The Evolution of Chloroplast RNA Editing

Michael Tillich,[1] *Pascal Lehwark,† Brian R. Morton,‡ and Uwe G. Maier**

*Cell Biology, Philipps–University of Marburg, Marburg, Germany; †Indiji Software, Marburg, Germany; and ‡Department of Biological Sciences, Barnard College, Columbia University

RNA editing alters the nucleotide sequence of an RNA molecule so that it deviates from the sequence of its DNA template. Different RNA-editing systems are found in the major eukaryotic lineages, and these systems are thought to have evolved independently. In this study, we provide a detailed analysis of data on C-to-U editing sites in land plant chloroplasts and propose a model for the evolution of RNA editing in land plants. First, our data suggest that the limited RNA-editing system of seed plants and the much more extensive systems found in hornworts and ferns are of monophyletic origin. Further, although some eukaryotic editing systems appear to have evolved to regulate gene expression, or at least are now involved in gene regulation, there is no evidence that RNA editing plays a role in gene regulation in land plant chloroplasts. Instead, our results suggest that land plant chloroplast C-to-U RNA editing originated as a mechanism to generate variation at the RNA level, which could complement variation at the DNA level. Under this model, many of the original sites, particularly in seed plants, have been subsequently lost due to mutation at the DNA level, and the function of extant sites is merely to conserve certain codons. This is the first comprehensive model for the evolution of the chloroplast RNA-editing system of land plants and may also be applicable to the evolution of RNA editing in plant mitochondria.

Introduction

RNA editing has been observed in the chloroplasts of all land plants analyzed to date with the exception of the marchantiid subclass of liverworts (Freyer et al. 1997). In the mRNAs of seed plant chloroplasts, cytidine is converted into uridine (C-to-U RNA editing) at about 30 different positions (Maier et al. 1995; Wakasugi et al. 1996; Hirose et al. 1999; Schmitz-Linneweber et al. 2002; Tillich et al. 2005). These editing events are nonsynonymous with respect to the encoded polypeptide sequences and often affect amino acids that play a role in proper protein function (Bock et al. 1994; Zito et al. 1997; Sasaki et al. 2001). The editing sites themselves are recognized by relatively short sequence stretches in their immediate upstream region ("*cis*-elements," about 15 nt in length) (Bock et al. 1996; Chaudhuri and Maliga 1996; Bock et al. 1997; Hirose and Sugiura 2001; Reed, Peeters, et al. 2001; Miyamoto et al. 2002). Little is known about the factors responsible for this recognition, but various indirect data indicate that each editing site, or in some cases a small set of sites, must be recognized by specific proteinaceous factors encoded in the plant nuclear genome (Chaudhuri et al. 1995; Bock et al. 1996, 1997; Reed and Hanson 1997; Hirose and Sugiura 2001; Reed, Lyi, et al. 2001; Reed, Peeters, et al. 2001; Miyamoto et al. 2002, 2004). To date, though, only 2 nuclear encoded proteins have been shown to influence chloroplast RNA editing, and a direct participation in the editing process has yet to be established for either. CP31, a general RNA-binding and highly abundant chloroplast protein, is required for editing of 2 different tobacco sites in vitro (Hirose and Sugiura 2001) and, therefore, potentially represents a general editing factor. The PPR (Pentatricopeptide) protein CRR-4 has been shown to be essential for editing of a specific site in the chloroplasts *ndhD* mRNA of *Arabidopsis thaliana* (Kotera et al. 2005) and may be the first representative of the long sought-after site-specific editing factors.

Outside the seed plants, one representative each of ferns and hornworts, *Adiantum capillus-veneris* and *Anthoceros formosae*, respectively, have been analyzed in detail for chloroplast RNA editing (Kugita et al. 2003; Wolf et al. 2004). In both species, extensive editing, both C-to-U and—in contrast to seed plants—also U-to-C, has been shown to occur. In the chloroplasts of *A. formosae*, both types of editing are found to a similar extent, 509 C-to-U and 433 U-to-C editing sites, whereas in the chloroplasts of *A. capillus-veneris* 315 C-to-U and only 35 U-to-C editing sites have been identified. These data indicate that editing is far more extensive in these lineages than in the seed plants, and it is unknown whether the editing machineries of hornwort, fern, and seed plants are of a monophyletic origin. Only 51 editing events occur at homologous sites in *A. formosae* and *A. capillus-veneris*, and 5 events occur at homologous sites in seed plants and *A. capillus-veneris* (Wolf et al. 2004). This leaves open the possibility that the processes evolved independently. Furthermore, if the "one site one factor" hypothesis established for seed plant chloroplasts is applicable to ferns and hornworts, then we would predict hundreds of site specificity factors encoded in their genomes.

Despite the progress toward understanding the molecular mechanisms of, in particular, chloroplast RNA editing, another challenge is to unravel its raison d'être, that is, to explain its existence in evolutionary terms. This goal is particularly intriguing because, in contrast to other RNA-editing systems, no function, such as regulation of gene expression or generation of functional protein isoforms with different properties (reviewed in Gott and Emeson 2000; Gott 2003), has been attributed to plastid RNA editing yet. For a few chloroplast editing sites, edited partially or in a tissue-dependent manner, a role for the regulation of gene expression has been suggested. Nonediting of internal codons might result into alternative polypeptides (Hirose et al. 1999; Karcher and Bock 2002), and editing events restoring cryptic translational start codons might regulate protein synthesis (Hirose and Sugiura 1997). However, for none of them, a regulating function like, for example, the production of a functional alternative ("unedited") gene product has been demonstrated yet. On the contrary, a study investigating editing efficiency and abundance of maize plastid RNAs in various tissues demonstrated that nearly

[1] Present address: Institute for Biology, Humboldt University Berlin, Berlin, Germany.

Key words: RNA editing, point mutation, chloroplast, plastid.

E-mail: maier@staff.uni-marburg.de.

Mol. Biol. Evol. 23(10):1912–1921. 2006
doi:10.1093/molbev/msl054
Advance Access publication July 11, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

all editing sites are edited in concert at high efficiency in chloroplasts (Peeters and Hanson 2002). And even if some of the sites are edited at a lower efficiency in some nonphotosynthetic tissues, a biological effect of reduced editing efficiency is expected to be superceded by the much stronger variations in abundance of the respective transcripts (Peeters and Hanson 2002). The main function of chloroplast RNA editing, therefore, appears to be the generation of codons important for protein function. This raises the possibility that a seemingly complex editing mechanism has evolved in the nuclear genome whose function is to neutralize point mutations, or to compensate for the lack thereof at certain sites, in the chloroplast genome. A 3-step model for the evolution of editing systems neutralizing mutations in organellar genomes as a consequence of genetic drift has been proposed by Covello and Gray (1993). In this study, we analyze editing sites from several taxa starting with an analysis of seed plants and then expanding the analysis to include the fern *A. capillus-veneris* and the hornwort *A. formosae*. Based on the similarity of context, or surrounding nucleotides, we propose a model for the evolution of RNA editing in land plant chloroplasts. Given the similarities across all land plants, we propose that the extant systems are of monophyletic origin. The context features suggest that the function of the original system was to generate sequence variation at the RNA level, in essence adding to variation generated at the DNA level. Over time, some of these original sites have been lost due to mutations occurring at the DNA level with this loss occurring to a lesser degree at sites with lower mutation rates. We also discuss the possibility that a similar model is applicable to RNA editing in seed plant mitochondria.

Materials and Methods

Lola Software

Lola Software accompanied by documentation is freely available for download at www.indiji.de.

Genome Sequences and Editing Sites

All chloroplast and plant mitochondrial genome sequences used in this study were downloaded from GenBank (chloroplast: *A. thaliana* [NC_000932], *Nicotiana tabacum* [NC_001879], *Atropa belladonna* [NC_004561], *Zea mays* [NC_001666], *Pinus thunbergii* [NC_001631], *A. capillis-veneris* [NC_004766], and *A. formosae* [NC_004543]; mitochondrial: *A. thaliana* [NC_001284]). The chloroplast editing sites of *A. formosae* and *A. capillus-veneris* and the mitochondrial editing sites of *A. thaliana* were extracted form the respective GenBank genome entries. Chloroplast editing sites of seed plants were collected from the indicated publications. Codon usages were taken from the Kazusa Codon Usage Database (http://www.kazusa.or.jp/codon/).

Results and Discussion

C-to-U RNA Editing in Seed Plant Chloroplast Occurs Predominately at Positions with Low C-to-T Point Mutation Rates on DNA

It is known from studies on chloroplast RNA editing in seed plants that many editing sites are flanked by a U upstream and an A downstream (what we will call a U_A context). To more fully investigate the context of RNA editing sites, all sites reported for the 5 seed plant species that have been analyzed in detail (*N. tabacum*, *Z. mays*, *P. thunbergii*, *A. belladonna*, and *A. thaliana*) were sampled. Altogether 155 C-to-U chloroplast editing events have been described for these 5 species (Maier et al. 1995; Wakasugi et al. 1996; Hirose et al. 1999; Schmitz-Linneweber et al. 2002; Tillich et al. 2005), and these 155 sites comprise 85 unique positions where RNA editing occurs (table 1). The contexts of these 85 editing sites are shown in figure

Table 1
Chloroplast RNA-Editing Sites of 5 Seed Plant Species

Gene	Codon Number	Codon	Source	Gene	Codon Number	Codon	Source
accD	265	uCg	At	*petD*	37	uCa	Pt
atpA	258	uCa	Pt		108	uCa	Pt
	264	cCc	Nt		133	uCc	Pt
	265	ucC	Nt		161	Caa	Pt
	335	gCa	Pt	*petG*	6	cCa	Pt
	345	uCa	Pt		27	uCa	Pt
	383	uCa	Zm	*petL*	1	aCg	Pt
atpF	22	uCa	Pt		2	cCu	At
	31	cCa	At		25	uCg	Pt
	34	uCa	Pt		32	Caa	Pt
	124	Ccc	Pt	*psaB*	620	uCa	Pt
clpP	187	Cau	At	*psbB*	464	uCu	Pt
matK	236	Cau	At	*psbE*	72	Ccu	At
	451	Cau	Zm	*psbF*	26	uCu	At
ndhA	17	uCg	Zm	*psbL*	1	aCg	Nt
	114	uCa	At	*psbl2*	31	cCa	Pt
	158	uCa	Zm		1	aCg	Zm
	189	uCa	Ab	*rpl20*	103	uCa	Nt
	358	uCc	Nt	*rpoA*	277	uCa	Nt
ndhB	50	uCa	At	*rpoB*	113	uCu	At
	156	cCa	At		158	uCa	Nt
	196	Cau	At		184	uCa	At
	204	uCa	Nt		187	uCg	Zm
	246	cCa	Nt		206	cCg	Zm
	249	uCu	At		667	uCu	Nt
	277	uCa	At		811	uCa	At
	279	uCa	At	*rpoC1*	21	uCa	Nt
	291	uCa	At	*rpoC2*	925	uCg	Zm
	419	Cau	At		1248	uCa	Nt
	494	cCa	At	*rps2*	45	aCa	Nt
ndhD	1	aCg	At		83	uCa	Nt
	128	uCa	At	*rps3*	61	uCa	Zm
	200	uCa	Nt	*rps8*	61	uCa	Zm
	225	uCa	At	*rps14*	27	uCa	At
	293	uCa	At		50	cCa	At
	296	cCc	At	*ycf3*	15	uCc	Zm
ndhF	21	uCa	Zm		62	aCg	Zm
	97	uCa	At	*ycf5*	235	uCa	Pt
ndhG	UTR	UTR	Zm		239	aCu	Pt
	17	uCc	At		301	Cgu	Pt
	116	cCa	Zm				
petB	7	Cgg	Pt				
	204	cCa	Nt				
	200	cCa	Pt				
	212	Cca	Pt				

NOTE.—RNA editing in the chloroplasts of 5 seed plant species occurs at 85 different, nonhomologous genome positions. This set was constructed by the consecutive extension of the plastid editotype of *Arabidopsis thaliana* (At) with additional sites of *Nicotiana tabacum* (Nt), *Atropa belladonna* (Ab), *Zea mays* (Zm), and *Pinus thunbergii* (Pt) (see text for references). The genes and codons affected by editing (uppercase ''C'' represents the editing sites) as well as the source species from which each editing site was retrieved are shown.

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

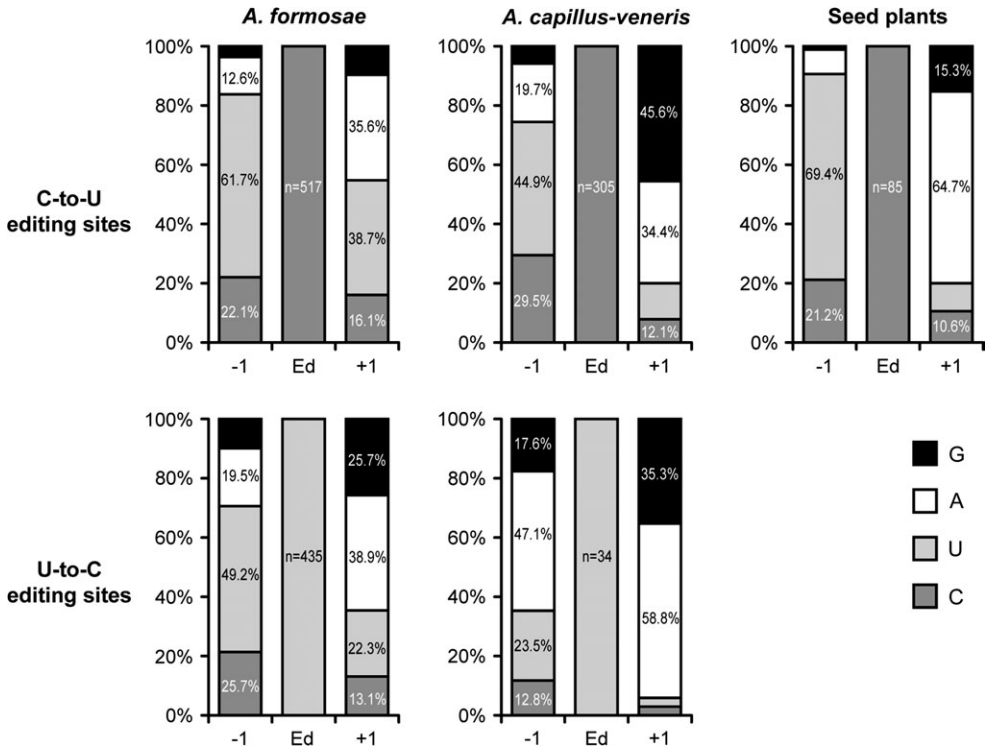

FIG. 1.—Sequence context of chloroplast RNA-editing sites of the investigated seed plants, the fern *Adiantum capillus-veneris* and the hornwort *Anthoceros formosae* (see text for references). The contexts are presented separately for C-to-U (top) and U-to-C editing sites (bottom). "Ed" depicts the editing site; "−1" and "+1," the nucleotide positions up- or downstream, respectively. The absolute number of editing events is shown in the central columns; values are indicated for adjacent nucleotides that exceed a frequency of 10%.

1. Overall, 69.4% of sites have a U upstream and 64.7% have an A downstream whereas a total of 40 editing events (47.0%) occur in a U_A context (fig. 2*A*). The contexts of C-to-U editing sites from *A. formosae* and *A. capillus-veneris* show a similar bias toward a U_A context, although it is weaker than in seed plants (figs. 1 and 2).

As is the case in the chloroplasts of *A. formosae* and *A. capillus-veneris* (Kugita et al. 2003; Wolf et al. 2004; Supplementary Table 1), the codons most frequently affected by RNA editing in seed plant chloroplasts are UCN serine codons (yielding UUR leucine or UUY phenylalanine codons after editing). In all 3 taxa, hornwort, fern, and seed plants, these conversions represent the majority of editing events (Supplementary Table 1). Serine-to-leucine/phenylalanine conversions are expected to have a strong impact on the encoded polypeptide because the involved amino acids differ strongly in their physicochemical properties. In contrast, editing events exchanging codons for amino acids with similar physicochemical properties (such as leucine-to-phenylalanine or alanine-to-valine) or editing at third-codon positions occur rarely (Supplementary Table 1).

Although there is a bias in amino acid exchanges, a few of lines of evidence suggest that the bias toward a U upstream and an A downstream from an edited site does not depend on the identity of the encoded amino acid. First, of the 50 serine codons affected by editing, 35 (70%) are UCA codons whereas the mean codon usage of UCA in the analyzed seed plant chloroplast genomes is only 25.4% of the UCN serine codons (fig. 3, top panel). Second, these 35 UCA edits make up 87.5% of the 40 UCR-to-UUR serine-to-leucine conversions, whereas UCA composes just 66.5% of UCR serine codons (fig. 3, top panel). Third, of the 15 CCN proline codons edited at second-codon position, 10 (66.7%) are CCA codons, which make up only 28.1% of CCN proline codons (fig. 3, middle panel). Finally, the U_A bias is also exhibited by editing sites located at first codon positions: 9 of the 12 codons (75.0%) edited at the first position exhibit a U at −1 and 5 of those are CAU codons (fig. 3, bottom panel). Overall, these lines of evidence suggest that the U_A context bias is independent of codon position and is not a function of any targeting of serine codons. Instead, the high proportion of edited serine codons in seed plant chloroplasts seems to be the result of additive selection for 1) replacements involving amino acids with profound differences in their physicochemical properties and 2) bias toward a U_A context.

The finding that editing sites in seed plant chloroplasts occur predominantly within a U_A context correlates with findings concerning context and variation of point mutation rates. It is well established that in flowering plant organelle DNA, the transition ratio of a given nucleotide depends on the identity of its immediate neighboring bases. In general, the transition rate decreases with an increase in the A/T context (the number of A/T base pairs, 0, 1, or 2, immediately flanking the site) as well as with an increase in the number of 5′ pyrimidines (Morton et al. 1997; Morton 2003). Overall, in flowering plant chloroplast DNA, the lowest transition rate is exhibited by a nucleotide residing in the T_A context, a context that has an A/T context of two and two 5′ pyrimidines (Morton 2003). Thus, in seed plant

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

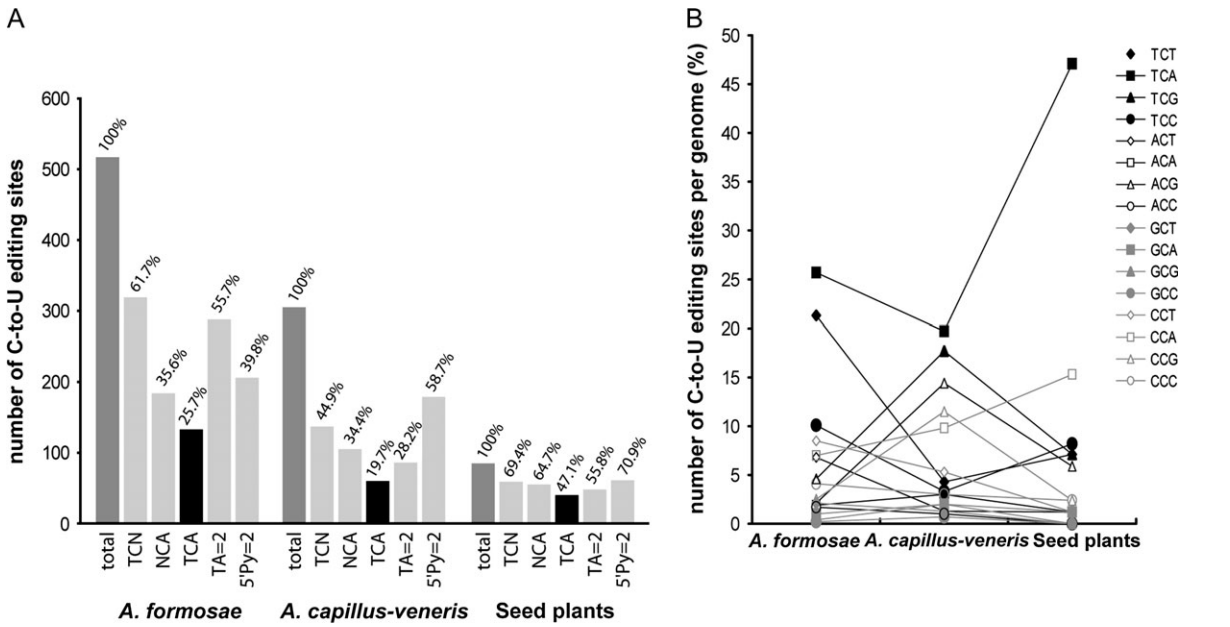

FIG. 2.—(*A*) Number and DNA sequence contexts of C-to-U RNA-editing sites described for the chloroplasts of *Anthoceros formosae*, *Adiantum capillus-veneris*, and the 5 seed plant species (see text). Columns represent the total number of described editing events (dark gray) and the number of those residing in the indicated nucleotide contexts (light gray and black columns). "C" indicates the editing sites; percentages refer to the total amount of editing sites in the respective subset. Note the relative increase of the T_A consensus context harboring editing sites for seed plants (black). (*B*) Relative abundance of dinucleotide contexts harboring C-to-U RNA-editing sites in the chloroplast genomes of seed plants, *A. capillus-veneris* and *A. formosae* (see text for references).

chloroplast RNA, 40 of 85 C-to-U editing sites are biased toward positions with a low C-to-T mutation rate.

A Comparison of Editing Sites in Hornwort, Fern, and Seed Plant Chloroplasts

Another unresolved issue concerning chloroplast RNA editing is the relationships between the systems that exist in various land plant lineages. These relationships are important for establishing when chloroplast RNA editing evolved. To investigate this, the chloroplast editotypes of *A. formosae* and *A. capillus-veneris* were compared to seed plants. A total of 18 of the 85 (21.0%) chloroplast editing sites in seed plants are shared with at least one of these two taxa and thus could be remnants of the original editing system of land plants (table 2). As in seed plants, RNA-editing sites seem to evolve rapidly within hornwort and ferns (Freyer et al. 1997; Duff and Moore 2005), so it is likely that even more such sites will be identified when more fern and hornwort species are analyzed for editing. Wolf et al. (2004) reported an overlap of 53 chloroplast editing sites between *A. capillus-veneris* and *A. formosae*. The observed distribution of shared editing sites is consistent with an early evolution of RNA editing followed by independent losses of editing sites during seed plant evolution from a common ancestor with many editing sites.

In the chloroplasts of seed plants, the recognition of editing sites by the corresponding *trans*-factors is thought to occur via a sequence-specific interaction between the *trans*-factors and the sequence region immediately upstream from the editing sites. These so-called "*cis*-elements" have been mapped to a region of about 20 nt upstream of several editing sites. It is probable that sets or clusters of editing sites are recognized by the same, or very similar, *trans*-factor(s) and that these sites exhibit sequence similarities within their upstream regions (Chateigner-Boutin and Hanson 2002, 2003; Tillich et al. 2005). We hypothesized that if editing occurs in the chloroplasts of hornwort and fern by a mechanism comparable to the one known from the chloroplasts of seed plants, then we should be able to detect putative *cis*-elements in *A. formosae* and *A. capillus-veneris* by clustering of editing sites with similar upstream elements.

To test this, we developed a bioinformatic tool, Lola, which extracts definable sequence regions from a genome and compares them pairwise. Lola was used in this study to extract the regions surrounding editing sites (−100 to +100 exclusive the editing position) and generate all possible pairwise combinations of the extracted sequences with the editing sites aligned at the midpoint. Subsequently, Lola ran a window of 15 nt in size along the aligned sequence pairs in one-nucleotide steps. At each position, the number of sequence fragments exceeding a certain similarity threshold was counted and the cumulative results for all window scans at each position were visualized in diagrams (fig. 4, Supplementary Fig. 1A). For both *A. formosae* and *A. capillus-veneris* sequence, similarity was higher in the region between −26 to +12 than it was elsewhere. To exclude the possibility that the observed increase in similarity between the regions upstream of editing sites was simply the result of a nucleotide bias surrounding editing sites (codons preferentially subject to editing, for instance), we performed the same analysis but excluded nucleotide positions −1 and +1 (fig. 4, Supplementary Fig. 1A). This resulted in a reduction, but not elimination, of the observed increase in similarity

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

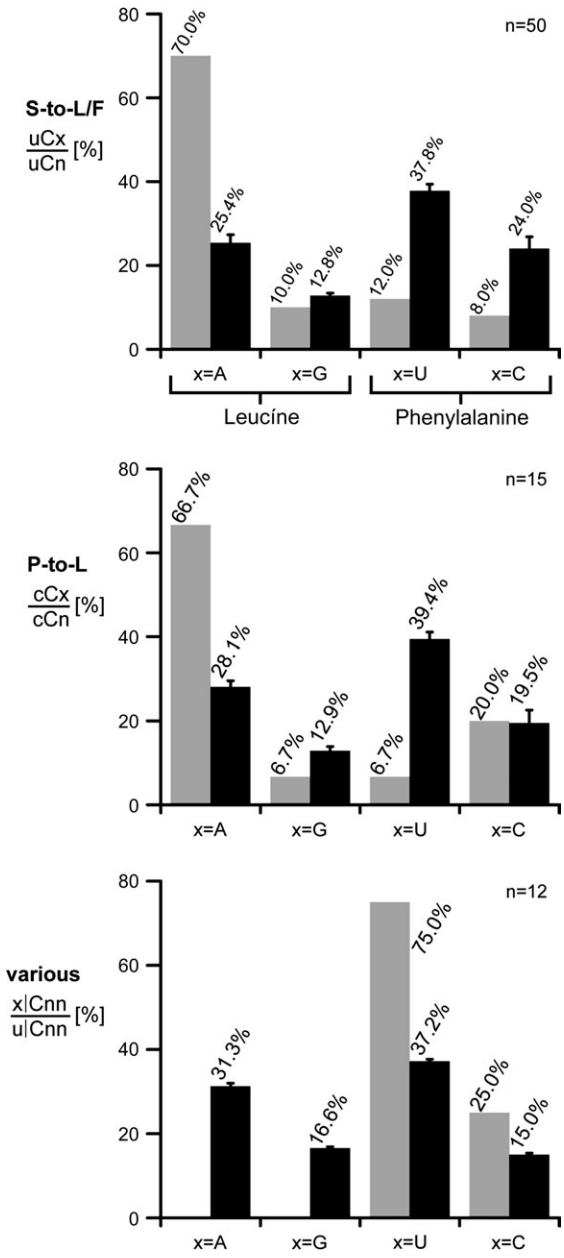

FIG. 3.—Comparison of nucleotide frequencies of third-codon positions adjacent to C-to-U editing sites and codon usages for the 5 analyzed seed plant species (see text). To analyze a bias at third-codon positions within edited codons, all UCN serine (top panel) and CCN proline codons (middle panel) edited at the second-codon position were extracted and compared to the codon usages for proline or serine codons, respectively. To investigate a nucleotide bias upstream of edited sites, all codons edited at first position were extracted and the nucleotides upstream of those were compared to the general third-codon position usage (bottom panel). The arrow bars in diagrams for seed plants represent the standard deviation of the codon usages for the 5 species analyzed. All numbers are in percent except the absolute number of the respective editing events shown in the upper right corner of each diagram. Note the stronger bias for an A downand a U upstream of editing sites in seed plants. This bias is also clearly visible for seed plants within UCR serine codons edited to UUR leucine codons (top panel).

upstream from −26 to +12 for both *A. formosae* and *A. capillus-veneris*. Further analyses revealed that these similarities do not result from a general consensus motif but from individual small sets or clusters of editing sites with similar upstream sequences (data not shown). These increased similarities in the upstream regions between editing sites indicate that, as is the case in seed plants, editing sites in the chloroplasts of hornwort and fern are defined by their upstream regions. Therefore, it is very likely that the recognition of chloroplast editing sites proceeds in a very similar manner in all land plants, further supporting the model that the chloroplast RNA-editing machinery is of a monophyletic origin.

RNA Editing in Plant Mitochondria

RNA editing is also observed in the mitochondria of land plants. In seed plants, the only clear difference between the editing systems of chloroplasts and mitochondria is the number of editing sites. Whereas seed plant chloroplasts contain about 30 editing sites, seed plant mitochondria usually posses several hundreds, an example being the 411 editing sites that have been reported in the mitochondria of *A. thaliana* (Giege and Brennicke 1999). Less is known about the absolute number of editing sites in the mitochondria of hornworts and ferns. However, available data, based on the analysis of single genes or gene segments, indicate that extensive editing may exist in the mitochondria of hornworts and ferns (Steinhauser et al. 1999; Groth-Malonek et al. 2005; Duff 2006).

Apart from this difference, editing in both compartments exhibits many similarities (Maier et al. 1996): First, editing in chloroplasts and mitochondria always coincides phylogenetically (Freyer et al. 1997; Steinhauser et al. 1999). Second, the same nucleotide conversions (solely C-to-U or C-to-U and U-to-C editing events together) always take place in both organelles. Third, available data suggest that *cis*-acting elements are similar in location and extension to those observed in seed plant chloroplasts (Farre et al. 2001; Choury et al. 2004; Takenaka et al. 2004; Neuwirt et al. 2005; van der Merwe et al. 2006). Finally, the mitochondrial editing sites of *Arabidopsis* display a bias toward having a U upstream and an A downstream (Giege and Brennicke 1999). These substantial similarities between the 2 plant organellar editing systems point to a common origin of both and hence, studying the 2 systems together may provide insights into the evolution of each.

In order to determine whether or not the regions upstream of mitochondrial RNA-editing sites share similarities, we applied Lola Software on the 411 RNA-editing sites described for the mitochondria of *Arabidopsis*. An increase in the similarity between the upstream regions was detected by our analysis (Supplementary Fig. 1B). As with the results described above for the chloroplasts of *A. formosae* and *A. capillus-veneris*, the similarities between the mitochondrial sites of *A. thaliana* are located within a sequence window from −23 to +10 with respect to the editing positions. Additionally, these similarities do not result from a general consensus motif but rather from small individual clusters of sites (data not shown). Notably, the position of the similarities detected by our in silico analyses coincides with

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

Table 2
Chloroplast RNA-Editing Sites Shared between Seed Plants and Hornwort or Fern

	Seed Plants										Fern		Hornwort	
	Dicots						Monocot		Gymnosperm					
	Ath		Nta		Abe		Zma		Pth		Aca		Afo	
Gene	Codon	Number	Codon	Number	Codon	Number	Codon	Number	Codon	Number	Codon	Number	Codon	Number
accD	uCg S>L	265	L	284	L	276	—		cCa P>L	88	uCa S>L	88	L	88
atpA	L	258	L	258	L	258	L	259	uCa S>L	258	L	259	uCa S>L	258
atpA	L	383	L	383	L	383	uCa S>L	383	L	383	L	384	cCu P>L	383
ndhA	L	159	L	159	L	159	uCa S>L	158		—	L	159	uCu S>F	160
ndhB	uca	204	uCa S>L	204	uCa S>L	204	uCa S>L	204		—	cCu P>L	189	L	186
ndhB	uCa S>L	277	uCa S>L	277	uCu S>F	277	uCa S>L	277		—	cCa P>L	260	L	259
ndhB	uCa S>L	279	uCa S>L	279	uCu S>F	279	L	279		—	uCa S>L	262	L	261
ndhD	aCg T>M	1	aCg T>M	1	aCg T>M	1	M	1		—	M	1	aCg T>M	1
ndhD	uCa S>L	293	L	293	uCa S>L	293	uCa S>L	293		—	L	294	uCa S>L	293
petB	L	200	L	200	L	200	L	219	cCa P>L	200	cCg P>L	200	uCa S>L	200
petD	L	37	L	37	L	37	L	37	uCa S>L	37	uCa S>L	36	uCa S>L	37
petD		*		*		*		*	Caa Q>*	161	Caa Q>*	160		*
petL		*		*		*		*	Caa Q>*	32	Caa Q>*	32		*
psbL	M	1	aCg T>M	1	aCg T>M	1	M	1	M	1	aCg T>M	1	M	1
psbl2	L	31	L	31	L	31	L	31	cCa P>L	32	cCu P>L	31	L	31
rpl20	L	103	uCa S>L	103	uCa S>L	103	uCa S>L	103	L	103	cCa P>L	101	L	101
rpoB	F	669	uCu S>F	667	uCu S>F	667	F	675	F	662	F	668	uCu S>F	668
rpoC2	L	776	L	778	L	778	uCg S>L	925		—	cCg P>L	768	F	773

NOTE.—Edited codons are indicated as gray boxes (editing site, uppercase C). The resulting changes for the encoded amino acids are shown. Homologous codons where no editing was detected are represented by the encoded amino acids. ''—'': gene or region of gene not present, ''*'': stop codon or beyond gene, Afo: *Anthoceros formosae*, Aca: *Adiantum capillus-veneris*, Ath: *Arabidopsis thaliana*, Nta: *Nicotiana tabacum*, Abe: *Atropa belladonna*, Zma: *Zea mays*, and Pth: *Pinus thunbergii*. Note that at virtually all positions the same or a similar (L and F) amino acid is encoded after editing. In contrast, pre-edited codons appear more variant (S and P) suggesting that selection on the pre-edited state is more relaxed than on the postedited state.

the location of experimentally determined *cis*-elements (Farre et al. 2001; Choury et al. 2004; Takenaka et al. 2004; Neuwirt et al. 2005; van der Merwe et al. 2006). These results strongly corroborate the interpretation that the corresponding similarities detected in the chloroplast genomes of *A. formosae* and *A. capillus-veneris* (fig. 4; Supplementary Fig. 1A) reflect real *cis*-elements. Overall, our results indicate a close relationship between chloroplast and mitochondrial RNA editing and suggest that the recognition of editing sites proceeds in a similar fashion in both compartments.

A Model for the Evolution of Plant Organelle RNA Editing

Based on the evidence presented above, the model we propose is that 1) plant chloroplast RNA editing is of monophyletic origin and 2) chloroplast and mitochondrial RNA editing evolved as systems that generated new variation in a manner similar to point mutations. The first aspect of the model is based on the comparison between seed plants, ferns, and hornworts discussed above. The number of edited sites has subsequently been reduced by the occurrence of independent C-to-T and/or T-to-C point mutations with a much larger reduction in seed plants. Independent losses of RNA-editing sites have generated the species-specific patterns of editing sites (the editotypes) found in extant plants. If chloroplast RNA editing is in fact monophyletic, then it would indicate that chloroplast U-to-C editing sites have been eliminated completely in seed plants.

The second aspect of the model is based on the observation that chloroplast RNA-editing sites are predominantly at positions with a low point mutation rate in seed plant chloroplast genomes. The current bias toward T_A contexts is probably the result of some combination of an initial bias

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

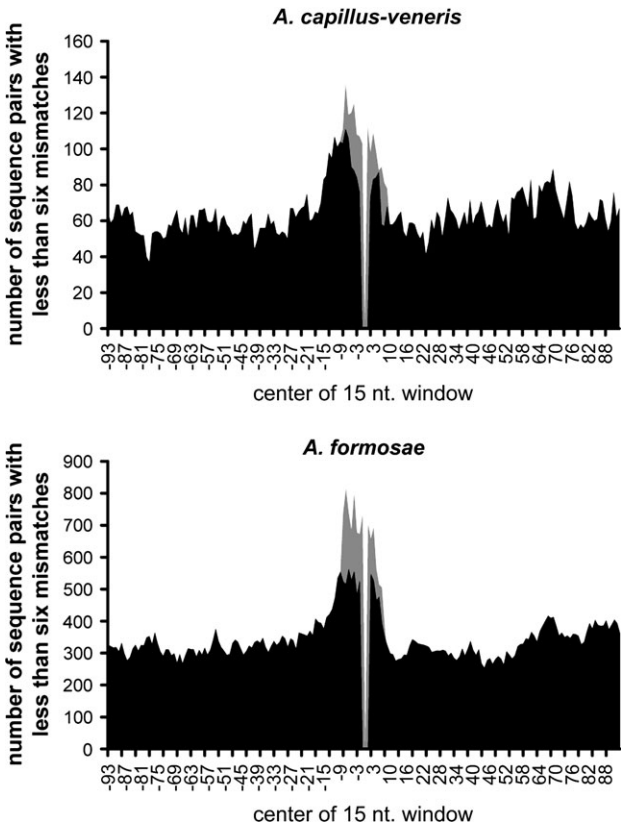

FIG. 4.—Increased similarities between the upstream regions of chloroplast C-to-U RNA-editing sites in *Adiantum capillus-veneris* and *Anthoceros formosae*, respectively as detected by Lola software. The regions surrounding all C-to-U editing sites from −100 to +100 (editing sites at 0; omitted from the subsequent analyses) were extracted from the corresponding genomic sequences and analyzed separately. Within each of the 2 subsets, all possible nonredundant pairwise combinations of the extracted sequence segments were generated with the editing sites at the center. Subsequently, Lola ran a window of 15 nt in size along all of the sequence pairs in one-nucleotide steps. At each step, the number of sequence pairs exhibiting less than 6 mismatches was counted visualized as a diagram (black graph plus gray tip; *x* axis values indicate the center of the window with respect to the editing site at 0 and the *y* axis values, the number of sequence pairs that exceeded the threshold at the corresponding window position). Increases in similarities between the sequence pairs are detectable directly upstream of editing sites for all 2 sets. These similarities do not result from nucleotides biases around editing sites alone because they are still detectable when also the nucleotides adjacent to editing sites (−1 and +1) are excluded from the analysis (black graph without gray tip). Analogous experiments carried out with the chloroplast U-to-C RNA-editing sites of *A. formosae* as well as with the mitochondrial C-to-U RNA-editing sites of *Arabidopsis thaliana* yielded similar results (Supplementary Fig. 1).

toward editing at sites where C-to-T mutations were less frequent, thus in essence compensating for a lack of variation at such sites, and a lower subsequent rate of loss of the need for editing at such sites due to genome mutation; it is probable that over time, some of the edited sites have been lost due to C-to-T point mutations at the appropriate genome site, something that would occur at a lower rate in T_A contexts. This model implies that there are no additional functions for editing such as gene regulation, leading to the prediction that C-to-U editing sites could be replaced by a transition mutation in the genome without any negative impact on plant fitness. This prediction is consistent with the variation of editing sites observed between seed plant species (Funk et al. 2005). Moreover, a C-to-T mutation has been recently set experimentally at an editing site (site *atpA*-264) in the chloroplast *atpA* gene of tobacco. Also consistent with the prediction is that mutant plants that are homoplastomic for a T at *atpA*-264 exhibit a wild-type phenotype (Schmitz-Linneweber et al. 2005). Overall, the data support the model that RNA editing in seed plant chloroplasts functions solely to generate at the RNA level the same effect as a C-to-T point mutation at the DNA level.

One implication of our model is that a decrease in the general mutation rate could result in an increase in the number of editing sites that are retained. Indeed, a correlation between the number of editing sites and the mutation rate is found in several instances in seed plant chloroplasts and mitochondria. First, roughly a quarter of seed plant chloroplast editing sites are located in the *ndhB* gene. This gene is located in the inverted repeat region of the chloroplast genomes, which evolves about 2.3 times slower than the 2 single-copy regions (Perry and Wolfe 2002). Second, in contrast to chloroplasts, the mitochondria of seed plants posses several hundreds of editing sites and the plant mitochondrial DNA is estimated to evolve about 3 times slower than the chloroplast DNA (reviewed in Palmer 1990). Interestingly, it was recently reported that the mitochondrial genome of Geraniaceae has undergone a period of increased mutation rate that appears to correlate with an unusually low number of editing sites (Parkinson et al. 2005).

No RNA-editing event has yet been identified in the chloroplasts and mitochondria of the marchantiid subclass of liverworts, represented by *Marchantia polymorpha* (Steinhauser et al. 1999). This suggests that RNA editing was lost in the marchantiid subclass and demonstrates that chloroplast RNA editing is not required in land plants. The chloroplast genome of *M. polymorpha* exhibits a higher A + T content (71.2%) than *A. formosae* (67.1%), *A. capillis-veneris* (58.0%), and the 5 seed plant species investigated here (61.5–63.7%). The increased A + T content of the chloroplast chromosome of *M. polymorpha* probably indicates different mutational or recombinatorial properties of the chloroplast genome of *M. polymorpha* that may have contributed to the elimination of RNA-editing sites. Further investigation of this genome should help shed light on this possibility.

At a first glance, it may be difficult to imagine that a complex machinery evolved to, in essence, imitate point mutations at the DNA level, as we have proposed. However, if we consider just the chloroplast genome, a typical leaf cell can contain more than 100 chloroplasts, each of which can contain several hundred copies of its genome. Thus, chloroplasts form a relatively small, asexual, and highly polyploid population within plant cells. Such a population is expected to fix mildly deleterious point mutations (Lynch and Blanchard 1998; Blanchard and Lynch 2000) such as has been described for other endocytobionts, *Buchnera* (Moran 1996). Also, in contrast to a point mutation occurring in one of the chloroplast chromosomes, an editing factor encoded by the nuclear genome for an organellar site immediately affects the transcripts from all genome copies of the whole organelle population. This would be an important advantage, particularly in cases when a mutation in the chloroplast genome is beneficial for the plant cell but has

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

a negative effect on the organelle population. Such a situation is exemplified by a human mitochondrial mutation associated with the mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes syndrome. Here, a mutation in a mitochondrial tRNA gene causes defects in mitochondrial protein synthesis and respiration. However, mutant genomes replace wildtype genomes due to a replication advantage (Yoneda et al. 1992). We propose that a combination of these properties and (possibly) an event inducing point mutations in the early land plant evolution opened the gates for organelle RNA editing to evolve. A similar hypothesis was proposed for the evolution of tRNA editing in metazoan mitochondria (Borner et al. 1997). Another important factor is that among all known RNA-editing systems that affect protein-coding regions, it is only in plants that the factors conferring site specificity to the editing machinery are encoded by a different genome than the target sites. This means that the nuclear encoded editing system might have provided a means to compensate for genetic drift in the chloroplast genome. As suggested recently by Lynch et al. (2006), the different speed of evolution found between the nuclear and organelle genomes of plants, with the organelle genomes evolving much slower than the nuclear genome, might have been a prerequisite for the evolution of plant organellar editing. The relative low mutation rates in the organelle genomes may prevent the mutational decay of *cis*-elements, what may in turn be required to permit the evolution of the corresponding recognition/editing factors in the nuclear genome (Lynch et al. 2006).

Mechanistically, the editing system could have arisen through proteins with other functions in RNA metabolism. These proteins might have been used to guide the yet unknown RNA-editing activity, a *trans*-aminase for instance. Once established, these proteins could have been duplicated and adopted to serve different editing sites or sets of sites. Interestingly, the recently identified editing factor for the *ndhD* message of *Arabidopsis*, CRR-4, belongs to the PPR family of RNA-binding proteins. The PPR family contains several hundred members in *Arabidopsis* and rice, the majority of them predicted to be involved in organellar RNA metabolism (Lurin et al. 2004).

The apparent reduction in the number of editing sites during the evolution of higher plants could indicate that editing is disappearing in seed plants as genome mutations eliminate the need for editing at certain sites. This rate of loss of editing sites will be correlated with mutation rate meaning that remaining sites should be predominantly at sites with low point mutation rates as we observe. The appearance of a few new editing sites late in the evolution of flowering plants (Drescher et al. 2002; Tillich et al. 2005) though indicates that this loss is not irreversible. Instead, it could be that the genome evolves toward an equilibrium state in the gain and loss of editing sites. In seed plant mitochondria, RNA-editing activities affecting third-codon positions show a higher evolutionary variability (edited vs. nonedited C) than editing activities affecting nonsynonymous codon position (Shields and Wolfe 1997). This is expected because editing at nonsynonymous codon positions in general generates conserved amino acids and hence should be under positive selection. However, the existence of a high variance of random-like edits at third-codon positions might result from modulations of editing activities that enable editing to extend to novel sites. To clearly distinguish between an "on the way out" model and a balanced gain and loss model, though, more detailed analyses of putative novel editing sites are required.

Supplementary Material

Supplementary Table 1 and Figure 1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was initiated and inspired by Rainer M. Maier, codiscoverer of chloroplast RNA editing, to whose memory it shall be dedicated. We thank Mischa Dieterle and Sebastian Risi for helpful discussions on Lola Software and Christian Schmitz-Linnerweber and Peter Poltnigg for critical reading of the manuscript. U.G.M. and M.T. are supported by the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich Transregio 1.

Literature Cited

Blanchard JL, Lynch M. 2000. Organellar genes: why do they end up in the nucleus? Trends Genet 16:315–20.

Bock R, Hermann M, Fuchs M. 1997. Identification of critical nucleotide positions for plastid RNA editing site recognition. RNA 3:1194–200.

Bock R, Hermann M, Kössel H. 1996. In vivo dissection of cis-acting determinants for plastid RNA editing. Embo J 15: 5052–9.

Bock R, Kössel 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 13:4623–8.

Borner GV, Yokobori S, Mörl M, Dorner M, Pääbo S. 1997. RNA editing in metazoan mitochondria: staying fit without sex. FEBS Lett 409:320–4.

Chateigner-Boutin AL, Hanson MR. 2002. Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. Mol Cell Biol 22:8448–56.

Chateigner-Boutin AL, Hanson MR. 2003. Developmental covariation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. Nucleic Acids Res 31:2586–94.

Chaudhuri S, Carrer H, Maliga P. 1995. Site-specific factor involved in the editing of the psbL mRNA in tobacco plastids. Embo J 14:2951–7.

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 15:5958–64.

Choury D, Farre JC, Jordana X, Araya A. 2004. Different patterns in the recognition of editing sites in plant mitochondria. Nucleic Acids Res 32:6397–406.

Covello PS, Gray MW. 1993. On the evolution of RNA editing. Trends Genet 9:265–8.

Drescher A, Hupfer H, Nickel C, Albertazzi F, Hohmann U, Herrmann RG, Maier RM. 2002. C-to-U conversion in the intercistronic ndhI/ndhG RNA of plastids from monocot plants: conventional editing in an unconventional small reading frame? Mol Genet Genomics 267:262–9.

Duff RJ, Moore FB. 2005. Pervasive RNA editing among hornwort rbcL transcripts except Leiosporoceros. J Mol Evol 61:571–8.

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

Duff R. 2006. Divergent RNA editing frequencies in hornwort mitochondrial nad5 sequences. Gene 366:285–91.
Farre JC, Leon G, Jordana X, Araya A. 2001. cis Recognition elements in plant mitochondrion RNA editing. Mol Cell Biol 21:6731–7.
Freyer R, Kiefer-Meyer MC, Kössel H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. Proc Natl Acad Sci USA 94:6285–90.
Funk HT, Poltnigg P, Schmitz-Linneweber C, Tillich M. 2005. Transcript polishing in higher plants plastids by RNA editing. Endocytobiosis Cell Res 15:491–503.
Giege P, Brennicke A. 1999. RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. Proc Natl Acad Sci USA 96:15324–9.
Gott JM. 2003. Expanding genome capacity via RNA editing. C R Biol 326:901–8.
Gott JM, Emeson RB. 2000. Functions and mechanisms of RNA editing. Annu Rev Genet 34:499–531.
Groth-Malonek M, Pruchner D, Grewe F, Knoop V. 2005. Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol Biol Evol 22:117–25.
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 262:462–7.
Hirose T, Sugiura M. 1997. Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast ndhD mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes. Embo J 16:6804–11.
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 20:1144–52.
Karcher D, Bock R. 2002. The amino acid sequence of a plastid protein is developmentally regulated by RNA editing. J Biol Chem 277:5570–4.
Kotera E, Tasaka M, Shikanai T. 2005. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433:326–30.
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 31:2417–23.
Lurin C, Andres C, Aubourg S, et al. 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16: 2089–103.
Lynch M, Blanchard JL. 1998. Deleterious mutation accumulation in organelle genomes. Genetica 102–103:29–39.
Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. Science 311: 1727–30.
Maier RM, Neckermann K, Igloi GL, 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 251:614–28.
Maier RM, Zeltz P, Kössel H, Bonnard G, Gualberto JM, Grienenberger JM. 1996. RNA editing in plant mitochondria and chloroplasts. Plant Mol Biol 32:343–65.
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 22:6726–34.
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 USA 101:48–52.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci USA 93:2873–8.
Morton BR. 2003. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. J Mol Evol 56:616–29.
Morton BR, Oberholzer VM, Clegg MT. 1997. The influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome. J Mol Evol 45:227–31.
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 11:1563–70.
Palmer JD. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet 6:115–20.
Parkinson CL, Mower JP, Qiu YL, Shirk AJ, Song K, Young ND, Depamphilis CW, Palmer JD. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol 5:73.
Peeters NM, Hanson MR. 2002. Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. RNA 8:497–511.
Perry AS, Wolfe KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J Mol Evol 55:501–8.
Reed ML, Hanson MR. 1997. A heterologous maize rpoB editing site is recognized by transgenic tobacco chloroplasts. Mol Cell Biol 17:6948–52.
Reed ML, Lyi SM, Hanson MR. 2001. Edited transcripts compete with unedited mRNAs for trans-acting editing factors in higher plant chloroplasts. Gene 272:165–71.
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 29:1507–13.
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 276:3937–40.
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 17:1815–28.
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 19:1602–12.
Shields DC, Wolfe KH. 1997. Accelerated evolution of sites undergoing mRNA editing in plant mitochondria and chloroplasts. Mol Biol Evol 14:344–9.
Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. 1999. Plant mitochondrial RNA editing. J Mol Evol 48:303–12.
Takenaka M, Neuwirt J, Brennicke A. 2004. Complex cis-elements determine an RNA editing site in pea mitochondria. Nucleic Acids Res 32:4137–44.
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 43:708–15.
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 580:268–72.
Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel 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 USA 93:8766–70.

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022

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.

Yoneda M, Chomyn A, Martinuzzi A, Hurko O, Attardi G. 1992. Marked replicative advantage of human mtDNA carrying a point mutation that causes the MELAS encephalomyopathy. Proc Natl Acad Sci USA 89:11164–8.

Zito F, Kuras R, Choquet Y, Kössel 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.

William Martin, Associate Editor

Accepted June 30, 2006

Downloaded from https://academic.oup.com/mbe/article/23/10/1912/1096958 by guest on 20 August 2022