UC Irvine
UC Irvine Previously Published Works

Title

Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes

Permalink

https://escholarship.org/uc/item/9f37r1w2

Journal

Current Genetics: Lower Eukaryotes and Organelles, 56(5)

ISSN

1432-0983

Authors

Salmans, Michael Lee
Chaw, Shu-Miaw
Lin, Ching-Ping
et al.

Publication Date

2010-10-01

DOI

10.1007/s00294-010-0312-4

Peer reviewed

RESEARCH ARTICLE

Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes

**Michael Lee Salmans · Shu-Miaw Chaw ·
Ching-Ping Lin · Arthur Chun-Chieh Shih ·
Yu-Wei Wu · R. Michael Mulligan**

Abstract Sequence analysis of organelle genomes and comprehensive analysis of C-to-U editing sites from flowering and non-flowering plants have provided extensive sequence information from diverse taxa. This study includes the first comprehensive analysis of RNA editing sites from a gymnosperm mitochondrial genome, and utilizes informatics analyses to determine conserved features in the RNA sequence context around editing sites. We have identified 565 editing sites in 21 full-length and 4 partial cDNAs of the 39 protein-coding genes identified from the mitochondrial genome of *Cycas taitungensis*. The information profiles and RNA sequence context of C-to-U editing sites in the *Cycas* genome exhibit similarity in the immediate flanking nucleotides. Relative entropy analyses indicate that similar regions in the 5′ flanking 20 nucleotides have information content compared to angiosperm mitochondrial genomes. These results suggest that evolutionary constraints exist on the nucleotide sequences immediately adjacent to C-to-U editing sites, and similar regions are utilized in editing site recognition.

Keywords RNA editing · Relative entropy · Organelle evolution

M. L. Salmans · R. M. Mulligan (✉)
Department of Developmental and Cell Biology,
University of California, Irvine, CA 92697-2300, USA
e-mail: rmmullig@uci.edu

S.-M. Chaw · C.-P. Lin
Biodiversity Research Center, Academia Sinica,
128 Academy Road, Section 2, Taipei 115, Taiwan

A. C.-C. Shih
Institute of Information Science, Academia Sinica,
128 Academy Road, Section 2, Taipei 115, Taiwan

Y.-W. Wu
School of Informatics and Computing, Indiana University,
Bloomington, IN 47405-7000, USA

Introduction

Comprehensive analysis of RNA editing sites has been reported for several chloroplast and mitochondrial genomes (mtDNAs) of non-flowering and flowering plants. Chloroplast genomes (cpDNAs) of flowering plants typically possess about 30 C-to-U editing sites and no U-to-C editing sites (Wakasugi et al. 2001). However, cpDNAs of some non-flowering plants exhibit much higher frequencies of RNA editing. For example, the cpDNA from the hornwort, *Anthoceros formosae*, exhibits extensive editing with 509 C-to-U and 433 U-to-C editing sites (Kugita et al. 2003a, b), but other non-flowering plants such as *Marchantia* have no RNA editing sites (Oda et al. 1992). Comprehensive analysis of editing sites in the cpDNA of the fern *Adiantum capillus-veneris* reported 350 C-to-U and 35 U-to-C editing sites (Wolf et al. 2003, 2004). The moss *Takakia lepidozioides* has been partially characterized for editing sites, and 302 C-to-U and no U-to-C editing sites were reported (Yura et al. 2008). Several angiosperm mtDNAs have between 350 and 500 C-to-U editing sites (Giege and Brennicke 1999; Kubo et al. 2000; Notsu et al. 2002; Handa 2003; Mower and Palmer 2006), and the *Cycas* mtDNA was predicted to include over 1,000 C-to-U editing sites (Chaw et al. 2008).

Thus, the prevalence of C-to-U editing sites in the cpDNAs and mtDNAs of diverse plants offers an excellent opportunity to evaluate the nucleotide distribution in different organelle systems and diverse taxa of land plants. Several computational studies have examined RNA editing sites (Cummings and Myers 2004; Tillich et al. 2006; Mulligan et al. 2007; Jobson and Qiu 2008; Yura et al. 2008). C-to-U editing sites in angiosperm mtDNAs have similar information profiles in the 5′ and 3′ flanking regions and a similar RNA context (Mulligan et al. 2007). These results are consistent with molecular analyses that demonstrated that 5′ flanking regions are required for editing site conversion (Takenaka et al. 2004; van der Merwe et al. 2006). Edited chloroplast RNA fragments for ndhB-9 and ndhF-1 share sequence similarity around the editing site, and are specifically bound by the same chloroplast protein (Kobayashi et al. 2008). In addition, CRR4, a PPR protein required for editing in *Arabidopsis* chloroplasts, is a sequence-specific RNA-binding protein that binds to sequences comprised of 25 nucleotides upstream and 10 nucleotides downstream of the ndhD-1 editing site (Okuda et al. 2006).

RNA editing sites in lower plant cpDNAs have also been analyzed with computational tools. The distribution of nucleotides around editing sites in lower and higher plant chloroplasts was examined within codons, and a detailed analysis of the effects on codon changes and usage was developed (Tillich et al. 2006), and a model for the evolution of RNA editing based on conservation of codon usage was proposed. The RNA sequences flanking 302 C-to-U editing sites from *Takakia* cpDNA were classified into eight groups with common patterns, and these patterns could be used to predict novel editing sites (Yura et al. 2008). Recently, Jobson and Qiu (2008) examined C-to-U and U-to-C editing patterns with respect to codon position and amino acid changes in the cpDNAs and mtDNAs of plants (Jobson and Qiu 2008). Editing was reported to increase the hydrophobicity and molecular size of the amino acid side chain, was most abundant in genes of membrane proteins, and was more frequent in T-rich sequences and in genes under positive selection (Jobson and Qiu 2008).

In this paper, we report the first extensive analysis of RNA editing from the mtDNA of a gymnosperm, *Cycas taitungensis* (Taitung cycad) and confirm 565 editing sites in 25 mitochondrial genes. The *Cycas* editing sites and sequence data from known flowering and non-flowering plant mtDNAs are analyzed with informatics tools. Common features of the editing sites from these diverse taxa suggest similar mechanisms of editing site recognition and conversion.

Methods

DNA sequence analysis

Five grams of fresh *Cycas* leaves were frozen in liquid nitrogen and pulverized with a mortar and pestle. Total RNA was extracted and purified by RNeasy® Plant Mini Kit (Qiagen, Hilden). For reverse transcriptase-polymerase chain reaction (RT-PCR) assay, total RNA was treated with DNAse I and then extracted with phenol–chloroform to eliminate DNA contamination. RNA was reverse transcribed to synthesize cDNA with Superscript II reverse transcriptase (Invitrogen, Indianapolis) and a gene-specific primer (Additional file 2) according to the manufacturer's protocol. cDNAs were PCR amplified with specific reverse and forward primer pairs (Additional file 2). The PCR products were purified by gel extraction (Gel-M, Viogene Inc., Taiwan) and directly sequenced with the BigDye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. DNA was sequenced with Applied Biosystems ABI 3700 sequencer.

DNA and RNA sequence data

cpDNA sequences and the identification of editing sites were obtained from the following Genbank accessions and citations: *Adiantum capillus-veneris*, AY178864 (Wolf et al. 2003, 2004); *Anthoceros formosae*, NC_004543 (Kugita et al. 2003a, b); *Takakia lepidozioides*, AB193121, AB254134, AB299142, AB367138, AB367138 (Yura et al. 2008); *Zea mays*, X86563 (Maier et al. 1995); and *Nicotiana tabacum*, Z00044 (Shinozaki et al. 1986; Tsudzuki et al. 2001). mtDNA sequences and the identification of editing sites were obtained from the following Genbank accessions and citations: *C. taitungensis*, AP009381 (Chaw et al. 2008); *Arabidopsis thaliana*, NC001284 (Giege and Brennicke 1999); *Beta vulgaris*, AP006444 (Kubo et al. 2000); *Brassica napus*, BA000009, DQ381444–DQ381465 (Handa 2003; Mower and Palmer 2006); *Oryza sativa*, BA000029 (Notsu et al. 2002). Additional file 3 provides accession numbers for cDNA sequence of *Cycas* mitochondrial genes.

Protein-coding sequences for each genome were annotated with edited nucleotides represented by an upper case C, and these sequences are available in Additional files 4, 5, 6, and 7. Thus, edited nucleotide positions are represented as the unedited nucleotide in the sequences analyzed in this study. The sequence files were limited to known protein-coding sequences larger than 100 nucleotides, and small or uncharacterized ORFs, introns, and other non-coding sequences were not analyzed.

Computational analyses

Computational analyses were performed as previously described (Mulligan et al. 2007). Briefly, the nucleotide distribution around all edited and unedited cytidines was analyzed in a one-, two-, or three-nucleotide sliding window. Each coding sequence was scanned for edited and unedited C, and the sequence was written to an array of edited or unedited sequences. Thus, the sequences flanking all edited or unedited cytidines were aligned in a matrix. The frequency of nucleotides around the edited or unedited nucleotide (P or Q, respectively) was used to calculate the selectivity ratio (P/Q). Thus, a nucleotide or series of nucleotides with a selectivity ratio of one has the same relative frequency around edited and unedited cytidines, while a nucleotide with a selectivity ratio greater than 1 is more frequently present around an edited cytidine. Relative entropy was calculated as the Kullback–Leibler distance by the equation $d = \sum P_k \log(P_k/Q_k)$ over k terms ($k = 4^n$) for the distribution of nucleotides in 1, 2, or 3 nucleotide windows.

Random editing site assignment

Random editing site assignment was used to produce coding sequences with randomly assigned editing sites. The editing site reassignment program scans each coding-sequence entry, and determines the number and codon position of each of the editing sites. This program randomly assigns a cytidine in the same codon position as an editing site and maintains the number and codon position of editing sites in a coding sequence. As a result, it is not "random". Statistics such as mean, standard deviation, variance, and confidence intervals were determined from 1,000 iterations of random editing site reassignment.

Results

Characteristics of abundant RNA editing in *Cycas* mtDNA

Table 1 shows the distribution of RNA editing sites within gene sequences of the *C. taitungensis* mtDNA. Five hundred and sixty-five editing sites are confirmed by cDNA sequence analysis of 21 genes and partial sequences for four additional genes. Using PREP-Mt with several cutoff scores (Mower 2005), Chaw et al. (2008) predicted that the *Cycas* mtDNA has more than 1,000 C-to-U editing sites in the 39 protein-coding genes. Table 2 shows the distribution of editing sites within codons, and the distribution in the first, second and third codon positions are 30, 65, and 5%,

respectively. This pattern of editing site distribution is very similar to that in angiosperm mtDNAs (Mulligan et al. 2007). Some start and stop codons are created by RNA editing in the *Cycas* mtDNA (Additional file 1). Start codons are created by editing ACG to ATG in three genes, including *atp1*, *cox1* and *sdh3*, while stop codons are created by CGA to UGA conversion in *atp6*, *atp9*, *ccmFC*, *nad4*, *rps12*, and *sdh3*, and by CAA to UAA conversion for *atp8*, *nad4L*, and *rps11*. Table 2 shows the distribution of observed editing sites within codons in *Cycas* mitochondrial genome.

Informatics analysis shows high information in the 5′ flanking sequences of editing sites in the *Cycas* mtDNA

The relative entropy around *Cycas* editing sites is analyzed in a sliding window of 1, 2, or 3 nucleotides (Fig. 1a, b, c, respectively). The profiles show large values at nucleotides −1 and −2 and small peaks in the 5′ flanking region (−9, −6/−5/−4) and at +1. The influence of codon position is analyzed by separate analyses of editing sites in the first or second codon position, and the relative entropy of these subsets also exhibits similar information profiles (Fig. 1d). The highest relative entropy is present in the −1 position, and profile is similar in the 5′ flanking region with a peak at −5. The major difference in the relative entropy around editing sites in codon positions 1 and 2 is the large peak at the +2 position in CPA1, and a similar result was observed in angiosperm mitochondrial genomes (Mulligan et al. 2007). This position represents the first downstream wobble position, and synonymous mutations may allow optimization of the editing site for efficient editing, and would result in increased entropy at these positions. The information content around editing sites in angiosperm mtDNAs is similar with very high information immediately 5′ of the editing site, a peak at nucleotides −6/−5/−4 and +1, and relatively little information in the 3′ flanking region (Mulligan et al. 2007).

RNA sequence context of C-to-U editing sites

The highest level of information around C-to-U editing sites resides in the nucleotides immediately upstream of the edited nucleotide. Table 3 compares the selectivity ratios (P/Q) for the distribution of dinucleotides in the −2/−1 position around editing sites in three mtDNAs and five cpDNAs. The analysis of C-to-U editing sites in *Cycas* mtDNA reveals a similar distribution in the angiosperm mtDNAs, *Arabidopsis* and *Oryza*. The dinucleotides UU, UC, CU, and AU are highly enriched at the −2/−1 position (Table 3). Dinucleotides with a purine at the −1 position were rarely observed at the −2/−1 position and exhibited very small selectivity ratios.

Table 1 RNA editing site distribution in *Cycas*, *Arabidopsis*, *Beta*, and *Oryza* genes and gene loss from the mitochondrial genomes

Function	Gene	*Cycas taitungensis*			# ES	Arab # ES	Beta # ES	Oryza # ES
		Est. editing sites						
		1.0	0.8	0.6				
Complex I	*nad1*	38	44	47	–	24	20	23
	nad2	37	44	52	–	31	24	30
	nad3	15	28	29	31	12	12	15
	nad4	63	77	89	–	32	19	20
	nad4L	10	14	15	17	9	10	10
	nad5	66	82	86	–	27	17	11
	nad6	21	25	32	–	10	11	18
	nad7	23	30	31	–	28	20	32
	nad9	8	16	17	24	7	5	12
Complex II	*sdh3*	6	8	11	27	(t)	(t)	(t)
	sdh4				(t)	(Ψ)	(Ψ) 4	(t)
Complex III	*Cob*	46	49	51	54	7	13	19
Complex IV	*cox1*	55	57	62	44 (p)	0	0	4
	cox2	18	21	21	–	15	9	19
	cox3	19	26	29	28	8	4	1
Complex V	*atp1*	26	34	39	22 (p)	5	3	5
	atp4	6	7	9	15	8	12	9
	atp6	38	41	47	54 (p)	1	12	17
	atp8	11	11	11	15	0	2	4
	atp9	10	12	12	14	4	5	8
Cytochrome biogenesis	*ccmB*	21	32	42	37 (p)	39	30	35
	ccmC	23	25	38	6	28	28	36
	ccmFC	27	32	34	–	16	13	27
	ccmFN	35	40	47	–	22	23	31
	ccmFN2					12	–	–
Ribosomal proteins	*rps1*	5	5	9	17	(t)	(t)	3
	rps2	6	8	11	17	(t)	(t)	10
	rps3	9	17	21	–	10	6	10
	rps4	15	26	31	–	15	11	15
	rps7	5	7	14	10	0	3	2
	rps10	4	5	6	–	(t)	(t)	(t)
	rps11	4	4	4	14	(t)	(t)	(Ψ) 4
	rps12	4	14	14	16	8	6	0
	rps13	2	7	9	8	(t)	2	8
	rps14	4	4	7	8	(Ψ)	(t)	(Ψ) 0
	rps19	6	7	8	11	(Ψ)	(t)	6
	rpl2	6	6	10	–	1		1
	rpl5	8	13	14	11	10	5	1
	rpl16	4	6	7	13	8	(t)	12
Other	*matR*	21	25	30	–	9	9	(nt)
	mttB	13	25	38	52	24	19	33
Total		738	934	1,084	565	441	357	491

The number of editing sites in *Cycas* mitochondria is estimated based on computational prediction utilizing cutoff scores of 1.0, 0.8 and 0.6 (Mower 2005), with 0.6 as the recommended score to balanced false negative and false positive. The actual number of editing sites for 21 *Cycas* mitochondrial genes is shown in column 6. Partial cDNAs are presented for *atp1*, *atp6*, *coxI*, and *ccmB*, and these are noted with "(p)". The number of editing sites for *Arabidopsis*, *Beta*, and *Oryza* mtDNAs is obtained from the literature (Giege and Brennicke 1999; Notsu et al. 2002; Mower and Palmer 2006). Pseudogenes are represented with "(Ψ)", and the number of reported editing sites. Genes that are not transcribed are indicated with "(nt)", and those have been transferred to the nuclear genome are indicated with "(t)"

Table 2 Editing site distribution within codons in the *Cycas* mitochondrial genome

Original codon		Edited codon		Editing sites confirmed
Codon	Amino acid	Codon	Amino acid	
ACA	Thr (T)	ATA	Ile (I)	2
ACC	Thr(T)	ACT	Thr (T)	4
ACC	Thr(T)	ATC	Ile (I)	2
ACG	Thr (T)	ATG	Met (M)	7
ACT	Thr (T)	ATT	Ile (I)	5
ATC	Ile (I)	ATT	Ile (I)	13
CAA	Gln (Q)	TAA	Stop	3
CAC	His (H)	TAC	Tyr (Y)	4
CAG	Gln (Q)	TAG	Stop	1
CAT	His (H)	TAT	Tyr (Y)	17
CCA	Pro (P)	CTA	Leu (L)	42
CCA	Pro (P)	TCA	Ser (S)	12
CCC	Pro (P)	CCT	Pro (P)	5
CCC	Pro (P)	CTC	Leu (L)	19
CCC	Pro (P)	TCC	Ser (S)	13
CCC	Pro (P)	TTC	Phe (F)	12
CCG	Pro (P)	CTG	Leu (L)	39
CCG	Pro (P)	TCG	Ser (S)	10
CCT	Pro (P)	CTT	Leu (L)	22
CCT	Pro (P)	TCT	Ser (S)	16
CCT	Pro (P)	TTT	Phe (F)	4
CGA	Arg (R)	TGA	Stop	3
CGC	Arg (R)	CGT	Arg (R)	1
CGC	Arg (R)	TGC	Cys (C)	8
CGG	Arg (R)	TGG	Trp (W)	33
CGT	Arg (R)	TGT	Cys (C)	23
CTC	Leu (L)	CTT	Leu (L)	1
CTC	Leu (L)	TTC	Phe (F)	5
CTG	Leu (L)	TTG	Leu (L)	3
CTT	Leu (L)	TTT	Phe (F)	6
GCA	Ala (A)	GTA	Val (V)	1
GCC	Ala (A)	GTC	Val (V)	1
GCG	Ala (A)	GTG	Val (V)	7
GCT	Ala (A)	GTT	Val (V)	1
GTC	Val (V)	GTT	Val (V)	1
TCA	Ser (S)	TTA	Leu (L)	60
TCC	Ser (S)	TCT	Ser (S)	11
TCC	Ser (S)	TTC	Phe (F)	35
TCG	Ser (S)	TTG	Leu (L)	47
TCT	Ser (S)	TTT	Phe (F)	55
TTC	Phe (F)	TTT	Phe (F)	8
Total				562

The distribution of confirmed editing sites in the *Cycas* mitochondria transcriptome is shown within codons

A scatter plot compares the selectivity ratios for dinucleotides upstream of *Cycas* and angiosperm mitochondrial editing sites (Fig. 2a); thus, each of the 16 points corresponds to the selectivity ratios for a specific dinucleotide. An extraordinary level of congruence exists between the selectivity ratios of the dinucleotides in the $-2/-1$ position of the *Cycas* mtDNA and the angiosperm mtDNAs. Linear regression analysis of these data indicates slope values near 1, Y-intercepts near zero, and coefficients of determination (R^2) greater than 0.9 (Fig. 2a).

The C-to-U editing sites in the cpDNAs of non-flowering plants show a similar trend in the pyrimidine-rich dinucleotides upstream of editing sites (Table 3). The selectivity ratios upstream of C-to-U editing sites in the cpDNAs of the hornwort (*Anthoceros*), the moss (*Takakia*), and the fern (*Adiantum*) are very similar to selectivity ratios observed in angiosperm mitochondria. Figure 2b compares the selectivity ratios of *Takakia* cpDNA editing sites with plant mtDNA editing sites, and the high degree of similarity is indicated by a coefficient of determination (R^2) greater than 0.85. Therefore, the distribution of nucleotides around C-to-U editing sites in non-flowering plant cpDNAs is very similar to those in plant mtDNAs, which strengthens a common origin and early evolution of RNA editing in the two organelle systems.

Discussion

C-to-U editing sites from diverse plant sources contain similar information profiles

Informatics analyses demonstrate strong similarities in the C-to-U editing sites across diverse taxa and organelle systems. The information profiles around C-to-U editing sites generally exhibit high relative entropies in the $-1/-2$ regions and smaller peaks in the 5′ flanking 20 nucleotides. In contrast, there is generally little information in the 3′ flanking nucleotides. Furthermore, RNA sequence context at C-to-U editing sites is very similar across these diverse taxa and both organelle systems. In plant mitochondria and in lower plant chloroplasts, pyrimidine-rich dinucleotides are highly enriched upstream of C-to-U editing sites, and a very low frequency of purines exists at -1. Thus, there is a strong selection of nucleotides immediately adjacent to C-to-U editing sites across eight organelle genomes including taxa that diverged at least 400 million years ago (mya) (Palmer et al. 2004). The conserved information profile and nucleotide context around C-to-U editing sites may result from constraints related to the editing mechanism. The peaks

Fig. 1 The relative entropy around C-to-U editing sites in *Cycas* mitochondria. The relative entropy for the distribution of nucleotides is plotted for 30 nucleotides flanking RNA editing sites in 1, 2, or 3 nucleotide sliding windows (**a**, **b**, **c**, respectively). Random editing site assignment is used to reassign editing sites in the same codon position, and relative entropy analysis of 1,000 editing site reassignments is used to determine a mean relative entropy value and a 95% confidence interval. **d** The effect of codon position on relative entropy in *Cycas* mitochondrial editing sites determined in a one nucleotide window for editing sites in the first or second codon position (CPA1, CPA2). The number of editing sites analyzed in the first and second codon position is 173 and 376. Only 29 editing sites are present in the third codon position and these data are not presented
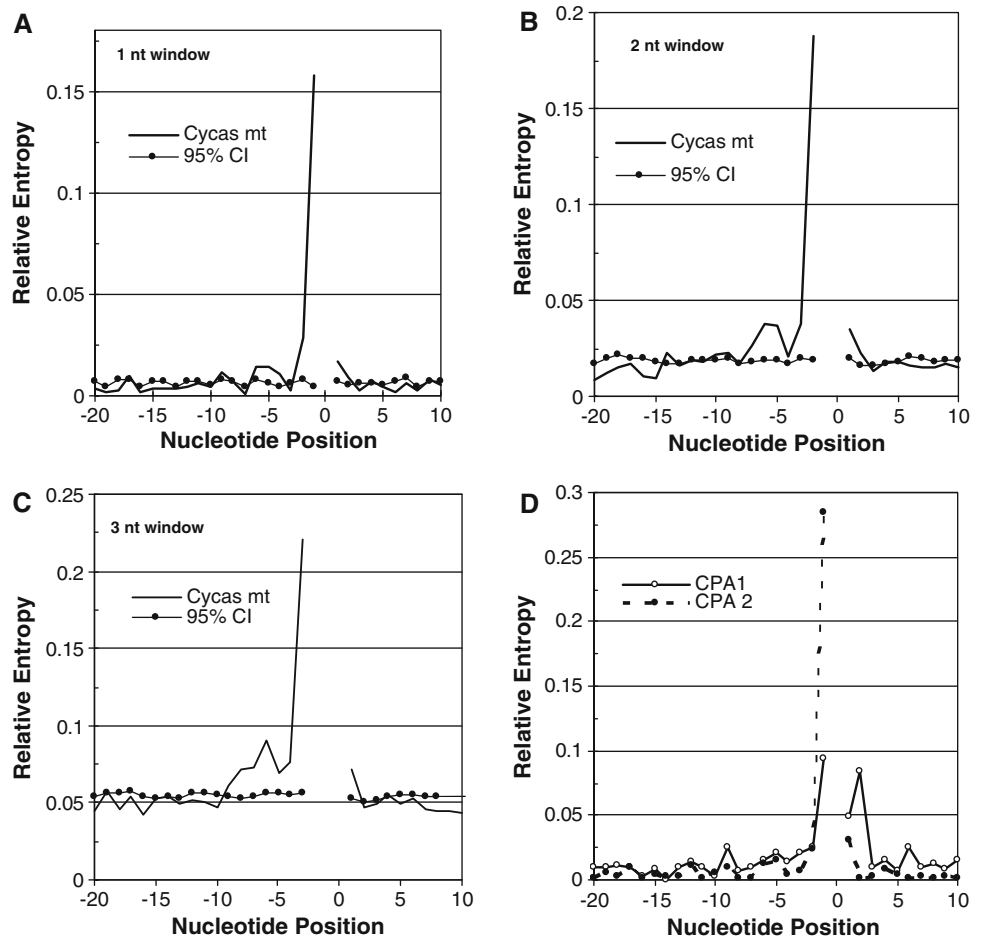

Table 3 Selectivity ratios at −2/−1 around C-to-U editing sites in cpDNAs and mtDNAs

	mtDNA			cpDNA				
	Cycas	*Arab*	*Oryza*	*Anthoceros*	*Takakia*	*Adiantum*	*Nicotiana*	*Zea*
UU	2.19	2.26	2.26	2.39	1.95	1.98	3.10	2.64
UC	2.12	1.96	2.38	1.59	2.29	1.89	1.96	1.14
CU	2.82	1.99	2.35	1.29	1.96	1.32	0.64	1.52
AU	1.46	1.82	1.50	1.48	1.28	1.05	2.26	2.23
CC	1.09	1.18	1.37	0.82	1.35	1.21	2.50	1.84
GU	1.28	0.97	1.23	0.72	1.18	1.71	0	2.02
AC	0.94	1.26	0.57	1.62	0.65	1.47	0.66	0.79
GC	0.81	0.82	0.47	0.57	0.39	0.76	0	0
UA	0.34	0.28	0.30	0.79	0.60	0.60	0	0.52
UG	0.28	0.31	0.28	0.27	0.33	0.51	0	0
CA	0.07	0.11	0.22	0.56	0.39	0.65	0	0
GA	0.17	0.05	0.06	0.05	0.40	0.79	0	0
CG	0.32	0.13	0.07	0.34	0.16	0.23	0	0
AA	0.08	0.12	0.04	0.09	0.14	0.36	0	0
AG	0.12	0.04	0	0.11	0.11	0.33	0	0
GG	0	0.05	0.16	0.12	0.21	0	0	0

The distribution of dinucleotides around C-to-U editing sites at position −2/−1 is compared in three mtDNAs and five cpDNAs. The frequency of dinucleotides adjacent to an edited or unedited cytidine (P or Q, respectively) is the number of times that a dinucleotide is observed divided by the total number of edited or unedited cytidines. The selectivity ratio is the ratio of the frequencies for the dinucleotide around an edited and unedited cytidines (P/Q)
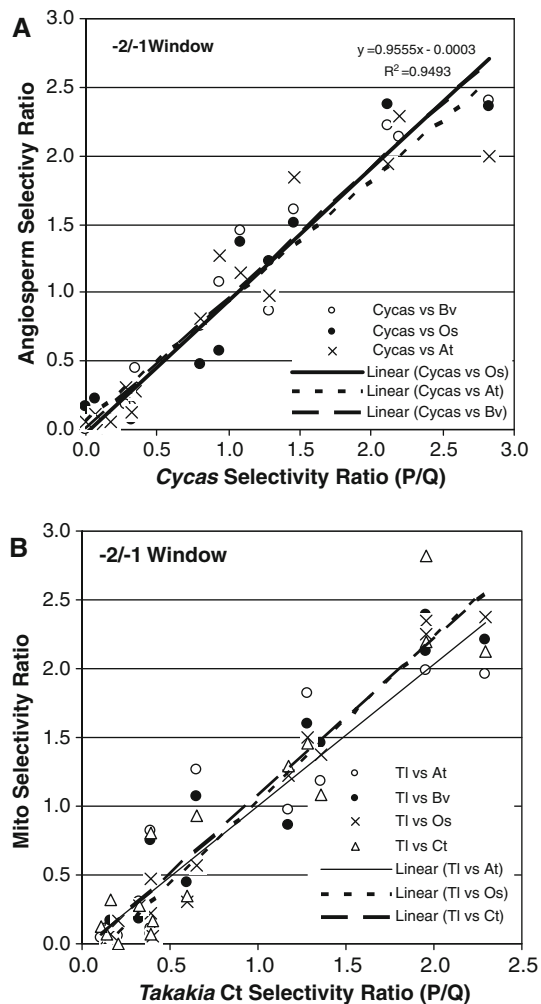
Fig. 2 Selectivity ratios around C-to-U editing sites are similar in chloroplasts and mitochondria in non-flowering and flowering plants. The selectivity ratios (P/Q) for dinucleotides in the $-2/-1$ window (**a**, **b**) are compared in a scatter plot. Each point represents the selectivity ratios for a specific dinucleotide in the two species. **a** Compares the selectivity ratios for C-to-U editing sites in *Cycas* and angiosperm mitochondria genomes in the $-2/-1$ window. Linear regression analysis for the *Cycas* selectivity ratios plotted against the *Arabidopsis*, *Oryza*, and *Beta* selectivity ratios indicates a strong congruence with slopes near 1, y-intercepts near zero, and large coefficients of determination (*Arabidopsis*: slope = 0.88, intercept = 0.06, R^2 = 0.91; *Beta*: slope = 0.96, intercept = 0.00, R^2 = 0.95; *Oryza*: slope = 0.0.97, intercept = 0.02, R^2 = 0.94). **b** Compares the selectivity ratios of C-to-U editing sites in *Takakia* chloroplasts with plant mitochondrial genomes in the $-2/-1$ window. Linear regression analysis *Takakia* selectivity ratios plot against the *Cycas*, *Arabidopsis*, *Oryza*, and *Beta* selectivity ratios indicates a strong congruence in the selectivity ratios (*Cycas*: slope = 1.14, intercept = -0.07, R^2 = 0.89; *Arabidopsis*: slope = 1.03, intercept = -0.03, R^2 = 0.85; *Beta*: slope = 1.12, intercept = -0.09, R^2 = 0.91; *Oryza*: slope = 1.20, intercept = -0.02, R^2 = 0.97)

in the information profile suggest that similar positions are utilized in editing site recognition. These results further substantiate the model that common editing site features may

exist immediately adjacent to C-to-U editing sites, and that *cis*-elements for individual editing sites reside in the 5′ flanking region.

The information profile around C-to-U editing sites usually exhibits small peaks in the 5′ flanking region. In contrast to the strong sequence conservation at $-1/-2$ positions of editing sites, relatively little sequence similarity is observed in the scatter plots in these upstream regions (data not shown). Molecular analyses of the sequences required for editing site conversion in angiosperm chloroplast and mitochondrial systems have demonstrated that the *cis*-element includes approximately 20 nucleotides of upstream sequence and relatively little downstream sequence (Shikanai 2006; van der Merwe et al. 2006; Hayes and Hanson 2008). Clusters of editing sites have been proposed for the recognition of editing sites in higher plant chloroplasts (Chateigner-Boutin and Hanson 2002), and some editing sites can be grouped into clusters that exhibit sequence similarity. In some cases, PPR genes are required for processing two or more editing sites, and the *cis*-elements share limited sequence similarity (Chateigner-Boutin 2008; Okuda et al. 2009, 2010; Zehrmann et al. 2009). The analysis of the *cis*-elements for 34 editing sites and 15 PPR proteins in *Arabidopsis* required for RNA editing indicated that the *cis*-elements for editing sites were not strikingly similar (Hammani et al. 2009).

Conclusions

Informatics analyses demonstrate that C-to-U editing sites share common features across diverse taxa and organelle systems. The information profiles around editing sites in these diverse systems show similar patterns. Furthermore, the nucleotides at $-1/-2$ show remarkable similarity across diverse taxa and different organelles systems. The conserved information profiles and nucleotide context around C-to-U editing sites across these broad taxa may be a constraint of common features of the editing mechanism.

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, Ramos-Vega M, Guevara-Garcia A, Andres C, de la Luz Gutierrez-Nava M, Cantero A, Delannoy E, Jimenez LF, Lurin C, Small I, Leon P (2008) CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. Plant J 56:590–602

Chaw SM, Shih AC, Wang D, Wu YW, Liu SM, Chou TY (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol 25:603–615

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

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

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

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 (2008) High conservation of a 5′ element required for RNA editing of a C target in chloroplast psbE transcripts. J Mol Evol 67:233–245

Jobson RW, Qiu YL (2008) Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? Biol Direct 3:43

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

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res 28:2571–2576

Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K (2003a) The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res 31:716–721

Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K (2003b) 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

Mower JP (2005) PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinform 6:96

Mower JP, Palmer JD (2006) Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. Mol Genet Genom 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 24:1971–1981

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 Genom 268:434–445

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T et al (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol 223:1–7

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

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

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

Palmer JD, Soltis DE, Chase MW (2004) The plant tree of life: an overview and some points of view. Am J Bot 91:1437–1445

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

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

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

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

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

Wakasugi T, Tsudzuki T, Sugiura M (2001) The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. Photosynth Res 70:107–118

Wolf PG, Rowe CA, Sinclair RB, Hasebe M (2003) Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res 10:59–65

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

Yura K, Miyata Y, Arikawa T, Higuchi M, Sugita M (2008) Characteristics and prediction of RNA editing sites in transcripts of the moss *Takakia lepidozioides* chloroplast. DNA Res 15:309–321

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