficulty in creation of an *in vitro* system, and suggest why only a limited set of editing sites have been described that are edited *in vitro* at high efficiency.

ACKNOWLEDGEMENTS

We thank Barbara Eaglesham for maintaining transplastomic plants in tissue culture and Pal Maliga for providing the chloroplast transformation vector. This work was supported by NSF MCB 0344007 and NIH GM 50723 grants to M.R.H and a NIH graduate traineeship to M.L.H.

REFERENCES

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

Chateigner-Boutin AL, Hanson MR. 2003. Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res 31*:2586-2594.

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

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

Covello PS, Gray MW. 1993. On the evolution of RNA editing. *Trends Genet 9*:265-268.

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

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

Hegeman CE, Halter CP, Owens TG, Hanson MR. 2005a. Expression of complementary RNA from chloroplast transgenes affects editing efficiency of transgene and endogenous chloroplast transcripts. *Nucleic Acids Res 33*:1454-1464.

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

Hermann M, Bock R. 1999. Transfer of plastid RNA-editing activity to novel sites suggests a critical role for spacing in editing-site recognition. *Proc Natl Acad Sci U S A 96*:4856-4861.

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

Maier RM, Zeltz P, Kossel H, Bonnard G, Gualberto JM, Grienenberger JM. 1996. RNA editing in plant mitochondria and chloroplasts. *Plant Mol Biol 32*:343-365.

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

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

Nakajima Y, Mulligan RM. 2005. Nucleotide specificity of the RNA editing reaction in pea chloroplasts. *J Plant Physiol 162*:1347-1354.

Neuhaus H, Scholz A, Link G. 1989. Structure and expression of a split chloroplast gene from mustard (*Sinapis alba*): ribosomal protein gene *rps16* reveals unusual transcriptional features and complex RNA maturation. *Curr Genet 15*:63-70.

Neuwirt J, Takenaka M, JA VDM, Brennicke A. 2005. An *in vitro* RNA editing system from cauliflower mitochondria: Editing site recognition parameters can vary in different plant species. *RNA 11*:1563-1570.

Nickelsen J, Link G. 1991. RNA-protein interactions at transcript 3' ends and evidence for *trnK-psbA* cotranscription in mustard chloroplast. *Mol Gen Genet*:89-96.

Nickelsen J, Link G. 1993. The 54 kDa RNA-binding protein from mustard chloroplasts mediates endonucleolytic transcript 3' end formation *in vitro*. *Plant J 3*:537-544.

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

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

Reed ML, Hanson MR. 1997. A heterologous maize *rpoB* editing site is recognized by transgenic tobacco chloroplasts. *Mol Cell Biol 17*:6948-6952.

Reed ML, Lyi SM, Hanson MR. 2001a. Edited transcripts compete with unedited mRNAs for trans-acting editing factors in higher plant chloroplasts. *Gene 272*:165-171.

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

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

Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, Herrmann RG, Maier RM. 2005. Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. *Plant Cell 17*:1815-

1828.

Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM. 2002. The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol 19*:1602-1612.

Smith HC, Gott JM, Hanson MR. 1997. A guide to RNA editing. *RNA 3*:1105-1123.

Svab Z, Maliga P. 1993. High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc Natl Acad Sci U S A 90*:913-917.

Takenaka M, Brennicke A. 2003. *In vitro* RNA editing in pea mitochondria requires NTP or dNTP, suggesting involvement of an RNA helicase. *J Biol Chem*.

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

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

Zito F, Kuras R, Choquet Y, Kossel H, Wollman FA. 1997. Mutations of cytochrome b6 in Chlamydomonas reinhardtii disclose the functional significance for a proline to leucine conversion by *petB* editing in maize and tobacco. *Plant Mol Biol 33*:79-86.

Zoubenko OV, Allison LA, Svab Z, Maliga P. 1994. Efficient targeting of foreign genes into the tobacco plastid genome. *Nucleic Acids Res 22*:3819-3824.

CHAPTER 3

∗INDENTIFICATION OF A SEQUENCE A MOTIF CRITICAL FOR EDITING OF A TOBACCO CHLOROPLAST TRANSCRIPT

ABSTRACT

Nucleotides are specifically and efficiently targeted for modification from C to U within transcripts of chloroplasts in higher plants. Although the enzymatic apparatus responsible for altering Cs to Us has not been identified, the sequences surrounding editing sites are known to contain information essential for efficient editing. We set out to determine the nucleotides that are critical for editing of a particular C, NTpsbE C214, in chloroplast transcripts in tobacco. Assay of editing of substrates with different lengths of 5′ and 3′ sequence around the target C was carried out to delimit the region of sequence critical for editing *in vitro*. Mutated substrates were then constructed with an altered nucleotide at each position within the previously defined region around NTpsbE C214. Individual nucleotides both 5′ and 3′ of the edited nucleotide were found to be important for editing. The sequence GCCGUU, which occurs 5′ of the editing site, was discovered to be critical for editing. Editing substrates mutated to alter the distance between the GCCGUU sequence and NTpsbE C214 resulted in the generation of a new editing target, the 3′ adjacent nucleotide. These data are consistent with a model in which the selection of the C target for editing is determined by its distance from a crucial 5′ sequence.

∗ Hayes ML and Hanson MR (2007) RNA 13:281-8, ©2007 by Cold Spring Harbor Laboratory Press.

INTRODUCTION

C-to-U RNA editing is a vital component of chloroplast gene expression in higher plant species. First observed in maize chloroplasts (Hoch et al. 1991), chloroplast RNA editing is now known to occur also in other seed plants as well as in hornworts, bryophytes, true ferns, and fern allies (Freyer et al. 1997). Though C-to-U RNA editing of chloroplasts transcripts occurs in all vascular plants that have been examined, the particular C targets of editing are not conserved across different plant species. Some species carry Ts at the homologous location where other plants encode a C that is modified to U at the transcript level. A typical number of editing targets in the chloroplast transcripts of an angiosperm species is 30 to 40, while many more Cs are targeted in mitochondria (Maier et al. 1995; Wakasugi et al. 1996; Corneille et al. 2000; Tsudzuki et al. 2001; Schmitz-Linneweber et al. 2002; Inada et al. 2004; Tillich et al. 2005; Sasaki et al, 2006; Kahlau et al. 2006).

When the sequences around all known chloroplast editing sites are aligned, no common consensus sequence can be detected. The most extensive information on the effect of cis-acting sequences on RNA editing has been obtained in tobacco, where chloroplast transformation and active chloroplast editing extracts have allowed analysis of mutated substrates both *in vivo* and *in vitro* (Chaudhuri and Maliga, 1996; Bock and Fuchs, 1997; Hirose and Sugiura, 2001; Hayes et al. 2006). Of primary importance for editing efficiency is the sequence 5′ of the C target of editing. Previous analyses of editing *in vitro* by Miyamoto et al. (2002) revealed the presence of important editing cis-elements in the sequence -15/-1 region 5′ to the single target of editing in tobacco *psbE* transcripts. The -15/-5 sequence affects the specific UV-crosslinking of a 56 kDa protein to *psbE* transcripts (Miyamoto et al. 2004). We set out to further define the sequence requirements within the critical 5′ and 3′ regions of the tobacco *psbE* editing site.

Not only has the identity of particular nucleotides around a C target been shown to affect editing efficiency, but also the particular C that is targeted for editing can be altered from one to another by insertion or deletion mutations in a substrate. Hermann and Bock (1999) found that mutated *ndhB* transcripts expressed *in vivo* were edited at novel sites if the distance between an unidentified 5′ cis-element and the normal C target of editing was altered. In *ndhB* transcripts, a "molecular ruler" might be critical to specify the selection of the C to be edited. To gain more information concerning the mechanism of selection of the C target in chloroplast transcripts, we have assayed the effect of nucleotide insertions and deletions on editing of tobacco *psbE* transcripts i*n vitro.*

MATERIALS AND METHODS

Preparation of tobacco chloroplast extracts

Tobacco chloroplast extracts were prepared as described in Hegeman et al. (2005). Tobacco (*Nicotiana tabacum* var. Petit Havana) was grown at 25C for 4-6 weeks under metal halide lights with a 16h light, 8h dark cycle. Plants were covered for 2 days before chloroplast isolation. Chloroplasts were isolated from fully expanded leaves as described by Miyamoto et al. (2002). Intact chloroplasts were separated on a continuous Percoll gradient and washed. Chloroplasts were lyzed using Triton X-100 and a hypertonic buffer to produce chloroplast extracts. Extracts were then dialyzed and contained about 20μg/μl of protein.

Creation of RNA substrates

Oligonucleotides (Integrated DNA Technologies, Coralville, IA) were used to produce DNA templates by PCR (Table 3.1). Templates containing Arabidopsis *psbE* sequences with different lengths of 5′ sequence from -150 to -13 were constructed as described in Hegeman et al. (2005). Templates with single nucleotide mutations were

constructed using mutant primers.

Synthetic RNA was created through *in vitro* transcription using the DNA template PCR product. The DNA template was removed by TurboDNAse digestion. The RNA was then cleaned up using the ZYMO cleanup-5 Kit.

Editing *in vitro*

Reaction conditions for editing *in vitro* were similar to those described in Hegeman et al. (2005). Optimized reaction conditions included the elimination of magnesium acetate and an increase in ATP concentration from 10mM to 100mM. Poisoned primer extension (PPE) was carried out as in Hegeman et al. (2005) to quantify the percent editing in templates. Three different primers were used for PPE: NTpsbE_PPE_G was used to quantify editing at C214 of all substrates without mutations within the 5′ region; NTpsbE_PPE_C was used to quantify editing at the C214 editing site in substrates with mutations in the 3′ region; and NTpsbE_AltC_G was used to quantify editing at C215.

Competition experiments

Self-competitor RNAs were created identically to the p*sbE* substrates except the flanking sequences were swapped. Primers were used to add the T7KS bacterial sequence 5′ of the *psbE* sequence and the SK sequence formed the 3′ end. RNA competitors were then produced by *in vitro* transcription from the PCR product template. RNA competitors, therefore, could not be amplified by using the same SK and KS primers as used for the editing substrates. Different concentrations of RNA competitors were then added to typical reaction conditions immediately after the addition of 10 fmol of substrate.

Table 3.1. Sequences of oligonucleotides (Integrated DNA Technologies, Coralville, IA) used for the following experiments.

Name	Sequence 5′-3′	Purpose
ATpsbE150For	CGCTCTAGAACTAGTGGATCTTCATAG CATTACTATACCT	ATpsbE150 template
ATpsbE99For	CGCTCTAGAACTAGTGGATCAGCACCG GTTTAGCTTACGA	ATpsbE99 template
ATpsbE54For	CGCTCTAGAACTAGTGGATCGAGTATT TTACAGAGAGCCG	ATpsbE54 template
ATpsbE31For	CGCTCTAGAACTAGTGGATCAGGCATT CCATTAATAACAG	ATpsbE 31 template
ATpsbE22For	CGCTCTAGAACTAGTGGATCATTAATA ACAGGCCGTTTTG	ATpsbE 22 template
ATpsbE13For	CGCTCTAGAACTAGTGGATCAGGCCGT TTTGATCCTTTGG	ATpsbE13 template
ATpsbE15Rev	TCGAGGTCGACGGTATCTCGAGTTGTT CCAAAGGA	ATpsbE templates
ATpsbE15_G	TCGAGGTCGACGGTATCTCGAGTTGTT CCAAAG	poisoned primer extension
T7SK	TAATACGACTCACTATAGGGCGCTCTA GAACTAGTGGATC	Template construction
KS	TCGAGGTCGACGGTATC	Template construction
NTpsbE100For	CGCTCTAGAACTAGTGGATCCAGCACC GGTTTAGCTTACG	NTpsbE templates
NTpsbE+15Rev	TCGAGGTCGACGGTATCCGAGTTGTTC CAAAGGATCA	NTpsbE+15 template
NTpsbE+10Rev	TCGAGGTCGACGGTATCTGTTCCAAAG GATCAAAACG	NTpsbE+10 template
NTpsbE-20Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCAGTTATTTATGGAATTCC	NTpsbE-20U→A template
NTpsbE-19Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCAGTTATAAATGGAATTC	NTpsbE-19A→U template
NTpsbE-18Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCAGTTAATAATGGAATT	NTpsbE-18A→U template
NTpsbE-17Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCAGTTTTTAATGGAAT	NTpsbE-17U→A template
NTpsbE-16Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCAGTAATTAATGGAA	NTpsbE-16A→U template
NTpsbE-15Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCAGATATTAATGGA	NTpsbE-15A→U template
NTpsbE-14Rev	CGACGGTATCTGTTCCAAAGGATCAAA ACGGCCACTTATTAATGG	NTpsbE-14C→G template

Table 3.1. (Continued)

NTpsbE-13Rev	TCGACGGTATCTGTTCCAAAGGATCA AAACGGCCTGTTATTAATG	NTpsbE-13U→A template
NTpsbE-12Rev	GTCGACGGTATCTGTTCCAAAGGATC AAAACGGCGAGTTATTAAT	NTpsbE-12G→C template
NTpsbE-11Rev	GTCGACGGTATCTGTTCCAAAGGATC AAAACGGGCAGTTATTAA	NTpsbE-11G→C template
NTpsbE-10Rev	GTCGACGGTATCTGTTCCAAAGGATC AAAACGCCCAGTTATTA	NTpsbE-10C→G template
NTpsbE-9Rev	GTCGACGGTATCTGTTCCAAAGGATC AAAACCGCCAGTTATT	NTpsbE-9C→G template
NTpsbE-8Rev	GGTCGACGGTATCTGTTCCAAAGGAT CAAAAGGGCCAGTTATTAA	NTpsbE-8G→C template
NTpsbE-7Rev	GGTCGACGGTATCTGTTCCAAAGGAT CAAACCGGCCAGTTATTAA	NTpsbE-7U→G template
NTpsbE-6Rev	GGTCGACGGTATCTGTTCCAAAGGAT CAATACGGCCAGTTATTAA	NTpsbE-6U→A template
NTpsbE-5Rev	GGTCGACGGTATCTGTTCCAAAGGAT CAGAACGGCCAGTTA	NTpsbE-5U→C template
NTpsbE-4Rev	AGGTCGACGGTATCTGTTCCAAAGGA TCGAAACGGCCAGTT	NTpsbE-4U→C template
NTpsbE-3Rev	AGGTCGACGGTATCTGTTCCAAAGGA TTAAAACGGCCAGT	NTpsbE-3G→A template
NTpsbE-2Rev	AGGTCGACGGTATCTGTTCCAAAGGA ACAAAACGGCCAGT	NTpsbE-2U→A template
NTpsbE-1Rev	GTCGACGGTATCTGTTCCAAAGGTTC AAAACGGCCAGTTA	NTpsbE-1U→A template
NTpsbE+1Rev	TCGAGGTCGACGGTATCTGTTCCAAA CGATCAAAACGGCC	NTpsbE+1C→G template
NTpsbE+2Rev	TCGAGGTCGACGGTATCTGTTCCAAT GGATCAAAACGGCC	NTpsbE+2U→A template
NTpsbE+3Rev	TCGAGGTCGACGGTATCTGTTCCATA GGATCAAAACGGCC	NTpsbE+3U→A template
NTpsbE+4Rev	TCGAGGTCGACGGTATCTGTTCCTAA GGATCAAAACGGCC	NTpsbE+4U→A template
NTpsbE+5Rev	TCGAGGTCGACGGTATCTGTTCGAAA GGATCAAAACGGCC	NTpsbE+5G→C template
NTpsbE+6Rev	TCGAGGTCGACGGTATCTGTTGCAAA GGATCAAAACGGCC	NTpsbE+6G→C template
NTpsbE+7Rev	CGAGGTCGACGGTATCTGTACCAAA GGATCAAAACGGCCA	NTpsbE+7A→U template

Table 3.1. (Continued)

NTpsbE+8Rev	TCGAGGTCGACGGTATCTGGTCCAAAGGATCAAAACGGCC	NTpsbE+8A→C template
NTpsbE+9Rev	TCGAGGTCGACGGTATCTCTTCCAAAGGATCAAAACGGC	NTpsbE+9C→G template
NTpsbE+10Rev	TCGAGGTCGACGGTATCAGTTCCAAAGGATCAAAACGGCC	NTpsbE+10A→U template
NTpsbE_PPE_C	GCCGACAAGGAATTCCATTAATAACTGGCC	Poisoned primer extension
NTpsbE_PPE_G	TCGAGGTCGACGGTATCTGTTCCAAAG	Poisoned primer extension
NTpsbE_AltC_G	TCGAGGTCGACGGTATCTGTTCCAAA	Poisoned primer extension
NTrpoBC338 For	GATACCGTCGACCTCGATTATATGTATCCGCGGGATTAAT	NTrpoB C338 template
NTrpoBC338 Rev	GATCCACTAGTTCTAGAGCGATTGACTATAGAAGTTCCCA	NTrpoB C338 template
NTndhBC467For	GATACCGTCGACCTCGAGAAATGGCTATAACAGAGTTTCT	NTndhB C467 template
NTndhBC467Rev	GATCCACTAGTTCTAGAGCGGAAACATTCTGGGGCTACAA	NTndhB C467 template
NTpsbE-1ΔRev	TCGAGGTCGACGGTATCTGTTCCAAAGGTCAAAACGGCCAGTT	NTpsbE-1Δ template
NTpsbE+2UIn	TCGAGGTCGACGGTATCTGTTCCAAAAGGATCAAAACGGCCAGT	NTpsbE+2U template
NTpsbE-1Δ,+2Uin	TCGAGGTCGACGGTATCTGTTCCAAAAGGTCAAAACGGCCAGTT	NTpsbE-1Δ,+2U template
NTpsbE+2AIn	TCGAGGTCGACGGTATCtgttccaaaTgGatcaaaacggcca	NTpsbE+2A template
NTpsbE-2ΔRev	TCGAGGTCGACGGTATCtgttccaaagGacaaaacggccagtta	NTpsbE-2Δ template
NTpsbE-3ΔRev	TCGAGGTCGACGGTATCtgttccaaagGataaaacggccagttat	NTpsbE-3Δ template
NTpsbE-4ΔRev	CGAGGTCGACGGTATCtgttccaaagGatcaaacggccagttatt	NTpsbE-4Δ template
NTpsbE+8URev	TCGAGGTCGACGGTATCtgAtccaaagGatcaaaac	NTpsbE+8A→U template
NTpsbE+8GRev	TCGAGGTCGACGGTATCtgCtccaaagGatcaaaac	NTpsbE+8A→G template

Table 3.1. (Continued)

NTpsbE+2CRev	TCGAGGTCGACGGTATCtgttccaaGgGatc aaaacggcc	NTpsbE+2U→C template
NTpsbE+2GRev	TCGAGGTCGACGGTATCtgttccaaCgGatc aaaacggcc	NTpsbE+2U→G template
NTpsbE+3CRev	TCGAGGTCGACGGTATCtgttccaGagGatc aaaacggcc	NTpsbE+3U→C template
NTpsbE+3GRev	TCGAGGTCGACGGTATCtgttccaCagGatc aaaacggcc	NTpsbE+3U→G template
NTpsbE-2GRev	AGGTCGACGGTATCtgttccaaagGaCcaaaa cggccagtta	NTpsbE-2A→G template
NTpsbE-2CRev	AGGTCGACGGTATCtgttccaaagGaGcaaaa cggccagtta	NTpsbE-2A→C template
NTpsbE-3CRev	AGGTCGACGGTATCtgttccaaagGatGaaaa cggccagttat	NTpsbE-3G→C template
NTpsbE-3URev	AGGTCGACGGTATCtgttccaaagGatAaaaa cggccagttat	NTpsbE-3G→U template
NTpsbE-4ARev	AGGTCGACGGTATCtgttccaaagGatcTaaa cggccagttatt	NTpsbE-4U→A template
NTpsbE-4GRev	AGGTCGACGGTATCtgttccaaagGatcCaaa cggccagttatt	NTpsbE-4U→G template
NTpsbE-5ARev	AGGTCGACGGTATCtgttccaaagGatcaTaa cggccagttatta	NTpsbE-5U→A template
NTpsbE-5GRev	AGGTCGACGGTATCtgttccaaagGatcaCaa cggccagttatta	NTpsbE-5U→G template
NTpsbE-6ARev	AGGTCGACGGTATCtgttccaaagGatcaaTa cggccagttattaa	NTpsbE-6U→A template
NTpsbE-6GRev	AGGTCGACGGTATCtgttccaaagGatcaaCa cggccagttattaa	NTpsbE-6U→G template
NTpsbE-7ARev	AGGTCGACGGTATCtgttccaaagGatcaaaT cggccagttattaat	NTpsbE-7U→A template
NTpsbE-7GRev	AGGTCGACGGTATCtgttccaaagGatcaaaC cggccagttattaat	NTpsbE-7U→G template
NTpsbE-8CRev	AGGTCGACGGTATCtgttccaaagGatcaaaa Gggccagttattaatg	NTpsbE-8G→C template
NTpsbE-8ARev	AGGTCGACGGTATCtgttccaaagGatcaaaa Tggccagttattaatg	NTpsbE-8G→A template
NTpsbE-9ARev	AGGTCGACGGTATCtgttccaaagGatcaaaac Tgccagttattaatgg	NTpsbE-9C→A template
NTpsbE-9URev	AGGTCGACGGTATCtgttccaaagGatcaaaac Agccagttattaatgg	NTpsbE-9C→U template

Table 3.1. (Continued)

NTpsbE-10ARev	AGGTCGACGGTATCtgttccaaagGatcaaaacgTccagttattaatgga	NTpsbE-10C→A template
NTpsbE-10URev	AGGTCGACGGTATCtgttccaaagGatcaaaacgAccagttattaatgga	NTpsbE-10C→U template
NTpsbE-11ARev	AGGTCGACGGTATCtgttccaaagGatcaaaacggTcagttattaatggaa	NTpsbE-11G→A template
NTpsbE-11URev	AGGTCGACGGTATCtgttccaaagGatcaaaacggAcagttattaatggaa	NTpsbE-11G→U template
NTpsbECompFor	GATACCGTCGACCTCGACAGCACCGGTTTAGCTTACG	NTpsbE_KS-SK template
NTpsbECompRev	GATCCACTAGTTCTAGAGCGTGTTCCAAAGGATCAAAACG	NTpsbE_KS-SK template

RESULTS

The sequence 13 nucleotides upstream and 10 nucleotides downstream of the tobacco *psbE* C214 is critical for editing.

The 5′ sequence requirements of Arabidopsis *psbE* had been examined using Arabidopsis extracts and 6 substrates we constructed to carry different lengths of 5′ sequence (Hegeman et al. 2005). Tobacco and Arabidopsis *psbE* sequences are very similar, with only 7 nucleotide differences between tobacco and Arabidopsis *psbE* in the -150/+15 region around the editing site. Tobacco extracts are known to efficiently edit Arabidopsis *psbE* editing substrates (Hegeman et al. 2005). Because we already had produced 6 Arabidopsis *psbE* substrates, we assayed their editing efficiencies in tobacco chloroplast extracts to obtain an initial indication of the 5' sequence essential for editing (Fig. 1).

All 6 substrates exhibited some level of editing, including templates with only 13 nucleotides 5′ of the edited nucleotide (Figure 3.1). The substrates containing only 23 or 13 nucleotides 5′ of the edited nucleotide were significantly less edited than the longer substrates. Because the -99/+15 substrate was edited most efficiently, two

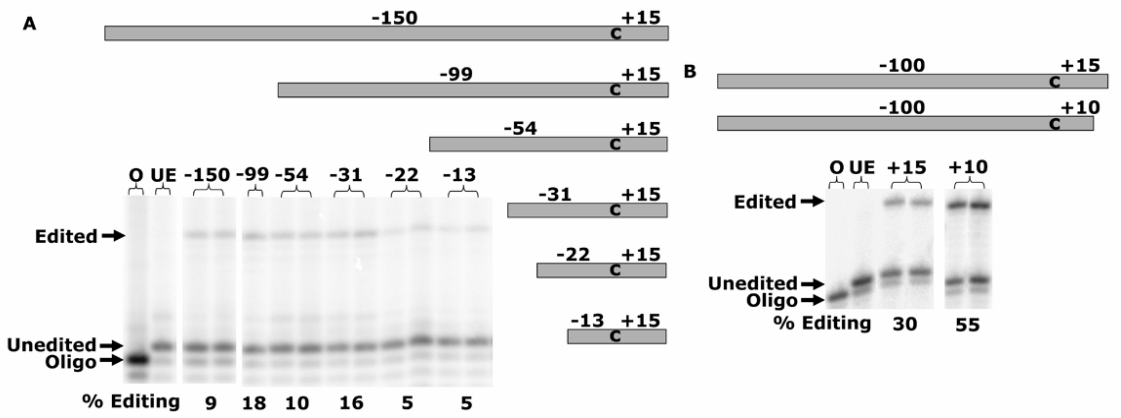


Figure 3.1. Editing in RNA substrates with various lengths of 5′ (A) and 3′ (B) sequence. Gray boxes represent the substrates containing *psbE* sequence from (**A**) Arabidopsis and (**B**) tobacco with the number of nucleotides 5′ and 3′ indicated. Electrophoretograms of the poisoned primer extension (PPE) reactions used to assay editing efficiencies are shown below the substrate diagrams, with the % editing calculated as described previously (Peeters and Hanson, 2002). Lane O, PPE without template, indicating the size of the labeled oligo; Lane UE, PPE of substrate incubated without competent editing extract. All other lanes are labeled according to the length of sequence (A) 5' and (B) 3'.

additional substrates were produced, each carrying 100 nucleotides of tobacco *psbE* sequence 5′ to the editing site (Figure 3.1). The shorter substrate (-100/+10) was edited more highly than the longer one (-100/+15), though both substrates are edited. Ten nucleotides 3′ of NTpsbE C214 are evidently sufficient for efficient editing *in vitro*.

Individual nucleotides 5′ and 3′ of the edited nucleotide are important for editing

Our initial experiments with substrates of different lengths stimulated us to focus more closely on the importance of specific nucleotides in the sequence from -20 to +10 surrounding the C214 editing target. Thirty substrates were constructed, each with one nucleotide substituted with the complementary nucleotide in the -20/+10 region. In order to easily observe the effect of mutations, we analyzed substrates containing the -100/+10 region around the edited nucleotide, due to the high editing efficiency of wild-type substrates carrying the -100/+10 sequence (Figure 3.1).

Substrates with changes in either 5′ or 3′ nucleotides were reduced in editing compared to substrates with wild-type sequences (Figure 3.2). Alterations in the sequence from -11/-2, +2/+4, and +8/+9 around C214 were all harmful to editing. However, changes within the -11/-7 region were the most inimical to editing *in vitro*. Alteration of the 5′ adjacent nucleotide to the C target from a T to an A did not significantly reduce the ability for the substrate to be edited, in agreement with prior observations using a different RNA substrate *in vitro* (Miyamoto et al. 2004). However, Miyamoto et al. (2004) found that RNA substrates with a mutation at the 5′ adjacent nucleotide from T to G or T to C exhibited greatly decreased editing efficiency *in vitro*. Therefore, we can conclude that the sequence from -11/-1 to NTpsbE C214, as well as sequences from +2/+4 and +8/+9, are critical for editing at this site.

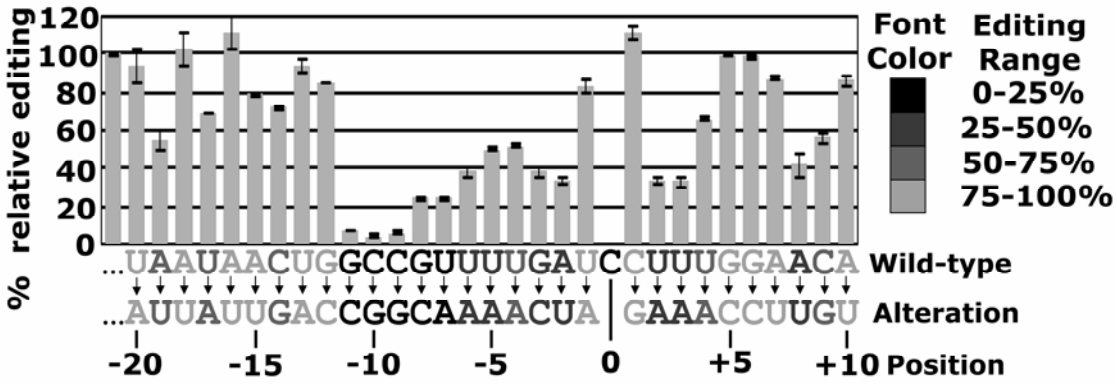


Figure 3.2. Effect of mutations in nucleotides surrounding NTpsbE C214 on editing efficiency. The gray bars represent percent editing relative to the wild-type substrate of various substrates that contain one nucleotide altered from wild-type tobacco sequence. The wild-type sequence at each position around the edited nucleotide is listed below each bar and the alteration contained in the mutant substrate is listed below the wild-type nucleotide. Letters representing the nucleotides are colored from light to dark gray, with darker color indicating those nucleotides that exhibit more detrimental effects on editing than nucleotides shown in lighter color.

Sequences in common between NTpsbE C214 and editing sites in *rpoB* and *ndhB* are not critical for editing of NTpsbE C214

The sequence 5′ to NTpsbE C214 exhibits some identity to sequences 5′ to NTndhB C467 and NTrpoB C338 (Hegeman et al. 2005; Figure 3.3). Editing of sites exhibiting some 5′ sequence identity has previously been found to be affected when one of the "cluster" of sites is overexpressed *in vivo*. For example, overexpressing NTrpoB C473 *in vivo* results in decreased editing of NTpsbL C2 and NTrps14 C80, presumably due to competition for a common factor required for editing of these sites (Chateigner-Boutin and Hanson, 2002). Because no chloroplast transgenic plant overexpressing NTpsbE C214 is available, we did not know whether the sequence similarity between the *psbE* site and NTndhB C467 and NTrpoB C338 was merely fortuitous or actually is significant to editing.

Inspection of editing of the mutated substrates shown in Fig. 2 reveals that some of the 5′ nucleotides in common between the three sites are not important for editing of NTpsbE C214, namely the TAATAAC sequence. To investigate further whether these three editing sites might share factors that interact with the common nucleotides, we decided to carry out competition experiments *in vitro* to test whether excess amounts of transcripts carrying either NTrpoB C338 or NTndhB C467 would be inimical to NTpsbE C214 editing, implying limiting amounts of a shared factor.

In order to carry out competition experiments that mimic the parameters of editing *in vitro*, we investigated how much NTpsbE C214 substrate we could add to the reaction *in vitro* before the percent of edited molecules decreased (Figure 3.4). At 10 fmol of RNA editing substrate, the editing percentage *in vitro* was equivalent to that observed at 1 fmol and 0.1 fmol. However, the amount of editing was greatly reduced at 100 fmol and undetectable at 1 pmol.

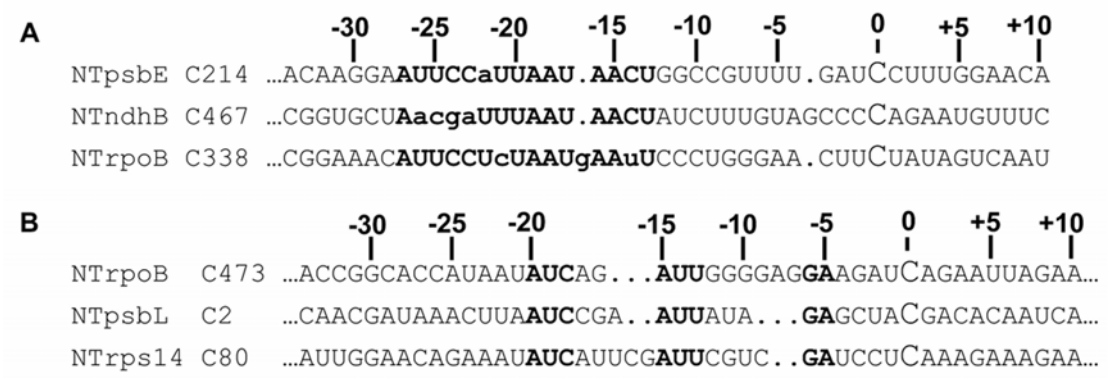


Figure 3.3. Similarity of 5′ regions of clusters of tobacco editing sites. (**A**) Alignment of editing sites with common sequences with NTpsbE C214. (B) Sequences common in 5′ region of three editing sites known to be affected when RpoB C473 is overexpressed (Chateigner-Boutin and Hanson, 2002). Bold characters indicate common sequences and the edited nucleotides are represented in a larger font.

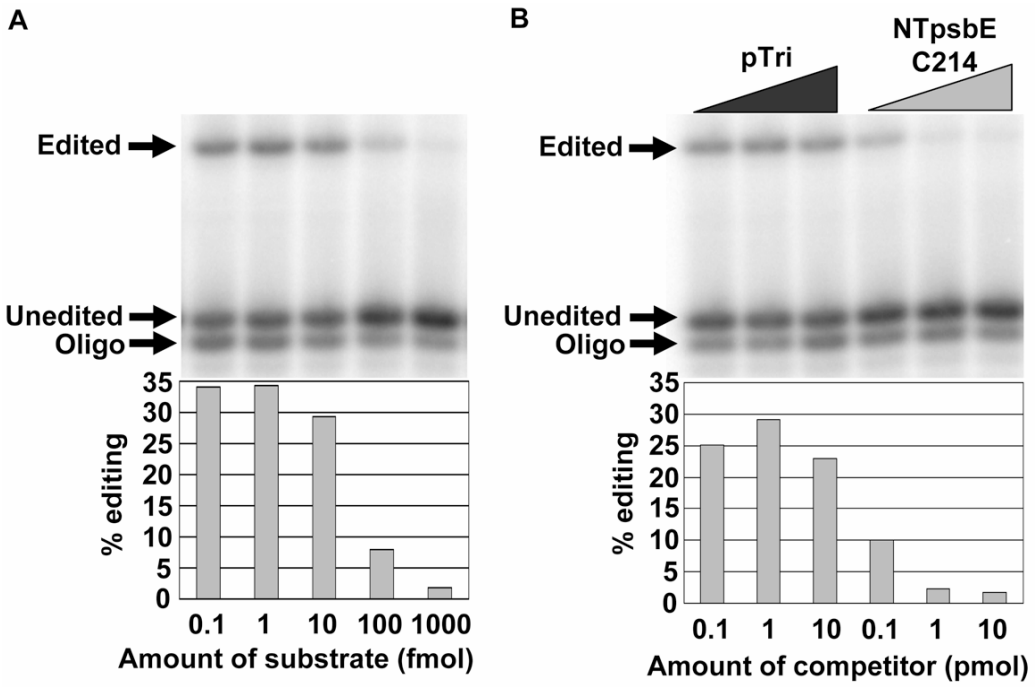


Figure 3.4. Increasing amounts of substrate (A) or self-competitor RNA (B) reduces the percentage of substrate edited *in vitro*. Competitor RNAs were added at 0.1, 1, and 10 pmol amounts to reactions in standard *in vitro* conditions with 10 fmol of editing substrate. Competitor RNA amounts correspond to 10X, 100X, and 1000X of the editing substrate respectively. pTri RNA is a non-specific competitor transcribed from the control plasmid from the T7 MegaShortscript kit (Ambion). NTpsbE C214 self-competitor contains the same -100/+10 region as the substrate, but SK and KS flanking sequences have been swapped to prevent amplification by RT-PCR.

We then determined the amount of additional *psbE* RNA needed to reduce editing activity when the initial concentration of editing substrate is 10 fmol (Figure 3.4). When either 100 fmol or 1 pmol of *psbE* RNA are added, the editing *in vitro* of the 10 fmol of input RNA is reduced. In other words, limitations on editing efficiency become evident when 10 times more substrate is present than can be handled by the editing apparatus present in the chloroplast extracts. Thus, if another editing site can compete for the same factors needed for *psbE* editing, we would expect that adding 10 or 100 times more of the competitor RNA should result in a decrease in *psbE* editing.

We created competitor RNAs containing -100/+10 regions of sequence around the editing sites NTrpoB C338 and NTndhB C467. Addition of these RNAs in amounts up to 1 pmol had no effect on editing of 10 fmol of NTpsbE C214 substrate (Figure 3.5). The lack of competitor effect suggests that the common region of sequence is not a critical editing cis-element recognized by a limiting editing factor *in vitro*.

Nucleotide preference for editing at key positions around the editing site

Both the competitor experiments (Figure 3.5) and the mutated substrate experiments (Figure 3.2) suggested that the 5′ and 3′ nucleotides of most importance around NTpsbE C214 are the 13 nucleotides located at -11/-1 , +2, +3, and +8. We therefore more intensively examined these nucleotide positions by making substrates in which each of the nucleotide positions was altered to carry each of the three other possible nucleotides. Editing efficiency *in vitro* of these 39 different substrates was assayed and then calculated relative to a wild-type substrate (Figure 3.6, Table 3.2). The editing apparatus at nucleotides positions -11 to -6 displayed a strong preference for the wild-type sequence GCCGUU; mutation to any other base reduced the ability for that substrate to be edited (Figure 3.6, Table 3.2). At positions -5, -4, and +3 a strong preference for pyrimidines is evident. At the -2 position a C nucleotide is

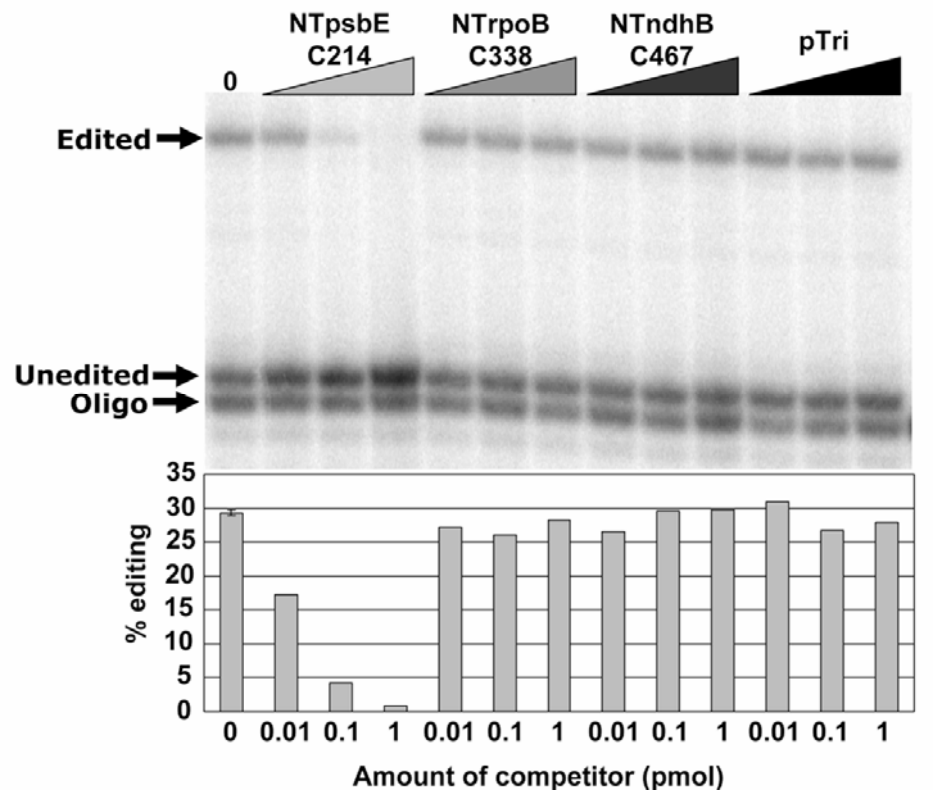


Figure 3.5. Effect of the presence of transcripts carrying NTndhB C467 and NTrpoB C338 on editing of NTpsbE C214 *in vitro*. Competitor RNAs were added to reactions in standard *in vitro* conditions at amounts of 0.01, 0.1, and 1 pmol, corresponding to 1X, 10X, and 100X of editing substrate respectively.

Figure 3.6. The nucleotide preference for the tobacco editing apparatus at important positions around the editing site. The editing percentages of RNA substrates with one nucleotide mutated to each of the three nucleotides at each important position were calculated relative to the nucleotide present in wild-type substrates containing tobacco sequences. The amount of editing in mutant substrates was then compared through generation of a sequence logo using the Weblogo program (Crooks et al. 2004). The size of the characters is based on the relative editing of each substrate. Larger characters represent nucleotides that are preferred at positions around the editing site for more efficient editing.

Table 3.2. Editing in RNA substrates with one nucleotide mutated to each of the three alternate bases at important positions around the editing site. The editing percentage relative to wild-type was calculated in duplicate reactions for each substrate with one standard deviation around the mean.

Mutation	% Relative Editing	Std dev.	Mutation	% Relative Editing	Std dev.	Mutation	% Relative Editing	Std dev.
-11C	7	0.4	-11A	48	0.9	-11U	9	0.7
-10G	4	0.4	-10A	4	0.3	-10U	12	0.6
-9G	6	0.5	-9A	27	0.8	-9U	33	2.3
-8C	24	0.3	-8A	27	0.6	-8U	20	0.5
-7A	24	0.4	-7G	1	0.1	-7C	1	0.1
-6A	38	3.3	-6C	29	0.1	-6G	12	1.4
-5A	50	0.6	-5G	6	1.5	-5C	64	0.4
-4A	52	1.7	-4G	3	0.2	-4C	86	0.2
-3C	38	2.0	-3A	62	0.5	-3U	47	2.1
-2U	33	1.8	-2C	145	3.7	-2G	10	4.2
+2A	33	2.0	+2C	62	0.03	+2G	60	0.2
+3A	33	2.9	+3C	50	1.4	+3G	20	0.4
+8C	1	0.1	+8G	63	1.2	+8U	42	5.5

preferred over the native A nucleotide. Transcripts with nucleotide changes to G residues in the regions -10/-9, -7/-4, and -2/-1 are extremely poor editing substrates. The editing apparatus appears to tolerate G residues in the positions 5′ of the editing site where there are G residues in the wild-type sequence.

The edited nucleotide is determined by a molecular ruler from a 5′ element

Although the editing site in *psbE* is flanked by a 3′ C, editing has been detected only at C214. This specificity could either be due to an editing mechanism that is processive, that edits the first C it encounters from a cis-acting element, or that utilizes a "molecular ruler", where the editing site is a precise distance from a cis-element and the edited C. Possibly the critical sequences detected 5′ of the editing site in *psbE* (Figure 3.7) could act as the "anchor" for a molecular ruler.

We created 7 substrates to test the effects of altering the local spacing around the editing site. We deleted nucleotides 5′ of the editing site and inserted nucleotides 3′ to determine whether the native editing site C214 was invariably edited in the substrates *in vitro* or whether editing could occur at the 3′ nucleotide C215 (Figure 3.7).

There was no editing at the native editing site C214 in substrates with nucleotides deleted immediately 5′ of the editing site. Editing was not significantly reduced when a U was inserted 3′ of C215. Editing was moderately reduced when an A was inserted at position +2. The A insertion influenced the spacing of the minor 3′ element UU at +2/+3, but UU at +2/+3 is maintained in the substrate with the U insertion. Shifting nucleotides one nucleotide 3′ by insertion of a U at +2 had no affect on editing, and therefore the location of the +8/+9 nucleotides, the identity of which is important in editing efficiency (Figure 3.2), can be shifted without affecting editing. Substrates -4Δ, -3Δ, -2Δ, and -1Δ were not edited *in vitro* at C214.

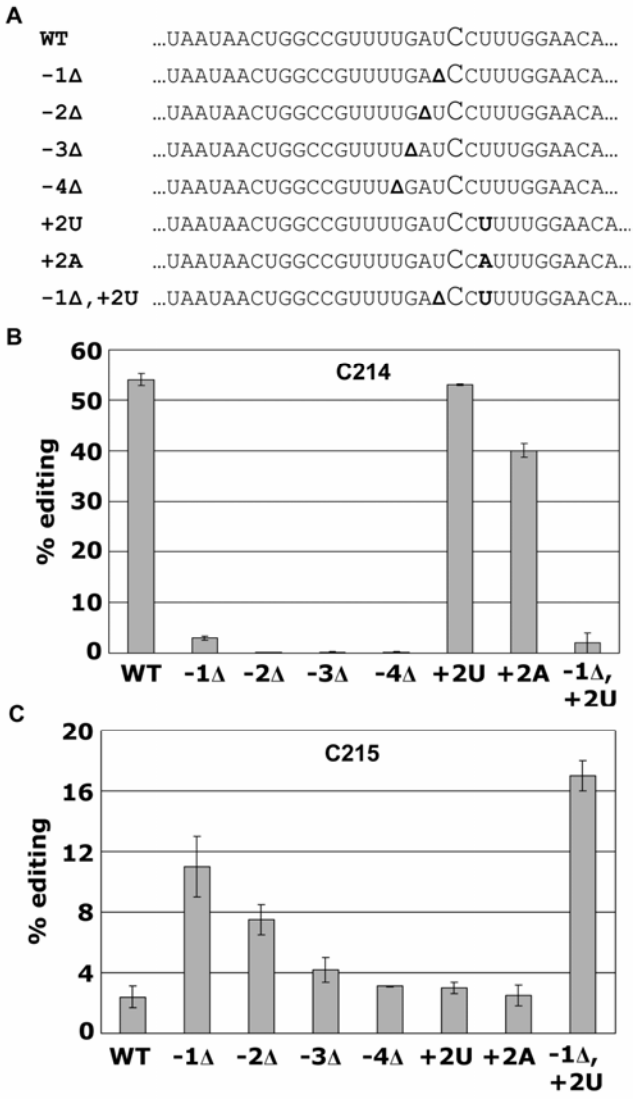


Figure 3.7. Editing in C214 and C215 in substrates with deletions or insertions compared to wild-type tobacco sequence. (**A**) Alignment of mutant substrates constructed to examine spacing constraints for editing. Bolded characters represent deleted or inserted nucleotides (**B**) Gray bars depict the percent editing of mutant substrates at the normal editing site C214, and (**C**) the 3′ adjacent nucleotide C215. (**B, C**) Error bars are calculated from duplicate reactions as one standard deviation from the mean.

Substrates -1Δ, -2Δ, and -1Δ, +2U all exhibited significant editing at the 3′ adjacent nucleotide C215. Therefore, the editing site is determined by the distance of a C nucleotide from the 5′ cis-element. The amount of editing at C215 is much lower than the wild-type editing at C214. This is likely to be due to the presence of a C immediately 5′ of C215, while the presence of A or U at -1 result in higher editing at C214 than a 5′ C (Miyamoto et al. 2004). Insertion of nucleotides 3′ of the editing site did not lead to a shift in the C target; therefore, the sequence 3′ of the editing site is not an important determinant for the location of editing. The -1Δ, +2U substrate was more efficiently edited at C215 than the substrate carrying only the -1Δ, suggesting a preference for greater distance between the C target and a downstream element.

DISCUSSION

The minimal size of surrounding sequence sufficient for editing of particular Cs has not been characterized for the vast majority of the 38 known tobacco editing sites in chloroplasts. The known minimal regions of surrounding sequence for editing *in vivo* are in close proximity to the editing sites; -16/+5 in NTpsbL C2, -12/+11 in NTndhB C683, -21/+2 in NTndhB C692, -20/+6 in NTrpoB C473. The smallest region identified to be sufficient for correct selection of a C target *in vitro* is the -13/+10 region around NTpsbE C214.

We observed that the editing efficiency of smaller *in vitro* substrates is much lower compared with larger substrates with more native sequence around the editing site. The reduced editing efficiency in smaller editing substrates may preclude determination of the minimal region for editing for many of the uncharacterized editing sites. The lower editing efficiency in smaller substrates may result from the absence of auxiliary sequences that enhance editing *in vitro*, or a reduction in the affinity of the editing components for small RNAs.

By determining the nucleotide preference for editing at important positions, we have identified a 6-nucleotide sequence, GCCGUU, that is critical for editing of *psbE* transcripts *in vitro*. A 56 kDa protein has been found to undergo specific UV-crosslinking to *psbE* transcripts (Miyamoto et al. 2004). *PsbE* transcripts with mutations in the -15/-6 region did not compete for binding of the 56kDa protein to a wild-type *psbE* transcript, indicating that the -15/-6 region contains some critical sequences for factor binding. It seems likely that the actual region important for binding of this factor is at -11/-6, where the GCCGUU motif we have detected is located. Our studies have correlated a region critical for efficient editing with a region known to be important for binding of a protein.

The GCCGUU motif is found near the *psbE* editing site but not proximal to other known C targets in chloroplasts. The GCCGTTnnnnnC sequence is present only once within the known tobacco chloroplast protein-coding sequences. Therefore, the six nucleotide motif may be sufficient to uniquely identify one cytidine within all chloroplast coding sequences. If a particular factor binds to this motif, then its role in editing is likely to be specific to *psbE* rather than to multiple editing sites. This suggests that some editing sites are independently regulated by one specificity factor. A single factor also appears important in editing of NTndhD C2, because mutation in a single gene encoding a putative RNA-binding protein results in abolition of editing only at NTndhD C2 (Kotera et al. 2005). However, the reduction of editing observed at multiple sites following overexpression transcripts encoded by a chloroplast transgene carrying a similar editing site suggests that there also are groups of sites that share cis-elements and factors (Chateigner-Boutin and Hanson, 2002; Hayes et al. 2006). Perhaps some chloroplast editing sites are regulated individually and other editing sites are regulated by shared factors.

At present our study is the only analysis of nucleotide preference surrounding a chloroplast editing site where substrates were tested in which the wild-type nucleotide was mutated to each of the other three nucleotides. This thorough analysis of nucleotide preference for editing reveals that sequence requirements are more complex than would be detected by merely testing substrates in which the complementary nucleotide has replaced the native nucleotide. Although the 6 nucleotide motif GCCGUU is critical for editing, other nucleotide preferences are less nucleotide-specific. At positions -5, -4, and +3 there is a preference for pyrimidines, and at -2, a C or A will suffice. The increase in editing in a substrate with a C at -2 instead of the native A may be due to a general preference for the dipyrimidines CT at the -2/-1 positions by the editing mechanism. Although the dipyrimidine is probably favored, mutation of an A at -2 to a C maybe constrained due to the requirement to maintain an Asp codon. G nucleotides at many positions reduce editing efficiency; unexpectedly, a C at +8 completely abolishes editing. We have not been able to create models involving secondary structures that might explain these sequence preferences and requirements.

We have determined that the edited nucleotide in *psbE* is a precise distance from the discovered 5′ cis-element. Therefore a "molecular ruler" is a more likely mechanism for the observed specificity of selection of C target than a processive enzyme. Because NTndhB C692 also appears to be sensitive to distance (Hermann and Bock, 1999), a measuring mechanism may be a feature common to all chloroplast editing sites. However, the impact of 5′ and 3′ nucleotide insertions or deletions has only been extensively studied for two sites in chloroplasts. More editing sites will have to be analyzed to make global conclusions about mechanisms for target selection.

ACKNOWLEDGMENTS

This work was supported by NSF grant MCB 0344007 to M.R.H and an NIH graduate traineeship to M.L.H.

REFERENCES

Bock R, Hermann M, Fuchs M. 1997. Identification of critical nucleotide positions for plastid RNA editing site recognition. *RNA 3*:1194-1200.

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

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

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

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**:1188-1190.

Freyer R, Kiefer-Meyer MC, Kossel H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A* **94**:6285-6290.

Hayes ML, Reed M. L., Hegeman C. E., 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**:3750-3762.

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

Hermann M, Bock R. 1999. Transfer of plastid RNA-editing activity to novel sites suggests a critical role for spacing in editing-site recognition. *Proc Natl Acad Sci U S A* **96**:4856-4861.

Hirose T, Sugiura M. 2001. Involvement of a site-specific trans-acting factor and a

common RNA- binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *EMBO J* **20**:1144-1152.

Hoch B, Maier RM, Appel K, Igloi GL, Kossel H. 1991. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* **353**:178-180.

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

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

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

Maier RM, Neckermann K, Igloi GL, Kossel H. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* **251**:614-628.

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

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

Sasaki T, Yukawa Y, Wakasugi T, Yamada K, 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**:802-10.

Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM. 2002. The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol* **19**:1602-1612.

Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM. 2005. Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J* **43**:708-715.

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

Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kossel H, Sugiura M. 1996. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. Proc Natl Acad Sci U S A 93:8766-8770.

CHAPTER 4

∗ASSAY OF EDITING OF EXOGENOUS RNAS IN CHLOROPLAST EXTRACTS OF ARABIDOPSIS, MAIZE, PEA, AND TOBACCO

ABSTRACT

Nucleotides within transcripts of chloroplasts and mitochondria are modified through C-to-U RNA editing in vascular plants. The specific protein components and enzymatic machinery required for editing have not been defined. A consensus sequence is not present around all editing sites, complicating the discovery of cis-sequence elements critical for editing. Chloroplast extracts capable of carrying out editing *in vitro* along with precise quantification of editing extent of exogenous transcripts will facilitate identification of both cis- and trans-factors. We have optimized an *in vitro* assay originally developed to study editing in tobacco and pea chloroplasts and have expanded the assay to include the study of chloroplast editing in the model species Arabidopsis and the monocot maize. The superior genetic resources in these two species can now be utilized in conjunction with biochemical analysis to dissect the editing apparatus. We have improved the assay conditions for editing *in vitro,* achieving efficient editing (as much as 92%) with certain RNA substrates. Unlike the initial assay that relied on qualitative analysis, we are able to achieve precise quantification of editing activity within 1% through a simple poisoned primer extension (PPE) assay with radiolabeled oligonucleotides.

∗ Hayes ML and Hanson MR (2007) Methods of Enzymology 424: 459-482, ©2007 by Elsevier.

INTRODUCTION

Vascular plant organellar transcripts undergo C-to-U RNA editing. Although editing occurs in both mitochondria and chloroplasts, this chapter will focus on editing in chloroplasts. The editing of transcripts has been shown to be indispensable for the synthesis of several functional chloroplast proteins (Bock et al., 1994; Zito et al., 1997; Sasaki et al., 2001; Schmitz-Linneweber et al., 2005). The critical components of the molecular apparatus required for organellar RNA editing in plant systems are poorly understood. The *crr4* gene in Arabidopsis is essential for the editing of one site in *ndhD* transcripts (Kotera et al., 2005), and the encoded chloroplast-targeted protein CRR4 is known to bind artificial RNAs containing *ndhD* sequences (Okuda et al., 2006). Unfortunately, neither the biochemical function of this protein nor any other editing factor has been deciphered. Chloroplast extracts competent to edit exogenously synthesized RNA substrates, along with a sensitive assay of editing extent, are critical for dissecting the plant editing apparatus.

Alignment of the sequences around all the editing sites in one species has not revealed a common sequence crucial for editing. Sequence requirements for each editing site therefore must be examined on an individual site-by-site basis. Substrates have been created that are capable of being edited through the expression of exogenous trangenes in transplastomic plants (Bock et al., 1994; Chaudhuri et al., 1995; Reed & Hanson, 1997). The creation of transplastomic plants for extensive study of specific sequence requirements is impractical, given the number of plants required to investigate thoroughly even one chloroplast editing site in higher plants, along with the low efficiency and species' constraints of current chloroplast transformation technology. Chloroplast extracts and RNA substrates that can be edited *in vitro* and assayed with precise quantification can more rapidly identify the critical cis-elements for editing at a particular site. However, a present drawback is that not

all chloroplast editing sites are represented by artificial RNA substrates that can be edited *in vitro*.

The first chloroplast *in vitro* editing system was developed by Hirose & Sugiura (2001) for tobacco (*Nicotiana tabacum*) editing sites NTpsbL C2[1] and NTndh C737. Editing analysis was determined qualitatively by inspection of autoradiograms of thin layer chromatographs of a C nucleotide labeled at the editing site that was converted to labeled U by RNA editing. Chloroplast transcript editing *in vitro* was extended by Miyamoto et al. (2002) for NTpsbE C214 and NTpetB C611 and for pea (*Pisum sativum*) extracts. A major shortcoming of the method was that small changes in a substrates' editing efficiency could not be detected due to the qualitative nature of the assay. The extracts, RNA substrates, and assay were developed for tobacco and pea, species from which high-quality chloroplasts are readily isolated. However, these species do not have the extensive genetic resources that are available for certain other plants.

Quantification of editing *in vivo* has been accomplished by direct sequencing, bulk sequencing and poisoned primer extension (PPE) assay (Bock et al., 1994; Reed et al., 2001; Peeters & Hanson, 2002). We developed an *in vitro* editing system based on tobacco chloroplast extracts as described by Hirose and Sugiura (2001), but we utilized RNA synthesized *in vitro* and we assayed editing extent of exogenously added substrates with the PPE method for the quantification of editing (Hegeman et al., 2005b). Incubation conditions for editing *in vitro* with pea extracts and substrates made in a similar manner were optimized by Nakajima & Mulligan (2005). We further extended the number of species studied by RNA editing to include the model plant

[1] The nomenclature used in this manuscript was established in Hayes et al. (2006). A two-letter abbreviation denotes the species that contains the editing site; the gene that contains the editing site is listed; the identity of the nucleotide modified is indicated; and the position in nucleotides from the A of the initiation codon is shown.

Arabidopsis thaliana and the monocot maize (Hegeman et al., 2005b; Hayes et al., 2006). These *in vitro* editing systems have proven useful for the determination of chloroplast editing site sequence requirements (Hayes et al., 2006; Hayes & Hanson, 2007).

In this article, we describe the methods used for analysis of editing *in vitro* for 4 distantly related plant species. We investigated the sensitivity, accuracy, and precision of the PPE assay used for quantification. We provide detailed protocols for the isolation of chloroplasts, the creation of editing competent chloroplast extracts, the synthesis of RNA editing substrates, the conditions for the *in vitro* assay, the quantification of editing, and the analysis of competing RNAs.

GROWTH CONDITIONS FOR THE ISOLATION OF INTACT CHLOROPLASTS FROM TOBACCO, PEA, MAIZE AND ARABIDOPSIS

Overview

Efficient isolation of intact plastids is critical for the *in vitro* editing system because an extract containing highly concentrated chloroplast proteins (5-20 μg/μL) is required. To obtain a high yield of intact chloroplasts, the quality of the leaf tissue is of utmost importance. Typically the best yields are from plants that are young and growing vigorously. The growth conditions for each plant species are listed below.

Plant Growth Protocol

Maize *(Zea mays)*

1. Imbibe maize seeds of the varieties Seneca Horizon (Seedway, Hall, NY), Jubilee (Seedway), or B73 overnight before planting for more uniform germination.

2. Sow seeds to a depth of approximately 2.5 cm, 2.5 cm apart in 54x28 cm flats containing Metro Mix 360 soil (Sungro Horticulture, Bellevue, WA). Three flats are sufficient per chloroplast preparation.

3. Grow maize seedlings at 25ºC under metal halide lights with a 16 hour light and 8 hour dark cycle for 10-14 days. The light intensity is approximately 250 W m^{-2}. Fertilize plants daily using a continuous liquid feed of 100 PPM nitrogen, 24 PPM phosphorous, 95 PPM potassium (21-5-20, J.R. Peters, Inc., Allentown, PA.), and 0.04 oz./gal Epsom salts for 5 days. Alternate this treatment with watering plants daily for 2 days using clear water.

4. Harvest leaf tissue when plants have three leaves and a fourth leaf is emerging. Homogenize around 80-100 g of leaf tissue per 400 ml of Buffer B. Starch is relatively less abundant in tissues grown in this manner compared to leaves of other plants that we have used to isolate chloroplasts. Therefore, we found that it is not necessary to cover plants with aluminum foil prior to harvest.

Arabidopsis thaliana

1. Imbibe seeds of *A. thaliana*, ecotype Columbia, for 48 hours at 4ºC.

2. Sow seeds in Metro Mix 360 soil in 54x28 cm flats.

3. Thin plants to about one plant every 7.6 cm. Three flats are planted per chloroplast preparation.

4. Grow Arabidopsis plants for 4-6 weeks in a growth chamber with 10 hours light at 21ºC and 14 hours dark at 16ºC as described in Hegeman et al. (2005b) and (Peltier et al., 2001). The light intensity is approximately 100 µmol photons m^{-2} sec^{-1}. Growth conditions are particularly important for obtaining good yields of intact chloroplasts from *A. thaliana.*

5. Harvest leaves and float them in an ice cold water bath for 30 minutes as suggested by Kunst (1998) before mincing in Buffer B. Homogenize around 50 g of leaves per 400 ml of Buffer B.

Tobacco (*Nicotiana tabacum*)

1. Sow tobacco seeds of the varieties Petit Havana or Samsun NN onto Metro

Mix 360 soil in 54x28 cm flats with 4x9 cell dividers. Three flats are planted per chloroplast preparation. Cover flats with flexible plastic sheeting such as Saran Wrap after planting for one week to promote uniform germination. Grow the tobacco plants at 25ºC under metal halide lights with a 16 hour light and 8 hour dark cycle. The light intensity is approximately 250 W m^{-2}.

2. After germination, remove the plastic and thin the plants to one seedling per cell. Grow the plants for 4-6 weeks. Fertilize and water plants as described for maize.

3. Cover the flats with aluminum foil 48 hours prior to harvest to reduce starch accumulation. Harvest around 80-100 g of leaves per 400 ml of Buffer B.

Pea (*Pisum sativum*)

1. Sow pea seeds of the variety Laxtons Progress (Page Seed Company, Greene, NY) about 2.5 cm deep and 7.6 cm apart in Metro Mix 360 soil in 54x28cm flats. Three flats are planted per chloroplast preparation.

2. Grow plants at 25ºC under metal halide lights with 16 hour light and 8 hour dark cycle for 4-6 weeks. The light intensity is approximately 250 W m^{-2}. Fertilize and water plants as described for maize.

3. Cover plants 48 hours prior to harvest plants with aluminum foil to reduce the accumulation of starch in leaves. Harvest around 80-100 g of leaf tissue per 400 ml of Buffer B.

ISOLATION OF CHLOROPLASTS AND PRODUCTION OF EDITING COMPETENT EXTRACTS

Overview

All published methods for isolation of editing competent chloroplast extracts utilize buffers based on the ones used in the first reported *in vitro* assay (Hirose & Sugiura, 2001). For convenience, we describe these in Table 4.1. We find that the

Table 4.1 Buffers Used for the isolation of editing competent chloroplast extracts

Solution	Composition
Buffer A	0.6 M mannitol (Fisher Scientific, Pittsburgh, PA) 100 mM HEPES-NaOH pH 8.0 (Fisher Scientific) 4 mM EDTA (Sigma-Aldrich, St. Louis, MO)
Buffer B[a]	0.3 M mannitol 50 mM HEPES-NaOH pH 8.0 2 mM EDTA 15 µM BSA (EMD Chemicals Inc, Gibbstown NJ) 15 µM PVP-40 (Sigma-Aldrich) 10.6 mM BME (Sigma-Aldrich)
Buffer C[a]	0.3 M mannitol 50 mM HEPES-NaOH pH 8.0 2 mM EDTA 15 µM BSA 10.7 mM BME
Buffer D[a]	0.315M mannitol 50mM HEPES-NaOH pH 8.0 2 mM EDTA
Buffer E	30 mM HEPES-KOH pH 7.7 10 mM MgOAc (J.T. Baker, Phillipsburg, NJ) 0.2% Triton-X (Sigma-Aldrich,) 2M KCl (Sigma-Aldrich) 2 mM DTT (Sigma-Aldrich)
Buffer F	30 mM HEPES-KOH pH 7.7 45 mM KOAc (Fisher Scientific) 30 mM NH_4OAc (Fisher Scientific) 10% glycerol (Fisher Scientific) 2 mM DTT

[a]For Arabidopsis chloroplast isolation, the EDTA concentration in these solutions is increased to 4 mM.

method for isolating tobacco chloroplasts as described by Hirose and Sugiura (2001) is suitable for the isolation of chloroplasts in Arabidopsis and maize with only minor modifications. Higher yields of intact Arabidopsis chloroplasts can be obtained using a greater concentration of EDTA as suggested by Somerville (1981) in Buffer B, C and

D, compared to the original isolation conditions by Hirose and Sugiura (2001). All other plants are prepared with the same buffers as listed in Hirose and Sugiura (2001). The procedure used for all four species is similar, with differences primarily based on the growing conditions listed above. The procedure for isolation of plastids is listed below.

Chloroplast Isolation and Extraction Protocol

Note: All solutions should be chilled on ice before use and steps should be performed as quickly as possible to alleviate premature lysis of chloroplasts and protein degradation.

1. Create the Percoll gradient solution by thoroughly mixing Percoll (Sigma-Aldrich), an equal volume of Buffer A, and to a final concentration of 10.5 mM β-mercaptoethanol (BME, Sigma-Aldrich). Add 30 mL of gradient solution to each of four clear 50 mL centrifuge tubes. The continuous gradient is formed by centrifugation of the Percoll solution in tubes at 43,140xg (Sorvall SS-34 rotor) for 30 minutes at 4ºC with the brake off.

2. Remove whole leaves using scissors and rinse them in cold water to remove any loose soil. Cover leaves with 400 ml of ice cold Buffer B and chop them into fine pieces using razor blades. Leaves are sufficiently chopped when pieces are smaller than 1 cm^2 and the Buffer B turns green in color. Add the minced leaves and Buffer B to a 500 ml bottle chilled on ice. Homogenize the leaf tissue with a Polytron (Brinkmann, Westbury, NY) at a speed between 4 and 5 until intact leaf pieces are no longer observed and the liquid is a thick soup. Strain the homogenate through 4 layers of cheesecloth into a 500 ml flask chilled on ice.

3. Centrifuge the strained homogenate until the rotor reaches 5858xg (Sorvall SGA rotor), then immediately stop the centrifugation and allow the rotor to come to rest with the brake on at 4ºC. The chloroplasts are now present in a dark green pellet at

the bottom of the tube. Discard supernatant. Gently resuspend the pelleted chloroplasts in Buffer C using a paintbrush. Layer the chloroplasts onto the prepared 50 % Percoll continuous gradients and centrifuge them at 8084xg (Sorvall HB-4 rotor) for 15 minutes at 4ºC with the brake off. The intact chloroplasts segregate as a lower dark green band while chloroplasts with ruptured membranes will remain near the top of the gradient. Starch accumulates as a white pellet at the bottom of the tube. Remove the lower dark green band carefully using a Pasteur pipette without disturbing the upper band and starch pellet. Dilute the chloroplasts with four volumes of Buffer D. Pellet the chloroplasts at 4124xg (Sorvall HB-4 rotor) for 1 minute at 4ºC with the brake on. Dispose of the supernatant. Resuspended the pellet in 30 ml Buffer D and centrifuge at 4124xg (Sorvall HB-4 rotor) for 40 seconds at 4ºC with the brake on. Remove supernatant.

4. Resuspend the chloroplast pellets in a minimal volume of Buffer E and incubate them on ice for 30 minutes for lysis. Centrifuge the lysate at 15,600xg in an Eppendorf 5414 microfuge for 10 minutes at 4ºC. Remove the supernatant carefully without disturbing the dark green membrane pellet to a Tube-O-Dialyzer (G-Biosciences, St. Louis, MO) 8000 MCO and dialyze in 200 ml of Buffer F for 1.5 hours. Add the Tube-O-Dialyzer containing the lysate to fresh Buffer F and dialyze for an additional 3.5 hours. Divide the extracts into aliquots, flash freeze them in liquid N_2, and store them at -80ºC. We have stored some extracts for 24 months without loss of activity.

EDITING OF EXOGENOUS RNA *IN VITRO*

Overview

The first *in vitro* editing system reported involved a series of complex steps for RNA substrate construction and thin layer chromatography separations (Hirose and

Sugiura, 2001). These made the system expensive, time-consuming, and cumbersome. Therefore we have simplified the production of substrates for our *in vitro* editing system. RNA substrates are constructed by two rounds of PCR followed by *in vitro* transcription (Fig. 1, Hegeman et al. 2005b), without the RNA ligation of radiolabeled oligonucleotides required by the Hirose and Sugiura (2001) protocol. Bacterial sequences SK and KS on the respective 5' and 3' ends of each substrate allow for the specific amplification of editing substrates through RT-PCR with SK and KS primers without gene specific regions. Since different substrates all share SK and KS sequences, amplification by RT-PCR following the editing reaction can be performed using the same primers. Thus, many substrates can be analyzed rapidly and easily in a cost-effective manner.

Exogenous RNA editing substrates vary considerably in their editing efficiency and the minimum amount of sequence they must contain around the C target. The sequences required in editing substrates both *in vivo* and *in vitro* are typically located 5' nearby an editing site (Chaudhuri & Maliga, 1996; Bock et al., 1997; Miyamoto et al., 2002; Hayes & Hanson, 2007). Editing substrates expressed from transgenes *in vivo* have been reported for four sites with little endogenous 5' and 3' sequences surrounding the C-editing target, suggesting the sequences required for editing are within 20 nt 5' and 6 nt 3' (Bock et al., 1996; Chaudhuri & Maliga, 1996; Reed et al., 2001).

In vitro RNA editing substrates have typically been constructed with at least around 100 nt of sequence 5' of the editing site and around 15 nt of 3' sequence in an attempt to ensure that RNAs contain the important cis-elements for editing (Table 4.2). Reported editing reactions *in vitro* vary in editing efficiency from 5-92%, depending on the substrate and chloroplast extract. We chose 5% calculated editing as the threshold to classify if RNAs are significantly edited, because we have observed

that read-through, extension past the dideoxy nucleotide, can be as high as 3% and the standard deviation for duplicate reactions is usually within 1%. Several *in vitro* substrates for NTrpoB C473 and ATpsbE C214 contain less native sequence than contained in other editing substrates but are edited as well as larger substrates for the same sites (Hayes et al., 2006; Hayes & Hanson, 2007). Results with these smaller substrates indicate that a region of at most 30 nt 5' and 10 nt 3' of native sequence may be sufficient for editing of other sites *in vitro*.

Several RNAs containing more than 100 nt 5' and ≥ 10 nt 3' of native sequence around many editing sites cannot be efficiently edited *in vitro* (Table 4.3). It is not yet clear why exogenous RNAs containing certain sites cannot be edited. *In vivo* editing substrates for sites NTndhB C737 and NTndhB C746 can be edited although exogenous RNAs containing more native sequence around the same sites cannot be edited efficiently *in vitro* (Bock et al., 1996; Hirose & Sugiura, 2001; Sasaki et al., 2006). It is unlikely, based on the analysis of a number of other sites *in vivo* and *in vitro*, that these RNAs do not contain sequences required for editing.

In comparing editing *in vivo* vs. *in vitro*, we must consider that depending on whether RNAs are expressed from a transgene *in vivo* or are synthesized *in vitro*, different sequences flank the region of sequence containing the editing site. The substrates we use for editing assays *in vitro* contain bacterial sequences such as SK and KS, whereas *in vivo* substrates contain sequences from chloroplast promoters and terminators. Editing of a substrate for NTrpoB C473 containing a -20A→U mutation compared to wild-type differed *in vitro* depending on the flanking sequences (Hayes et al., 2006). The editing efficiency of some RNAs containing required sequence elements may be affected by flanking sequences *in vitro*.

Table 4.2. Wild-type Substrates that Exhibit 5% or Greater Editing i*n vitro*

Species[a]	Gene[b]	Position[c]	5'[d]	3'[e]	Reference
Maize	NTrpoB	473	31	22	Hayes et.al., 2006
Tobacco	NTrpoB	473	31	22	Hayes et.al., 2006
Tobacco	NTpsbE	214	128	10	Miyamoto et al., 2002
Tobacco	NTpsbE	214	120	20	Sasaki et al, 2006
Tobacco	NTpsbE	214	100	15	Hayes and Hanson, 2006
Tobacco	NTpsbE	214	120	10	Hayes and Hanson, 2006
Tobacco	NTndhA	341	120	21	Sasaki et al, 2006
Tobacco	NTndhB	467	120	20	Sasaki et al, 2006
Tobacco	NTndhB	586	120	20	Sasaki et al, 2006
Tobacco	NTndhB	1481	120	20	Sasaki et al, 2006
Tobacco	NTndhD	2	120	20	Sasaki et al, 2006
Tobacco	NTndhF	290	120	20	Sasaki et al, 2006
Tobacco	NTndhG	50	120	20	Sasaki et al, 2006
Tobacco	NTndhG	347	120	20	Sasaki et al, 2006
Tobacco	NTpetB	611	121	11	Miyamoto et al., 2002
Tobacco	NTpsbL	2	150	15	Hirose and Sugiura, 2001
Tobacco	ATpsbE	214	150	15	Hayes and Hanson, 2006
Tobacco	ATpsbE	214	99	15	Hayes and Hanson, 2006
Tobacco	ATpsbE	214	54	15	Hayes and Hanson, 2006
Tobacco	ATpsbE	214	31	15	Hayes and Hanson, 2006
Tobacco	ATpsbE	214	22	15	Hayes and Hanson, 2006
Tobacco	ATpsbE	214	13	15	Hayes and Hanson, 2006
Arabidopsis	ATpsbE	214	150	15	Hegeman at al., 2005
Arabidopsis	ATpsbE	214	99	15	Hegeman at al., 2005
Arabidopsis	ATpsbE	214	54	15	Hegeman at al., 2005
Arabidopsis	ATpsbE	214	31	15	Hegeman at al., 2005
Pea	PSpetB	611	150	21	Nakajima and Mulligan, 2005
Pea	NTpetB	611	121	11	Miyamoto et al., 2002

[a] Species used to create the competent chloroplast extracts

[b] Gene where the editing site and flanking sequences are derived

[c] Position in nucleotides from the A of the initiation codon within the coding region containing the editing site

[d] Number of nucleotides of sequence flanking 5' of the edited nucleotide contained in the *in vitro* substrate

[e] Number of nucleotides of sequence flanking 3' of the edited nucleotide contained in the *in vitro* substrate

Table 4.3 RNAs that contain wild-type sequences around an editing site that exhibit less than 5% editing *in vitro*

Species [a]	Gene [b]	Position [c]	5' [d]	3'e [e]	Reference
Arabidopsis	ATndhB	149	150	15	in this report
Arabidopsis	ATndhB	467	150	15	in this report
Arabidopsis	ATndhB	586	150	15	in this report
Arabidopsis	ATndhB	611	150	15	in this report
Arabidopsis	ATndhB	1481	150	15	in this report
Arabidopsis	ATpsbE	214	22	15	Hegeman at al., 2005
Arabidopsis	ATpsbE	214	13	15	Hegeman at al., 2005
Maize	ZMrpoB	545	30	30	in this report
Maize	ZMrpoB	617	30	30	in this report
Maize	ZMrpl2	2	30	30	in this report
Maize	ZMrps8	182	30	30	in this report
Maize	ZMpetB	667	30	30	in this report
Maize	ZMndhB	467	30	30	in this report
Tobacco	NTndhB	149	120	21	Sasaki et al, 2006
Tobacco	NTndhB	611	120	21	Sasaki et al, 2006
Tobacco	NTndhB	737	120	21	Sasaki et al, 2006
Tobacco	NTndhB	737	156	10	Hirose and Sugiura, 2001
Tobacco	NTndhB	746	120	21	Sasaki et al, 2006
Tobacco	NTndhB	830	120	21	Sasaki et al, 2006
Tobacco	NTndhB	836	120	21	Sasaki et al, 2006
Tobacco	NTndhD	383	120	21	Sasaki et al, 2006
Tobacco	NTndhD	599	120	21	Sasaki et al, 2006
Tobacco	NTndhD	674	120	21	Sasaki et al, 2006
Tobacco	NTrpoB	338	100	10	in this report
Tobacco	NTndhB	467	100	10	in this report
Tobacco	NTndhA	1073	120	21	Sasaki et al, 2006

[a] Species used to create the competent chloroplast extracts

[b] Gene where the editing site and flanking sequences are derived

[c] Position in nucleotides from the A of the initiation codon within the coding region containing the editing site

[d] Number of nucleotides of sequence flanking 5' of the edited nucleotide contained in the *in vitro* substrate

[e] Number of nucleotides of sequence flanking 3' of the edited nucleotide contained in the *in vitro* substrate

One possibility that would explain why RNAs containing some editing sites cannot be edited is that they are folded in a way that does not permit editing. Editing of NTrpoB C473 has been shown to be eliminated or reduced by expression of sequences that could create double-stranded structures within the editable substrate (Hegeman et al., 2005a; Hayes et al., 2006). A structural motif has not been identified in the RNAs that would explain the lack of editing. Perhaps such sites cannot be edited because critical factors are not stable or lost during isolation of chloroplast extracts. Another possibility is that the usual *in vitro* editing conditions are suboptimal for certain sites. For editing of ATpsbE C214 using Arabidopsis chloroplasts extracts, elimination of MgOAc and addition of 10 mM ATP to the *in vitro* editing assay conditions increased editing (Hegeman et al., 2005b). This contrasts with the 3 mM MgOAc and 3 mM ATP found optimal for NTpsbL C2 using tobacco extracts in Hirose and Sugiura (2001) and the 3 mM MgOAc and 1 mM ATP reported in Miyamoto et al. (2002) and (Miyamoto et al., 2004) for NTpsbE C214. Possibly by altering the conditions of the *in vitro* reaction, some RNAs in Table III will be editable *in vitro*. Unfortunately, at present the editing sites that can be analyzed *in vitro* are limited in number.

The current optimized conditions used for the *in vitro* editing assay are based on the ones first described in Hirose and Sugiura (2001). MgOAc was found to be inhibitory by Hegeman et al. (2005b) and is excluded from our assay conditions, unlike the 3 mM reported to be optimal by Hirose and Sugiura (2001) and Miyamoto et al. (2002). Also concentrations of HEPES-KOH pH 7.7, KOAc, DTT, and NH_4OAc have been increased and the concentration of protease inhibitors has been reduced compared to Hirose and Sugiura (2001) and Miyamoto et al. (2002). The concentration of RNAse inhibitors is lower than Hirose and Sugiura (2001) but equal to the optimized conditions in Miyamoto et al. (2002). The procedure for the editing

assay is listed below.

In Vitro Editing Assay Protocol

1. Create *in vitro* transcription templates for the production of RNA editing substrates using two rounds of PCR (Figure 4.1). In the first round, amplify a gene fragment including an editing site and surrounding sequence using total genomic DNA as template and primers designed to flank the 5' end of the gene fragment with the bacterial sequence SK and the 3' end with KS (Figure 4.1). Add the T7 sequence 5' of the SK sequence using the amplification products from the first round of PCR as template. Purify the PCR products using the Qiaquick PCR purification kit (Qiagen Inc., Valencia CA).

2. Produce RNAs by using the purified PCR products from the second round of PCR as template and the T7 Megashortscript kit (Ambion Inc., Austin, TX) to set up *in vitro* transcription reactions (20 μL). Incubate *in vitro* transcription reactions for 2 hours at 37 ºC. Remove DNA templates by adding 2 U of TURBODNase (Ambion Inc.) to *in vitro* transcription reactions and incubating 15 minutes at 37ºC. After synthesis, purify RNAs using the RNA Cleanup Kit (Zymo Research, Orange, CA). Quantify RNAs using absorbance readings from a spectrophotometer at 260 nm and 280 nm. Dilute RNAs to a working concentration of 100 pM.

3. To edit RNAs, create editing reaction mixtures containing 45 mM HEPES-KOH pH 7.7 (Fisher Scientific), 67.5 mM KOAc (Fisher Scientific), 45 mM NH_4OAc (Fisher Scientific), 5% glycerol (Fisher Scientific), 1% Polyethylene Glycol 6000 (USB, Cleveland, OH), 1 mM ATP[2] (Sigma-Aldrich), 6 mM DTT (Sigma-Aldrich),

[2] 10mM ATP is optimal for ATpsbE C214 and NTpsbE C214. (Hayes and Hanson, 2007)

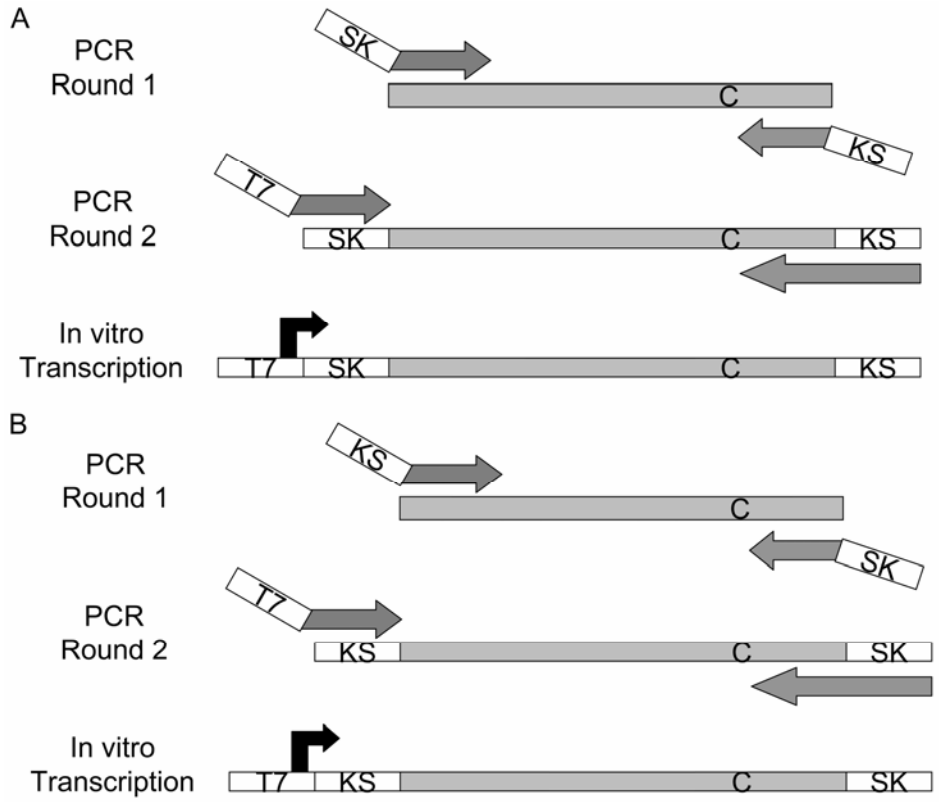


Figure 4.1. A diagram depicting the construction of (A) RNA editing substrates and (B) competitor RNAs. Grey bars represent sequences that flank an endogenous editing site. The position of the edited nucleotide is denoted by a large bolded C. Arrows and/or boxes signify primer sequences. Grey arrows indicate portions of the primer complementary to template. Black arrows symbolize transcription start sites for T7 RNA polymerase. Oligonucleotides were obtained from IDT (Coralville, IA). Universal amplification sequences expressed in the RNA substrate are: SK: 5'-CGCTCTAGAACTAGTGGATC-3'; KS: 5'-GATACCGTCGACCTCGA -3'; T7SK: 5'-TAATACGACTCACTATAGGGCGCTCTAGAACTAGTGGATC.

0.8X Complete, Mini, EDTA-free, Protease Inhibitor Cocktail (Roche Applied Science, Indianapolis, IN), 2.4U/μL RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen, Carlsbad, CA), 6.4 μg/μL chloroplast extract[3], and 8 pM RNA substrate. We find 12.5 μL reactions are most economical when performed in 0.2 mL semi-skirted 96-well PCR plates by Abgene (Epsom, United Kingdom). Incubate the editing reactions for 2 hours at 30ºC, then for 5 minutes at 65ºC. Chill reactions in an ice water bath for 5 minutes and centrifuge at 1509xg in a Heraeus Labofuge 400 for 1 minute at room temperature to pellet precipitated proteins.

4. Use 1 μL of supernatant from each reaction as template in a reverse transcription (RT) reaction (10 μL) using the Sensiscript RT kit (Qiagen Inc.). In order to specifically amplify *in vitro* substrates and not endogenous nucleic acids, a primer containing KS sequence is used as the gene specific primer in the RT reaction. Incubate the RT reactions for 1 hour at 37ºC.

5. Amplify ssDNAs copied from edited exogenous RNAs by using 2.5 μL of the RT reaction as template, SK and KS primers, and the Taq PCR Master Mix Kit (Qiagen Inc.) in PCR reactions (25 μL). Cycle the PCR reactions 30 times at 95ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 30 seconds. Clean the PCR product from primers and nucleotides by using the Exosap-IT PCR Clean-up (USB) reaction before use as template in the PPE assay.

QUANTIFICATION OF EDITING OF EXOGENOUS RNAS

Overview

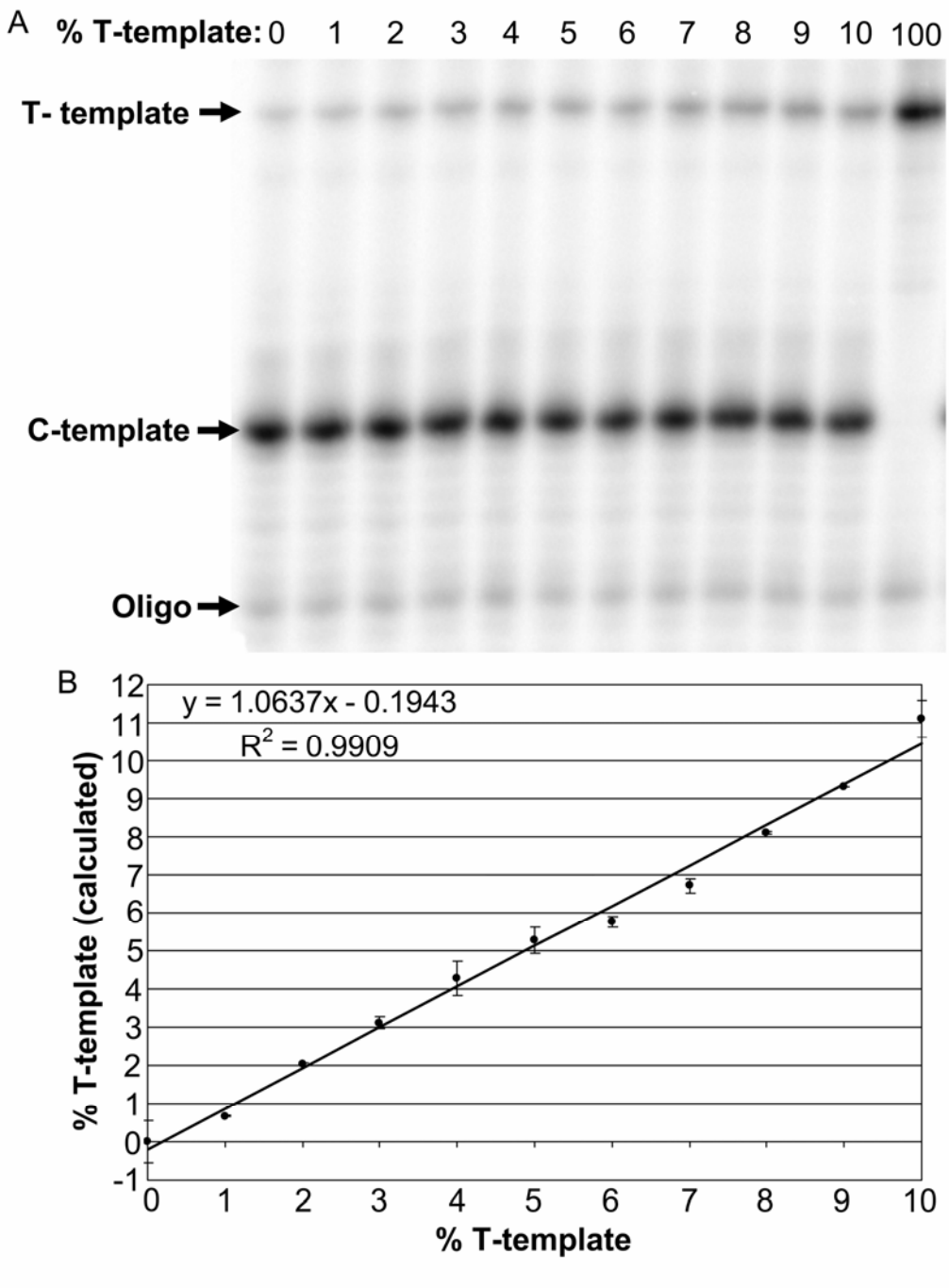
Direct sequencing, bulk sequencing, and radiolabeling the edited nucleotide, followed by TLC separation, thymine-DNA glycosylase (TDG) assay, or PPE using

[3] For editing reactions using maize chloroplasts extracts use 1.6 μg/μL as in Hayes et al. (2006)

either radiolabeled or fluorescently labeled oligonucleotides have all been used for the analysis of editing in organelles both *in vivo* and *in vitro* (Bock et al., 1994; Hirose & Sugiura, 2001; Reed et al., 2001; Peeters & Hanson, 2002; Takenaka & Brennicke, 2003; Sasaki et al., 2006). Each method has its own unique advantages and disadvantages. We have found that PPE using radiolabeled oligonucleotides is sufficiently economical, expedient, sensitive, accurate, and precise for the needs of our assay.

The PPE assay was examined to ensure that the sensitivity and accuracy of the method exist even with low editing extents encountered in some *in vitro* editing reactions (Figure 4.2). A PPE template containing the sequence around NTrpoB C473 with a C at the editing site position was constructed through PCR as in Figure 4.1 without *in vitro* transcription. A template with a T at the position of the editing site was created using a specific primer that overlapped the editing site containing an A nucleotide at the complementary position. The DNA concentrations of the two templates were calculated. The templates were then diluted to equal concentrations. The two templates were mixed in specific ratios and used in PPE reactions. With relative concentrations of T-template from 0-10%, the calculated % T-template is within around 1% (Figure 4.2). Each reaction was also performed in duplicate reactions and the standard deviation was within 1% for independent samples. Therefore, the linear relationship of calculated percent T-template versus percent T-template in mixed templates is clear for the range 0-10% (Figure 4.2). In a similar reconstruction experiment, the linear relationship was observed in the range 0-100% (Peeters and Hanson, 2002).

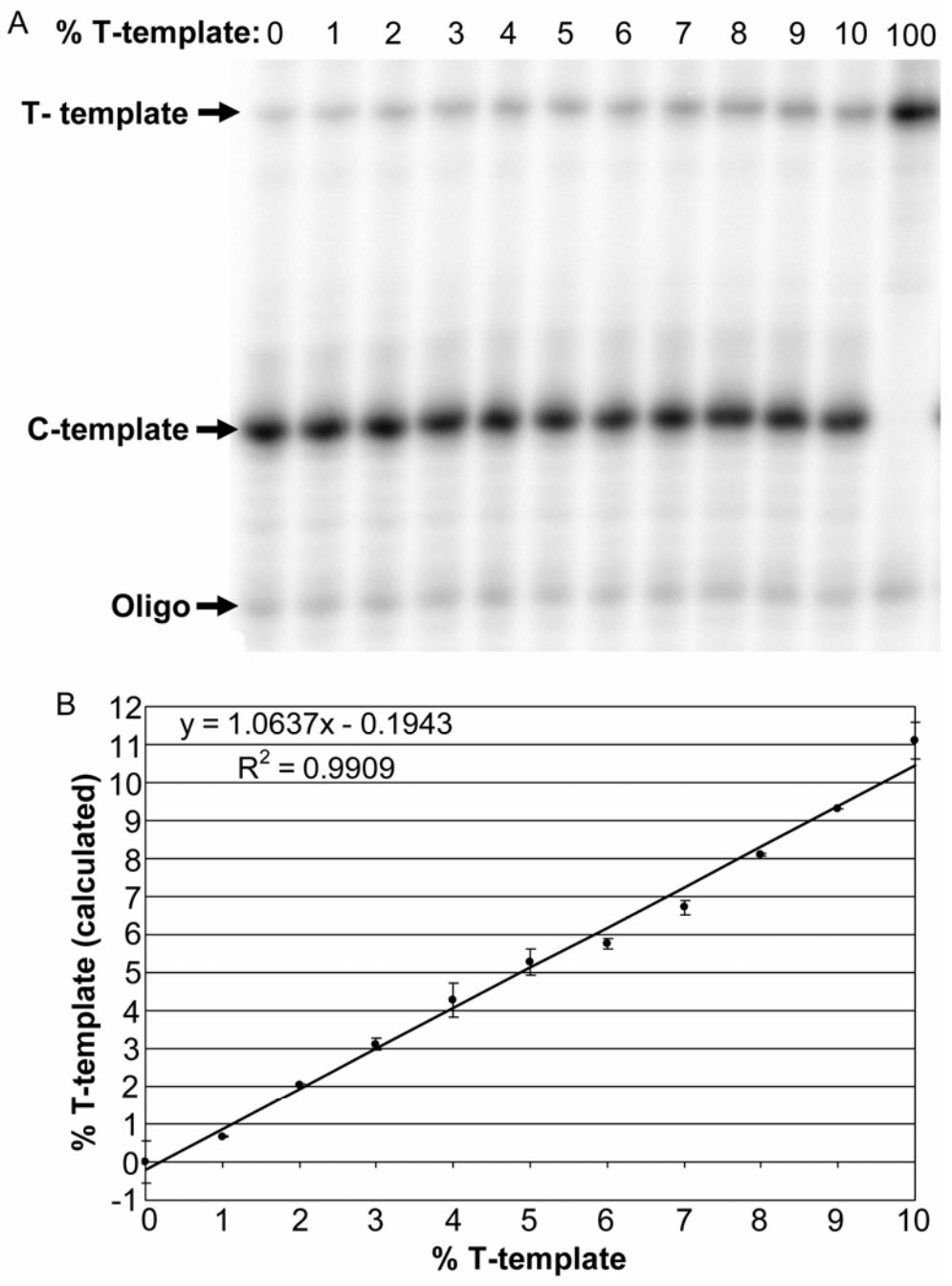
Figure 4.2. Accuracy and precision of the PPE assay. PPE templates were constructed expressing the -31/+22 region around the editing site NTrpoB C473 with either a C or at T nucleotide at the position of the editing site through PCR. Templates were mixed with different percentages of T-template from 1-10% in 1% intervals. (A) An electrophoretogram shows the size separated products from the PPE assay. The bands are labeled T-template and C-template according to the expected size of extension products from priming off each template contained in the PPE reaction. The labeled oligonucleotide not extended during the reaction is labeled Oligo. The % T-template used in each PPE reaction is indicated above the corresponding lane. (B) A graph was produced to determine the accuracy and precision of the PPE assay. The x-axis shows the % T-template contained in each PPE reaction. The intensity of each band of the electrophoretogram was calculated and the amount of read through calculated in lane 0% T-template was subtracted from experimental samples. The y-axis displays the % T-template calculated for each PPE reaction. The equation for the regression line and the R^2 value are shown in the top left of the graph.


A
% T-template: 0 1 2 3 4 5 6 7 8 9 10 100
T- template
C-template
Oligo
B
y = 1.0637x - 0.1943
R2 = 0.9909
% T-template (calculated)
% T-template

One concern with the PPE assay is the tendency of sequencing polymerases to run through and not stop at the appropriate nucleotide (Roberson & Rosenthal, 2006). To minimize these effects, it is crucial to use ThermoSequenase (USB) as the PPE polymerase, eliminate nucleotide contamination of templates, and choose the best oligonucleotide and ddNTP for the editing site. Many polymerases were compared in a PPE fluorescent assay reported by Roberson and Rosenthal (2006). By far the ThermoSequenase enzyme had the lowest amount of read-through. Nucleotide contamination can also lead to significant read–through, so it is necessary to remove dNTPs and primers that might interfere with PPE. For this purpose we treat all PPE templates with Exosap-IT (USB), which contains alkaline phosphatase activity, before assay. The enzymes in Exosap-IT are easily deactivated by heating the reaction to 80ºC for 15 minutes. The choice of oligonucleotide and ddNTP is based on the native sequences around the editing site. Primers may be chosen so that they anneal either on the + or – strand of the PCR template. The dideoxy nucleotide to be used should be chosen so that bands can be clearly resolved on polyacrylamide gels (Figure 4.3). C and G nucleotides are acceptable dideoxy nucleotides for terminating primer extension at most editing sites. We have experienced a low amount of read-through, about 0-3%, which can easy be measured and accounted for by subtracting the read-through signal from sample measurements.

Recently, two fluorescent PPE assays that rely on fluorescently labels have been reported as alternatives to radioactive methods for quantifying editing (Roberson & Rosenthal, 2006; Sasaki et al., 2006). The difference between these two fluorescent assays is that Sasaki et al.'s (2006) assay used a DNA sequencer to detect fluorescently labeled ddNTP incorporation by a one-nucleotide extension reaction. In contrast, Roberson and Rosenthal (2006) used labeled oligonucleotides in a reaction resulting in extension by several nucleotides. Extension products were separated on

Figure 4.3. The Poisoned Primer Extension assay (PPE) is represented using either (A) ddGTP or (B) ddCTP as the poisoned nucleotides. Characters depict nucleotides around the editing site in a PPE template, PPE products, 5' labeled oligonucleotides, and nucleotides used in the PPE reaction. The symbol (*) signifies the 5' radiolabeled phosphate. The editing site and the next nucleotide that can terminate the PPE reaction are indicated under large arrows. Small arrows point to the expected PPE products from the template shown above.


A
% T-template: 0 1 2 3 4 5 6 7 8 9 10 100
T- template
C-template
Oligo
B
y = 1.0637x - 0.1943
R2 = 0.9909
% T-template (calculated)
% T-template

polyacrylamide gels, and the intensity of the bands was calculated by a Typhoon 9200 imager. Results by Roberson and Rosenthal (2006) are analogous to what we observe using radiolabeled oligonucleotides with a very high accuracy and precision. The method used in Sasaki et al. (2006) is less accurate than either our assay or Roberson and Rosenthal (2006), due most likely to the method of detection. Neither fluorescent assay has been evaluated to determine whether they are accurate within 1% in the 0-10% range. The procedure for our PPE assay for examining editing efficiency is described below.

Poisoned Primer Extension Assay Protocol

1. Set up PPE reactions (15 μL) using the ThermoSequenase Cycle Sequencing kit (USB) containing 50 ng of RT-PCR product, 1X ThermoSequenase buffer (USB), 4 mM each of three dNTPs and one ddNTP, 0.42 nM 5' labeled oligo, and 4U ThermoSequenase (USB). Incubate PPE reactions for 5 cycles at 90ºC for 5 seconds, 50ºC for 30 seconds, and 72ºC for 10 seconds. Add 5 μL of Stop Solution (USB) to each reaction.

2. Produce a 12% acrylamide sequencing gel 42x33 cm using the Sequagel Sequencing System (National Diagnostics, Atlanta, GA). Incubate stopped PPE reactions at 72ºC for 5 minutes to denature extension products. Load 5 μL of each PPE reaction into each lane of the polyacrylamide gel. Run the gel at 65 mA for 2 hours. Remove one gel plate and wrap the gel in a flexible transparent plastic support such as Saran wrap. Expose a phosphorimaging cassette (Amersham Biosciences) to the gel either for 1 hour or overnight depending on the intensity of the signal.

3. Develop electrophoretograms using a Storm 860 phosphorimager (Amersham Biosciences, Piscataway, NJ). Determine the intensity of bands using ImageQuant V. 5.2 (Amersham Biosciences) software. Calculate the percent editing of RNA substrates for each reaction and one standard deviation around the mean for

duplicate reactions. Establish the amount of read-through from the intensity of bands resulting from a PPE reaction, using a substrate that is not edited as template. Subtract the signal from the other measurements.

REDUCTION OF EDITING THROUGH THE ADDITION OF COMPETITOR RNAS

Overview

The percentage of substrates edited decreases when excessive amounts of substrates are added greater than the capacity of the reaction (Figure 4.4). This reduction is thought to be due to competition within the reactions for editing factors. Competition experiments are useful for examining the cis-elements within the editing substrates. The first report of competition *in vitro* was reported with the initial *in vitro* assay (Hirose and Sugiura, 2001). In the initial assay, oligoribonucleotides were used as competitors that contained sequences around the editing site. For competition reactions, 2000X more competitor than editing substrate was required (Hirose and Sugiura, 2001). There was a reduction in editing in many reactions with only vector RNA added as competitor. This was most likely due to nonspecific interactions stemming from using such a high amount of competitor compared to substrate. The conditions for the competition assay were improved in Miyamoto et al. (2002) when addition of 100X and 1000X competitor RNA versus substrate substantially reduced editing compared to reactions with only vector RNA added as competitor. They did not observe a large reduction in editing reactions with vector RNA added as competitor compared to reactions without competitor RNAs. In Arabidopsis extracts, the same absolute amount of competitor, 1 pmol (Hegeman et al., 2005b), reduced editing as in Miyamoto et al. (2002). This suggests a similar amount of competitor is required to reduce *in vitro* editing in Arabidopsis and tobacco extracts. We have

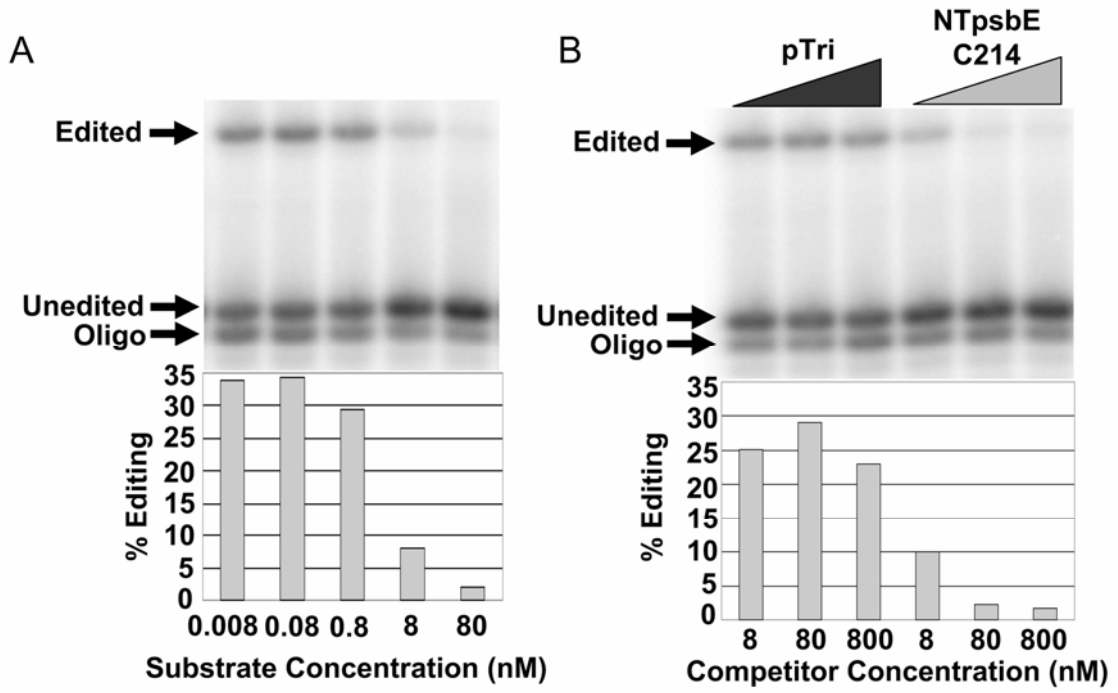


Figure 4.4. Increasing amounts of substrate (A) or self-competitor RNA (B) reduces the percentage of substrate edited *in vitro*. (A) Editing and (B) competition reactions use a substrate expressing the tobacco editing site NTpsbE C214 in tobacco chloroplast extracts. (A and B) Graphs indicating the calculated percent editing are located below corresponding electrophoretograms. Arrows point to the expected sizes of PPE products from edited and unedited templates as well as unextended labeled oligo. Shaded triangles signify concentrations of added competitor RNAs added to editing reactions. (B) Competitor, labeled NTpsbE C214, contains the same region of *psbE* as the editable substrate. Competitor pTri is a RNA of similar length as the substrate but does not share common sequences with the editing substrate. This figure was reproduced in modified form from Hayes and Hanson (2007).

further refined the competition assay to reduce nonspecific interactions by the addition of only 10X to 100X competitor for specific competition of the RNA substrate (Hayes and Hanson, 2007).

Competitor RNA substrates are constructed similarly as RNA substrates (Figure 4.1). Unlike the initial competition experiments which utilized editing substrates that are not radiolabeled at the editing site (Hirose and Sugiura, 2001), we developed an alternative type of competitor for our *in vitro* system. The first competitors in our *in vitro* assay were identical to the RNA substrates except they did not contain the SK and KS bacterial sequences (Hegeman et al., 2005b). This prevented amplification of the competitor RNAs during RT-PCR amplification of the substrate. One caveat about such competitors is that the sequences SK and KS sometimes make up a considerable portion of the editing substrate, so that competitors without these sequences could have different structural features compared to the editing substrate. Currently we construct competitor RNAs so that the SK and KS bacterial sequences are in swapped positions (Hayes and Hanson, 2007). This prevents them from being amplified by RT-PCR along with the substrate, but allows similar sequences to be represented. The procedure for the competition experiments is shown below.

In Vitro Competition Assay Protocol

1. Construct competitor RNAs similarly as RNA substrates except in comparison the flanking sequences SK and KS are swapped (Figure 4.1). Create *in vitro* transcription templates for the production of RNA editing substrates through two rounds of PCR. In the first round, amplify a gene fragment including an editing site and surrounding sequence using total genomic DNA as template. Use primers designed to flank the 5' end of the gene fragment with the bacterial sequence KS and the 3' end with SK. Use the amplification products from the first round as template for

a second round of PCR. In the second round, add the T7 sequence 5' of the KS sequence and purify PCR products using the Qiaquick PCR purification kit (Qiagen Inc.).

2. Determine the maximum concentration of RNA substrate that can be edited efficiently in the *in vitro* editing reaction. Titrate substrate RNAs in editing reactions at concentrations between 8 pM to 80 nM. Use the maximum concentration of substrate that is edited as highly after 2 hours at 30ºC as the 8 pM reaction in later competition reactions. We have observed that 0.8 nM is maximal for many different substrates in extracts from different species.

3. Add competitor RNAs at concentrations 0.8 nM, 8 nM, or 80 nM to 0.8 nM of substrate RNA in the optimized editing reactions (12.5 μL). Conditions for competition assays are identical to optimized editing reactions except for changes to the RNA added.

4. Reverse-transcribe RNA substrates specifically in 10 μL reactions using the bacterial primer KS and 1 μL of the competition reaction as template. Perform PCR as in normal editing reactions with bacterial primers SK and KS and 2.5 μL of the RT reaction as template. Treat RT-PCR products with Exosap-IT (USB), and use as templates in the PPE assay.

QUANTIFICATION OF EDITING OF ENDOGENOUS RNAS

Overview

It has been the focus of this article to describe the *in vitro* editing assay. The amount of editing in endogenous transcripts can also be determined with the PPE assay. The procedure is based on Peeters and Hanson (2002) with the primary change being in the type of primers used for reverse transcription. Instead of using random hexamers we find it is optimal to use gene-specific primers. The procedure for assay

of editing *in vivo* is listed below.

***In Vivo* Editing Quantification Protocol**

1. Isolate total RNA from 100 mg of leaf tissue using Trizol (Invitrogen).

2. Remove contaminating DNA through a DNAse treatment using the TURBO DNA-free Kit (Ambion Inc.).

3. Quantify RNA using A 260/280 readings from a spectrophotometer.

4. Reverse-transcribe transcripts of interest using the Sensiscript Kit (Qiagen Inc.) and a gene-specific primer. Use cDNAs as template in a PCR reaction to amplify regions of interest. Treat PCR amplification products with Exosap-IT (USB) and use as templates in the PPE assay.

ACKNOWLEDGMENTS

This work was supported by NSF grant MCB 0344007 to M.R.H.

REFERENCES

Bock R, Hermann M, Fuchs M. 1997. Identification of critical nucleotide positions for plastid RNA editing site recognition. *RNA 3*:1194-1200.

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

Bock R, Kossel H, 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.

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

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

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

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

Hegeman CE, Halter CP, Owens TG, Hanson MR. 2005a. Expression of complementary RNA from chloroplast transgenes affects editing efficiency of transgene and endogenous chloroplast transcripts. *Nucleic Acids Res 33*:1454-1464.

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

Hirose T, Sugiura M. 2001. Involvement of a site-specific trans-acting factor and a

common RNA- binding protein in the editing of chloroplast mRNAs: development of a chloroplast *in vitro* RNA editing system. *EMBO J 20*:1144-1152.

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

Kunst L. 1998. Preparation of physiologically active chloroplasts from Arabidopsis. *Methods Mol Biol 82*:43-48.

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

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

Nakajima Y, Mulligan RM. 2005. Nucleotide specificity of the RNA editing reaction in pea chloroplasts. *J Plant Physiol 162*:1347-1354.

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

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

Peltier JB, Ytterberg J, Liberles DA, Roepstorff P, van Wijk KJ. 2001. Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *J Biol Chem 276*:16318-16327.

Reed ML, Hanson MR. 1997. A heterologous maize *rpoB* editing site is recognized by

transgenic tobacco chloroplasts. *Mol Cell Biol 17*:6948-6952.

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

Roberson LM, Rosenthal JJ. 2006. An accurate fluorescent assay for quantifying the extent of RNA editing. *RNA 12*:1907-1912.

Sasaki T, Yukawa Y, Wakasugi T, Yamada K, 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, Nagano Y. 2001. Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem 276*:3937-3940.

Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, Herrmann RG, Maier RM. 2005. Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. *Plant Cell 17*:1815-1828.

Somerville CR, Sommerville, S.C. and Ogren, W.L. 1981. Isolation of photosynthetically active protoplasts and chloroplasts from *Arabidopsis thaliana*. *Plant Science Letters 21*:89-96.

Takenaka M, Brennicke A. 2003. *In vitro* RNA editing in pea mitochondria requires NTP or dNTP, suggesting involvement of an RNA helicase. *J Biol Chem*.

Zito F, Kuras R, Choquet Y, Kossel H, Wollman FA. 1997. Mutations of cytochrome b6 in *Chlamydomonas reinhardtii* disclose the functional significance for a proline to leucine conversion by *petB* editing in maize and tobacco. *Plant Mol Biol 33*:79-86.

CHAPTER 5

INDEPEDENT EMERGENCE OF CHLOROPLAST RNA EDITING SITES IN HOMOLOGOUS GENES

ABSTRACT

C-to-U editing modifies 30-40 distinct nucleotides within higher plant chloroplast transcripts. Many C targets are located at the same position in homologous genes from different plants; these may either have emerged independently or share a common origin. The cis-element GCCGUU is known to be necessary for the editing *in vitro* of templates containing the site NTpsbE C214. We investigated *psbE* sequences from many plant species to determine in what lineage editing of psbE C214 emerged and whether the motif is conserved in plants with a C214. Plants with a C214 frequently contain the intact motif. The C214 site in *psbE* genes is represented in members of four branches of spermatophytes but not in gnetophytes, suggesting editing of psbE C214 is likely to have been present in the ancestor of spermatophytes. *Sciadopitys verticillata* appears to have evolved editing of C214 independently, based on the presence of nucleotide differences in the motif and its phylogeny. Extracts from chloroplasts from a species that has a difference in the motif and lacks the C target are incapable of editing tobacco psbE C214 substrates, implying the critical protein factors are not retained without a C target. We examined the possibility of multiple evolutionary events for two additional editing sites. Putative editing cis-elements were uncovered in the 5′ UTRs near editing sites psbL C2 and ndhD C2. As in *psbE*, plants from different lineages that share the same C target conserve different sequences, suggesting some of the common C targets emerged independently.

INTRODUCTION

C-to-U RNA chloroplast RNA editing alters the sequence of transcripts in all plants except *Marchantia*. A common ancestor to *Marchantia polymorpha* likely lost the editing mechanism secondarily (Groth-Malonek et al., 2007). Due to the distribution of editing in plants, the general mechanism probably evolved in a common ancestor to all extant embryophytes (land plants) (Tillich et al., 2006). Although all spermatophytes (seed plants) most likely share a general mechanism for editing, the Cs that are targeted for editing differ considerably between species (Freyer et al., 1997). The collection of Cs that are edited in a particular species, its "editotype," can vary even between closely related *Nicotiana* species (Sasaki et al., 2003). Typically, the editotype of chloroplasts from each spermatophyte includes only about 30-40 sites (Maier et al., 1995; Wakasugi et al., 1996; Corneille et al., 2000; Tsudzuki et al., 2001; Schmitz-Linneweber et al., 2002; Tillich et al., 2005). Some of the C targets in the chloroplasts of one species are at the same position as ones in homologous genes (Tsudzuki et al., 2001). Thirty different edited Cs in homologous chloroplast genes are common to the related species *Nicotiana tabacum*, *Atropa belladonna*, and *Solanum lycopersicum* (Kahlau et al., 2006). A C target common to all spermatophytes is not known. We examined three sites to gain information about the evolutionary origin of C targets in homologous genes.

RNA editing at C214 in *psbE* chloroplast transcripts was first identified through the analysis of cDNA sequences from *Pinus thunbergii* (Wakasugi et al., 1996). A homologous editing site at the same position in *psbE* has been experimentally verified using the same techniques in *Nicotiana tabacum*, *Ginkgo biloba*, and *Arabidopsis thaliana* (Hirose et al., 1999; Kudla & Bock, 1999; Tillich et al., 2005). Using our proposed nomenclature for editing sites (Hayes et al., 2006), the C214 in tobacco *psbE* is designated NTpsbE C214 (species initials, gene name,

number of nucleotides from the start codon where the A is counted as 1. Editing at C214 changes a genomically encoded proline codon at position 72 to a serine codon that is conserved in other plants. The entire photosystem II cytochrome b559 alpha subunit encoded by *psbE* is highly conserved in plants, with 90% identity between the predicted amino acid sequence of the encoded proteins between *Nicotiana tabacum* and *Marchantia polymorpha*. Conservation of the serine codon through editing at amino acid position 72 suggests a strong selection against the presence of a proline, which would be encoded by unedited transcripts of *psbE*. We propose that this selection results in a maintenance of the editing factor and the corresponding cis-element. While maintenance is likely to be due to selection, editing sites might emerge from neutral drift, according to the model of Covello and Grey (1993). If Cs at the same position in homologous position are found to arise independently during evolution, then perhaps there is some evolutionary advantage to a plant to undergo editing at a particular position in the transcripts of certain genes.

In tobacco, an unknown protein that is estimated to be 56 kDa can be UV-crosslinked to an editable *psbE* template but not a template with nucleotide differences in a critical sequence for editing (Miyamoto et al., 2004). A trans-factor essential for editing psbE C214 has not yet been identified in any plant. The sequence GCCGTT, which precedes the NTpsbE C214, was found to be critical for editing *in vitro* (Hayes & Hanson, 2007b). The minimal sequence required for *in vitro* editing in ATpsbE C214 is similar to that discovered in tobacco (Hegeman et al., 2005). Since the sequence requirements are similar, a homologous or related factor could be required for editing of C214 in both species. Conservation of the motif in species with C214 would suggest the presence of a similar factor. However, C214, like most editing sites, is present in the coding region, thus possibly limiting the evolutionary divergence of nearby sequences. Also, the mutation rate of the genome encoded by plant

chloroplasts is exceptionally low compared to many mitochondrial and nuclear genomes of other phylogenetic groups (Lynch et al., 2006). Fortunately, a large number of *psbE* sequences are available from many species because of past interest in this gene for understanding phylogenetic relationships. We were therefore able to use this data to determine whether the motif is maintained in plants that exhibit a C target in *psbE*.

Another type of editing site useful for study of cis-acting elements are those few that are within the initiation codon of chloroplast genes. At these sites, editing modifies transcripts with a genomically encoded ACG to AUG, allowing for efficient protein synthesis to occur. Because the sequences immediately 5' are non-coding, there should be less selective pressure acting on these sequences. Two of these sites, ndhD C2 and psbL C2, are found in the chloroplasts of diverse species. We analyzed sequences around both sites to determine putative editing cis-elements.
Previously published studies have provided some information about the cis-elements near the editing sites situated within initiation codons. ATndhD C2 is not edited in Arabidopsis lines with mutations in the *crr4* gene (Kotera et al., 2005). C*rr4* is one of only two nuclear genes that have currently been linked to defects in specific editing sites (Kotera et al., 2005; Okuda et al., 2007). The CRR4 protein binds to *ndhD* transcripts, suggesting it participates as a site-specificity factor of the editing machinery (Okuda et al., 2006). Functional orthologs of CRR4 have not been reported; the evolution of site specificity factors is currently unresolved. The cis-element critical for CRR4 protein binding has been delimited to the –25/+10 region around ATndhD C2 (Okuda et al., 2006). Recognition of NTpsbL C2 has also been studied: an unidentified tobacco 25 kDa protein can be UV-crosslinked to a tobacco substrate containing NTpsbL C2 (Hirose & Sugiura, 2001). The cis-element required for editing has been limited to the -16/+5 region around the editing site (Chaudhuri &

Maliga, 1996). We therefore examined the conservation of sequences near ndhD C2 and psbL C2 to determine whether or not conserved sequences, possibly representing cis-elements for C target recognition, could be detected.

MATERIALS AND METHODS

Plant Material

Leaves were collected from *Cycas revoluta*, *Dioon endule*, and *Zamia latifolia* from mature plants growing in the horticulture greenhouses at Cornell University. Stem tissue of *Cuscuta sp*. was acquired from local sources and leaves of *Nelumbo lutea* came from a native population on the East side of Cayuga Lake near Union Springs, NY. Leaf tissue of *Nelumbo nucifera* was gathered from the Ithaca Farmers market (Ithaca, NY). Needles of *Sciadopitys verticillata* were collected from the Cornell Plantations (Ithaca, NY).

DNA Isolation

DNA was isolated using a CTAB protocol. 100 µL of 2X CTAB buffer (2% hexadecyltrimethyl ammonium bromide, 1.4 M NaCl, 100 mM TrisHCl ph 8, 20 mM EDTA) was added to 100 mg of leaf tissue. The tissue was then ground in the CTAB buffer. Following grinding 600 µL of 2X CTAB was added to the macerated leaf tissue and the solution was incubated at 65ºC for 30 minutes. Next 700 µL of chloroform was added and mixed for 5 minutes; the mixture was then centrifuged at 10,000 rpm in a microfuge for 10 minutes. The aqueous supernatant was pipeted into a new Eppendorf tube and mixed with 700 µL of chloroform. The aqueous supernatant was removed to a new Eppendorf tube and 700 µL of isopropanol was added. The DNA was precipitated at 4ºC for 30 minutes and centrifuged at 10,000 rpm for 10

minutes. The DNA pellet was washed in 500 µl of 70% ethanol and allowed to air dry for 10 minutes. The DNA pellet was resuspended in 30 µL of nuclease-free water.

RNA Isolation and cDNA Synthesis

Around 100 mg of leaf tissue was pulverized in liquid nitrogen with a mortar and pestle. RNA was extracted following the RNeasy protocol provided in the RNeasy Kit (Qiagen, Valencia, CA). The amount of RNA was quantified using the absorbance at 260/280 nm in a spectrophotometer. Reverse transcription reactions (10 µL) were set up using the protocol from the Omniscript Kit (Qiagen) with 500 ng of RNA as template.

Polymerase Chain Reaction Conditions and Sequencing

PCR reactions (25 µL) were prepared using 2X Master Mix (Qiagen) and 10 pmol each oligonucleotide. Either 2.5 µL from a reverse transcription reaction or 100 ng of genomic DNA was added as template. For sequencing 2.5 µL of the PCR reaction was mixed with 1 µL of Exosap-IT (USB, Cleveland, OH) and the reaction was incubated for 15 minutes at 37ºC followed by 15 minutes at 80ºC. Sequencing reactions (18 µL) contained 10 pmol of oligonucleotide and 3.5 µL of the Exosap-IT reaction and were submitted to the Biotechnology Resource Center at Cornell University.

In vitro Editing Assay

Assays for *in vitro* editing were performed as in Hegeman et al. (2005) with minor modifications. Editing reaction mixtures (12.5 µL) contained 45 mM HEPES-KOH pH 7.7 (Fisher Scientific), 67.5 mM KOAc (Fisher Scientific), 45 mM NH_4OAc (Fisher Scientific), 5% glycerol (Fisher Scientific), 1% Polyethylene Glycol 6000

(USB, Cleveland, OH), 10 mM ATP (Sigma-Aldrich), 6 mM DTT (Sigma-Aldrich), 0.8X Complete, Mini, EDTA-free, Protease Inhibitor Cocktail (Roche Applied Science, Indianapolis, IN), 2.4 U/μL RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen, Carlsbad, CA), 1.6 μg/μL chloroplast extract, and 8 pM RNA substrate (Hayes & Hanson, 2007a). Each reaction was incubated for 2 hours at 30ºC. The RNA was amplified by RT-PCR as in Hegeman et al. (2005). An RNA editing substrate for the site NTrpoB C473 was identical to R54KS and assayed by poisoned primer extension as in (Hayes et al., 2006). R54KS contains 31 nucleotides upstream and 22 nucleotides downstream of the editing site. The RNA substrate containing NTpsbE C214 contained sequences 100 nt upstream and 10 nt downstream of the editing site (Hayes & Hanson, 2007b).

Oligonucleotides

Oligonucleotides (Integrated DNA Technologies, Coralville, IA) were used in experiments reported here. NtpsbEFor 5'-ATTCGATACTGGGTCATTCATAGC-3' and NtpsbERev 5'-CTAAAACGATCTACTAAATTCATCGAG-3' were used to amplify *psbE* sequences using both genomic DNA and cDNAs as templates from *Cycas revoluta*, *Dioon endule*, *Zamia latifolia*, *Sciadopitys verticillata*, *Nelumbo lutea*, and *Nelumbo nucifera*. CspsbEFor 5'-ATTCGATATTGGGTCATTCATAGT and CspsbERev 5'-TTGGGACCTCCTACTAAATTCATTCTG was used for amplification of *psbE* sequences from *Cuscuta* species.

RESULTS

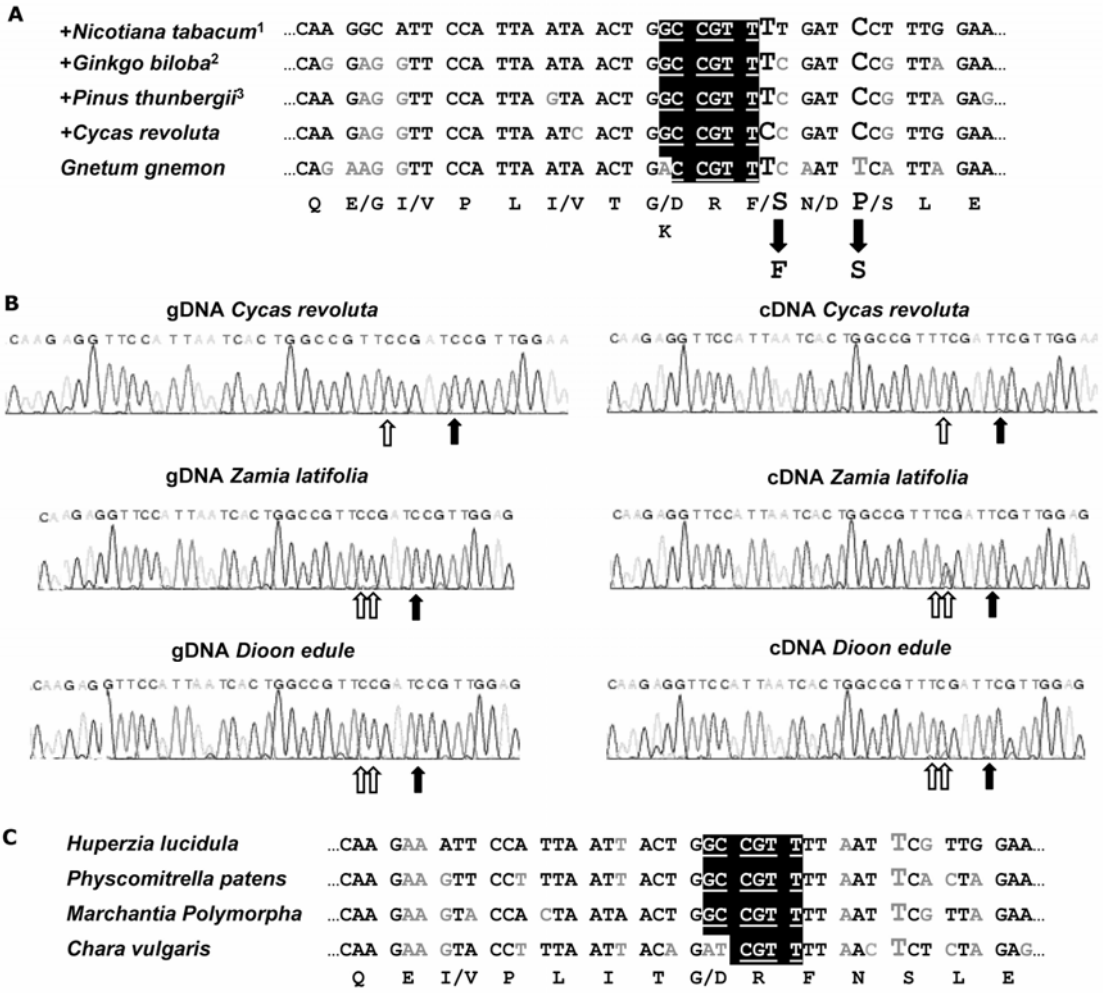
The common ancestor of spermatophytes probably edited C214 in psbE

A C214 in chloroplast transcripts is edited to a T214 in several angiosperms as

well as species from two, Ginkgoales and Coniferales, of the four main extant groups of gymnosperms (Figure 5.1). In order to establish in what lineage editing of a psbE C214 likely emerged, we investigated whether a C or T was present in the genome of the remaining two groups, Cycadales and Gnetales, as well as in groups that diverged before the emergence of spermatophytes. The chloroplast genome of all examined cycads has a C214 within *psbE* and would therefore be expected to undergo editing. Conversely, we found a T214 is present in all analyzed Gnetales. A T214 was discovered to be present in all examined members of groups that diverged before the emergence of seed plants anthocerophytes, marchantiophytes, pteridophytes, and lycopodiophytes. Due to the presence of a C214 in 3 of the 4 major branches of gymnosperms and angiosperms, the editing site likely emerged in a common ancestor of Spermatophyta. The T214 in Gnetales is likely due to a C to T mutation in a common ancestor of the lineage.

To determine if the C214 in cycads is truly edited in chloroplast transcripts, we generated cDNAs from *psbE* transcripts isolated from leaf tissues of three cycads *Cycas revoluta*, *Dioon endule*, and *Zamia latifolia*. We discovered that, in addition to editing C214, all three cycads investigated have an editing site at C209 that is unique to Cycadophyta (Figure 5.1). Modification of C209 to U209 changes the genomically encoded serine codon to phenylalanine. A phenylalanine codon is conserved at the same position in all other higher plants examined (data not shown). *Dioon endule* and *Zamia latifolia* partially edit C210 of *psbE.* The editing site is likely to be specific to the cycad family Zamiaceae (Figure 5.1). Editing at C210 does not alter the encoded amino acid and is a rare example of a silent site in chloroplasts. Possibly the partial editing is due to an imperfect "molecular ruler," which has been shown to be important in correct selection of several C targets (Hayes and Hanson, 2007).

Figure 5.1. Evolution of the psbE C214 and its cis-element. (A) A sequence alignment indicates the sequence around and including the editing site of the major branches of seed plants. A species with or without experimental evidence of editing is signaled by + or – around the editing site respectively; [1](Hirose et al. 1999), [2](Kudla and Bock 1999), [3](Wakasugi et al. 1996). (A and C) A black box highlights the critical editing cis-element from tobacco. Predicted amino acids are shown below DNA sequences. Gray characters depict nucleotides that differ from tobacco sequences. Arrows point to predicted amino acids encoded by edited transcripts. (B) Sequence traces displaying *psbE* sequences from cDNA and genomic DNA from three cycads. Empty and filled arrows mark cycad and seed plant-specific editing sites respectively. (C) A sequence alignment of plants and algae illustrating a region of *psbE* containing the motif.


A
+Nicotiana tabacum[1] ...CAA GGC ATT CCA TTA ATA ACT GGC CGT TTT GAT CCT TTG GAA...
+Ginkgo biloba[2] ...CAG GAG GTT CCA TTA ATA ACT GGC CGT TTC GAT CCG TTA GAA...
+Pinus thunbergii[3] ...CAA GAG GTT CCA TTA GTA ACT GGC CGT TTC GAT CCG TTA GAG...
+Cycas revoluta ...CAA GAG GTT CCA TTA ATC ACT GGC CGT TCC GAT CCG TTG GAA...
Gnetum gnemon ...CAG AAG GTT CCA TTA ATA ACT GAC CGT TTC AAT TCA TTA GAA...
Q E/G I/V P L I/V T G/D R F/S N/D P/S L E
K
F S
B
gDNA Cycas revoluta
cDNA Cycas revoluta
gDNA Zamia latifolia
cDNA Zamia latifolia
gDNA Dioon edule
cDNA Dioon edule
C
Huperzia lucidula ...CAA GAA ATT CCA TTA ATT ACT GGC CGT TTT AAT TCG TTG GAA...
Physcomitrella patens ...CAA GAA GTT CCT TTA ATT ACT GGC CGT TTT AAT TCA CTA GAA...
Marchantia Polymorpha ...CAA GAA GTA CCA CTA ATA ACT GGC CGT TTT AAT TCG TTA GAA...
Chara vulgaris ...CAA GAA GTA CCT TTA ATT ACA GAT CGT TTT AAC TCT CTA GAG...
Q E I/V P L I T G/D R F N S L E

The GCCGTT motif was likely present in a common ancestor of Embryophyta

The GCCGTT nucleotide motif is critical for *in vitro* editing NTpsbE C214. We examined many *psbE* sequences from plants to determine the evolution of the motif and its relationship with C214. The motif is observed without alteration in several lycopodiophytes, hepaticophytes, bryophytes, and anthocerotophytes (Figure 5.1). *Chara vulgaris*, a green alga, has *psbE* sequences with different nucleotides at positions that align with the motif. The motif is also present in 84% of the spermatophytes present in our initial dataset (Table 5.1). This suggests that the motif was present in a common ancestor to all embryophytes and the ancestor from which editing of C214 emerged.

Species with C214 frequently share a common motif in psbE

We analyzed three data sets to discover whether the presence of the GCCGTT motif correlated with the presence of a C214. In the first set sequences of all *psbE* genes present in the NCBI databases were gathered and analyzed (Table 5.1). Of the 332 spermatophyte species examined, 190 (57 %) were found to have a C214. Only 8 (4 %) have a nucleotide difference in the region that aligns with the motif and a C214 (Figure 5.2). There are 4 different changes within the 6 nt motif (Figure 5.2). Five of the 8 species with differences are from the Convolvulaceae family and 2 are within the genera *Cuscuta*.

A genomically encoded psbE T214 is present in 140 (43 %) spermatophyte species examined. A greater proportion of the species represented in the database, 43 (31 %), have differences in the motif compared with species with a C214. This includes 10 distinct changes within the 6 nt motif (Figure 5.2). We hypothesize that species with a T214 have reduced selection pressure to maintain the motif.

Since many of the 332 plants belong to genera that are over-represented in the database, we analyzed a second data set containing only one representative member of

Table 5.1. Editing sites present in three homologous chloroplast genes from a collection of diverse species

			psbE 214		*psbL* 2		*ndhD* 2		Accession #
Species	Family	Motif	Nuc.	Edit	Nuc.	Edit	Nuc.	Edit	
Nicotiana tomentosiformis	Solanaeae	Y	C	$+^{1}$	C	$+^{1}$	C	$-^{1}$	NC_007602
Nicotiana sylvestris	Solanaeae	Y	C	$+^{1}$	C	$+^{1}$	C	$+^{1}$	NC_007500
Nicotiana tobacum	Solanaeae	Y	C	$+^{2}$	C	$+^{2}$	C	$+^{2}$	NC_001879
Nicotiana glutinosa	Solanaeae	Y	C	$+^{12}$					AM177388
Solanum tuberosum	Solanaeae	Y	C		C		C		NC_008096
Lycopersicon esculentum	Solanaeae	Y	C		C		C		NC_007898
Solanum bulbocastanum	Solanaeae	Y	C		C		C		NC_007943
Atropa belladonna	Solanaeae	Y	T		C	$+^{13}$	C		NC_004561
Schizanthus pinnatus	Solanaeae	Y	C		C				AY100949
Capsicum annuum	Solanaeae				C	$+^{8}$			Ref: 8
Calystegia sepium	Convolvulaceae	Y	C		C				AY100887
Calystegia macrostegia	Convolvulaceae	Y	C		C				AY100886
Convolvulus sagittatus	Convolvulaceae	Y	C		C				AY100889
Convolvulus mauritanicus	Convolvulaceae	Y	C		C				AY100891
Convolvulus assyricus	Convolvulaceae	Y	C		C				AY100890
Convolvulus arvensis	Convolvulaceae	Y	C		C				AY100888
Polymeeria pusilla	Convolvulaceae	Y	C		C				AY100892
Argyreia nervosa	Convolvulaceae	Y	C		C				AY100872
Argyreia splendens	Convolvulaceae	Y	C		C				AY100871
Ipomoea coccinea	Convolvulaceae	Y	C		C				AY100853
Ipomoea batatas	Convolvulaceae	Y	C		C				AY100860
Ipomoea quamoclit	Convolvulaceae	Y	C		C				AY100854
Ipomoea arborescens	Convolvulaceae	Y	C		C				AY100858
Ipomoea pes-tigridis	Convolvulaceae	Y	C		C				AY100873
Ipomoea aquatica	Convolvulaceae	Y	C		C				AY100856
Ipomoea setosa	Convolvulaceae	Y	C		C				AY100857
Ipomoea obscura	Convolvulaceae	Y	C		C				AY100866
Ipomoea tiliacea	Convolvulaceae	Y	C		C				AY100859
Ipomoea wrightii	Convolvulaceae	Y	C		C				AY100855
Ipomoea alba	Convolvulaceae	Y	C		C				AY100861
Stictocardia tilifolia	Convolvulaceae	Y	C		C				AY100868
Stictocardia incomta	Convolvulaceae	Y	C		C				AY100869
Astripomoea malvacea	Convolvulaceae	Y	C		C				AY100863
Astripomoea grantii	Convolvulaceae	Y	C		C				AY100862
Turbina oenotheroides	Convolvulaceae	Y	C		C				AY100870
Turbina corymbosa	Convolvulaceae	Y	C		C				AY100865
Lepistemon owariensis	Convolvulaceae	Y	C		C				AY100867
Maripa repens	Convolvulaceae	Y	C		T				AY100936
Maripa paniculata	Convolvulaceae	Y	C		T				AY100937
Maripa glabra	Convolvulaceae	Y	C		T				AY100935
Dicranostyles ampla	Convolvulaceae	Y	C		T				AY100933
Dicranostyles mildbraediana	Convolvulaceae	Y	C		T				AY100934
Rapona tiliifolia	Convolvulaceae	Y	C		T				AY100927
Calycobolus nutans	Convolvulaceae	Y	C		T				AY100918
Calucobolus glaber	Convolvulaceae	Y	C		T				AY100924
Porana velutina	Convolvulaceae	Y	C		T				AY100919
Porana volubilis	Convolvulaceae	Y	C		T				AY100920
Porana commixta	Convolvulaceae	Y	C		C				AY100946
Metaporana parvifolia	Convolvulaceae	Y	T		T				AY100921
Falkia repens	Convolvulaceae	Y	C		T				AY100917
Dichondra occidentalis	Convolvulaceae	Y	C		T				AY100916
Dipteropeltis poranoides	Convolvulaceae	Y	C		T				AY100926
Iseia luxurians	Convolvulaceae	Y	C		C				AY100896

Table 5.1. (Continued)

Aniseia argentina	Convolvulaceae	Y	C		C	AY100895
Aniseia cernua	Convolvulaceae	Y	C		C	AY100894
Aniseia martinicensis	Convolvulaceae	Y	C		C	AY100893
Odonellia hirtiflora	Convolvulaceae	Y	C		T	AY100897
Tetralocularia pennellii	Convolvulaceae	Y	C		C	AY100898
Merremia peltata	Convolvulaceae	Y	C		C	AY100885
Merremia umbellata	Convolvulaceae	Y	C		C	AY100884
Merremia aegyptia	Convolvulaceae	Y	C		C	AY100875
Merremia hastata	Convolvulaceae	Y	C		C	AY100880
Merremia vitifolia	Convolvulaceae	Y	C		C	AY100876
Merremia dissecta	Convolvulaceae	Y	C		C	AY100874
Operculina sp. Romero 1701	Convolvulaceae	Y	C		C	AY100882
Operculina turpethum	Convolvulaceae	Y	C		C	AY100883
Operculina pteripes	Convolvulaceae	Y	C		C	AY100881
Xenostegia tridentata	Convolvulaceae	Y	C		C	AY100879
Hewittia scandens	Convolvulaceae	Y	C		C	AY100878
Hewittia sublobata	Convolvulaceae	Y	C		C	AY100877
Jacquemontia reclinata	Convolvulaceae	Y	C		T	AY100932
Jacquemontia tamnifolia	Convolvulaceae	Y	C		T	AY100929
Jacquemontia blanchetii	Convolvulaceae	Y	C		T	AY100931
Jacquemontia pentanthos	Convolvulaceae	Y	C		T	AY100928
Jacquemontia ovalifolia	Convolvulaceae	Y	C		T	AY100930
Tridynamia megalantha	Convolvulaceae	Y	C		C	AY100945
Cordisepalum thorelii	Convolvulaceae	Y	C		C	AY100940
Cordisepalum phalanthopetalum	Convolvulaceae	Y	C		C	AY100941
Poranopsis paniculata	Convolvulaceae	Y	C		C	AY100942
Cardiochlamys madagascariensis	Convolvulaceae	N	C		C	AY100943
Dinetus truncatus	Convolvulaceae	Y	C		C	AY100944
Erycibe hellwigii	Convolvulaceae	Y	C		C	AY100938
Erycibe glomerata	Convolvulaceae	Y	C		C	AY100939
Hildebrandtia africana	Convolvulaceae	Y	C		T	AY100902
Hildebrandtia sp	Convolvulaceae	Y	C		T	AY100901
Hildebrandtia promontorii	Convolvulaceae	Y	C		T	AY100900
Hildebrandtia valo	Convolvulaceae	Y	C		T	AY100899
Neuropeltis acuminata	Convolvulaceae	Y	C		T	AY100925
Bonamia media	Convolvulaceae	Y	C		T	AY100922
Bonamia thunbergiana	Convolvulaceae	N	C		T	AY100912
Bonamia spectabilis	Convolvulaceae	N	T		T	AY100911
Seddera hirsuta	Convolvulaceae	Y	C		T	AY100905
Sabaudiella aloysii	Convolvulaceae	Y	C		T	AY100903
Wilsonia humilis	Convolvulaceae	Y	C			AY100914
Wilsonia backhousei	Convolvulaceae	Y	C		T	AY100915
Evolvulus nuttalianus	Convolvulaceae	Y	C		T	AY100907
Evolvulus glomeratus	Convolvulaceae	Y	C		T	AY100906
Breweria rotundifolia	Convolvulaceae	Y	C		T	AY100908
Cressa depressa	Convolvulaceae	Y	C		T	AY100910
Cressa truxillensis	Convolvulaceae	Y	C		T	AY100909
Cladostigma hildebrandtioides	Convolvulaceae	Y	C		T	AY100904
Stylisma patens	Convolvulaceae	N	C		T	AY100913
Itzaea sericea	Convolvulaceae	Y	C		T	AY100923
Cuscuta japonica var. japonica	Convolvulaceae	Y	C		C	DQ852348
Cuscuta europaea	Convolvulaceae	Y	C		T	AY100951
Cuscuta gronovii	Convolvulaceae	N	C	-[14]	C	AY100954
Cuscuta sp. RGO 90-12	Convolvulaceae	Y	C	-	C	AY100955
Cuscuta pentagona	Convolvulaceae	N	C		C	AY100952
Cuscuta sandwichiana	Convolvulaceae	N	G		C	AY100953
Cuscuta sp.	Convolvulaceae	N	G	-		
Humbertia madagascariensis	Convolvulaceae	Y	C		C	AY100948
Montinia caryophyllacea	Convolvulaceae	N	T		T	AY100950
Columnea sp. Linqvist and Albert 30	Gesneriaceae	Y	C			AY128599
Pinguicula gracilis	Lentibulariaceae	N	T		T	AY128605

Table 5.1. (Continued)

Pinguicula grandiflora	Lentibulariaceae	N	T		T			AY128603
Pinguicula ehlersiae	Lentibulariaceae	N	T		T			AY128604
Utricularia pubescens	Lentibulariaceae	N	T		T			AY128602
Utricularia geminiscapa	Lentibulariaceae	N	T		T			AY128601
Utricularia alpine	Lentibulariaceae	N	T		T			AY128600
Panax ginseng	Araliaceae	Y	C		T		C	NC_006290
Daucus carota	Apiaceae	Y	C		T		C	NC_008325
Helianthus annuus	Asteraceae	Y	T		C		C	NC_007977
Lactuca sativa	Asteraceae	Y	T		C		C	NC_007578
Cornus mas	Cornaceae	Y	C		T			AF528870
Hydrangea macrophylla	Hydrangeaceae	Y	C		C			AF528876
Stewartia pseudocamellia	Theaceae	Y	C		T			AF528886
Vitus vinifera	Vitaceae	Y	C		C		C	NC_007957
Gossypium hirsutum	Malvaceae	Y	T		T		C	NC_007944
Gossypium barbadense	Malvaceae	Y	T		T		C	NC_008641
Arabidopsis thaliana	Brassicaceae	Y	C	$+^{3}$	T		C	$+^{9}$ NC_000932
Brassica rapa subsp. Pekinensis	Brassicaceae	Y	T		T			DQ231548
Nasturtium officinale	Brassicaceae	Y	C		T		C	AP009376
Lobularia maritima	Brassicaceae	Y	C		T		C	AP009375
Lepidium virginicum	Brassicaceae	Y	C		T		C	AP009374
Draba nemorosa	Brassicaceae	Y	C		T		C	AP009373
Crucihimalaya wallichii	Brassicaceae	Y	C		T		C	AP009372
Capsella bursa-pastoris	Brassicaceae	Y	C		T		C	AP009371
Barbarea verna	Brassicaceae	Y	C		T		C	AP009370
Arabis hirsuta	Brassicaceae	Y	C		T		C	AP009369
Olimarabidopsis pumila	Brassicaceae	Y	C		T		C	AP009368
Aethionema grandiflorum	Brassicaceae	Y	C		T		C	AP009367
Aethionema cordifolium	Brassicaceae	Y	C		T		C	AP009366
Cicer arientinum	Fabaceae	Y	T		C			AC145820
Medicago truncatula	Fabaceae	Y	T		C			NC_003119
Pisum sativum	Fabaceae	Y	T		T			AY007482
Lotus japonicus	Fabaceae	Y	C		C		C	NC_002694
Glycine max	Fabaceae	Y	C		C		C	NC_007942
Phaseolus vulgaris	Fabaceae	Y	C		T		C	DQ886273
Cucumis sativus	Cucurbitaceae	Y	T		T		C	NC_007144
Populus deltoides	Salicaceae	Y	C		T			X89651
Populus alba	Salicaceae	Y	C		T		C	NC_008235
Euonymus alatus	Celastraceae	Y	C		C			AF528872
Pelargonium hortorum	Geraniaceae	Y	C		C		C	NC_008454
Eucalyptus globulus	Myrtaceae	Y	C		C		C	NC_008115
Oenothera elata	Onagraceae	Y	T		T		C	NC_002693
Oenothera berteroana	Onagraceae	N	T					X07951
Rheum x hybridum	Polygonaceae	Y	T		C			AY007483
Mesembryanthemum crystallinum	Aizoaceae	Y	T		T			U04314
Antirrhinum majus	Plantaginaceae						C	$+^{10}$ Ref: 10
Beta vulgaris	Amaranthaceae	Y	T		C			AB242560
Spinacia oleracea	Amaranthaceae	Y	T		C	$+^{7}$	C	$+^{10}$ NC_002202
Phytolacca americana	Phytolaccaceae	N	T		C			AF528880
Cercidiphyllum japonicum	Cercidiphyllaceae	Y	C		C			AF123833
Jasminum nudiflorum	Oleaceae	Y	C		T		C	NC_008407
Ribes aureum	Grossulariaceae	Y	T		C			AF528883
Gunnera chilensis	Gunneraceae	Y	C		C			AY007478
Nelumbo lutea	Nelumbonaceae	N	C	-	C			AF528877
Nelumbo nucifera	Nelumbonaceae	N	C	-				
Platanus occidentalis	Platanaceae	Y	C		C		C	NC_008335
Morus indica	Moraceae	Y	C		C		C	NC_008359
Citrus sinensis	Rutaceae	Y	T		T		C	NC_008334
Coffea arabica	Rubiaceae	Y	C		T		C	NC_008535
Euptelea polyandra	Eupteleaceae	Y	C		T			AF528873
Nandina domestica	Berberidaceae	N	T		T		C	NC_008336
Mahonia aquifolium	Berberidaceae	Y	T		T			AF528867

Table 5.1. (Continued)

Hydrastis canadensis	Ranunculaceae	Y	T	T		AY007479
Ranunculus macranthus	Ranunculaceae	Y	T	T	C	NC_008796
Pachysandra terminalis	Buxaceae	Y	C	C		AF528878
Trochodendron aralioides	Trochodendraceae	Y	C	C		AF123842
Liriodendron tulipifera	Magnoliaceae	Y	C	C	C	NC_008326
Magnolia stellata	Magnoliaceae	Y	C	C		AY007481
Houttuynia cordata	Saururaceae	Y	C	C		AF528875
Saururus cernuus	Saururaceae	Y	C	C		AF123841
Piper betle	Piperaceae	Y	C	C		AF528881
Piper cenocladum	Piperaceae	Y	C	C	C	NC_008457
Saruma henryi	Aristolochiaceae	Y	C	C		AF528884
Asarum canadense	Aristolochiaceae	Y	C	C		AF123829
Aristolochia macrophylla	Aristolochiaceae	Y	C	T		AF528866
Lactoris fernandeziana	Lactoridaceae	Y	C	C		AF123839
Canella winterana	Canellaceae	Y	C	C		AF528868
Tasmannia lanceolata	Winteraceae	Y	C	C		AF528887
Drimys winteria	Winteraceae	Y	C	C		AF123835
Drimys granadensis	Winteraceae	Y	C	C	C	NC_008456
Calycanthus floridus	Calycanthaceae	Y	C	C	C	NC_004993
Hernandia peltata	Hernandiaceae	N	T	C		AF528874
Petermannia cirrosa	Petermanniaceae	Y	C	T		AY465612
Wurmbea pygmaea	Colchicaceae	Y	T	T		AY465611
Trillium grandiflorum	Melanthiaceae	Y	T	T		AY465610
Tricyrtis sp.	Colchicaceae	Y	T	T		AY465609
Smilax rotundifolia	Smilacaceae	Y	C	T		AY465608
Ripogonum elseyanum	Ripogonaceae	Y	C	T		AY465607
Prosartes trachycarpa	Ruscaceae	Y	T	T		AY465606
Philesia magellanica	Philesiaceae	Y	C	T		AY465605
Tripladenia cunninghamii	Colchicaceae	Y	T	T		AY465604
Medeola virginiana	Liliaceae	Y	C	T		AY465603
Luzuriaga radicans	Luzuriagaceae	Y	T	T		AY465602
Campynema lineare	Campynemataceae	Y	T	T		AY465601
Calochortus apiculatus	Liliaceae	Y	T	T		AY465600
Alstroemeria aurea	Alstroemeriaceae	N	T	T		AY465599
Carludovica drudei	Cyclanthaceae	N	T	T		AY465598
Pandanus copelandii	Pandanaceae	N	T	T		AY465597
Doryanthes palmeri	Doryanthaceae	Y	C	T		AY465596
Agapanthus africanus	Agapanthaceae	Y	T	T		AY465595
Xyris jupicai	Xyridaceae	N	T	T		AY465594
Strelitzia reginae	Strelitziaceae	Y	T	T		AY465593
Sparganium eurycarpum	Sparganiaceae	Y	T	T		AY465592
Kingia australis	Dasypogonaceae	Y	T	T		AY465591
Flagellaria indica	Flagellariaceae	Y	T	T		AY465590
Elegia fenestrata	Restionaceae	N	T	T		AY465589
Ecdeiocolea monostachya	Ecdeiocoleaceae	N	T	T		AY465588
Cyperus papyrus	Cyperaceae	N	T	T		AY465587
Helmholtzia glaberrima	Philydraceae	Y	T	T		AY465586
Sagittaria latifolia	Alismataceae	Y	C	T		AY007484
Scheuchzeria palustris	Scheuchzeriaceae	Y	T	T		AY147547
Butomus umbellatus	Butomaceae	Y	T	T		AY147546
Tofieldia glutinosa	Tofieldiaceae	Y	C	T		AY147548
Spathiphyllum wallisii	Araceae	Y	T	T		AY007487
Acorus calamus	Acoraceae	Y	C	T	C	NC_007407
Acorus americanus	Acoraceae	Y	C	T		DQ069631
Xeronema callistemon	Xeronemataceae	Y	T	T		AY147582
Yucca schidigera	Agavaceae	Y	T	T		DQ069636
Yucca glauca	Agavaceae	Y	T	T		AY147592
Chlorophytum comosum	Agavaceae	Y	T	T		AY147586
Muscari comosum	Hyacinthaceae	Y	T	T		AY147590
Phormium tenax	Hemerocallidaceae	Y	T	T		AY147579
Hermerocallis littorea	Hemerocallidaceae	N	T	T		AY147574

Table 5.1. (Continued)

Asphodelus albus	Asphodelaceae	Y	T	T			AY147567
Phyllostachys pubescens	Poaceae	N	T				DQ908935
Allium cepa	Liliaceae				C	+[9]	AJ278350
Allium porrum	Liliaceae				C	-[9]	AJ278352
Allium sativum	Liliaceae				C	+[9]	AJ278351
Aloe vera	Liliaceae				C	+[9]	AJ278353
Astelia alpina	Liliaceae	Y	C	T			AY147568
Narcissus elegans	Liliaceae	Y	T	T			AY147591
Curculigo capitulata	Liliaceae	Y	C	T			AY147571
Smilacina racemosa	Liliaceae	Y	T	T			AY147588
Lilium superbum	Liliaceae	Y	C	T			AY007480
Anticlea elegans	Liliaceae	N	T	T			AY147553
Stemona tuberosa	Stemonaceae	N	T	T			AY147552
Talbotia elegans	Velloziaceae	Y	T	T			AY147563
Narthecium ossifragum	Nartheciaceae	Y	T	T			AY147550
Dioscorea bulbifera	Dioscoreaceae	Y	T	T			AF123834
Burmannia Capitata	Burmanniaceae	N	T	T			AY147549
Aphyllanthes monspeliensis	Aphyllanthaceae	Y	T	T			AY147584
Cyanastrum cordifolium	Tecophilaeaceae	Y	C	T			AY147572
Phalaenopsis aphrodite	Orchidaceae	Y	T	T	C[α]	-[11]	NC_007499
Coelogyne cristata	Orchidaceae	Y	T	T			AY147570
Orchis rotundifolia	Orchidaceae	Y	T	T			AY147578
Cypripedium passerinum	Orchidaceae	Y	T	T			AY147573
Ixiolirion tataricum	Ixioliriaceae	Y	T	T			AY147576
Alania endlicheri	Boryaceae	Y	T	T			AY147566
Xanthorrhoea resinosa	Xanthorrhoea	Y	T	T			AY147581
Sisyrinchium montanum	Iridaceae	Y	T	T			AY147580
Iris missouriensis	Iridaceae	Y	T	T			AY147575
Blandfordia punicea	Blandfordiaceae	Y	T	T			AY147569
Lanaria lanata	Lanariaceae	Y	T	T			AY147577
Asparagus officinalis	Asparagaceae	Y	T	T			AY147585
Allium textile	Alliaceae	Y	T	T			AY147583
Muilla martima	Alliaceae	Y	T	T			AY147589
Lomandra longifolia	Laxmanniaceae	N	T	T			AY147587
Oryza sativa	Poaceae	N	T	T	T		NC_008155
Oryza nivara	Poaceae	N	T	T	T		NC_005973
Triticum aestivum	Poaceae	N	T	T	T		NC_002762
Secale cereale	Poaceae	N	T	T			X13326
Hordeum vulgare	Poaceae	N	T	T	T		NC_008590
Hordeum murinum	Poaceae				T		AJ278355
Zea mays	Poaceae	N	T	T	T		NC_001666
Saccharum officinarum	Poaceae	N	T	T	T		NC_006084
Sorghum bicolor	Poaceae	N	T	T	T		NC_008602
Agrostis stolonifera	Poaceae	N	T	T	T		NC_008591
Typha latifolia	Typhaceae	Y	T	T			DQ069635
Typha angustifolia	Typhaceae	Y	T	T			AY147564
Ananas comosus	Bromeliaceae	Y	T	T			AY147554
Mayaca fluviatilis	Mayacaceae	N	T	T			AY147559
Roystonea princeps	Arecaceae	Y	T	T			AY147562
Xiphidium caeruleum	Haemodoraceae	Y	T	T			AY147565
Hydrothrix gardneri	Pontederiaceae	Y	T	T			AY147558
Philydrum lanuginosum	Philydraceae	Y	T	T			AY147561
Palisota bogneri	Commelinaceae	Y	T	T			AY147560
Cartonema philydroides	Commelinaceae	N	T	T			AY147555
Dasypogon hookeri	Dasypogonaceae	Y	T	T			AY147556
Ensete ventricosum	Musaceae	Y	T	T			AY147557
Japonolirion osense	Petrosaviaceae	Y	T	T			AY147551
Chloranthus japonicus	Chloranthaceae	Y	T	T			AY007476
Ascarina lucida	Chloranthaceae	N	T	C			AY007474
Amborella trichopoda	Amborellaceae	Y	T	C	C		NC_005086
Trimenia moorei	Trimeniaceae	Y	T	C			AY116652

Table 5.1. (Continued)

Austrobaileya scandens	Austrobaileyaceae	Y	C		C			AY007475
Schisandra chinensis	Schisandraceae	Y	C		C			AY007485
Illicium parviflorum	Schisandraceae	Y	C		C			AF123838
Nymphaea alba	Nymphaeaceae	Y	T		C		T	NC_006050
Nymphaea odorata	Nymphaeaceae	Y	T		C			AF188852
Nuphar advena	Nymphaeaceae	Y	T		C		C	NC_008788
Cabomba caroliniana	Cabombaceae	Y	T		C			AF123830
Ceratophyllum demersum	Ceratophyllaceae	Y	T		C			AF123832
Stangeria eriopus	Stangeriaceae	Y	C		T			AF469721
Ceratozamia miqueliana	Zamiaceae	Y	C		T			AF469715
Zamia furfuracea	Zamiaceae	Y	C		T			AF188847
Zamia latifolia	Zamiaceae	Y	C	+				
Encephalartos barteri	Zamiaceae	Y	C		T			AF469718
Dioon purpusii	Zamiaceae	Y	C		T			AF469717
Dioon endule	Zamiaceae	Y	C	+				
Cycas revoluta	Cycadaceae	Y	C	+	T			AF469716
Bowenia serrulata	Cycadaceae	Y	C		T			AF469713
Ginkgo biloba	Ginkgoaceae	Y	C	+[4]	T			AF123836
Abies lasiocarpa	Pinaceae	Y	C		T			AY664843
Pinus koraiensis	Pinaceae	Y	C		T			NC_004677
Pinus thunbergii	Pinaceae	Y	C	+[5]	T		T	NC_001631
Pseudotsuga menziesii	Pinaceae	Y	C		T			AY664844
Cedrus deodara	Pinaceae	Y	T		T			AF469714
Picea abies	Pinaceae	Y	T					AJ001023
Widdringtonia cedarbergensis	Cupressaceae	Y	T		T			AF528891
Thuja plicata	Cupressaceae	Y	T		T			AF528890
Cunninghamia lanceolata	Cupressaceae	Y	T		T			AF528871
Metasequoia glyptostroboides	Cupressaceae	Y	T		T			AF469719
Juniperus communis	Cupressaceae	N	T		T			AY664848
Taxodium distichum	Cupressaceae	Y	T		T			AF528888
Cephalotaxus harringtonia	Cephalotaxaceae	Y	T		T			AF528869
Sciadopitys verticillata	Sciadopityaceae	N	C	+	T			AY007486
Torreya californica	Taxaceae	Y	T		T			AY664847
Taxus brevifolia	Taxaceae	N	T		T			AF528889
Agathis australis	Araucariaceae	Y	T		T			AF528865
Araucaria bidwillii	Araucariaceae	Y	T		T			AY664846
Saxegothaea conspicua	Podocarpaceae	Y	T		T			AY664845
Phyllocladus alpinus	Podocarpaceae	Y	T		T			AF528879
Podocarpus chinensis	Podocarpaceae	N	T		T			AF469720
Gnetum gnemon	Gnetaceae	N	T		T			AF123837
Welwitschia mirabilis	Welwitschiaceae	Y	T		T			AY116660
Ephedra sinica	Ephedraceae	Y	T		T			AY007477
Adiantum capillus-veneris	Pteridaceae	N	T		C	+[6]	T	NC_004766
Angiopteris evecta	Marattiaceae	Y	T		T		T	NC_008829
Psilotum nudum	Psilotaceae	Y	T		T		T	NC_003386
Huperzia lucidula	Lycopodiaceae	Y	T		T		C	NC_008590
Selaginella uncinata	Selaginellaceae	Y	T		C		C[β]	AB197035
Physcomitrella patens	Funariaceae	Y	T		T		T	NC_005087
Anthoceros formosae	Anthocerotaceae	Y	T		T		T	NC_004543
Marchantia polymorpha	Marchantiaceae	Y	T		T		T	NC_001319
Chara vulgaris	Characeae	N	T		T		T	NC_008097

Note. The nucleotide at the same position as an editing site in tobacco is indicated by a character. Experimental evidence of editing if available is depicted as a + for a confirmed editing site and a – if significant editing was not observed. An accession number from a GenBank submission used to determine the nucleotide at the editing

site position is listed for each species. For psbE the presence of an intact or diverged motif is represented by a Y or N, respectively.

[1]Sasaki et al. 2003

[2]Hirose et al. 1999

[3]Tillich et al. 2005

[4]Kudla and Bock 1999

[5]Wakasugi et al. 1996

[6]Wolf, Rowe, and Hasebe 2004

[7]Bock et al. 1993

[8]Kuntz et al. 1992

[9]Lopez-Serrano et al. 2001

[10]Neckermann et al. 1994

[11]Zeng, Liao, and Chang 2007

[12]Tillich et al. 2006

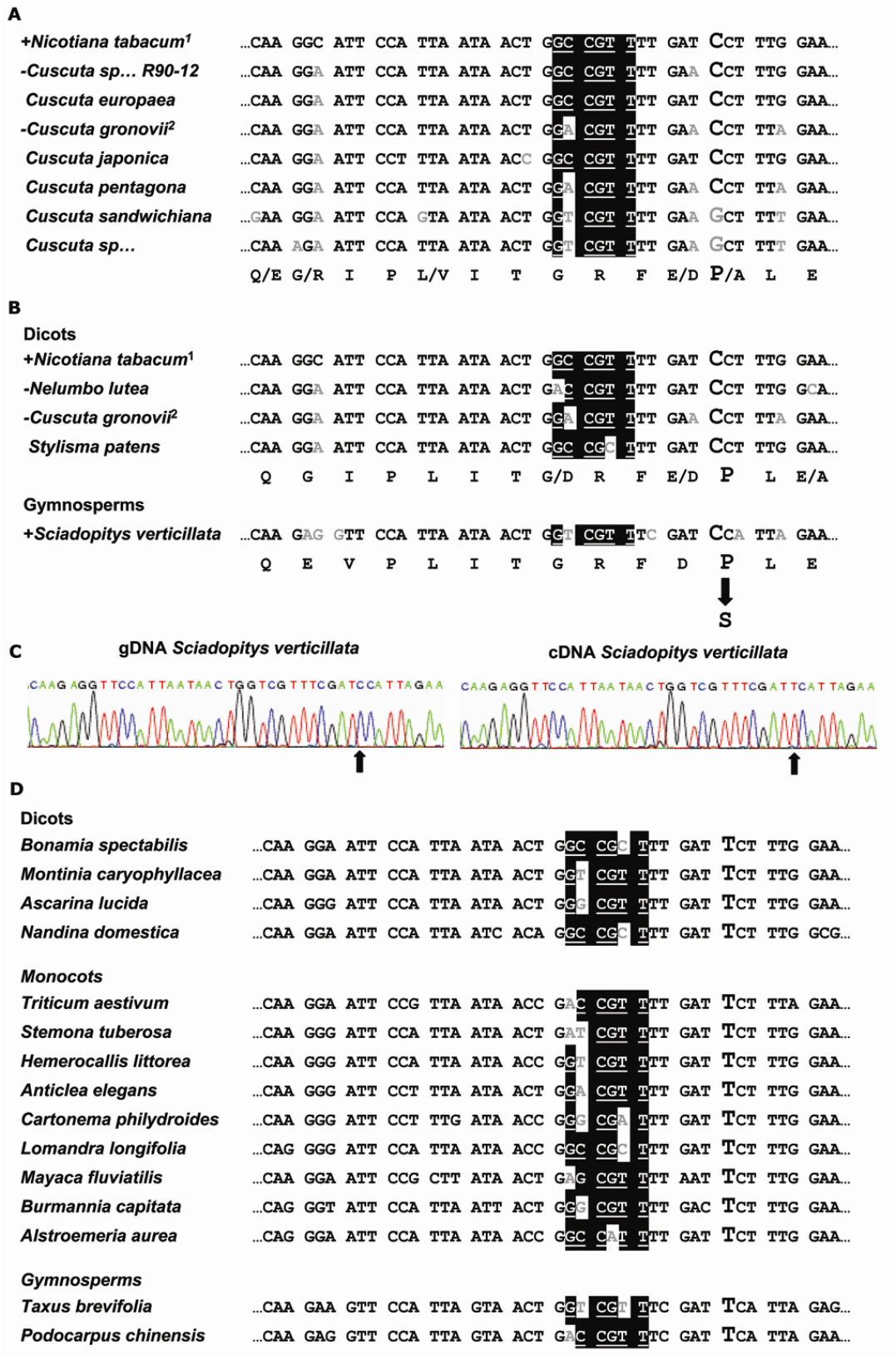
[13]Schmitz-Linneweber et al. 2005

[14] U. Maier, personal communication

[α]putative pseudogene

[β]ATG immediatedly upsteam of the ACG codon

Figure 5.2. Comparison of *psbE* sequences in selected seed plants with differences in the motif. (*A*) Sequences from tobacco and Cuscuta *psbE* are aligned to show the differences in the motif and editing site. (*A* and *B*) A species with or without experimental evidence of editing is signaled by + or – around the editing site respectively; [1](Hirose et al. 1999), [2](U. Maier, personal communication). (*A*, *B*, and *D*) A black box highlights the critical editing cis-element from tobacco. Predicted amino acids are show below DNA sequences. Gray characters depict nucleotides that differ from tobacco sequences. Arrows point to predicted amino acids encoded by edited transcripts. (*B*) A sequence alignment illustrating the different sequences in the motif observed in a few species with a C214. (*C*) Sequence traces displaying *psbE* sequences from cDNA and genomic DNA from *Sciadopitys verticillata*. Filled arrows mark editing sites. (*D*) An alignment of psbE sequences from species with differences in the motif and a T214.


A
+Nicotiana tabacum1 ...CAA GGC ATT CCA TTA ATA ACT GGC CGT TTT GAT CCT TTG GAA...
-Cuscuta sp... R90-12 ...CAA GGA ATT CCA TTA ATA ACT GGC CGT TTT GAA CCT TTG GAA...
Cuscuta europaea ...CAA GGA ATT CCA TTA ATA ACT GGC CGT TTT GAT CCT TTG GAA...
-Cuscuta gronovii2 ...CAA GGA ATT CCA TTA ATA ACT GGA CGT TTT GAA CCT TTA GAA...
Cuscuta japonica ...CAA GGA ATT CCT TTA ATA ACC GGC CGT TTT GAT CCT TTG GAA...
Cuscuta pentagona ...CAA GGA ATT CCA TTA ATA ACT GGA CGT TTT GAA CCT TTA GAA...
Cuscuta sandwichiana ...GAA GGA ATT CCA GTA ATA ACT GGT CGT TTT GAA GCT TTT GAA...
Cuscuta sp... ...CAA AGA ATT CCA TTA ATA ACT GGT CGT TTT GAA GCT TTT GAA...
Q/E G/R I P L/V I T G R F E/D P/A L E
B
Dicots
+Nicotiana tabacum1 ...CAA GGC ATT CCA TTA ATA ACT GGC CGT TTT GAT CCT TTG GAA...
-Nelumbo lutea ...CAA GGA ATT CCA TTA ATA ACT GAC CGT TTT GAT CCT TTG GCA...
-Cuscuta gronovii2 ...CAA GGA ATT CCA TTA ATA ACT GGA CGT TTT GAA CCT TTA GAA...
Stylisma patens ...CAA GGA ATT CCA TTA ATA ACT GGC CGC TTT GAT CCT TTG GAA...
Q G I P L I T G/D R F E/D P L E/A
Gymnosperms
+Sciadopitys verticillata ...CAA GAG GTT CCA TTA ATA ACT GGT CGT TTC GAT CCA TTA GAA...
Q E V P L I T G R F D P L E
S
C
gDNA Sciadopitys verticillata
cDNA Sciadopitys verticillata
D
Dicots
Bonamia spectabilis ...CAA GGA ATT CCA TTA ATA ACT GGC CGC TTT GAT TCT TTG GAA...
Montinia caryophyllacea ...CAA GGA ATT CCA TTA ATA ACT GGT CGT TTT GAT TCT TTG GAA...
Ascarina lucida ...CAA GGG ATT CCA TTA ATA ACT GGG CGT TTT GAT TCT TTG GAA...
Nandina domestica ...CAA GGA ATT CCA TTA ATC ACA GGC CGC TTT GAT TCT TTG GCG...
Monocots
Triticum aestivum ...CAA GGA ATT CCG TTA ATA ACC GAC CGT TTT GAT TCT TTA GAA...
Stemona tuberosa ...CAA GGG ATT CCA TTA ATA ACT GAT CGT TTT GAT TCT TTG GAA...
Hemerocallis littorea ...CAA GGG ATT CCA TTA ATA ACC GGT CGT TTT GAT TCT TTG GAA...
Anticlea elegans ...CAA GGG ATT CCT TTA ATA ACT GGA CGT TTT GAT TCT TTG GAA...
Cartonema philydroides ...CAA GGG ATT CCT TTG ATA ACC GGG CGA TTT GAT TCT TTG GAA...
Lomandra longifolia ...CAG GGG ATT CCA TTA ATA ACC GGC CGC TTT GAT TCT TTG GAA...
Mayaca fluviatilis ...CAA GGA ATT CCG CTT ATA ACT GAG CGT TTT AAT TCT TTG GAA...
Burmannia capitata ...CAG GGT ATT CCA TTA ATT ACT GGG CGT TTT GAC TCT TTG GAA...
Alstroemeria aurea ...CAG GGA ATT CCA TTA ATA ACC GGC CAT TTT GAT TCT TTG GAA...
Gymnosperms
Taxus brevifolia ...CAA GAA GTT CCA TTA GTA ACT GGT CGT TTC GAT TCA TTA GAG...
Podocarpus chinensis ...CAA GAG GTT CCA TTA GTA ACT GAC CGT TTC GAT TCA TTA GAA...

each genus. To prevent bias that could occur if a species had undergone a recent loss of an editing site, if there was an equal number of species from a genus with a C and a T at 214 we selected a member with a C for our second data set. Likewise, we added a species to the database that has the motif in those cases where equal numbers of species in the genus have the precise motif as have nucleotide differences in the motif. Using this data set we examined 238 different genera (data not shown). Of these 116 (49 %) have a C214 and 122 (51 %) have a T214. Of the members that had a C214 only 6 (5 %) have changes in the motif. From this second data set, 34 (28 %) genera have both a T14 and exhibit differences in the motif.

Many of the genera represented in existing databases are from the same plant family. Therefore to examine whether this fact might cause bias, we created a third data set containing one representative member from each plant family (Table 5.2). If the family had equal numbers of members with a C214 and a T214 we included a member with a C214. If the family had an equal number of species with changes in the motif as species with an intact motif, a species was selected that had the motif. Of the species chosen from each family, 51 (39 %) have a C214 and 80 (61 %) have a T214. Only 2 (4 %) of members that have a C214 also have a difference in the motif compared to 16 (20 %) from members with a T214.

There may be some intrinsic bias even in the plant family dataset due to the sequences available and varying mutation rates. To limit this bias, we investigated if the species in the database comprised of one representative member of a plant family conserved a different 6 nt sequence (Table 5.2). We chose the GAGCCG sequence at positions 174-179 from the initiation codon in tobacco chloroplasts because of its similarity with the previous motif and its location upstream of the editing cis-element. We found that of the 51 plants with a C target of editing, 18 (35 %) had a change in the GAGCCG sequence. Thirty (37 %) plants that have a T 214 also have a difference

Table 5.2. Editing sites present in homologous chloroplast *psbE* genes from a selected member of each plant family.

		psbE		
Family	species	motif	214	GAGCCG
Solanaeae	*Nicotiana tabacum*	Y	C	Y
Convolvulaceae	*Ipomoea batatas*	Y	C	Y
Gesneriaceae	*Columnea sp.*	Y	C	Y
Lentibulariaceae	*Pinguicula gracilis*	N	T	Y
Araliaceae	*Panax ginseng*	Y	C	N
Apiaceae	*Daucus carota*	Y	C	N
Asteraceae	*Helianthus annuus*	Y	T	N
Cornaceae	*Cornus mas*	Y	C	Y
Hydrangeaceae	*Hydrangea macrophylla*	Y	C	Y
Theaceae	*Stewartia pseudocamellia*	Y	C	Y
Vitaceae	*Vitus vinifera*	Y	C	Y
Malvaceae	*Gossypium hirsutum*	Y	T	Y
Brassicaceae	*Arabidopsis thaliana*	Y	C	Y
Fabaceae	*Glycine max*	Y	C	Y
Cucurbitaceae	*Cucumis sativus*	Y	T	Y
Salicaceae	*Populus alba*	Y	C	Y
Celastraceae	*Euonymus alatus*	Y	C	N
Geraniaceae	*Pelargonium hortorum*	Y	C	N
Myrtaceae	*Eucalyptus globulus*	Y	C	Y
Onagraceae	*Oenothera elata*	Y	T	Y
Polygonaceae	*Rheum x hybridum*	Y	T	Y
Aizoaceae	*Mesembryanthemum crystallinum*	Y	T	Y
Amaranthaceae	*Beta vulgaris*	Y	T	Y
Phytolaccaceae	*Phytolacca americana*	N	T	Y
Cercidiphyllaceae	*Cercidiphyllum japonicum*	Y	C	Y
Oleaceae	*Jasminum nudiflorum*	Y	C	Y
Grossulariaceae	*Ribes aureum*	Y	T	Y
Gunneraceae	*Gunnera chilensis*	Y	C	N
Nelumbonaceae	*Nelumbo lutea*	N	C	Y
Platanaceae	*Platanus occidentalis*	Y	C	Y
Moraceae	*Morus indica*	Y	C	Y
Rutaceae	*Citrus sinensis*	Y	T	Y
Rubiaceae	*Coffea Arabica*	Y	C	N

Table 5.2. (Continued)

Eupteleaceae	*Euptelea polyandra*	Y	C	Y
Berberidaceae	*Mahonia aquifolium*	Y	T	Y
Ranunculaceae	*Ranunculus macranthus*	Y	T	N
Buxaceae	*Pachysandra terminalis*	Y	C	Y
Trochodendraceae	*Trochodendron aralioides*	Y	C	N
Magnoliaceae	*Liriodendron tulipifera*	Y	C	Y
Saururaceae	*Saururus cernuus*	Y	C	Y
Piperaceae	*Piper betle*	Y	C	Y
Aristolochiaceae	*Asarum canadense*	Y	C	Y
Lactoridaceae	*Lactoris fernandeziana*	Y	C	Y
Canellaceae	*Canella winterana*	Y	C	Y
Winteraceae	*Tasmannia lanceolata*	Y	C	Y
Calycanthaceae	*Calycanthus floridus*	Y	C	Y
Hernandiaceae	*Hernandia peltata*	N	T	Y
Acoraceae	*Acorus calamus*	Y	C	N
Agapanthaceae	*Agapanthus africanus*	Y	T	Y
Boryaceae	*Alania endlicheri*	Y	T	Y
Alliaceae	*Allium textile*	Y	T	N
Alstroemeriaceae	*Alstroemeria aurea*	N	T	Y
Bromeliaceae	*Ananas comosus*	Y	T	N
Aphyllanthaceae	*Aphyllanthes monspeliensis*	Y	T	N
Asparagaceae	*Asparagus officinalis*	Y	T	Y
Asphodelaceae	*Asphodelus albus*	Y	T	Y
Blandfordiaceae	*Blandfordia punicea*	Y	T	Y
Burmanniaceae	*Burmannia Capitata*	N	T	N
Butomaceae	*Butomus umbellatus*	Y	T	Y
Campynemataceae	*Campynema lineare*	Y	T	Y
Cyclanthaceae	*Carludovica drudei*	N	T	Y
Tecophilaeaceae	*Cyanastrum cordifolium*	Y	C	N
Cyperaceae	*Cyperus papyrus*	N	T	N
Dasypogonaceae	*Dasypogon hookeri*	Y	T	Y
Dioscoreaceae	*Dioscorea bulbifera*	Y	T	Y
Doryanthaceae	*Doryanthes palmeri*	Y	C	N
Ecdeiocoleaceae	*Ecdeiocolea monostachya*	N	T	N
Restionaceae	*Elegia fenestrata*	N	T	N
Musaceae	*Ensete ventricosum*	Y	T	Y
Flagellariaceae	*Flagellaria indica*	Y	T	N

Table 5.2. (Continued)

Philydraceae	*Helmholtzia glaberrima*	Y	T	Y
Pontederiaceae	*Hydrothrix gardneri*	Y	T	Y
Iridaceae	*Iris missouriensis*	Y	T	Y
Ixioliriaceae	*Ixiolirion tataricum*	Y	T	Y
Petrosaviaceae	*Japonolirion osense*	Y	T	Y
Lanariaceae	*Lanaria lanata*	Y	T	Y
Liliaceae	*Lilium superbum*	Y	C	Y
Laxmanniaceae	*Lomandra longifolia*	N	T	Y
Luzuriagaceae	*Luzuriaga radicans*	Y	T	Y
Mayacaceae	*Mayaca fluviatilis*	N	T	N
Hyacinthaceae	*Muscari comosum*	Y	T	Y
Nartheciaceae	*Narthecium ossifragum*	Y	T	N
Orchidaceae	*Orchis rotundifolia*	Y	T	Y
Commelinaceae	*Palisota bogneri*	Y	T	N
Pandanaceae	*Pandanus copelandii*	N	T	Y
Petermanniaceae	*Petermannia cirrosa*	Y	C	Y
Philesiaceae	*Philesia magellanica*	Y	C	Y
Hemerocallidaceae	*Phormium tenax*	Y	T	Y
Ruscaceae	*Prosartes trachycarpa*	Y	T	Y
Ripogonaceae	*Ripogonum elseyanum*	Y	C	Y
Arecaceae	*Roystonea princeps*	Y	T	Y
Alismataceae	*Sagittaria latifolia*	Y	C	Y
Scheuchzeriaceae	*Scheuchzeria palustris*	Y	T	Y
Smilacaceae	*Smilax rotundifolia*	Y	C	Y
Sparganiaceae	*Sparganium eurycarpum*	Y	T	N
Araceae	*Spathiphyllum wallisii*	Y	T	N
Stemonaceae	*Stemona tuberosa*	N	T	Y
Strelitziaceae	*Strelitzia reginae*	Y	T	Y
Velloziaceae	*Talbotia elegans*	Y	T	Y
Tofieldiaceae	*Tofieldia glutinosa*	Y	C	Y
Melanthiaceae	*Trillium grandiflorum*	Y	T	Y
Colchicaceae	*Tripladenia cunninghamii*	Y	T	Y
Poaceae	*Triticum aestivum*	N	T	N
Typhaceae	*Typha latifolia*	Y	T	N
Xanthorrhoea	*Xanthorrhoea resinosa*	Y	T	Y
Xeronemataceae	*Xeronema callistemon*	Y	T	Y
Haemodoraceae	*Xiphidium caeruleum*	Y	T	Y

Table 5.2. (Continued)

Xyridaceae	*Xyris jupicai*	N	T	N
Agavaceae	*Yucca glauca*	Y	T	Y
Chloranthaceae	*Chloranthus japonicus*	Y	T	Y
Amborellaceae	*Amborella trichopoda*	Y	T	Y
Trimeniaceae	*Trimenia moorei*	Y	T	N
Austrobaileyaceae	*Austrobaileya scandens*	Y	C	N
Schisandraceae	*Schisandra chinensis*	Y	C	N
Nymphaeaceae	*Nymphaea alba*	Y	T	N
Cabombaceae	*Cabomba caroliniana*	Y	T	N
Ceratophyllaceae	*Ceratophyllum demersum*	Y	T	N
Stangeriaceae	*Stangeria eriopus*	Y	C	N
Zamiaceae	*Dioon purpusii*	Y	C	N
Cycadaceae	*Cycas revoluta*	Y	C	N
Ginkgoaceae	*Ginkgo biloba*	Y	C	N
Pinaceae	*Pinus thunbergii*	Y	C	N
Cupressaceae	*Thuja plicata*	Y	T	N
Cephalotaxaceae	*Cephalotaxus harringtonia*	Y	T	N
Sciadopityaceae	*Sciadopitys verticillata*	N	C	N
Taxaceae	*Torreya californica*	Y	T	N
Araucariaceae	*Araucaria bidwillii*	Y	T	N
Podocarpaceae	*Phyllocladus alpinus*	Y	T	N
Gnetaceae	*Gnetum gnemon*	N	T	N
Welwitschiaceae	*Welwitschia mirabilis*	Y	T	N
Ephedraceae	*Ephedra sinica*	Y	T	N

Note. A C or T represents the nucleotide at the same position as the editing site in tobacco. The presence of an intact or diverged cis-element, GCCGTT is represented by a Y or N, respectively. The presence of an intact or different 6 nt sequence, GAGCCG upstream of the cis-element at positions 174-179 is represented by a Y or N respectively.

in the sequence. Thus there is no correlation with absence or presence of the C214 with differences in this particular sequence.

Cuscuta sp. frequently have differences in the motif and do not edit a C214

Of the species identified with differences in the motif 4 (50 %) are in the genus *Cuscuta*. RNA and DNA were obtained from two *Cuscuta* species to determine whether members of the genera of parasitic plants edit a psbE C214. PCR products were sequenced that contained *psbE* sequences from cDNA and genomic DNA templates. One *Cuscuta sp.* had a *psbE* sequence identical to one listed as *Cuscuta sp. R90-12* from (Stefanovic et al., 2002). This species of *Cuscuta* has a *psbE* sequence with the motif and C214. Sequences of cDNAs from *psbE* transcripts also have C214 indicating that the analyzed transcripts were not edited. The genomic DNA and cDNA sequence of *psbE* from the other *Cuscuta sp.* was found to have a G214.

Sciadopitys but not Nelumbo edits a C214 despite a difference in the cis-element

In order to determine whether other species with a change in the motif can edit C214, cDNA and genomic DNA were isolated from leaf tissues of *Nelumbo nucifera*, *Nelumbo lutea*, and *Sciadopitys verticillata*. cDNA sequences obtained from both *Nelumbo* species were found to contain a C214, indicating that neither species is capable of editing the C214. This is consistent with the hypothesized critical nature of the motif for tobacco based on observations of *in vitro* editing in *psbE* substrates. *PsbE* sequences from *Sciadopitys verticillata* cDNAs have a T214, indicating editing of C214 despite a difference in the sequence of the motif. Either editing of C214 in this species requires a sequence dissimilar from what is required in tobacco or the factor can tolerate a change in the motif, unlike the factor in tobacco.

Phylogeny of psbE C214

We explored reports of molecular phylogeny of spermatophytes to model the evolution of psbE C214 editing. The editing status and the presence and absence of the

motif was fitted to a phylogenic tree containing members from each of the major branches of Spermatophyta (Bowe et al., 2000). The distribution of editing sites and intact motifs suggests two independent origins for editing of *psbE* C214 (Figure 5.3). We discovered two independent losses of the editing site, presumably by genomic C-to-T mutation, and four independent alterations in the motif within gymnosperms (Figure 5.3). Many angiosperms have a T214 and/or differences in the motif but were not fitted on the phylogenic tree due to the lack of firmly established molecular phylogeny for many of the species (Table 5.1). Many plant families and genera contain members that have a C to T mutation at the editing site. A faster C to T mutation rate at editing targets compared to other Cs has been observed in mitochondria (Shields & Wolfe, 1997). This supports the hypothesis that many individual editing sites are dispensable.

Phylogenetic analysis places *Sciadopitys verticillata* within the conifer II clade described by (Bowe et al., 2000) in which all *psbE* sequences from other studied members have a T214 (Figure 5.4). *Sciadopitys verticillata* has a change in the motif that was found to be inimical for editing *in vitro* of NTpsbE C214 (Hayes & Hanson, 2007b). *Taxus brevifolia* is also within this clade and has the same nucleotide difference in the motif as does *Sciadopitys verticillata*, possibly indicating a common origin (Figure 5.3). It is likely that a common ancestor for the conifer II clade had a T214 in *psbE*. The different sequence constraints for editing in *Sciadopitys verticillata* C214 are probably due to an independent origin of editing at this site (Figure 5.3).

All Gnetophyta have a T214, indicating that a common ancestor to the clade probably acquired a C-to-T mutation in the chloroplast genome. Both *Welwitschia mirabilis* and *Ephedra sinica* have no changes in the motif, even though they have a T214, suggesting that presence of the motif is the ancestral state of the Gnetophyta. *Gnetum gnemon* not only has a T214 but also has a change in the 6 nucleotide motif.

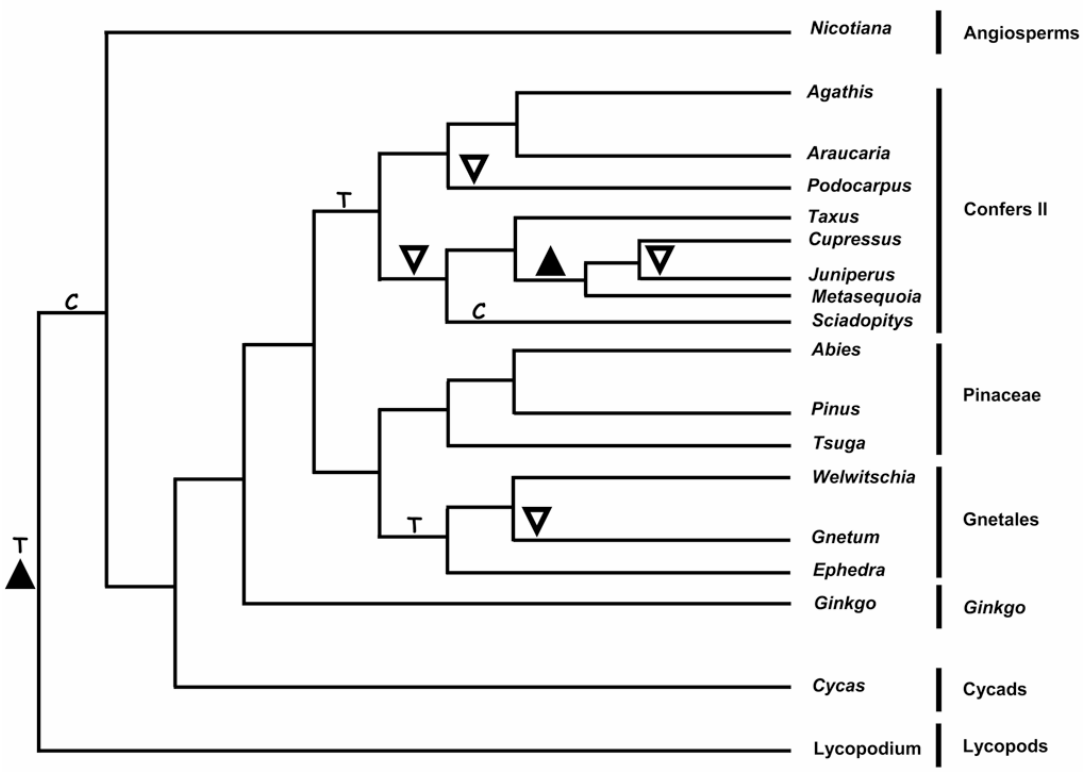


Figure 5.3. A diagram depicts a theoretical model of editing site and motif co-evolution. A T or C symbolizes a common ancestor that had a T214 or a C214 in *psbE* respectively. The presence or absence of the unchanged motif is denoted by ▲ or ∇ respectively. The phylogenic tree was based on (Bowe, Coat, and dePamphilis 2000) and the lengths of branches do not correspond with phylogenetic differences.

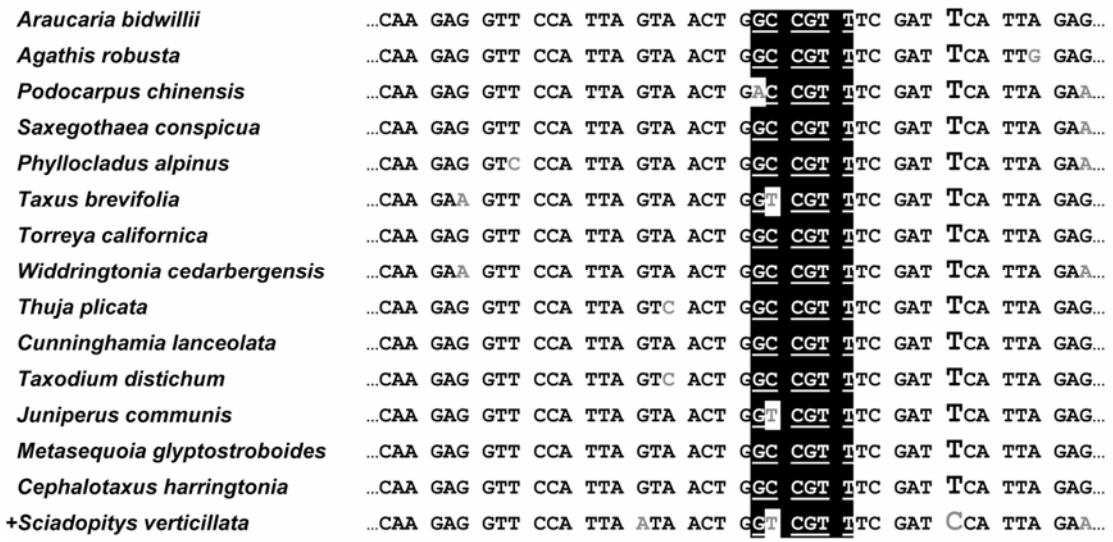


Figure 5.4. An alignment of *psbE* sequences around the editing site in *Sciadopitys verticillata* and homologous sequences in other members of the Conifer II clade defined by Bowe, Coat, and dePamphilis, (2000). The nucleotide at the editing site position is indicated by a character in a larger font. Nucleotides that are different from *Araucaria bidwillii* are represented by grey characters. The 6 nucleotide motif necessary for editing of a tobacco substrate *in vitro* is depicted by a black box and is underlined.

Therefore it is likely that a common ancestor to *Gnetum gnemon* had a T14 and accumulated a change in the motif (Figure 5.3).

The presence of different conserved sequences suggest independent emergence of editing of two sites in psbL and ndhD

We wanted to determine whether other editing sites have critical conserved sequences around C targets of editing. If different species exhibit different conserved sequences near the same editing site, then the editing events may have evolved independently. We chose psbL C2 and ndhD C2 for detailed examination because they are edited in a diverse group of species and they are near the 5′ UTR so that the cis-element would likely be present in a non-coding region with less constraint on sequence changes than cis-elements within a coding region.

PsbL sequences available from plants were aligned and grouped according to the presence of a C or T at the editing site. The sequence TAnAnAnCTA from -10 to -1 is conserved in all spermatophytes examined and could be involved in processes that are distinct from editing. We discovered a putative editing cis-element ATnnnTnGnG from -12 to -4 as a region of absolute conservation within the 5′ UTR in spermatophytes with a C2 (Figure 5.5). Neither *Adiantum capillus-veneris* nor *Selaginella uncinata*, both of which have a C2, contain the putative cis-element but share some common sequences with each other.

We also examined ndhD C2 sequences for putative cis-elements implicated in editing. The sequence TnnnTTTA from -7 to -4 is conserved in both the lycopodiophyte, *Huperzia lucidula*, which has a C2 as well as several angiosperms that have a T2 (Figure 5.6), and may not be relevant to editing. *Selaginella uncinata*, another lycopodiophyte, does not contain the same sequence from -7 to -4 in *ndhD*. The sequence TATnTnnT from -16 to -9 is conserved in seed plants with ndhD C2, but is not present in *Selaginella uncinata* or *Huperzia lucidula*. Therefore editing in

Figure 5.5. Sequence alignments showing conserved sequences in the 5′ UTR of *psbL*. (A) Alignment of *psbL* sequences from a lycopod and a fern. (A and B) A species with or without experimental evidence of editing is signaled by + or – around the editing site respectively; [1](Wolf, Rowe, and Hasebe 2004), [2](Hirose et al. 1999), [3](Bock et al. 1993). Black and gray characters represent absolutely conserved and divergent nucleotides, respectively, between the lycopod and fern (A), angiosperms with a C214 (B), or angiosperms (C). (A, B, and C) The nucleotide at the editing site position is indicated by a character with larger font. (B) Alignment of *psbL* sequences from angiosperms. Nucleotides are numbered above the sequence according from position from the editing site. Characters representing nucleotides conserved only in species with a C214 or all angiosperms are signaled by * and ^ respectively. A box highlights a region of sequence that contains the cis-element identified by (Chaudhuri and Maliga 1996). (C) Alignment of *psbL* sequences from angiosperms with a T214.

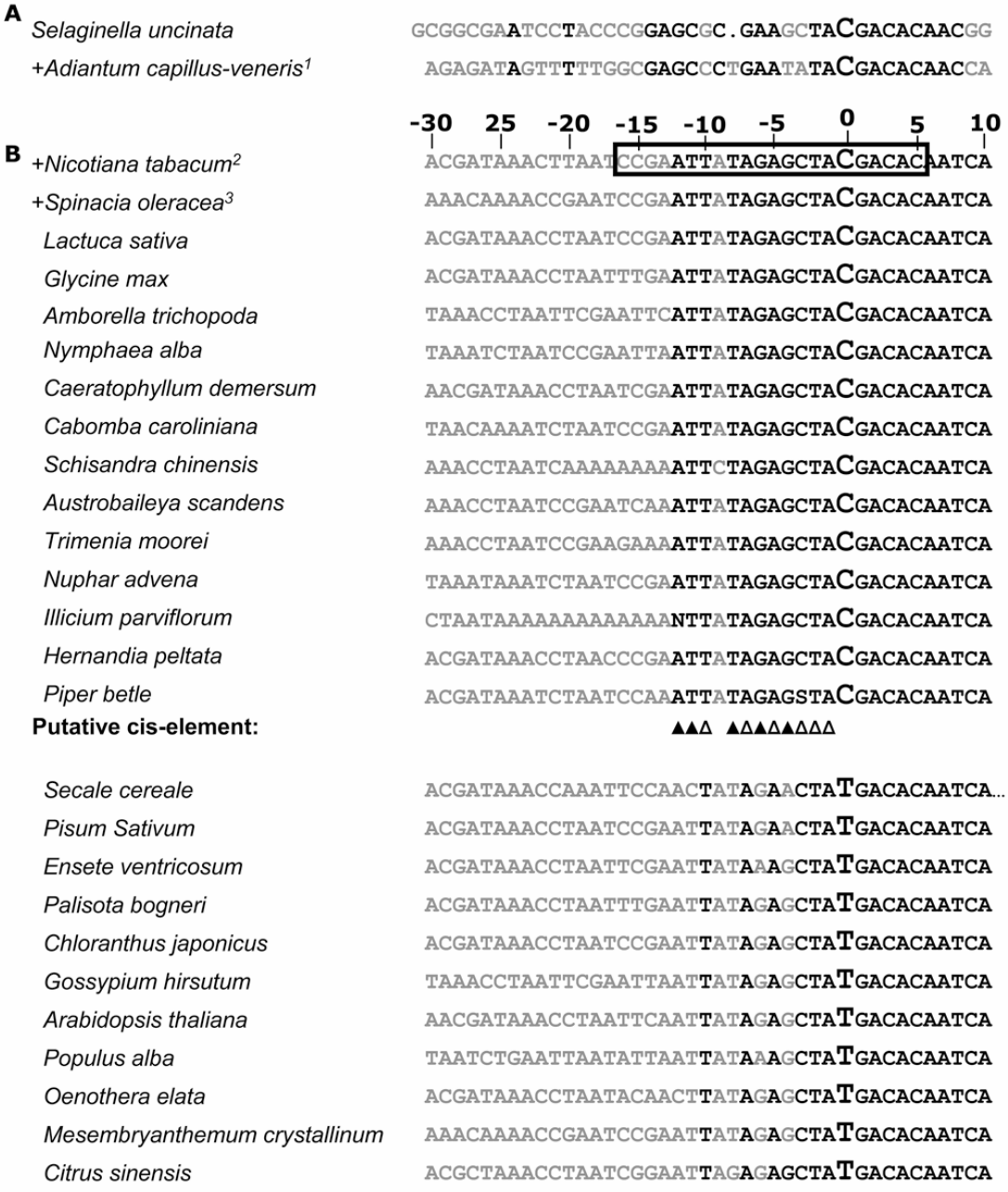
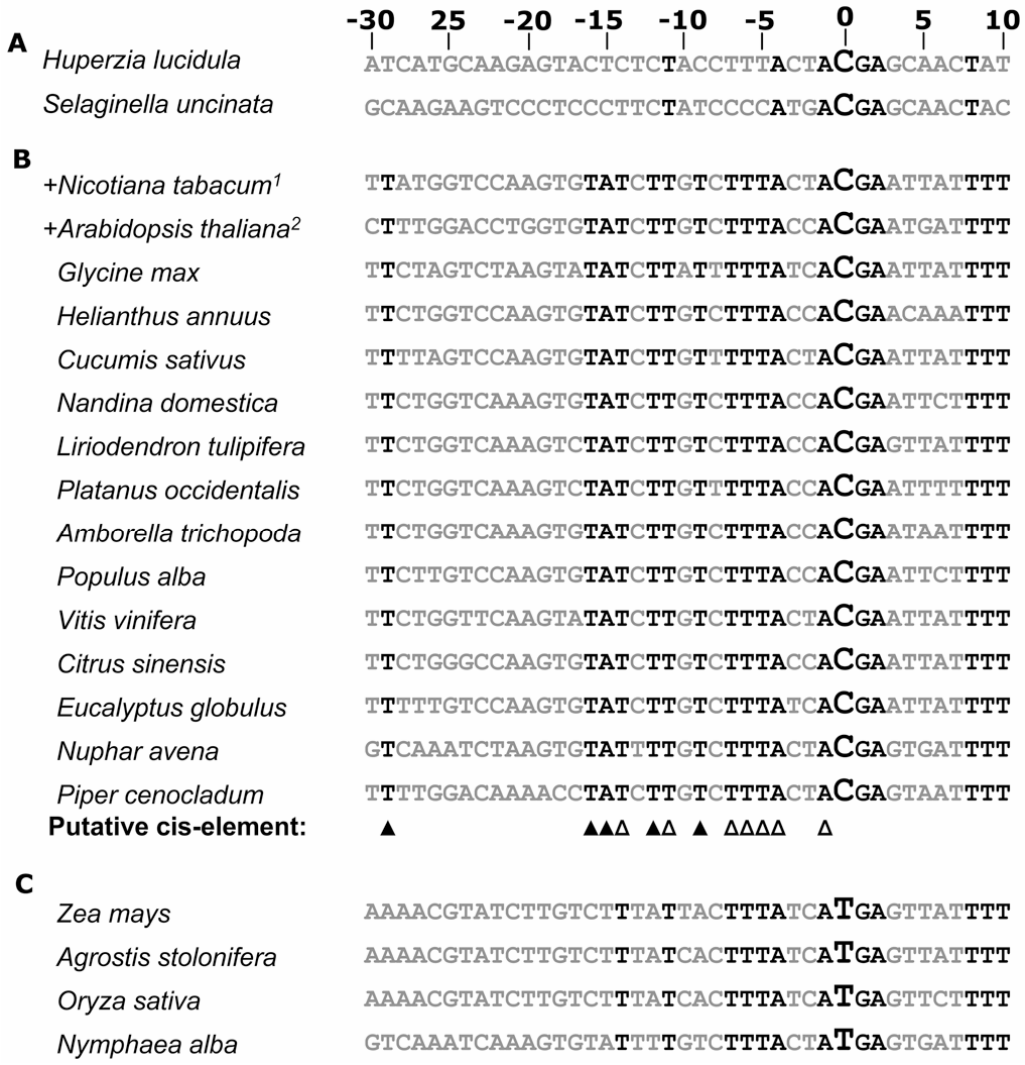

A
Selaginella uncinata
+Adiantum capillus-veneris[1]
-30 25 -20 -15 -10 -5 0 5 10
B
+Nicotiana tabacum[2]
+Spinacia oleracea[3]
Lactuca sativa
Glycine max
Amborella trichopoda
Nymphaea alba
Caeratophyllum demersum
Cabomba caroliniana
Schisandra chinensis
Austrobaileya scandens
Trimenia moorei
Nuphar advena
Illicium parviflorum
Hernandia peltata
Piper betle
Putative cis-element:
Secale cereale
Pisum Sativum
Ensete ventricosum
Palisota bogneri
Chloranthus japonicus
Gossypium hirsutum
Arabidopsis thaliana
Populus alba
Oenothera elata
Mesembryanthemum crystallinum
Citrus sinensis

Figure 5.6. Sequence alignments showing conserved sequences in the 5′ UTR of *ndhD*. (A) Alignment of *ndhD* sequences from two lycopods. Nucleotides are numbered above the sequence according from position from the editing site. A species with or without experimental evidence of editing is signaled by + or – around the editing site respectively; [1](Hirose et al. 1999), [2](Lopez-Serrano et al. 2001). Black and gray characters represent absolutely conserved and divergent nucleotides, respectively, between lycopods (A), angiosperms with a C214 (B), or angiosperms (C). (A, B, and C) The nucleotide at the editing site position is indicated by a character with larger font. (B) Alignment of *ndhD* sequences from angiosperms. Characters representing nucleotides conserved only in species with a C214 or all angiosperms are signaled by * and ^ respectively. (C) Alignment of *ndhD* sequences from angiosperms with a T214.


-30 25 -20 -15 -10 -5 0 5 10
A
Huperzia lucidula ATCATGCAAGAGTACTCTCTACCTTTACTACGAGCAACTAT
Selaginella uncinata GCAAGAAGTCCCTCCCTTCTATCCCCATGACGAGCAACTAC
B
+Nicotiana tabacum1 TTATGGTCCAAGTGTATCTTGTCTTTACTACGAATTATTTT
+Arabidopsis thaliana2 CTTTGGACCTGGTGTATCTTGTCTTTACCACGAATGATTTT
Glycine max TTCTAGTCTAAGTATATCTTATTTTTATCACGAATTATTTT
Helianthus annuus TTCTGGTCCAAGTGTATCTTGTCTTTACCACGAACAAATTT
Cucumis sativus TTTTAGTCCAAGTGTATCTTGTTTTTACTACGAATTATTTT
Nandina domestica TTCTGGTCAAAGTGTATCTTGTCTTTACCACGAATTCTTTT
Liriodendron tulipifera TTCTGGTCAAAGTCTATCTTGTCTTTACCACGAGTTATTTT
Platanus occidentalis TTCTGGTCAAAGTCTATCTTGTTTTTACCACGAATTTTTTT
Amborella trichopoda TTCTGGTCAAAGTGTATCTTGTCTTTACCACGAATAATTTT
Populus alba TTCTTGTCCAAGTGTATCTTGTCTTTACCACGAATTCTTTT
Vitis vinifera TTCTGGTTCAAGTATATCTTGTCTTTACTACGAATTATTTT
Citrus sinensis TTCTGGGCCAAGTGTATCTTGTCTTTACCACGAATTATTTT
Eucalyptus globulus TTTTTGTCCAAGTGTATCTTGTCTTTATCACGAATTATTTT
Nuphar avena GTCAAATCTAAGTGTATTTTGTCTTTACTACGAGTGATTTT
Piper cenocladum TTTTGGACAAAACCTATCTTGTCTTTACTACGAGTAATTTT
Putative cis-element: ▲ ▲▲Δ ▲Δ ▲ ΔΔΔΔ Δ
C
Zea mays AAAACGTATCTTGTCTTTATTACTTTATCATGAGTTATTTT
Agrostis stolonifera AAAACGTATCTTGTCTTTATCACTTTATCATGAGTTATTTT
Oryza sativa AAAACGTATCTTGTCTTTATCACTTTATCATGAGTTCTTTT
Nymphaea alba GTCAAATCAAAGTGTATTTTGTCTTTACTATGAGTGATTTT

Selaginella and *Huperzia* probably requires a different cis-element than C targets in spermatophytes.

Lack of editing in vitro of exogenous Nicotiana tabacum psbE C214 in maize

Many monocots do not have a T214 and are therefore would not be expected to have the capacity to edit an exogenous psbE C214 (Supplementary Table 1). We examined whether the NTpsbE C214 substrate could be edited *in vitro* using chloroplast extracts from *Zea mays*. We found that the maize extracts were not competent for editing of NTpsbE C214 although they can edit a substrate containing NTrpoB C473 *in vitro* (Figure 5.7). Maize chloroplasts contain a C target of editing that aligns with NTrpob C473. The maize chloroplasts extracts do not contain critical factors for editing C214 in *psbE* transcripts, either due to the lack of factors in maize chloroplasts necessary for editing of psbE C214 or a loss of functional factors during the extract preparation (Table 5.3).

DISCUSSION

The general C-to-U editing mechanism made its debut in an ancestor common to all Embryophyta. The chloroplasts of pteridophytes, hepaticophytes, bryophytes, and anthocerotophytes have a greater number of editing sites than examined spermatophytes (Tillich et al., 2006). Only a fraction of those sites are in locations that align with sites discovered in spermatophytes. Our data supports that sites in some species that occur at the same location have different conserved sequences around the site, indicating they are likely the result of multiple evolutionary events. Different nucleotides are conserved 5' of the C targets in the start codon of *psbL* and *ndhD* compared with homologous genes in members of Lycopodiophyta and Pteridophyta. Therefore the editing targets shared between two species are probably not all derived from a common origin. This supports that the cis-elements and corresponding editing

Figure 5.7. Maize chloroplast extracts do not edit tobacco *psbE* substrates *in vitro*. (A) Alignment of tobacco and four species where editing of an exogenous NTpsbE C214 has been investigated. A black box highlights the critical editing cis-element from tobacco. Gray characters depict nucleotides that differ from tobacco sequences. Predicted amino acids are shown below DNA sequences. Arrows point to predicted amino acids encoded by edited transcripts. (B) Image of an electrophoretogram from a poisoned primer extension reaction depicting editing *in vitro* of an NTrpoB C473 substrate using both maize and tobacco chloroplast extracts in the same conditions. (B and C) Bands are labeled according to the template extended. Lanes O and U are free oligonucleotide and unedited control reactions, respectively. Lanes labeled Tobacco and Maize are duplicate reactions using the chloroplast extracts of the indicated species. Editing percentages are indicated below lanes. Polymerase read-through in Lane U was subtracted from experimental samples and editing fixed to 0%. (C) Picture of an electrophoretagram showing editing in NTpsbE C214 substrates.

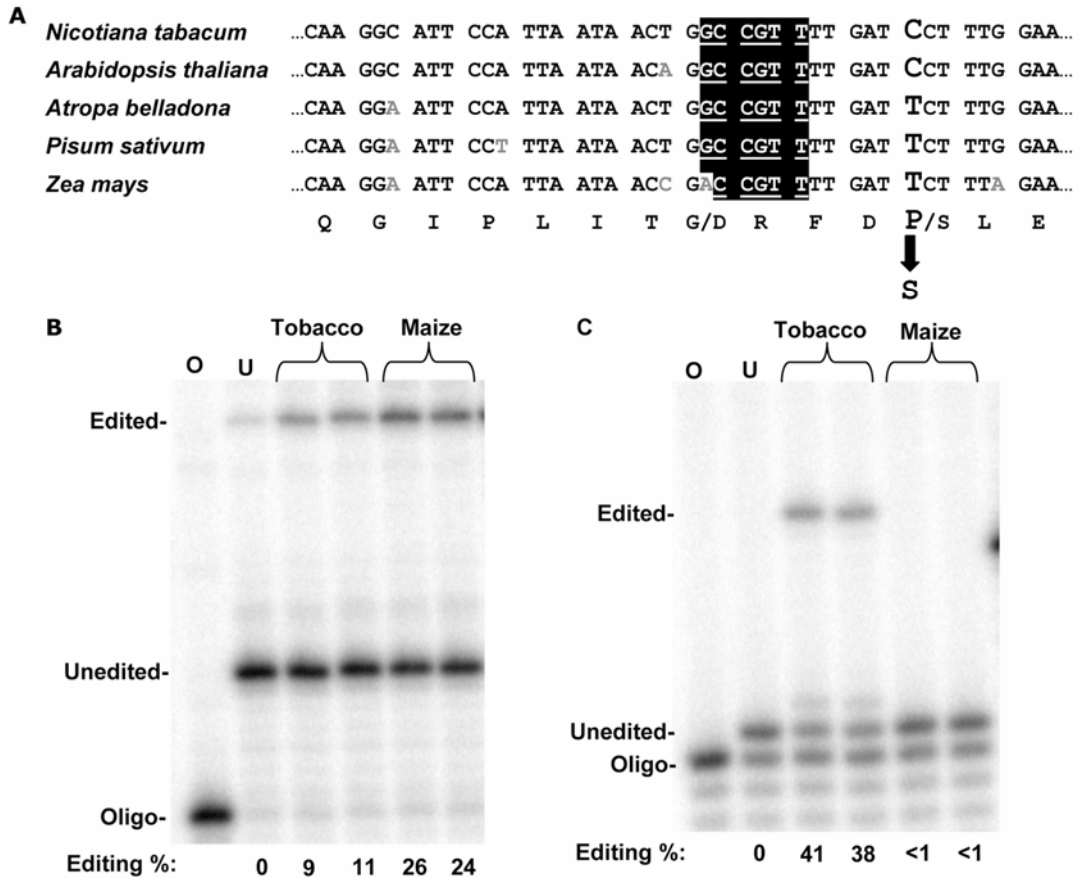

A
Nicotiana tabacum ...CAA GGC ATT CCA TTA ATA ACT GGC CGT TTT GAT CCT TTG GAA...
Arabidopsis thaliana ...CAA GGC ATT CCA TTA ATA ACA GGC CGT TTT GAT CCT TTG GAA...
Atropa belladona ...CAA GGA ATT CCA TTA ATA ACT GGC CGT TTT GAT TCT TTG GAA...
Pisum sativum ...CAA GGA ATT CCT TTA ATA ACT GGC CGT TTT GAT TCT TTG GAA...
Zea mays ...CAA GGA ATT CCA TTA ATA ACC GAC CGT TTT GAT TCT TTA GAA...
Q G I P L I T G/D R F D P/S L E
S
B
Tobacco
Maize
O
U
Edited-
Unedited-
Oligo-
Editing %: 0 9 11 26 24
C
Tobacco
Maize
O
U
Edited-
Unedited-
Oligo-
Editing %: 0 41 38 <1 <1

Table 5.3. Editing of a tobacco psbE C214-containing substrate in four diverse species. Species indicated represent the nuclear background or the source of chloroplast extracts in which editing was examined. The nucleotide at 214 shown is encoded by the native state in the chloroplast genome. A + or - indicates a T214 or C214 found in sequenced cDNAs from exogenous substrates, respectively.

Species	Nucleotide at 214	NTpsbE C214 editing
Nicotiana tobaccum	C	+[1]
Arabidopsis thaliana	C	+[2]
Atropa belladonna	T	+[3]
Pisum sativum	T	-[4]
Zea mays	T	-

[1] (Hirose et al. 1999); [2] (Tillich et al. 2005); [3] (Schmitz-Linneweber et al. 2005); [4] (Miyamoto, Obokata, and Sugiura 2002)

factors for most sites observed in spermatophytes emerged after the divergence of seed and seedless plants.

Sites appear to be continually evolving since many sites have arisen in different phyla. The site psbE C214 most likely emerged in an ancestor of all seed plants and again within the conifer II clade of Pinophyta. Also sites, ndhD C2 and psbL C2 emerged in certain lycopodiophytes and most likely independently in flowering plants. It is likely that specific editing sites are lost repeatedly throughout evolution, emphasizing their expendable nature. In order to maintain the 30-40 editing sites observed within chloroplasts, the rate of new site creation must be approximately as rapid as loss. New sites can be generated through a T-to-C mutation where a new, vestigial, or shared specificity factor can recognize the nearby cis-element and recruit the editing machinery consistent with the three-step model proposed by (Covello &

Gray, 1993).

Independent emergence of shared C targets in homologous genes may be due either to serendipity or convergent evolution. Sites surrounded by different conserved sequences were observed twice in psbE C214, ndhD C2, and psbL C2 suggesting new sites are created at the same position in homologous genes. Neutral drift is the generally accepted mechanism that allows for the emergence of editing factors and targets (Covello & Gray, 1993). Without selective advantage it would not be expected that editing sites would re-occur at the same position except by unusual chance events, yet we found evidence for several evolutionary recurrences of C targets even though we examined only 3 editing sites in detail. However, if a strong positive selection for a C target exists, editing sites should not be lost so frequently as observed. The best model that fits our observations is a weak positive selection that is strong enough to make re-emergence more likely than neutral drift but not powerful enough to prevent loss of C targets by C-to-T mutation. The main function of most editing sites is probably to compensate for genomic mutations, explaining why they are dispensable after a C-to-T mutation

Cuscuta species are parasitic and exhibit limited photosynthetic abilities. They also have a high frequency of changes in the motif and display the only known transition mutation at the editing site. Relaxed selection for functional photosynthetic genes would allow for the accumulation of changes in the motif observed in *Cuscuta* and the loss of critical editing components. Consistent with this idea, *Cuscuta sp. R90-12* has the critical motif but was not found to edit C214, possibly due to loss of a nuclear recognition factor. Further, *Cuscuta gronovii* cannot edit psbE C214 and contains a difference within the motif (U. Maier, personal communication). It is likely that other *Cuscuta* species that have not been examined do not edit C214.

Although Cuscuta sp. are obligate parasites, they vary in the degree to which

they have lost the ability to photosynthesize. In some members of Cuscuta species there is evidence for ongoing selection for *psbE* genes. *Cuscuta reflexa* has the largest genome of the examined parasitic Cuscuta and has similar gene content to tobacco chloroplasts (Revill et al., 2005). *Cuscuta reflexa* can edit psbE C214 (U. Maier, personal communication), perhaps reflecting a requirement for C214 editing for photosynthetic ability. In one local Cuscuta sp. and in *Cuscuta sandwichiana*, the G214 in *psbE* forms part of a GCT alanine codon. An alanine is a similar amino acid to the serine that is encoded either genomically or through RNA editing in other plants. The predicted PsbE amino acid sequence of the analyzed Cuscuta sp. does not diverge greatly from typical psbE-encoded subunits, suggesting a recent loss of selection for photosynthetic genes in some lineages.

Although a relaxed need for photosynthesis could explain the ability for parasitic species to survive despite a proline at position 72 in the photosystem II cytochrome b559 alpha subunit, it is unclear how Nelumbo species tolerate the mutation. The serine conserved in embryophytes is at the same position as an alanine in most cyanobacteria and algae. *Thermosynechococcus elongatus*, a cyanobacterium in which the structure of cytochrome b559 alpha subunit has been characterized has an alanine at position 72 that is part of a transmembrane α-helix domain (Ferreira et al., 2004). It is unlikely that a proline at the same position would allow the same folding of the helical domain and might negatively impact the function of the protein. We entered the tobacco amino acid sequence into SIFT version 2.1.2., a program designed to predict whether an amino acid substitution affects protein function (Ng & Henikoff, 2006). We found that substitution from S to P, which would be the result of translation from unedited mRNAs, is predicted to affect protein function with a score of 0.01. The Median sequence conservation was 3.18 with 393 sequences represented. Possibly the functional *psbE* gene has been transferred to the nucleus in *Nelumbo*.

Editing of psbL C2 and ndhD C2 in angiosperms is likely to be monophyletic because of the conservation of similar putative cis-elements. However, editing at psbL C2 in angiosperms, *Adiantum capillus-veneris*, and *Selaginella uncinata* may have arisen from at least two independent evolution events. Similarly, editing of ndhD C2 in angiosperms, *Huperzia lucidula*, and *Selaginella uncinata* may also have uncommon origins. In addition to sequence conservation, they are also separated phylogenetically. Functional orthologs for specificity factors such as CRR4 are likely to be present in all angiosperms that edit homologous sites that share cis-elements.

Heterologous editing of psbE C214 may be restricted to very closely related species such as *Atropa belladonna* and *Nicotiana tabacum*. Pea, *Pisum sativum*, has the motif but it does not edit an exogenous NTpsbE C214 (Miyamoto et al., 2002). This is likely to be due to the loss of a factor critical for editing in this species. *Glycine max* and *Lotus corniculatus*, two Fabaceae related to pea, both have a C214 and the motif. Pea and *Medicago truncatula* have a T214 and the motif. *Medicago truncatula* and pea are both in the IRLC clade and are more closely related to each other than to *Glycine max* or *Lotus corniculatus* (Wojciechowski et al., 2004). The closest relative to pea with an available *psbE* sequence and a C214 is *Lotus corniculatus*. The sequence of *psbE* in pea compared to *Lotus corniculatus* has 6 nucleotide differences other than the putative editing site. *Atropa belladonna* and tobacco are very closely related with only 2 differences in the *psbE* region other than the editing site. If we assume similar mutation rates in the two plant families then *Atropa belladonna* is more closely related to *Nicotiana tabacum* than pea is to *Lotus corniculatus*. This suggests that preservation of vestigial specificity factors for psbE C214 despite loss of target could be correlated with the time of divergence.

Members of the protein complex critical for editing a C target may be shared with other C targets. Over-expression of a substrate containing an editing-site leads to

the reduction of editing in the endogenous copy as well as a "cluster" of other sites (Chateigner-Boutin & Hanson, 2002). This has been hypothesized to be due to sequestration of critical editing components on over-expressed templates. The loss of editing factors specific to psbE C214 might be permissible because of the lack of allied editing sites that also require the same factor for efficient editing. In tobacco, no sites carrying cis-elements similar to psbE C214 are currently known. Therefore, the psbE C214 specificity factor is unlikely to have a function other than editing that is conserved in other higher plants.

Analysis of the chloroplast genome from *Cycas taitungesnsis* identified only one editing site (Wu et al., 2007). This would suggest that cycads have few chloroplast editing sites. On the contrary, we discovered in *psbE* two different editing sites in *Cycas revoluta* and three in two Zamiacae species. Most of the higher plants examined only have one site in *psbE*, making cycads comparatively editing-site "rich". Many editing sites have been found in the *cox3* transcript from *Cycas revoluta* mitochondria (Malek et al., 1996). Under further examination, cycads as a whole may contain many more unidentified sites with several unique to the group.

ACKNOWLEDGMENTS

This work was supported by NSF grant MCB 0344007 to M.R.H. The authors thank Charles Bullerwell for critical reading of this manuscript.

REFERENCES

Bock R, Hagemann R, Kossel H, Kudla J. 1993. Tissue- and stage-specific modulation of RNA editing of the psbF and psbL transcript from spinach plastids--a new regulatory mechanism? *Mol Gen Genet 240*:238-244.

Bowe LM, Coat G, dePamphilis CW. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc Natl Acad Sci U S A 97*:4092-4097.

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

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

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

Covello PS, Gray MW. 1993. On the evolution of RNA editing. *Trends Genet 9*:265-268.

Ferreira KN, Iverson TM, Maghlaoui K, Barber J, Iwata S. 2004. Architecture of the photosynthetic oxygen-evolving center. *Science 303*:1831-1838.

Freyer R, Kiefer-Meyer MC, Kossel H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A 94*:6285-6290.

Groth-Malonek M, Wahrmund U, Polsakiewicz M, 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*.

Hayes ML, Hanson MR. 2007a. Assay of Editing of Exogenous RNAs in Chloroplast Extracts of Arabidopsis, Maize, Pea, and Tobacco. *Methods Enzymol 424*:459-482.

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

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

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

Hirose T, Kusumegi T, Tsudzuki T, Sugiura M. 1999. RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. *Mol Gen Genet 262*:462-467.

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

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

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

Kudla J, Bock R. 1999. RNA editing in an untranslated region of the Ginkgo chloroplast genome. *Gene 234*:81-86.

Lopez-Serrano M, Del Campo EM, Sabater B, Martin M. 2001. Primary transcripts of ndhD of Liliaceae and Aloaceae require editing of the start and 20th codons. *J Exp Bot 52*:179-180.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science 311*:1727-1730.

Maier RM, Neckermann K, Igloi GL, Kossel H. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol 251*:614-628.

Malek O, Lattig K, Hiesel R, Brennicke A, Knoop V. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *Embo J 15*:1403-1411.

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

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

Ng PC, Henikoff S. 2006. Predicting the effects of amino Acid substitutions on protein function. *Annu Rev Genomics Hum Genet 7*:61-80.

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

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

Revill MJ, Stanley S, Hibberd JM. 2005. Plastid genome structure and loss of

photosynthetic ability in the parasitic genus Cuscuta. *J Exp Bot 56*:2477-2486.

Sasaki T, Yukawa Y, Miyamoto T, Obokata J, Sugiura M. 2003. Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (Nicotiana tabacum): comparative analysis shows the involvement of distinct trans-factors for ndhB editing. *Mol Biol Evol 20*:1028-1035.

Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, Herrmann RG, Maier RM. 2005. Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. *Plant Cell 17*:1815-1828.

Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM. 2002. The plastid chromosome of Atropa belladonna and its comparison with that of Nicotiana tabacum: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol 19*:1602-1612.

Shields DC, Wolfe KH. 1997. Accelerated evolution of sites undergoing mRNA editing in plant mitochondria and chloroplasts. *Mol Biol Evol 14*:344-349.

Stefanovic S, Krueger L, Olmstead RG. 2002. Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *Am J Bot 89*:1510-1522.

Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM. 2005. Editing of plastid RNA in Arabidopsis thaliana ecotypes. *Plant J 43*:708-715.

Tillich M, Lehwark P, Morton BR, Maier UG. 2006. The evolution of chloroplast RNA editing. *Mol Biol Evol 23*:1912-1921.

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

Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kossel H, Sugiura M. 1996. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc Natl Acad Sci U S A 93*:8766-8770.

Wojciechowski M, Lavin M, Sanderson M. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *American Journal of Botany 91*:1846-1862.

Wolf PG, Rowe CA, 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.

Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast Genome (cpDNA) of Cycas taitungensis and 56 Cp Protein-coding Genes of Gnetum parvifolium: Insights into CpDNA Evolution and Phylogeny of Extant Seed Plants. *Mol Biol Evol*.