



Identification of RNA editing sites in cotton (*Gossypium hirsutum*) chloroplasts and editing events that affect secondary and three-dimensional protein structures

Y. Jiang¹, S.L. Fan², M.Z. Song², J.N. Yu¹ and S.X. Yu²

¹College of Life Sciences, Shaanxi Normal University, Xi'an, Shaanxi, P.R. China

²Cotton Research Institute of the Chinese Academy of Agricultural Sciences, Anyang, Henan, P.R. China

Corresponding authors: J.N. Yu / S.X. Yu
E-mail: jnyu@snnu.edu.cn / yu@cricaas.com.cn

Genet. Mol. Res. 11 (2): 987-1001 (2012)

Received May 17, 2011

Accepted December 22, 2011

Published April 19, 2012

DOI <http://dx.doi.org/10.4238/2012.April.19.4>

ABSTRACT. RNA editing can alter individual nucleotides in primary transcripts, which can cause the amino acids encoded by edited RNA to deviate from the ones predicted from the DNA template. We investigated RNA editing sites of protein-coding genes from the chloroplast genome of cotton. Fifty-four editing sites were identified in 27 transcripts, which is the highest editing frequency found until now in angiosperms. All these editing sites were C-to-U conversion, biased toward *ndh* genes and U_A context. Examining published editotypes in various angiosperms, we found that RNA editing mostly converts amino acid from hydrophilic to hydrophobic and restores evolutionary conserved amino acids. Using bioinformatics to analyze the effect of editing events on protein secondary and three-dimensional structures, we found that 21 editing sites can affect protein secondary structures and seven editing sites can alter three-dimensional protein structures.

These results imply that 24 editing sites in cotton chloroplast transcripts may play an important role in their protein structures and functions.

Key words: Chloroplast; *Gossypium hirsutum* (cotton); RNA editing; Protein structures; RESOPS

INTRODUCTION

RNA editing is one of the post-transcriptional processes in which the nucleotide sequences of transcripts are changed by substitution, insertion or deletion of nucleotides. In higher plants, this process mainly occurs in chloroplast and mitochondria. Most of the editing events alter a specific C-to-U and less frequently from U-to-C (Bock, 2001; Shikanai, 2006). Plastid RNA editing was first reported in the maize *rpl2* transcript, in which the ACG codon is changed to the start codon AUG (Hoch et al., 1991). So far, RNA editing has been systematically investigated for the chloroplast protein-coding transcripts in the following species: *Anthoceros formosae* (Kugita et al., 2003), *Adiantum capillus-veneris* (Wolf et al., 2004), *Pinus thunbergii* (Wakasugi et al., 1996), *Pisum sativum* (Inada et al., 2004), *Nicotiana tabacum* (Sasaki et al., 2003, 2006), *Arabidopsis thaliana* (Lutz and Maliga, 2001), *Atropa belladonna* (Schmitz-Linneweber et al., 2002), *Solanum lycopersicum* (Kahlau et al., 2006), four species of Cucurbitaceae family (Guzowska-Nowowiejska et al., 2009), *Phalaenopsis aphrodite* (Zeng et al., 2007), *Zea mays* (Maier et al., 1995), *Oryza sativa* (Corneille et al., 2000), and *Saccharum officinarum* (Calsa et al., 2004). Most angiosperms have a relatively constant number of chloroplast RNA editing sites, anywhere from 21 to 44. In closely related taxa, the number of shared editing sites increases (Guzowska-Nowowiejska et al., 2009).

To date, more and more scholars focus on RNA editing research. Some hypotheses about evolution, significance and mechanisms of plastids RNA editing have been proposed (Fiebig et al., 2004; Miyata and Sugita, 2004; Schmitz-Linneweber et al., 2002, 2005; Shikanai, 2006). Previous studies thought the distribution of residues converted by RNA editing without any rules in amino acid sequences (Gray and Covello, 1993). Very few cases of RNA editing have been found to be related with protein active sites (Bock, 2000). More recently, many editing events were thought to influence the expression of chloroplast genes by changing the amino acid sequences, generating a new start codon or a stop codon, repairing a stop codon, modifying the open reading frame, etc. (Wakasugi et al., 1996; Lutz and Maliga, 2001). Some studies have shown that RNA editing has its rules and probably has important physiological functions. The vast majority of plastid editing events occurred in protein-coding regions, biased to the second codon position (Bock, 2000), and generally restored amino acid conservation compared to other species (Bock, 2001). Furthermore, residues encode by edited codons biased in helices and protein structural cores, which contribute to form stable three-dimensional (3-D) structures. Unedited products were generally unstable, folded incorrectly or disturbed the subunit assembly (Yura and Go, 2008). For example, the unedited *psbF* transcript of *A. thaliana* seriously affected the efficiency of the assembly of PSII complexes (Cai et al., 2009); unedited acetyl-CoA carboxylase carboxyl transferase β of pea lost function (Sasaki et al., 2001) and unedited *petB* mRNA of tobacco chloroplast caused a defect in heme attachment to cytochrome b_6 (Zito et al., 1997).

Cotton is the most important natural textile fiber source and economic crop in the world. In 2006, the gene map of the *Gossypium hirsutum* chloroplast genome was completely identified (Lee et al., 2006). Cotton's complete chloroplast genome is 160 and 301 bp in length. There are 112

unique genes within the cotton chloroplast genome, and 19 of these are duplicated in the inverted repeat, which yields a total of 131 genes, including 78 protein-coding genes. Besides, Lee et al. reported that there were 11 non-synonymous nucleotide substitutions, resulting in a total of nine amino acid changes, which were identified within *ndhC*, *rpl23*, *rpl20*, *rps3*, and *clpP*, but chloroplast RNA editing of cotton has not been systematically examined. In this paper, we analyzed RNA editing sites of cotton chloroplast protein-coding genes, and identified a total of 54 editing sites in 27 transcripts, of which 18 sites are specific in cotton. We also compared 27 transcripts with RNA editing sites of cotton with other homologous genes in different species. The result indicates that they have the same trend in distribution within the editing site and amino acid conversion patterns. Using bioinformatics we predicted protein secondary and tertiary structures of edited and unedited products. As a result, 24 editing sites could affect protein secondary structures and/or 3-D structures. These results imply that the above-mentioned 24 editing sites in 27 transcripts in cotton may have a close relationship with the protein folding correctly and executing their function effectively.

MATERIAL AND METHODS

Plant materials and growth conditions

Cotton plants (*G. hirsutum* cv. Coker 310FR) were grown in soil pots in the growth chamber under long-day conditions (16-h light/8-h dark cycle) at a constant temperature of 28°C. Cotton leaves were harvest from 4-week-old plants.

DNA isolation

Genomic DNA of cotton was isolated by a modified SDS-CTAB method. Leaves (0.1 g) were ground into powder in liquid nitrogen, then 0.6 mL CTAB was added as an extraction buffer (100 mM Tris-HCl, pH 8.0; 50 mM EDTA, pH 8.0; 1.0 M NaCl; 2% SDS-CTAB (w/v); 2% PVP (w/v); 2% β-mercaptoethanol) and incubated at 65°C for 30 min. The slurry was extracted twice with equal volume of chloroform:isoamyl alcohol (24:1). Supernatant was added to 2/3 volume isopropanol and centrifuged. The precipitate was washed with 75% ethanol twice and dissolved in 300 μL sterile water. NaAc (1/10 volume of 3 M, pH 5.2) and 2 volumes ethanol were added, placed at -20°C for 10 min, and centrifuged at 8000 rpm for 5 min. The precipitate was then washed with 75% ethanol twice and dissolved in 20 μL sterile water.

RNA isolation and cDNA synthesis

Total cellular RNA was isolated by a modified CTAB method. About 0.1 g leaves was frozen in liquid nitrogen and ground to a fine powder. The powders were mixed well with 0.6 mL extraction buffer (100 mM Tris-HCl, pH 8.0; 20 mM EDTA, pH 8.0; 2.0 M NaCl; 2% CTAB (w/v); 2% PVP (w/v); 0.5 g/L spermidine; 2% β-mercaptoethanol (v/v)) and incubated at 65°C for 3 min. The slurry was extracted with an equal volume of chloroform:isoamyl alcohol (24:1) twice. LiCl (1/4 volume of 10 M) was added, and the RNA was precipitated at -20°C for at least 6 h. Then, RNA was centrifuged and suspended in 0.5 mL SSTE (0.5% SDS; 2.0 M NaCl; 20 mM EDTA, pH 8.0), and further extracted sequentially with phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1). RNA samples were finally precipitated

by alcohol, dissolved in 20 μ L DEPC-treated H₂O and stored at -80°C.

The RNA samples were then treated with Rnase-free Dnase I (TaKaRa) at 37°C for 80 min to eliminate DNA contamination. cDNA synthesis was carried out according to the instruction manual of the PrimeScript™ RT reagent Kit (TaKaRa) using 3 μ L total RNA as template at 37°C for 15 min and stop reaction at 85°C for 5 s.

Polymerase chain reaction (PCR) and sequencing

The primers were designed according to protein-coding genes of the chloroplast genome of cotton (DQ345959) and were used for PCR and RT-PCR amplification. PCR and RT-PCR were carried out with the following profiles: initial denaturation for 3 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 52-55°C depending on the oligo nucleotides used, 1 min at 72°C, and a final extension for 10 min at 72°C. Each amplification product was electrophoresed on a 1% agarose gel and purified with a gel extraction kit (BioFlux), then sequenced in both directions by Shanghai Sangon Biological Engineering Technology & Services Company.

Identification of RNA editing sites

DNA and cDNA sequences were both spliced by the Seqman software and the RNA editing sites were identified by comparing each cDNA to DNA sequences using the ClustalW program.

Bioinformatic analysis

We performed transmembrane segment and N-terminal signal peptide sequence analysis of amino acid sequences predicted from edited mRNA and cDNA in TMHMM (<http://genome.cbs.dtu.dk/services/TMHMM/>), SignalP (<http://www.cbs.dtu.dk/services/SignalP>) and SIG-Pred (http://bmbpcu36.leeds.ac.uk/prot_analysis/Signal.html), respectively. We assigned a protein secondary structure and protein-specific domain before and after editing using SOPMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=NPSA/npsa_sopma.html) and SMART (<http://smart.embl-heidelberg.de/>); we predicted the localization of the editing site in the protein 3-D structure in RESOPS (<http://cib.cf.ocha.ac.jp/RNAEDITING/>); we used PDB (<http://www.rcsb.org/pdb/home/home.do>) for estimating protein 3-D structure alteration before and after RNA editing.

RESULTS

Cotton chloroplast has more RNA editing sites

According to our previous research, 58 RNA editing sites have been predicted in 29 transcripts in the cotton chloroplast genome (Jiang et al., 2010). In this paper, we isolated DNA and cDNA from the same sample of cotton and did PCR and RT-PCR for 29 protein-coding genes, respectively. For each of the 29 genes, the cDNA sequence aligned with the DNA sequence using ClustalW. A total of 54 editing sites were identified in 27 of 29 transcripts, which is the highest number of RNA editing sites found in angiosperms to date. All the RNA editing sites are C-to-U conversions (Table 1). Among the 54 editing sites, we detected 10 partially edited sites, *rpoB*-113, -184, -189 and -809, *matk*-153, -212, and -235, *rpl23*-30, *psbN*-10, and *rps18*-74 (Table 1).

Table 1. Comparison of RNA editing sites in *Atropa belladonna*, *Arabidopsis thaliana*, *Nicotiana tabacum*, and *Gossypium hirsutum*.

Gene	Codon position	Codon and amino acid change	<i>A. belladonna</i>	<i>A. thaliana</i>	<i>N. tabacum</i>	<i>G. hirsutum</i>
<i>atpA-1</i>	264	P(cCc)→L(cUc)	T	-	+	T
<i>atpA-2</i>	265	S(ucC)→S(ucU)	+	+	+	-
<i>atpA-3</i>	305	S(uCa)→L(uUa)	T	T	T	+
<i>atpA-4</i>	383	S(uCa)→L(uUa)	T	T	T	+
<i>atpF</i>	31	P(cCa)→L(cUa)	+	+	+	+
<i>ndhA-1</i>	114	S(uCa)→L(uUa)	+	+	+	+
<i>ndhA-2</i>	189	S(uCa)→L(uUa)	+	-(uga)	T	+
<i>ndhA-3</i>	358	S(uCc)→F(uUc)	+	T	+	T
<i>ndhB-1</i>	50	S(uCa)→L(uUa)	+	+	+	+
<i>ndhB-2</i>	156	P(cCa)→L(cUa)	+	+	+	+
<i>ndhB-3</i>	181	T(aCg)→M(aUg)	T	T	T	+
<i>ndhB-4</i>	196	H(Cau)→Y(Uau)	+	+	+	+
<i>ndhB-5</i>	204	S(uCa)→L(uUa)	+	+	+	+
<i>ndhB-6</i>	246	P(cCa)→L(cUa)	+	-	+	+
<i>ndhB-7</i>	249	S(uCu)→F(uUu)	+	+	+	+
<i>ndhB-8</i>	277	S(uCg)→L(uUg)	+	+	+	+
<i>ndhB-9</i>	279	S(uCa)→L(uUa)	+	+	+	+
<i>ndhB-10</i>	291	S(uCa)→L(uUa)	-	+	-	-
<i>ndhB-11</i>	419	H(Cau)→Y(Uau)	T	+	-	+
<i>ndhB-12</i>	494	P(cCa)→L(cUa)	+	+	+	+
<i>ndhC</i>	108	S(uCa)→L(uUa)	T	T	T	+
<i>ndhD-1</i>	1	T(aCg)→M(aUg)	+	+	+	+
<i>ndhD-2</i>	128	S(uCa)→L(uUa)	+	+	+	+
<i>ndhD-3</i>	200	S(uCa)→L(uUa)	T	-	+	T
<i>ndhD-4</i>	225	S(uCg)→L(uUg)	T	+	+	+
<i>ndhD-5</i>	293	S(uCa)→L(uUa)	+	+	T	+
<i>ndhD-6</i>	296	P(cCc)→L(cUc)	-	+	-	-
<i>ndhD-7</i>	433	S(uCa)→L(uUa)	+	T	+	+
<i>ndhD-8</i>	437	S(uCa)→L(uUa)	+	T	+	+
<i>ndhE</i>	78	S(uCa)→L(uUa)	-(cug)	T	-(cug)	+
<i>ndhF</i>	97	S(uCa)→L(uUa)	+	+	+	+
<i>ndhG-1</i>	17	S(uCg)→L(uUg)	+	+	+	+
		S(uCc)→F(uUc)	+	+	+	+
<i>ndhG-2</i>	116	S(uCa)→L(uUa)	-	T	+	T
<i>petB-1</i>	4/24	V(guC)→V(guU)	-(gua)	T	-(gua)	+
<i>petB-2</i>	160	R(cGg)→W(Ugg)	T	T	T	+
<i>petB-3</i>	204	P(cCa)→L(cUa)	+	T	+	T
<i>psbE</i>	72	P(cCa)→L(cUa)	T	+	+	-
<i>psbF</i>	26	S(uCu)→F(uUu)	T	+	-	+
<i>psbL</i>	1	T(aCg)→M(aUg)	+	T	+	T
<i>rpl20</i>	103	S(uCa)→L(uUa)	+	T	+	+
<i>rps2-1</i>	45	T(aCa)→I(aUa)	+	T	+	+
<i>rps2-2</i>	83	S(uCg)→L(uUg)	+	T	+	+
<i>accD-1</i>	265/271	S(uCg)→L(uUg)	T	+	T	+
<i>accD-2</i>	474	P(cCu)→L(cUu)	T	T	T	+
<i>clpP</i>	187	H(Cau)→Y(Uau)	T	+	-	+
<i>matK-1</i>	153/157	H(Cau)→Y(Uau)	T	T	T	+/-
<i>matK-2</i>	212/214	H(Cau)→Y(Uau)	T	+	T	+/-
<i>matK-3</i>	235/237	S(uCu)→F(uUu)	-	-	-	+/-
<i>petL</i>	2	P(cCu)→L(cUu)	T	+	T	+
<i>rps14-1</i>	27	S(uCa)→L(uUa)	+	+	+	+
<i>rps14-2</i>	50	P(cCa)→L(cUa)	T	+	+	+
		S(uCa)→L(uUa)	+	+	+	+
<i>rpl23-1</i>	24	S(uCu)→F(uUu)	-	T	-	+
<i>rpl23-2</i>	30	S(uCa)→L(uUa)	-	+	-	+/-
<i>rps18</i>	74	S(uCg)→L(uUg)	T	T	T	+/-
<i>rpoA-1</i>	67	S(uCu)→F(uUu)	T	+	T	-
<i>rpoA-2</i>	277	S(uCa)→F(uUa)	+	T	+	-
<i>rpoB-1</i>	113	S(uCu)→F(uUu)	+	+	+	+/-
<i>rpoB-2</i>	158	S(uCa)→L(uUa)	+	T	+	T
<i>rpoB-3</i>	184	S(uCa)→L(uUa)	+	+	+	+/-
<i>rpoB-4</i>	189	S(uCg)→L(uUg)	T	T	T	+/-
<i>rpoB-5</i>	667	S(uCu)→F(uUu)	+	T	+	T
<i>rpoB-6</i>	809/811	S(uCa)→L(uUa)	+	+	T	+/-
<i>rpoC1-1</i>	14/21	S(uCa)→L(uUa)	+	T	+	+
<i>rpoC1-2</i>	163	S(uCa)→L(uUa)	T	+	T	-
<i>rpoC2</i>	1248	S(uCa)→L(uUa)	+	T	+	-
<i>psbJ</i>	20	P(cCa)→L(cUa)	T	T	T	+

Continued on next page

Table 1. Continued.

Gene	Codon position	Codon and amino acid change	<i>A. belladonna</i>	<i>A. thaliana</i>	<i>N. tabacum</i>	<i>G. hirsutum</i>
<i>atp1</i>	207	S(uCa)→L(uUa)	T	T	T	+
<i>psa1</i>	28	S(uCu)→F(uUu)	T	T	T	+
<i>psbN</i>	10	S(uCu)→F(uUu)	T	T	T	+/-
<i>rps12</i>	74	S(uCa)→L(uUa)	T	T	T	+

Capital letters in codon triplets = nucleotides before or after editing; + = editing sites from experiment; T = no editing as T is already in the genome; - = no editing although C is in the genome; +/- = partial editing.

petB-24 is a silent editing site that cannot change amino acid type. The other 43 sites are fully edited. *ndh* genes have the highest editing frequency. Twenty-three editing sites, more than half in 54 editing sites, are shared by *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, and *ndhG*. *ndhB* has 11 editing sites and all of them are totally edited. *ndhD-1* also produces a new start codon that is fully edited, instead of the partial editing of many other species. Compared to *N. tabacum*, *A. thaliana* and *A. belladonna*, cotton has 18 unique editing sites (Table 1). Except for *ndhE-78*, *petB-24* and *rpl23-24*, the 15 sites restore evolutionarily conserved amino acids, which are similar to the codons of homologous proteins in other three species that had reversed C-to-T at the DNA level. *atpA-265* is a silent editing site in three species, which does not cause amino acid change after editing, whereas in cotton, this site maintains the unedited Ser(ucC).

Cotton RNA editing mainly happens in the second codon position and bias to U_A context

We further analyzed all 54 editing sites found in cotton. Of the sites, 11.1% are located in the first position of the codon, 87.0% sites are in the second position, and only 1 (1.9%) site occurs in the third position. We also investigated all RNA editing sites which have been reported in another three dicotyledons and four monocotyledons (Table 2). The result indicates that the second cotton position has the highest editing frequency, up to 80% or even more. Wolf et al. (2004) also reported that the most frequent RNA editing occurs in the second codon position, followed by the first one, and only a few cases occur in the third position in spermatophytes.

Table 2. Summary data of RNA editing in 4 dicotyledons and 4 monocotyledons.

Species	Dicotyledons				Monocotyledons			
	<i>Gohi</i>	<i>Nita</i>	<i>Atbe</i>	<i>Arth</i>	<i>Phap</i>	<i>Zema</i>	<i>Orsa</i>	<i>Saof</i>
Number of edited sites	54	38	35	35	44	27	26	24
C→U	100%	100%	100%	100%	100%	100%	100%	100%
1st codon edits	11.1%	5.4%	2.9%	16.7%	9.1%	7.4%	8.0%	4.3%
2nd codon edits	87.0%	91.9%	94.2%	80.0%	86.4%	92.6%	92.0%	95.7%
3rd codon edits	1.9%	2.7%	2.9%	3.3%	0%	0%	0%	0%
New starts	1	2	2	1	1	1	1	1
New stops	0	0	0	0	0	0	0	0
AA changes	53	37	34	32	42	26	25	23
Silent edits	1	1	1	1	0	0	0	0
Untranslated edits	0	0	0	2	2	1	1	1

Gohi = *Gossypium hirsutum*; *Nita* = *Nicotiana tabacum*; *Atbe* = *Atropa belladonna*; *Arth* = *Arabidopsis thaliana*; *Phap* = *Phalaenopsis aphrodite*; *Zema* = *Zea mays*; *Orsa* = *Oryza sativa*; *Saof* = *Saccharum officinarum*.

In addition, we further compared the characteristics of RNA editing sites in eight species. The 54 editing sites in cotton, editing obviously prefers an uridine_adenine (U_A)

context and takes up to 50% (Figure 1). In the other three dicotyledons and four monocotyledons, a similar consistency is found that the U_A context also gets to 40% or more (Table 2). It indicates that there is a bias in the distribution of RNA editing sites in spermatophytes.

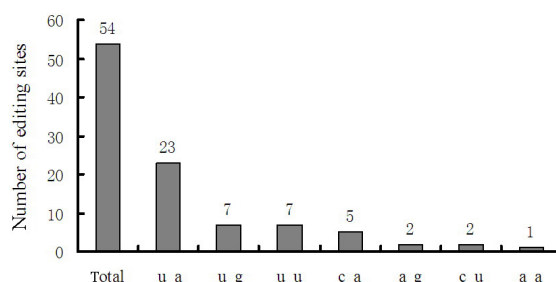


Figure 1. Number and sequence context of editing sites in the chloroplast of cotton.

RNA editing causes increase in hydrophobic amino acids

RNA editing sites are not randomly distributed in the genome. Most of them occur in the protein-coding region. As a result, editing often causes corresponding amino acid changes. Of the 54 sites in cotton, except for *petB-24*, which is a silent editing site, 53 sites can cause a codon to change. Among them, 45 sites convert amino acids from hydrophilic to hydrophobic. Eight sites switch amino acids from hydrophobic to hydrophobic. Interestingly, 37 (68.5%) events focus on the conversion of amino acids to leucine. We analyzed the conversion patterns of amino acid after RNA editing in eight spermatophytes that are mentioned above. The highest conversion rate is serine to leucine, which makes up 54.3% of all editing sites, followed by proline to leucine and serine to phenylalanine (16.7 and 12.0%, respectively) in a total of 276 editing sites (Table 3).

Table 3. The number of amino acid alterations by editing in 8 angiosperms.

AA changes	Gohi	Nita	Atbe	Arth	Phap	Zema	Orsa	Saof	Total
Ser→Leu	30	21	22	16	17	15	14	15	150
Pro→Leu	7	7	5	6	8	5	4	4	46
Ser→Phe	7	4	3	4	10	2	2	1	33
His→Tyr	5	1	1	4	4	2	2	1	20
Thr→Met	2	2	2	1	2	2	2	2	15
Thr→Ile	1	1	1	0	1	0	1	0	5
Ser→Ser	0	1	1	1	0	0	0	0	3
Pro→Ser	0	1	0	1	0	0	0	0	2
Val→Val	1	0	0	0	0	0	0	0	1
Arg→Trp	1	0	0	0	0	0	0	0	1
Total	54	38	35	33	42	26	25	23	276

For abbreviations, see legend to Table 2.

RNA editing affects protein structure

RNA editing impact on protein secondary structures

We analyzed 27 protein secondary structures before and after editing using SignalP, SIG-Pred, TMHMM, SOPMA, and SMART. Twenty-one editing sites may change corresponding protein secondary structure (Table 4).

Table 4. Prediction subunits secondary and three-dimensional (3-D) structure alteration by RNA editing in 27 protein coding genes.

Gene	Codon position [†]	Signal peptides	Transmembrane	The composition of second structure	3-D	Probably effect on function
<i>accD</i>	271	-	-	No change	Core	Low solubility ^a ; albino phenotype ^b ; no effect ^c
	474	-	-	No change	Non-core	ND
<i>atpA</i>	305	-	-	A→B and R→E	Core	ND
	383	-	-	No change	Core	ND
<i>atpF</i>	31	-	No change	A→E	ND	ND
<i>atpI</i>	207	No change	No change	No change	ND	ND
<i>clpP</i>	187	-	-	E→A	Surface	Unclear ^d
<i>matK</i>	153	-	-	No change	ND	ND
	212	-	-	R→A	ND	ND
	235	-	-	No change	ND	ND
<i>ndhA</i>	114	-	No change	No change	ND	ND
	189	-	No change	No change	ND	ND
<i>ndhB</i>	50	Cleavage sites changed from 2 to 1	No change	A→E and R	ND	ND
	156	-	Appear another two transmembrane	No change	ND	NDH partial loss of activity ^e
	181	-	No change	No change	ND	ND
	196	-	Appear another one transmembrane	A→E	ND	ND
	204	-	No change	No change	ND	ND
	246	-	No change	No change	ND	ND
	249	-	No change	No change	ND	NDH partial loss of activity ^e
	277	-	No change	No change	ND	ND
	279	-	No change	No change	ND	No effect ^f
	419	-	No change	E→A and A→R	ND	ND
	494	-	No change	A→E	ND	Chlorophyll fluorescence decreased; NDH partial loss of activity and stability ^g
<i>ndhC</i>	108	No change	No change	A→R	ND	ND
<i>ndhD</i>	1	-	No change	R→A	ND	ND
	128	-	No change	No change	ND	NDH partial loss of activity ^h
	225	-	No change	R→E		No effect ^g
	293	-	No change	No change	ND	NDH partial loss of activity ^e
	433	-	No change	No change	ND	ND
	437	-	No change	No change	ND	ND
<i>ndhE</i>	78	No change	No change	A→E	ND	ND
<i>ndhF</i>	97	No change	No change	No change	ND	Chlorophyll fluorescence decreased; NDH partial loss of activity and stability ^g
<i>ndhG</i>	17	Cleavage sites changed from 1 to 2	No change	E→A and R	ND	No effect ^f
<i>psaI</i>	28	No change	No change	R→A and B	ND	ND
<i>psbF</i>	26	Become longer	No change	No change	ND	Photosystem II-deficient phenotype; assembly of PSII complexes was impaired ^f
<i>psbJ</i>	20	Become longer	No change	R→E	ND	ND
<i>psbN</i>	10	No change	No change	E→A and R→B	ND	ND
<i>rpl20</i>	103	-	-	A→E and R	Core	ND
<i>rpl23</i>	24	-	-	No change	ND	ND
	30	-	-	No change	ND	ND
<i>rps2</i>	45	-	-	No change	Surface	ND
	83	-	-	No change	Non-core	ND
<i>rps12</i>	74	-	-	No change	ND	ND
<i>rps14</i>	27	-	-	No change	Non-core	ND
	50	-	-	No change	Non-core	ND

Continued on next page

Table 4. Continued.

Gene	Codon position [†]	Signal peptides	Transmembrane	The composition of second structure	3-D	Probably effect on function
<i>rps18</i>	74	-	-	No change	Surface	ND
<i>petB</i>	24	-	No change	No change	Surface	ND
	160	-	Structure of within and outside the membrane exchanged	R→E and B→R	Core	ND
<i>petL</i>	2	No change	No change	No change	ND	ND
<i>rpoB</i>	113	-	-	No change	Non-core	Partial loss of activity [†]
	184	-	-	No change	Surface	ND
	189	-	-	No change	Surface	ND
	809	-	-	R→A	Core	ND
<i>rpoC1</i>	14	-	-	No change	ND	ND

[†]Sasaki et al. (2001); ^bYu et al. (2009); ^cRobbins et al. (2009); ^dChateigner-Boutin et al. (2008); ^eOkuda et al. (2009); ^fOkuda et al. (2010); ^gHammani et al. (2009); ^hOkuda et al. (2007); ⁱCai et al. (2009); ^jZhou et al. (2009); [†]Codon position is the codon number of the edited gene from the start codon. - = genes that do not have signal peptides or transmembrane. No change = the residue has predicted structure and no change after editing. ND = the residues cannot be found in RESOPS database and their probably effect on function is unknown. A, B, E, and R = alpha helix, beta turn, extended strand, and random coil, respectively.

ndhB-50 and *ndhG-17* causes serine to convert to leucine, which might change the cleavage positions of signal peptides (Figure 2). The length of signal peptides of PsbF and PsbJ maybe increase after 26 and 20 codon positions edited, respectively (Figure 3). NdhB, with 156 sites edited, reveals two novelty transmembrane regions at codon 149 to 168 and 183 to 202. *ndhB-196* editing also appears in a new transmembrane structure in the region of codon 183 to 202. *petB* encodes a cytochrome *b₆* subunit and locates in the thylakoid membrane. *petB-160* editing has an obviously impact on transmembrane structures. This causes amino acids from 1 to 49, 127 to 135, and 226 to 235 to reverse to inside from outside of the thylakoid membrane, while the amino acids of 73 to 103 and 159 to 202 turn inside to outside (Figure 4). In addition, we found that 19 editing sites, *atpA-305*, *atpF-31*, *clpP-187*, *matK-212*, *ndhB-50*, *ndhB-196*, *ndhB-419*, *ndhB-494*, *ndhC-108*, *ndhD-1*, *ndhD-225*, *ndhE-78*, *ndhG-17*, *psaI-28*, *rpl20-103*, *psbJ-20*, *psbN-10*, *petB-160*, and *rpoB-809*, could affect the composition of secondary structures around editing sites (Figure 5). More than half of the editing sites (28 sites) prefer to form new α -helix structure at up- and down-stream regions around the editing codon (data not shown).

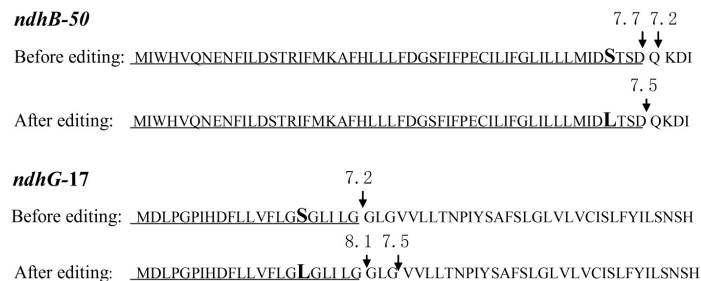


Figure 2. The signal peptide cleavage sites of *ndhB-50* and *ndhG-17* before and after editing. Letters underlined denote signal peptide sequence; the bold and capital letters in signal peptide indicate the amino acid before and after editing; the arrows denote the putative cleavage sites, and the numbers above the arrows show cutting frequency according to SIG-Pred analysis.

psbF-26Before editing: MTIDRTYPIFTVRWLAVHGLAVPTVSFLGSISAMQFIQRAfter editing: MTIDRTYPIFTVRWLAVHGLAVPTVFFLGSISAMQFIQR***psbJ-20***Before editing: MADTTGRIPLWIIGTVTGIPVIGLIGIFFYGSYSGLGSSLAfter editing: MADTTGRIPLWIIGTVTGILVIIGLIGIFFYGSYSGLGSSL

Figure 3. The signal peptide lengths of *psbF* and *psbJ* were changed with both genes edited in 26 and 20 sites, respectively. Letters underlined denote signal peptide sequence; the bold and capital letters in the signal peptide indicate the amino acid before and after editing.

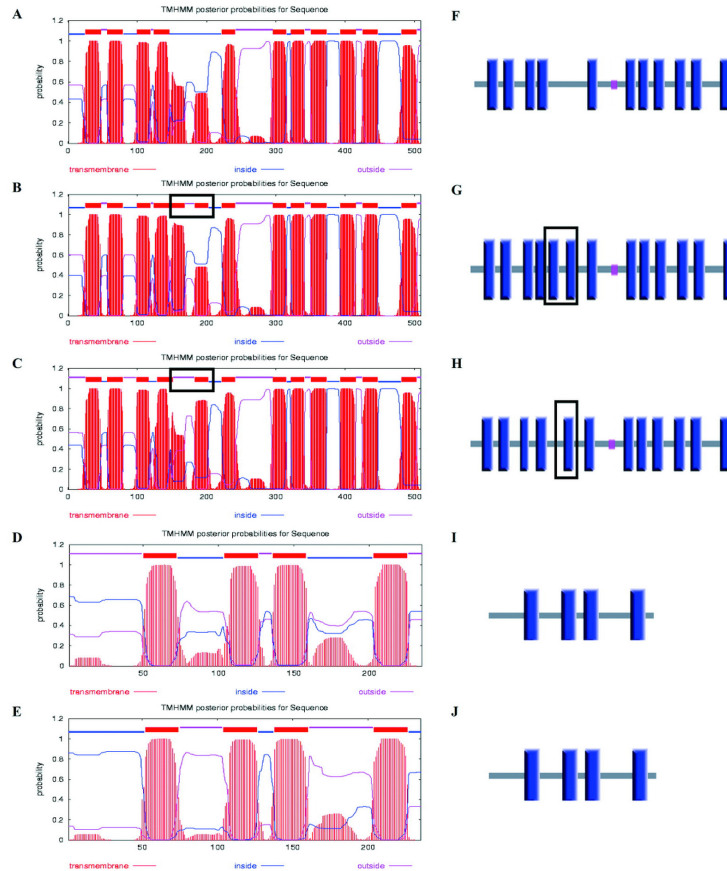


Figure 4. The editing in *ndhB-156*, *ndhB-196* and *petB-160* would affect the transmembrane structures for each protein. A-E are predicted by TMHMM; F-J are predicted transmembrane regions by SMART. A. and F. indicate NdhB transmembrane segments with *ndhB-156* sites unedited. B. and G. indicate NdhB transmembrane segments with *ndhB-156* sites edited. C. and H. indicate NdhB transmembrane segments with *ndhB-196* sites edited. D. and I. indicate PetB transmembrane segments with *petB-160* sites unedited. E. and J. indicate PetB transmembrane segments with *petB-160* sites edited. Black boxes indicate the changing transmembrane regions.

	<i>atpA-305</i>	<i>ndhB-494</i>	<i>psbJ-20</i>
Before editing	LERAAK S SSQLGEGSM TALPI hhhhhhhhhh h ttccceee c ce	SMIVCVI A ST I P G ISMNP I IA hhhhhhhhhh E etcccc e ee	LW I IGTVT G I P VIGLIGIFFY eeee cc cccc e eeeeeeee
After editing	LERAAK L SSQLGEGSM TALPI hhhhhhhh t ttccceee e ee	SMIVCVI L ST L G I SMNP I IA hh ee hhhh H etcccc e ee	LW I IGTVT L VIGLIGIFFY eeee cc t ee eeeeeeee
	<i>atpF-31</i>	<i>ndhD-1</i>	<i>rpl20-103</i>
Before editing	FGVNTDILATNP I NLSVVLGV cccchhhhhhhhhhh h hhhh	T NYFPWLT I IVFLPISAGSLL C cc cccc ee eechhhhh ee	K I L A Q I A I S N R N C L Y M IS N E I hhhhhhhh h ct th ee eh hh h
After editing	FGVNTDILAT N LNSVVLGV cccchhhhhhhhhhh ee hh h	M NYFPWLT I IVFLPISAGSLL C cc ch h h ee eechhhhh ee	K I L A Q I A L NRNCLY M IS N E I hhhhhhhh h ct tt ee ee cc cc
	<i>clpP-187</i>	<i>ndhD-225</i>	<i>petB-160</i>
Before editing	D V FMSATEAQ H GIVDLVAVK h he echhhhhhh te eeee ec	I G FL I A F AV K S P I I PLHTW L P hhhhhhhh h cccc cc cccc cc	A S F G V T G Y S L P R D Q I G Y W AV K hhhh h tt cc ct th hhhhhh h
After editing	D V FMSATEAQ Y GIVDLVAVK h he echhhhhhh te hhhh h	I G FL I A F AV K L P I I PLHTW L P hhhhhhhh h cccc cc ee ec	A S F G V T G Y S L P W D Q I G Y W AV K hhhh h t ee cc ch hhhhhh h
	<i>matK-212</i>	<i>ndhE-78</i>	<i>rpoB-809</i>
Before editing	R L FLFLY N S H T C E Y E S I F L F L heeeee cc chhhhh ee eee	A I A A A E A A I G S A I V S S I Y R N R hhhhhhhhhh h hhhhhh h tt	S S Y N P E T I R V Y I S Q K R E I K V G c cc chhhhh ee hh h t cc ee t
After editing	R L FLFLY S Y T C E Y E S I F L F L heeeee h hhhhhhhh ee eee	A I A A A E A A I G L A I V S S I Y R N R hhhhhhhhhh ee chhhhh ct	S S Y N P E T I R V Y I L Q K R E I K V G c h hhhhhh ee hh h t cc ee t
	<i>ndhB-50</i>	<i>ndhG-17</i>	<i>psaI-28</i>
Before editing	F G L I L L L M I D S T S D Q K D I P W L hh h hh ee eh h h ct cc cc ee	I H D F L L V F L G S G L I L G L G V V hh ee eeeh cc ce ee t cc ee	L V F P A I A M A S L S L Y V Q K T K I F he ch hhhhhhhh h cc cc ce
After editing	F G L I L L L M I D L T S D Q K D I P W L hh ee ee ee cc cc ct cc cc ee	I H D F L L V F L G L G L I L G L G V V hh ee ee ee cc cc ct ch ee	L V F P A I A M A S L F L Y V Q K T K I F he ch hhhhhhhh h h h t cc ce
	<i>ndhB-196</i>	<i>ndhB-419</i>	<i>psbN-10</i>
Before editing	M G G A S S I L V H F S W L Y G S S G G eh cc ch he eh hh ee et cc c	P P L A G F F G K L H L F W C G W Q A G L c cc h h hhhh ee ehhh h tt h	E T A T L V A I S I S G L L V S F T G Y A c ch ee ee hhhh h ee ee cc ee
After editing	M G G A S S I L V Y F S W L Y G S S G G eh cc ch he eh ee ee ee cc cc	P P L A G F F G K L Y L F W C G W Q A G L c cc ct h hhhh h hhhh h th	E T A T L V A I F I S G L L V S F T G Y A c ch hhhh h hhhh h h h tt ee
	<i>ndhC-108</i>		
Before editing	M S F D V L G V P V F I E A F I F V L I L I V G S V Y A W R e eh hhhhhhhhhhhhhhhhhhhhhhhhhhhh		
After editing	M S F D V L G V P V F I E A F I F V L I L I V L V Y A W R e ee cc cc chhhhhhhhhhhhhhhhhhhhhhh		

Figure 5. Editing would affect adjacent secondary structures in 19 editing sites. The blue capital letters indicate the edited amino acids; red capital letters indicate the changing secondary structures.

RNA editing impact on protein 3-D structures

The PDB online software was used to predict 3-D structures of proteins encoded by edited and unedited transcripts. The results indicated that seven editing sites in 4 proteins (*ndhA-114*, *ndhD-128*, *ndhD-225*, *ndhD-433*, *ndhC-108*, *ndhB-204*, and *ndhB-419*) are involved in their corresponding protein 3-D structures. We also analyzed protein 3-D structures of edited transcripts using RESOPS, which is a database for analyzing the correspondence of RNA editing sites to protein 3-D structure (Yura et al., 2009). To date, 41 3-D structures of chloroplast proteins with editing are listed in RESOPS. In 27 transcripts with 54 RNA editing

sites of cotton, nine genes (*atpA*, *clpP*, *rpoB*, *accD*, *petB*, *rpl20*, *rps14*, *rps18*, and *rps2*) match corresponding 3-D structures in the RESOPS database. We matched the above nine proteins with 17 editing sites to their 3-D structures. The results showed that 6 sites in the five proteins (*accD*-271, *atpA*-305, *atpA*-383, *petB*-160, *rpl20*-103, and *rpoB*-809) are located in structural cores, whereas 5 sites in the three proteins (*rpoB*-113, *accD*-474, *rps2*-83, *rps14*-27, and *rps14*-50) are located in non-core areas, and 6 sites (*rpoB*-184, *rpoB*-189, *rps2*-45, *rps18*-74, *petB*-24, and *clpP*-187) are located in the protein surfaces (Table 4).

DISCUSSION

In higher plant chloroplasts, RNA editing has been systematically determined for protein-coding transcripts from *Anthoceros formosae* to angiosperms. *A. formosae* has more than 900 editing sites, while angiosperms have a relatively constant number, ranging from 21 to 44. In this paper, we report that cotton has 54 RNA editing sites in 27 protein-coding transcripts. To date, this is the highest number of editing sites everfound in spermatophytes. Comparing all editing sites with those of three other seed plant species, cotton has 18 unique editing sites. Most of the 18 sites in the other three species have reversed mutation and maintained T at the DNA level. Cotton shares 28, 26, 25, and 13 editing sites with *Arabidopsis*, tobacco, *belladonna*, and *Phalaenopsis*, respectively. This suggests that cotton has a close relationship with *Arabidopsis*, which is consistent with the Lee et al. (2006) report.

RNA editing usually happens at particular sites with bases bias. In cotton, editing sites also prefer the U_A context and the second and first position codon in most. Häder and Sinha (2005) suggested that C-to-U RNA editing was a special protection method in plant evolution. Aquatic plants are protected by water from the hazardous effects of UV light (i.e., the formation of thymine dimers on DNA). When plants migrated to land, T-to-C mutation occurred at the genomic DNA level to protect them. In order to restore this mutation and form functional proteins, RNA editing appeared at the post-transcriptional level. With ozone layer formation, UV light gradually decreases. Some dispensable editing sites are lost, whereas the essential sites are maintained (Yura and Go, 2008). The maintained RNA editing sites have their codon preference, which leads to a different conversion frequency of amino acid residues. The top three conversion residues are Ser (UCA) to Leu (UUA), Pro (CCA) to Leu (CUA), and Ser (UCU) to Phe (UUU). Moreover, proline is a helix-breaker (Chou and Fasman, 1978), which could cause a defect in protein 3-D structure. When proline converts to leucine by RNA editing, the truncated helix will recover the ones that are appropriate for protein secondary structure formation or result in stable protein appearance (Yura and Go, 2008).

Comparing all 54 RNA editing sites in cotton with the seven other spermatophytes, we found that all the editing sites are conservative. RNA editing cases in cotton mostly restore the conserved amino acid of homologous protein among angiosperms, which further proves that RNA editing is a transcription repair process (Bock, 2000). Up to date, a number of experiments were performed to test the function of RNA editing. There is growing evidence that most of the unedited proteins have less function than the edited ones (Zito et al., 1997; Sasaki et al., 2001; Yura et al. 2009). We used bioinformatics to analyze protein secondary structures of all 27 transcripts in cotton before and after editing. The result shows that 21 editing sites would change the signal peptide, transmembrane structure or secondary structural composition (Figures 2-5). All the altered signal peptide and transmembrane structure may play an

important role in protein localization, interaction with membrane lipid or other subunits. In addition, more than half of the editing sites in cotton are located in the α -helix, which is consistent with Yura and Go (2008).

Using the PDB online software, we predicted the impact of 54 RNA editing sites on the corresponding protein 3-D structures. The results demonstrate that seven editing sites, *ndhA*-114, *ndhD*-128, *ndhD*-225, *ndhD*-433, *ndhC*-108, *ndhB*-204, and *ndhB*-419, in four genes change the corresponding protein 3-D structures because of RNA editing. Interestingly, all these sites focus on *ndh* dehydrogenase subunit genes, of which four sites also have predicted secondary structural changes (Table 4). These changes imply that the above-mentioned editing sites may affect the function of corresponding proteins. In *Arabidopsis*, *ndhD*-128 editing has been verified to maintain NDH complex activity and increase chlorophyll fluorescence after post-illumination (Okuda et al., 2007). Hammani et al. (2009) also reported that the NDH complex has lower stability and activity without editing at *ndhB*-494 and/or *ndhF*-97 in *Arabidopsis*.

We further used the RESOPS database to analyze the localization of 17 residues of 9 transcripts in the protein 3-D structures. The data show that six, five and six residues are targeted in protein structural cores, non-cores and surfaces, respectively. All the former six residues, four of them also have predicted changes in secondary structures (Figure 5), convert amino acids from hydrophilic to hydrophobic by RNA editing. The hydrophobic residues always tend to be buried inside of protein, which suggests that the change may be beneficial to form stability in the protein structural core. For example, in *Arabidopsis*, *accD*-265 residue forms a structural core of AccD. Sasaki et al. (2001) and Sasaki and Nagano (2004) proved that *accD*-265 was needed for carboxyltransferase and acetyl-CoA carboxylase activity *in vivo*. Yu et al. (2009) also confirmed that *accD*-265 decreases editing, acetyl-CoA carboxylase was dysfunctional and causes defects in the early chloroplast biogenesis, and even leads to an albino phenotype. *rpoA*-67 encoded Phenylalanine, which located in protein structural core according to the RESOPS database. Chateigner-Boutin et al. (2008) thought that the unedited *rpoA* transcript has a critical effect on the translation or stability of the RpoA protein. These results suggest that the editing of six residues (*accD*-271, *atpA*-305, *atpA*-383, *petB*-160, *rpl20*-103, and *rpoB*-809) is necessary for the stability and/or activity of the corresponding protein. Comparison with data in RESOPS, *rpoB*-113, *accD*-474, *rps2*-83, *rps14*-27, and *rps14*-50 are just embedded in proteins and not in a protein structural core. Zhou et al. (2009) has confirmed that the lack of editing of *rpoB*-113 in *Arabidopsis* possibly reduces RpoB activity. As a result, we deduced that the other 4 sites may also have a close relationship with the activity of each protein. The residues of *rpoB*-184, *rpoB*-189, *rps2*-45, *rps18*-74, *petB*-24, and *clpP*-187 are located on the each protein surface. Chateigner-Boutin et al. (2008) could not clearly determine the impact of the unedited *clpP*-187. Then, Yura et al. (2009) found that the Tyrosine of *clpP*-187 located on the ClpP surface according to RESOPS, and unedited residue of *clpP*-187 have almost no effect on the stability and function of ClpP. Similar to *clpP*-187, the other five sites may not be very important for the stability or function of each corresponding protein. In addition, all the residues that change protein 3-D structures focus on *ndh* genes, but this group gene has not been collected in the RESOPS database. This means that those residues with editing in NDH cannot currently be analyzed by RESOPS.

Taken together, our results imply that chloroplast RNA editing may influence protein secondary and 3-D structures, but the degree of the effect is different for each editing site. Edited codons located in the protein structure core or buried inside, such as *accD*-271, *atpA*-

305, *atpA*-383, *petB*-160, *rpl20*-103, *rpoB*-809, *rpoB*-113, *accD*-474, *rps2*-83, *rps14*-27, and *rps14*-50, may have a great impact on protein 3-D structures. Sites such as *rpoB*-184, *rpoB*-189, *rps2*-45, *rps18*-74, *petB*-24, and *clpP*-187 located on the surface may have a slight effect or no impact at all. All the conclusions still need experimental methods to analyze, and more editing sites are expected to be verified.

ACKNOWLEDGMENTS

Research supported by the National Key Basic Research Program of China “973” (#2010CB126000). We are thankful to the National Laboratory of Protein Engineering and Plant Genetic Engineering of Beijing University for providing help and technical support.

REFERENCES

- Bock R (2000). Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochimie* 82: 549-557.
- Bock R (2001). RNA Editing in Plant Mitochondria and Chloroplasts. In: RNA Editing (Bass BL, ed.). Oxford University Press, Oxford, 38-60.
- Cai W, Ji D, Peng L, Guo J, et al. (2009). LPA66 is required for editing *psbF* chloroplast transcripts in *Arabidopsis*. *Plant Physiol.* 150: 1260-1271.
- Calsa JT, Carraro DM, Benatti MR, Barbosa AC, et al. (2004). Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Curr. Genet.* 46: 366-373.
- Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andres C, et al. (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J.* 56: 590-602.
- Chou PY and Fasman GD (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* 47: 45-148.
- Corneille S, Lutz K and Maliga P (2000). Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol. Gen. Genet.* 264: 419-424.
- Fiebig A, Stegemann S and Bock R (2004). Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Res.* 32: 3615-3622.
- Gray MW and Covello PS (1993). RNA editing in plant mitochondria and chloroplasts. *FASEB J.* 7: 64-71.
- Guzowska-Nowowiejska M, Fiedorowicz E and Plader W (2009). Cucumber, melon, pumpkin, and squash: are rules of editing in flowering plants chloroplast genes so well known indeed? *Gene* 434: 1-8.
- Hammani K, Okuda K, Tanz SK, Chateigner-Boutin AL, et al. (2009). A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21: 3686-3699.
- Häder DP and Sinha RP (2005). Solar ultraviolet radiation-induced DNA damage in aquatic organisms: potential environmental impact. *Mutat. Res.* 571: 221-233.
- Hoch B, Maier RM, Appel K, Igloi GL, et al. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353: 178-180.
- Inada M, Sasaki T, Yukawa M, Tsudzuki T, et al. (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.
- Jiang Y, Yu J, Yao Y, Song M, et al. (2010). Research progress of cotton chloroplast genome. *Cotton Sci.* 22: 495-500.
- Kahlau S, Aspinall S, Gray JC and Bock R (2006). Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J. Mol. Evol.* 63: 194-207.
- Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, et al. (2003). RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Res.* 31: 2417-2423.
- Lee SB, Kaittanis C, Jansen RK, Hostetler JB, et al. (2006). The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. *BMC Genomics* 7: 61.
- Lutz KA and Maliga P (2001). Lack of conservation of editing sites in mRNAs that encode subunits of the NAD(P)H dehydrogenase complex in plastids and mitochondria of *Arabidopsis thaliana*. *Curr. Genet.* 40: 214-219.
- Maier RM, Neckermann K, Igloi GL and Kössel 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.
- Miyata Y and Sugita M (2004). Tissue- and stage-specific RNA editing of *rps 14* transcripts in moss (*Physcomitrella patens*) chloroplasts. *J. Plant Physiol.* 161: 113-115.
- Okuda K, Myouga F, Motohashi R, Shinozaki K, et al. (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, Chateigner-Boutin AL, Nakamura T, Delannoy E, et al. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* 21: 146-156.
- Okuda K, Hammami K, Tanz SK, Peng L, et al. (2010). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. *Plant J.* 61: 339-349.
- Robbins JC, Heller WP and Hanson MR (2009). A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast *accD* transcript. *RNA* 15: 1142-1153.
- Sasaki T, Yukawa Y, Miyamoto T, Obokata J, et al. (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, et al. (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 and Nagano Y (2004). Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci. Biotechnol. Biochem.* 68: 1175-1184.
- Sasaki Y, Kozaki A, Ohmori A, Iguchi H, et al. (2001). Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J. Biol. Chem.* 276: 3937-3940.
- Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, et al. (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.
- Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, et al. (2005). Pigment deficiency in nightshade/tobacco hybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. *Plant Cell* 17: 1815-1828.
- Shikanai T (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol. Life Sci.* 63: 698-708.
- Wakasugi T, Hirose T, Horihata M, Tsudzuki T, et al. (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.
- Wolf PG, Rowe CA and Hasebe M (2004). High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* 339: 89-97.
- Yu QB, Jiang Y, Chong K and Yang ZN (2009). AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. *Plant J.* 59: 1011-1023.
- Yura K and Go M (2008). Correlation between amino acid residues converted by RNA editing and functional residues in protein three-dimensional structures in plant organelles. *BMC Plant Biol.* 8: 79.
- Yura K, Sulaiman S, Hatta Y, Shionyu M, et al. (2009). RESOPS: A database for analyzing the correspondence of RNA editing sites to protein three-dimensional structures. *Plant Cell Physiol.* 50: 1865-1873.
- Zeng WH, Liao SC and 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.
- Zhou W, Cheng Y, Yap A, Chateigner-Boutin AL, et al. (2009). The *Arabidopsis* gene YS1 encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J.* 58: 82-96.
- Zito F, Kuras R, Choquet Y, Koessel H, et al. (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.