Nucleic Acids Research, 2003, Vol. 31, No. 20 ***5907–5916***
DOI: 10.1093/nar/gkg795

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*

Hirokazu Handa[1,2,*]

[1]Laboratory of Plant Genecology, National Agricultural Research Center for Hokkaido Region, Sapporo 062-8555, Japan and [2]Department of Plant Biotechnology, National Institute of Agrobiological Sciences, Tsukuba 305-8602, Japan

Received July 9, 2003; Revised and Accepted August 29, 2003

DDBJ/EMBL/GenBank accession no. AP006444

ABSTRACT

The entire mitochondrial genome of rapeseed (*Brassica napus* L.) was sequenced and compared with that of *Arabidopsis thaliana*. The 221 853 bp genome contains 34 protein-coding genes, three rRNA genes and 17 tRNA genes. This gene content is almost identical to that of *Arabidopsis*. However the *rps14* gene, which is a pseudo-gene in *Arabidopsis*, is intact in rapeseed. On the other hand, five tRNA genes are missing in rapeseed compared to *Arabidopsis*, although the set of mitochondrially encoded tRNA species is identical in the two Cruciferae. RNA editing events were systematically investigated on the basis of the sequence of the rapeseed mitochondrial genome. A total of 427 C to U conversions were identified in ORFs, which is nearly identical to the number in *Arabidopsis* (441 sites). The gene sequences and intron structures are mostly conserved (more than 99% similarity for protein-coding regions); however, only 358 editing sites (83% of total editings) are shared by rapeseed and *Arabidopsis*. Non-coding regions are mostly divergent between the two plants. One-third (about 78.7 kb) and two-thirds (about 223.8 kb) of the rapeseed and *Arabidopsis* mitochondrial genomes, respectively, cannot be aligned with each other and most of these regions do not show any homology to sequences registered in the DNA databases. The results of the comparative analysis between the rapeseed and *Arabidopsis* mitochondrial genomes suggest that higher plant mitochondria are extremely conservative with respect to coding sequences and somewhat conservative with respect to RNA editing, but that non-coding parts of plant mitochondrial DNA are extraordinarily dynamic with respect to structural changes, sequence acquisition and/or sequence loss.

INTRODUCTION

Mitochondria are semiautonomous organelles whose universally recognized function is to produce cellular ATP by the process of oxidative phosphorylation. This mitochondrial function is conserved in all eukaryotic cells, i.e. animal, fungi and plant cells, but nevertheless, the mitochondrial genomes of higher plants exhibit a number of unique features compared to their counterparts in animals or fungi. Higher plant mitochondrial genomes are not only larger in size, but they also contain structural rearrangements due to homologous intra- or inter-molecular recombination events. Moreover, specific modes of gene expression (e.g. *cis*- and *trans*-splicing, RNA editing, etc.) complicate the analysis of the information encoded by plant mitochondrial genomes.

The sizes of mitochondrial genomes vary widely even among higher plant species. With 208 kb, *Brassica hirta* has the smallest mitochondrial genome in higher plants (1), while the mitochondrial genome of muskmelon is estimated to be over 2400 kb (2). This size variation can occur relatively rapidly in evolution as exemplified within the cucumber family, where the mitochondrial genome size varies by more than 6-fold (2). Similarly, the plant mitochondrial genome organization is known to be very dynamic. It would be of great interest to understand such rapid evolutionary changes in size and structure, and their consequences with respect to gene content. To date, the complete mitochondrial genome sequences in higher plants have been determined in two dicot plants, *Arabidopsis thaliana* and sugar beet (3,4) and in one monocot, rice (5). Although significant structural differences are observed among the mitochondrial genomes of these three plant species, most of the genes are conserved aside from the occasional transfer of a mitochondrial gene to

*To whom correspondence should be addressed at Department of Plant Biotechnology, National Institute of Agrobiological Sciences, 2-1-2, Kan-non-dai, Tsukuba 305-8602, Japan. Tel: +81 29 838 8374; Fax: +81 29 838 8397; Email: hirokazu@affrc.go.jp

the nuclear genome in one angiosperm clade or the other. However, these species are only distantly related to each other, and not suitable for investigating rapid evolutionary changes over the entire mitochondrial genome.

In this study, rapeseed (*Brassica napus* L.) was chosen for the analysis of the tempo and mode of evolutionary changes for several reasons. First, rapeseed belongs to Cruciferae, the same family as *Arabidopsis*. It occupies a good phylogenetic position for making comparisons with the already sequenced *Arabidopsis* mitochondrial genome. Second, physical mapping revealed that the mitochondrial genome in rapeseed is about 220 kb in size (6), which is only two-thirds of that in *Arabidopsis* (367 kb) (3). This implies that these two plants can be used to study the evolution of genome size. Third, it is well known that *Brassica* species, including rapeseed, have the smallest mitochondrial genomes among higher plants. Investigation of the rapeseed mitochondrial genome can thus reveal the minimum sequence requirement for the mitochondrial genome of higher plants.

Here the complete nucleotide sequence and RNA editing content of the mitochondrial genome in rapeseed (*Brassica napus* L.) are presented and compared with those of *A.thaliana.*

MATERIALS AND METHODS

Mitochondrial DNA and RNA isolation

Mitochondria were isolated from green leaves of 8-week-old rapeseed plants (cv. Wester). Mitochondrial DNA (mtDNA) and mitochondrial RNA (mtRNA) were prepared using procedures described previously (7).

Sequencing procedures

Based on the physical map of rapeseed mitochondria (6) and available sequences of rapeseed mitochondrial genes, 28 oligonucleotides were synthesized. Using mitochondrial DNA as a template, long PCR was carried out using these nucleotides as primers to amplify 14 parts of the mitochondrial genome. The amplified fragments were completely digested with BamHI, EcoRI and HindIII, or were partially digested with Sau3AI. A recombinant DNA library containing BamHI, EcoRI, HindIII and Sau3AI fragments in pBluescript II SK+ vector (Stratagene, USA) provided templates for sequencing. PCR-derived amplification products were also used to close gaps between restriction fragments. Each base of the rapeseed mtDNA was covered at least three times (6.8 times on average). The DNA sequencer used was a Li-COR 4200L2 (Li-COR, USA).

Data analysis

Sequences were evaluated and assembled using GENETYX-MAC ver11.2.3 (Software Development, Japan). A database search was done at the National Center for Biotechnology Information (NCBI) using the BLAST network service (http://www.ncbi.nlm.nih.gov/BLAST/; 8). A tRNA gene search was carried out with the tRNA scan-SE service (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/; 9). The Pairwise BLAST program on our local server was used for the comparison between the whole genome sequences of rapeseed and the mitochondrial genomes of other plants.

RNA editing determination

Mitochondrial RNA was reverse transcribed with random hexamers using Superscript II reverse transcriptase (Invitrogen, USA). PCR amplification was done with primer pairs specific to each open reading frame (ORF), and the resultant PCR products were sequenced directly. Fluorescent charts for cDNA and genomic sequences were compared to determine RNA editing sites.

RESULTS AND DISCUSSION

Gene organization

The rapeseed mtDNA sequence forms a circle of 221 853 bp (accession number: AP006444) (Fig. 1), two-thirds the size of the *Arabidopsis* mtDNA (366 924 bp) (3). The overall G + C content of rapeseed mtDNA (45.2%) is comparable to that of *Arabidopsis* mtDNA (44.8%). Homology searches using the BLASTN and BLASTX programs and a tRNA gene search using the tRNA scan-SE program detected a total of 54 genes in the rapeseed mitochondrial genome: 34 known protein-coding genes, three ribosomal RNA genes (*rrn5*, *rrn18*, *rrn26*), and 17 tRNA genes (Tables 1 and 2), which altogether account for 17.4% (38 662 bp) of the genome. The positions of these genes in the rapeseed mitochondrial genome are illustrated in Figure 1. The content of protein-coding genes is completely identical between rapeseed and *Arabidopsis* with the exception of a gene for ribosomal protein S14 (*rps14*) (Table 1). Due to a stop codon and a deleted nucleotide, the *rps14* gene is a pseudo-gene in *Arabidopsis* (10) but an intact ORF in rapeseed. In rapeseed and *Arabidopsis*, *ccmFN* genes are divided into two reading frames (*ccmFN1* and *ccmFN2*) (3,11), a known peculiarity of the mitochondrial genomes of these two Cruciferae species, which distinguishes them from other plants (5). Seventeen tRNA genes specifying 15 species of amino acids were identified in the rapeseed mitochondrial genome (Table 2). Of these, 11 tRNAs are of mitochondrial origin and six are considered to be of plastid origin. *Arabidopsis* mtDNA encodes 22 tRNA genes, five more than rapeseed mtDNA. However, in both plant species the genes for at least six tRNAs (for amino acids G, A, V, L, T and R) are absent from the mitochondrial genome and these tRNAs have to be imported from the nucleus. The DNA sequences of the genes described above are highly conserved between rapeseed and *Arabidopsis*. Most of the protein-coding genes show more than 99% sequence similarity; the lowest similarity (97.3%) is found between *rps12* genes and the highest (99.9%) between *rpl16* genes.

In addition to previously identified genes, 45 ORFs larger than 100 codons in size were annotated in the rapeseed mitochondrial genome. None of these ORFs could be assigned a function based on sequence similarity at either the nucleotide or protein level. Moreover, none of these rapeseed ORFs has a counterpart in the *Arabidopsis* mitochondrial genome, in which 85 ORFs larger than 100 codons were described, although limited similarity is observed between some ORFs of rapeseed and *Arabidopsis*. Also, most ORFs do not show any significant homology to other genes found in the nuclear or plastid genome, although some have similarity to nuclear transposon or plastid sequences. However, all of these ORFs of

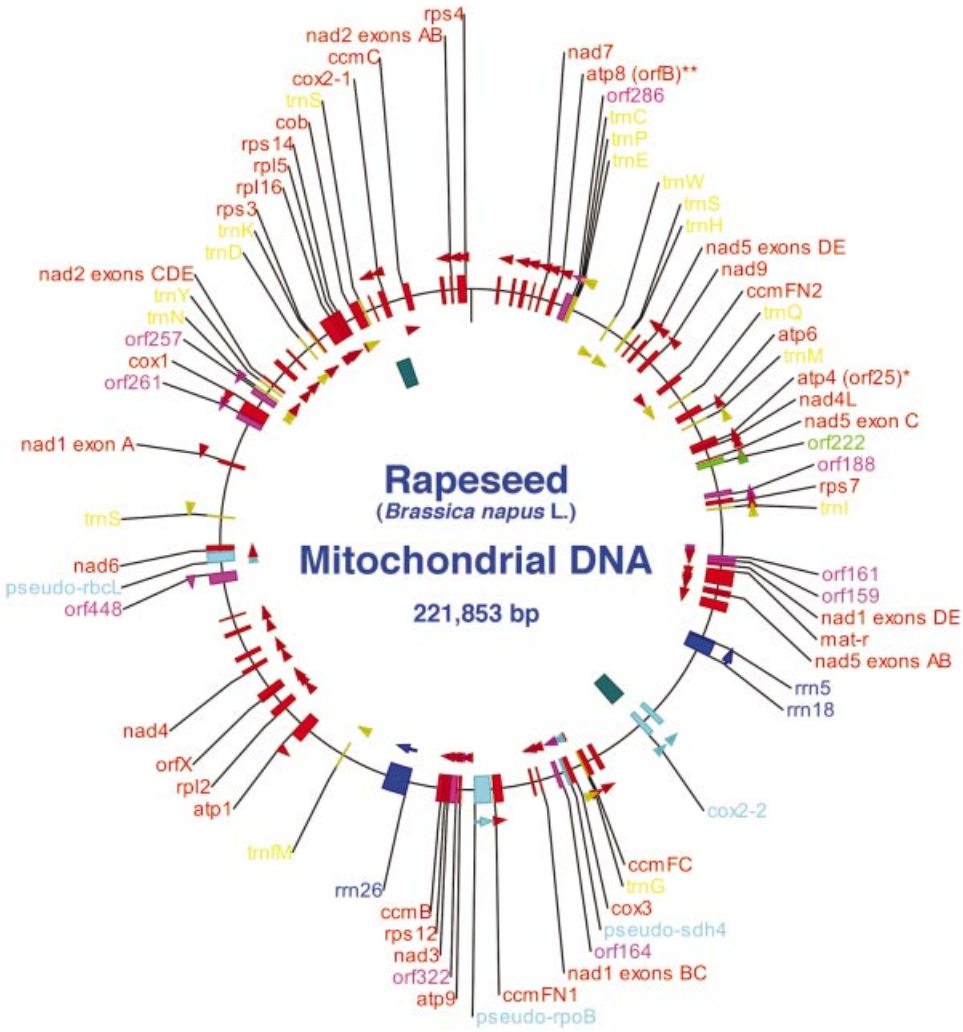

Figure 1. Gene organization of the rapeseed mitochondrial genome. Genes homologous to known protein-coding genes are indicated by red boxes. The blue boxes represent rRNA genes. Pink boxes represent unidentified ORFs longer than 150 amino acids. tRNA genes are represented by yellow boxes. Pseudo genes including plastid gene segments are shown in pale green. *orf222*, a cms-related gene (20), is shown by a green box. Arrowheads indicate the direction of reading frames. Dark green boxes located inside the circle represent 2 kb repeat regions. *From Heazlewood *et al.* (32); **From Sabar *et al.* (18).

possible transposon or plastid origin seem to be non-functional because of their fragmented and/or truncated nature.

In summary, despite their large size difference, the mtDNAs of the two species share nearly the same set of functional genes (protein genes, rRNA genes and tRNA genes). The additional ORFs are not shared between these two closely related plants, which means that these ORFs could not code for additional important protein information and may instead, although it seems unlikely, fulfil a species-specific function.

Large repeat sequences for tripartite structure

In this study, a 2427 bp sequence was found to be present as a direct repeat in the rapeseed mitochondrial genome, as reported previously (6). This sequence included the first exon, the intron, and part of the second exon of the *cox2* gene (Fig. 1). Due to this duplication, two copies of *cox2* genes exist in rapeseed, although these copies diverge from each other at the point 55 bp upstream of the stop codon. One copy (*cox2-1* in Fig. 1) is homologous to other plant mitochondrial *cox2* genes, but the other (*cox2-2* in Fig. 1) has an extension that shows no homology to any other sequences. The presence of these 2427 bp repeats implies a tricircular structure for the rapeseed mitochondrial genome, as postulated before (6). Via these 2427 bp repeats, the rapeseed mitochondrial genome could recombine into two subgenomic circles (124 908 and 96 945 bp circles).

Large repeat sequences are also found in the *Arabidopsis* and sugar beet mitochondrial genomes (6.5 and 4.2 kb, and 6.2 kb, respectively; 3,4), and have been suggested to be active in intramolecular recombination. However, the sequences of these repeats (*atp6* and *orf139* regions in *Arabidopsis*, *rrn26* region in sugar beet, and *cox2* region in rapeseed) are completely different from each other. Intramolecular recombination to generate the complexity of the plant mitochondrial genome is a common feature of higher plant mitochondrial genomes (12), except for one known example, *B.hirta* (1). However, the DNA sequences involved in the intramolecular recombination are highly species-specific, and different

Table 1. Protein-coding and ribosomal RNA gene content of mitochondrial genome of rapeseed compared to *Arabidopsis*, sugar beet and rice

Genes		Rapeseed	*Arabidopsis*	Sugar beet	Rice	Genes		Rapeseed	*Arabidopsis*	Sugar beet	Rice
Complex I	*nad1*	+	+	+	+	Ribosomal RNAs	*rrn5*	+	+	+	+
	nad2	+	+	+	+		*rrn18*	+	+	+	+
	nad3	+	+	+	+		*rrn26*	+	+	+	+
	nad4	+	+	+	+						
	nad4L	+	+	+	+	Ribosomal proteins	*rpl2*	+	+	-	+
	nad5	+	+	+	+		*rpl5*	+	+	+	+
	nad6	+	+	+	+		*rpl16*	+	+	-	+
	nad7	+	+	+	+		*rps1*	-	-	-	+
	nad9	+	+	+	+		*rps2*	-	-	-	+
							rps3	+	+	+	+
Complex II	*sdhB*	-	-	-	-		*rps4*	+	+	+	+
	sdhC	-	-	-	-		*rps7*	+	+	+	+
	sdhD	Ψ	Ψ	-	-		*rps11*	-	-	-	Ψ
							rps12	+	+	+	+
Complex III	*cob*	+	+	+	+		*rps13*	-	-	+	+
							rps14	+	Ψ	-	Ψ
Complex IV	*cox1*	+	+	+	+		*rps19*	-	Ψ	-	+
	cox2	+	+	+	+						
	cox3	+	+	+	+	Cytochrome-c-biogenesis	*ccmB*	+	+	+	+
							ccmC	+	+	Ψ	+
Complex V	*atp1*	+	+	+	+		*ccmFN*	-	-	+	+
	atp4 (orf25)[a]	+	+	+	+		*ccmFN1*	+	+	-	-
	atp6	+	+	+	+		*ccmFN2*	+	+	-	-
	atp8 (orfB)[b]	+	+	+	+		*ccmFC*	+	+	+	+
	atp9	+	+	+	+						
						Other ORFs	*tatC (orfX)*	+	+	+	+
							matR	+	+	+	+
							orf222	+	-	-	-

+, present; Ψ, pseudogene; –, absent.
[a]From Heazlewood *et al.* (32).
[b]From Sabar *et al.* (18).

sequences act as recombination repeats even within related species such as rapeseed and *Arabidopsis*.

Sequences shared among rapeseed and other plant genomes

Using the Pairwise BLAST software, the whole sequence of the rapeseed mitochondrial genome was compared with those of the *Arabidopsis* mitochondrial genome (3), sugar beet mitochondrial genome (4), *Arabidopsis* plastid genome (13) and rapeseed mitochondrial linear plasmid (14). The E (expectation) value threshold was set at $1e^{-10}$ for these comparative analyses among rapeseed and other genomes.

When the mitochondrial genome of rapeseed is compared to that of *Arabidopsis*, the sequences shared by these two related species total 143 126 bp, which represents 64.5% of the rapeseed genome. On the other hand, the sequences shared between rapeseed and sugar beet total 64 972 bp, representing 29.3% of the rapeseed genome. It has been reported that 78 057 bp of the mtDNA sequences are shared between *Arabidopsis* and sugar beet (4). These values are in good agreement with the phylogenetic relationship among these three plant species: rapeseed and *Arabidopsis* belong to the same family, Cruciferae, while sugar beet is a member of another family, Chenopodiaceae. However, one-third (about 78.7 kb) and two-thirds (about 223.8 kb) of the rapeseed and *Arabidopsis* mitochondrial genomes, respectively, were not aligned with each other. These non-homologous sequences are scattered through the rapeseed mitochondrial genome, and a total of 31 such regions longer than 1000 bp were identified. These data suggest that after divergence of these two species, in addition to the multiple recombination events, the mitochondrial genomes of both plants were very quickly modified by sequence acquisition and/or sequence loss encompassing up to two-thirds of the genome in the case of *Arabidopsis*. The large size variation between these two plants might be due partly to such rapid processes of sequence acquisition and/or sequence loss.

A BLAST search was carried out for the 31 non-homologous regions longer than 1 kb (a total of 51 517 bp) to explore the origins of these sequences. Of these sequences, 13.2% showed similarity to sequences in other plant mitochondrial genomes (not to *Arabidopsis* sequences) and 5.9% appeared to be of plastid origin. The fraction of sequences showing similarity to nuclear genome sequences was only 0.3%. The remaining sequences (80.6%) did not show any similarity to the registered sequences in the DNA sequence databases. Identifying the origin of these non-homologous sequences is of great importance and interest for understanding the evolution of higher plant mitochondrial genomes.

The mtDNA sequences shared by three dicot plants, rapeseed, *Arabidopsis* and sugar beet, totaled 61 870 bp, which corresponds to 27.9% of the rapeseed genome. This value is comparable to the sum total (68 343 bp) of coding sequences (38 662 bp) and *cis*-splicing introns (29 681 bp) of the rapeseed mitochondrial genome. These data suggest that the gene-coding sequences may be highly conserved among the mitochondrial genomes of higher plants, but other regions,

Table 2. tRNA gene content of mitochondrial genome of rapeseed compared to *Arabidopsis*, sugar beet and rice

	Rapeseed	*Arabidopsis*	Sugar beet	Rice
Native				
trnC-GCA	+	+	Ψ	Ψ
trnD-GUC	–	–	–	+
trnE-UUC	+	+	+	+
trnF-GAA	–	–	+	–
trnG-GCC	+	+	+	–
trnI-CAU	+	+	+	+
trnK-UUU	+	+ (2)	+	+
trnM-CAU	–	–	–	+
trnfM-CAU	+	+	+ (4)	+
trnP-UGG	+	+	+	+
trnQ-UUG	+	+	+	+
trnS-GCU	+	+ (2)	+	+
trnS-UGA	+	+ (2)	+	+
trnY1-GUA	+	+	+	+
Plastid-like				
trnC-GCA	–	–	–	+
trnD-GUC	+	+	+	–
trnF-GAA	–	–	–	+
trnH-GUG	+	+	+	+
trnI-CAU	–	–	–	Ψ
trnN-GUU	+	+	+	+
trnM-CAU	+	+	+	+
trnP-UGG	–	Ψ	Ψ	Ψ
trnR-UCU	–	–	–	Ψ
trnS-GGA	+	+	+	+
trnV-GAC	–	–	Ψ	Ψ
trnW-CCA	+	+	+	+
Converted				
trnY2-GUA	–	+	–	–
Unknown				
trnC2-GCA	–	–	+ (2)	–
No. of genes	17	22	25	23

+, present; Ψ, pseudogene; –, absent. Gene copy numbers are shown in parentheses.

i.e. non-coding sequences, may have diverged very rapidly during evolution, as they can no longer be aligned with each other.

In the rapeseed mitochondrial genome, 14 stretches of plastid-like sequences ranging from 43 to 2181 bp were identified. These sequences totaled 7950 bp, comprising 3.6% of the mitochondrial genome. In *Arabidopsis*, about 1% of the mitochondrial genome contains sequences imported from the plastid. In rapeseed, nine of the 14 plastid sequences found contain tRNA genes, seven of which are well conserved between the two plants. The fragments of *rbcL* and *rpoB* are also conserved in both plant mitochondria. Thus, these sequences may have been transferred from the plastid genome to the mitochondrial genome before the divergence of rapeseed and *Arabidopsis*. However, the rapeseed mitochondrial genome contains pieces of the *psaA* and *ycf2* genes, which are not found in *Arabidopsis* mtDNA. On the other hand, the pieces of the *psbD* and *ndhB* genes in the *Arabidopsis* mitochondrial genome are not present in rapeseed mtDNA. These results also suggest that functional sequences may be highly conserved, but sequence acquisition and/or sequence loss may be very frequent for non-functional gene pieces. Indeed, plastid tRNA genes in the mitochondrial genomes are intact and most likely functional, but protein-coding gene pieces seem to have no function in the mitochondrial genome of either plant.

It has been reported that a rapeseed mitochondrial linear plasmid had no sequence similarity to the mitochondrial genome (14). However, a 19 bp sequence, CTCTYCTTT-CAGTYGAGTT, was found to be common to these two DNA molecules. This sequence is located at three positions in the upstream regions of ORF2, 3 and 4 of the linear plasmid but is scattered at 41 positions in the rapeseed mitochondrial genome. Since there are no known functions (promoter, binding, etc.) for this sequence motif, it is unclear what role it has in rapeseed mitochondria. Interestingly, this motif is found 22 times in the *Arabidopsis* mitochondrial genome, although there have been no reports showing the presence of linear plasmids in *Arabidopsis* mitochondria. On the other hand, the sugar beet mitochondrial genome contains no such sequence motif. This sequence may be related to the co-evolution of the mitochondrial linear plasmid and mitochondrial genome specific to the Cruciferae species.

Potential promoters

By comparing sequences surrounding transcription start sites of plant mitochondria, the conserved nonanucleotide motif (CNM) has been identified in dicotyledonous plant species (15). Since no experimental information about the 5′ termini of primary transcripts is available for rapeseed, identification of potential promoter regions was carried out by screening of the rapeseed mitochondrial genome sequence with the CNM motif. In this search, only six CNM sequence variations were used, following the criteria of Dombrowski *et al.* (16).

Twenty-nine CNM motifs were found in the rapeseed mitochondrial genome. The number of potential promoters was reduced to 21 by specifying other characteristics of promoter sequences of dicot plants: the presence of an AT-box upstream of CNM and at least one purine at nucleotide positions +3 and +4. Fifteen promoters are located within a 1 kb region upstream of known genes and ORFs. These potential promoters may promote the transcription of the coding information *in vivo*. For instance, two of these potential promoters are found in the upstream region of the 18S rRNA genes (*rrn18*). In *Arabidopsis*, the transcription initiation site upstream of *rrn18* was experimentally investigated by primer extension analysis, and the CNM was found to reside at the transcription initiation site (17). Potential promoter sequences of rapeseed *rrn18* were identified at the same position in the homologous region of *Arabidopsis rrn18*, except that the rapeseed sequence is duplicated. In total, seven potential promoters upstream of identified genes, including *atp8* [formerly named *orfB*; (18)], *atp9*, *rrn26* and some tRNA genes, are located at the same positions as in the homologous regions of *Arabidopsis*. These genes may be transcribed in the same manner as in *Arabidopsis* using the same promoters. However, the four potential promoters for *rps7*, *rpl2* and two tRNA genes are either absent or present at a different position in the corresponding regions of the *Arabidopsis* mitochondrial genome. For these genes, there may be different transcriptional controls between rapeseed and *Arabidopsis*.

Table 3. RNA editing sites in mitochondria of rapeseed compared to those of *Arabidopsis*

Genes		Rapeseed	*Arabidopsis*	Common	Ref.	Genes		Rapeseed	*Arabidopsis*	Common	Ref.
Complex I	*nad1*	23	24	22		Ribosomal proteins	*rpl2*	2	1	0	
	nad2	25	31	24	(34)		*rpl5*	9	10	9	
	nad3	10	12	8	(7)		*rpl16*	6	8	5	
	nad4	35	32	29			*rps3*	8	10	7	
	nad4L	9	9	7			*rps4*	19	15	13	(38)
	nad5	29	27	24			*rps7*	1	0	0	
	nad6	11	10	10			*rps12*	7	8	7	(7)
	nad7	28	28	24			*rps14*	0	1	0	
	nad9	8	7	6							
						Cytochrome-c-biogenesis	*ccmB*	39	39	35	(7)
Complex III	*cob*	8	7	7			*ccmC*	25	28	23	
							ccmFN1	15	22	14	(11)
Complex IV	*cox1*	1	0	0			*ccmFN2*	10	12	9	(11)
	cox2-1	13	15	12			*ccmFC*	13	16	12	
	cox2-2	10	–	–							
	cox3	7	8	7		ORFs	*orf286*	0	–	–	
							orf188	0	–	–	
Complex V	*atp1*	5	5	4			*orf161*	0	–	–	
	atp4 (orf25)[a]	8	8	8			*orf159*	0	–	–	
	atp6	1	1	1	(35)		*orf164*	0	–	–	
	atp8 (orfB)[b]	3	0	0	(36)		*orf322*	0	–	–	
	atp9	4	4	4	(37)		*orf448*	0	–	–	
							orf261	–	–	–	
Other ORFs	*tatC (orfX)*	27	24	21			*orf257*	0	–	–	
	matR	8	9	6							
	orf222	0	–	–	(20)	Total		427	431	358	

Data for individual mRNA species previously analyzed have been included. The number of the reference (Ref.) is shown for those.
[a]From Heazlewood *et al.* (32).
[b]From Sabar *et al.* (18).

Some of the genes without a CNM in their 5′ UTR show only limited similarity to the corresponding regions of the *Arabidopsis* mitochondrial genome. For example, the rapeseed *nad6* and *atp1* 5′ UTR sequences diverged 2 and 53 bp upstream of the initiation codon, respectively, compared with the *Arabidopsis* sequences. The transcription of the *nad6* and *atp1* genes is likely to be initiated from different non-CNM promoters in the two plants. The actual number of promoters in the rapeseed mitochondrial genome should be investigated experimentally.

Introns and RNA editing sites

Before translation, mitochondrial messages in higher plant mitochondria may undergo RNA splicing and/or RNA editing, two post-transcriptional processing events that may critically alter the coding message.

In the rapeseed mitochondrial genome, 24 introns ranging in size between 839 and 3448 bp were identified. All of these introns belong exclusively to group II, as in other higher plants. Nineteen introns were *cis*-spliced sequences, and comprised 29 681 bp or 13.4% of the rapeseed mitochondrial genome. A total of five *trans*-spliced introns were found in the three *nad* genes, *nad1*, *nad2*, *nad5*. The intron numbers and positions are completely identical to those of *Arabidopsis* mitochondrial genome (3). Intron structures are also highly conserved between these two species.

RNA editing leads to change of a C to a U residue, sometimes creating an initiation or termination codon, but more often creating an internal codon with strong functional relevance. The editotype of *Arabidopsis*, that is, all editing sites found in the systematic transcriptome analysis of *Arabidopsis*, has been determined (19) and shown to consist of 441 sites scattered throughout the mitochondrial genome. There are 34 known protein-coding genes in rapeseed mitochondria as stated above. In total, 24 genes and nine ORFs longer than 150 amino acids were subjected to the editing analysis in this study. The remaining 10 genes had already been analyzed for RNA editing (Table 3). In addition, editing of *orf222*, the homologue of *orf224* associated with Polima male sterile cytoplasm, was detected (20).

In summary, 427 C to U RNA editing modifications were detected in genes and ORFs of rapeseed mitochondrial transcripts (Table 3). None of nine unidentified ORFs examined had editing sites in their transcripts, except for one partial editing event in the *orf188* transcript. The total number of editing sites (427 sites) in rapeseed is nearly identical to that (441 sites) in *Arabidopsis* (19). However, the number of editing sites shared by both plant mitochondria was only 358, which correspond to 83.8% and 81.2% of the total editing sites in rapeseed and *Arabidopsis* transcripts, respectively (Table 3). These percentages seem to be quite low considering that the average similarity in the primary DNA sequences between rapeseed and *Arabidopsis* is 99.2% for protein coding regions. This means that the tempo of diversification of RNA editing is more rapid than that of coding information.

Two RNA editing sites are located within the overlap region of *rps3* and *rpl16*. The second of these two edits creates a stop codon 21 amino acids downstream from the *rpl16* ATG codon. This RNA editing event is also observed in the *Petunia* and

Figure 2. Alignment of protein sequence deduced for *rps14* of rapeseed mitochondria with other plant RPS14 protein sequences. Rapeseed, rapeseed *rps14* (this study); *Oenothera*, deduced from *Oenothera rps14* cDNA (26); *Arabidopsis*, deduced from *Arabidopsis* nuclear *rps14* (28); Rice, deduced from rice nuclear *rps14* (30); *Marchantia*, deduced from *Marchantia* mitochondrial *rps14* (33). Amino acids that are conserved in at least three species are highlighted. Editing positions on *Oenothera rps14* transcripts are indicated by triangles, and changed amino acids are shaded.

Arabidopsis rpl16 transcripts (21,22), which suggests that *rpl16* gene might be a pseudo-gene. However, no ATG codon is encoded in the *rpl16* reading frames of *Oenothera* and *Marchantia* mitochondria, and a valine GTG codon is postulated to be used for translation initiation (23,24). This GTG codon is conserved in rapeseed mitochondria, 15 amino acids downstream from the stop codon introduced by the editing event. If GTG is used for translation initiation, RNA editing would have no effect on the translation of the *rpl16* gene.

The rapeseed *rps14* gene transcript is not edited. The *rps14* gene is one example of the gene-transfer events from the mitochondrion to the nucleus often found in plants. In fact, *rps14* is located in the mitochondrial genomes of broad bean, *Oenothera* and pea (25–27). In contrast *rps14* has been translocated to the nucleus in *Arabidopsis*, maize and rice (28–30). Two RNA editing events were observed in *Oenothera rps14* transcripts (26). The first editing event altered a codon from TCC (Ser) to TTC (Phe), and the second one modified a CCT (Pro) triplet to TCT (Ser). *Arabidopsis* Ψ*rps14* transcripts have retained the first editing site in the *Oenothera* transcript (26). Alignment of the deduced amino acid sequences from rapeseed and other plant *rps14* genes allowed us to determine that the second editing event in *Oenothera* occurred at a non-conserved position (Fig. 2). However the first RNA editing event, leading to the change of a Ser residue to a conserved Phe, seems to be essential for producing mature RPS14 protein (Fig. 2). It is possible that the rapeseed mitochondrial copy of *rps14* might be non-functional, although it is an intact reading frame and is transcribed. Further work is needed to examine whether a functional *rps14* gene exists in the rapeseed nuclear genome.

During the systematic analysis of RNA editing sites in rapeseed mitochondria, it was found that *ccmC* cDNA could not be amplified using the primer pair ccmC-F and ccmC-R, which are located before the start codon and after the stop codon of *ccmC*, respectively (Fig. 3A for the primer positions and Fig. 3B for the PCR products). Two other 3′-primers, ccmC-R1 and ccmC-R2 were also tested (Fig. 3A). Only one primer pair, ccmC-F and ccmC-R1, could amplify *ccmC* cDNAs (Fig. 3B). These results suggest that the *ccmC* transcripts are interrupted within the 50 bp-sequence before the stop codon. Twenty-five RNA editing sites were identified in the truncated *ccmC* transcript (ccmC-R1-primed cDNA); however, no editing creating a stop codon has been detected. This situation is also observed in *Arabidopsis ccmC* transcripts. The end of the *ccmC* transcript in *Arabidopsis* was mapped between primers ccmC-R1 and ccmC-R2 (Fig. 3A) (S.Binder and P.Giegé, personal communications). There are no data demonstrating the presence of a nuclear copy of the *ccmC* gene or the presence of the CcmC protein itself in rapeseed mitochondria at present. However, the present finding raises the question of the existence of CcmC and its mode of translation. Previously we reported that the structure of the *ccmFN* gene in rapeseed mitochondria was quite unusual (11). The *ccmFN* sequence is divided into two parts, which are about 74 kb apart (*ccmFN1* and *ccmFN2*, see Fig. 1). These two parts are transcribed and their transcripts are edited in a similar manner as the homologous transcripts of wheat (31). However, no mature transcript covering the whole coding region has been detected, a result which excludes a trans-splicing event. No other copy of this gene was found in either the nuclear or mitochondrial genome. Nevertheless, the protein product of *ccmFN* was identified in rapeseed mitochondria by western blot analysis. The mode of expression of these CcmFN and CcmC proteins in rapeseed remains mysterious.

CONCLUSIONS

The comparative analysis presented here allows a more comprehensive understanding of mitochondrial genome evolution in higher plants. It had previously been noted that plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence (6), which was concluded based on comparisons among *Brassica* species. Palmer and Herbon found numerous internal rearrangements of mitochondrial genomes by homologous recombination events through short dispersed repeats as the points of crossover. This picture is still true in part. In this study, rapeseed and *Arabidopsis* were shown to have nearly the same sets of genes for mitochondrial proteins, rRNAs and tRNAs, although the gene orders are different. The intron contents and locations are also completely conserved in the two plants. These results support the idea that internal genome rearrangement is one of the major

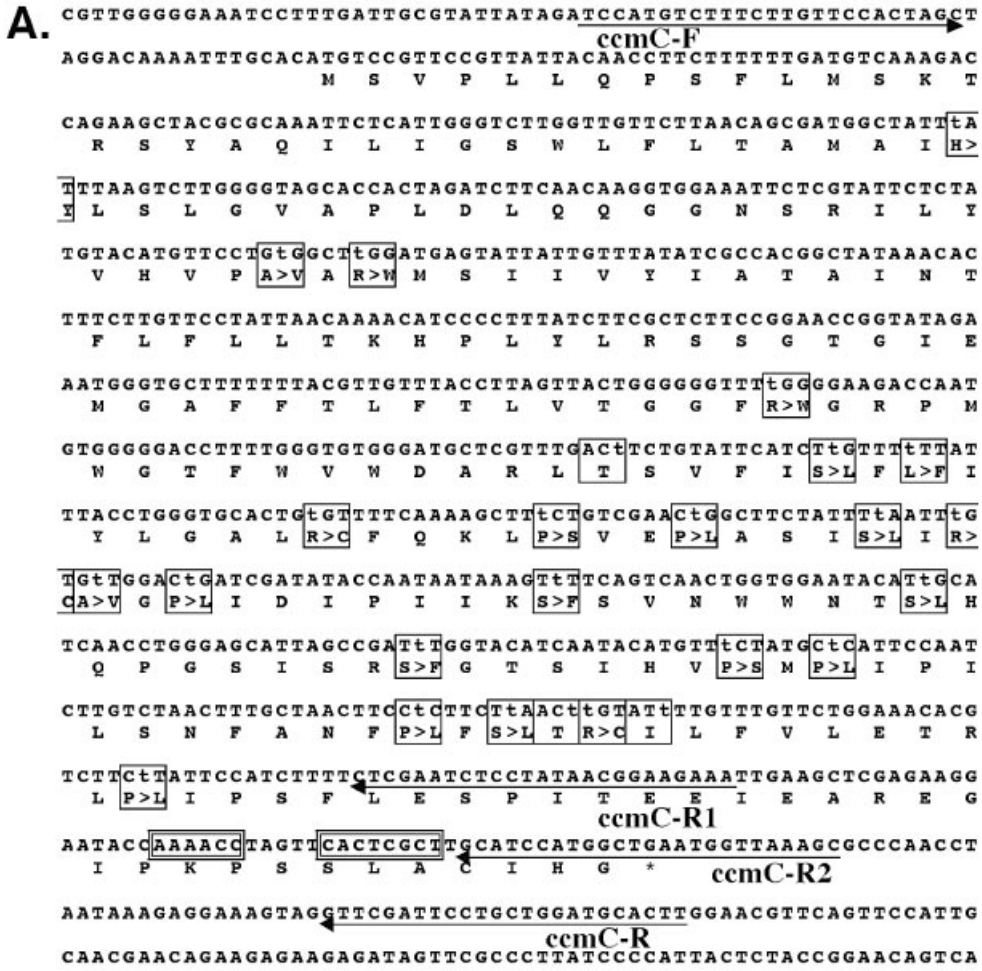

Figure 3. (**A**) Nucleotide sequence of the *ccmC* locus in rapeseed mitochondria. The polypeptide sequence is given below the genomic sequence. Codons altered by RNA editing are boxed together with the corresponding amino acid alterations. The edited nucleotides are shown in lower case letters. The ends of *ccmC* transcripts in *Arabidopsis* are boxed with double lines (S.Binder and P.Giegé, personal communications). The positions of oligonucleotides used for cDNA synthesis and PCR amplification are indicated by arrows. (**B**) Agarose gel electrophoresis of RT–PCR and PCR amplification products. The names of the oligonucleotides used as reverse primers are shown above the gel. Lane 1, PCR products using mitochondrial DNA as a template; lane 2, RT–PCR products with reverse transcriptase; lane 3, RT–PCR products without reverse transcriptase. M, DNA marker.

driving forces for the mitochondrial genome evolution of higher plants.

However, this study showed that large parts of mitochondrial genomes in higher plants are species-specific and show no homology to other plant mitochondrial genomes. One-third and two-thirds of the rapeseed and *Arabidopsis* mitochondrial genomes, respectively, were not aligned with each other. This means that after divergence of these species, the mitochondrial genomes of both plants were very quickly modified by sequence acquisition and/or sequence loss, accounting for up to two-thirds of the genome in the case of *Arabidopsis*. At this level, rapid sequence transfer, sequence acquisition and/or sequence loss is thought to have been a more important factor in the evolution than internal genome rearrangement. Parts of these non-alignable sequences originated from plastid or nuclear genomes, but most of them show no homology to registered sequences. It will be necessary to clarify the origin of these sequences and how they were transferred to mitochondria. Furthermore, the mtDNA sequences shared by three dicot plants, rapeseed, *Arabidopsis* and sugar beet, are limited to the coding sequences and introns. These data suggest that the plant mitochondrial genome is very dynamic and fluid, except for the primary gene information.

Some potential promoter regions are conserved between rapeseed and *Arabidopsis*, but the upstream regions of some genes are completely divergent from each other. On the other hand, RNA editing sites shared by both plant mitochondria constitute about 80% of the total editing sites. The remaining 20% of editing sites are species-specific. This percentage is quite high considering that the primary DNA sequences for the protein-coding regions are highly conserved between rapeseed and *Arabidopsis*. These data suggest that the evolutionary speed is higher at the level of gene regulation than at the primary gene sequence level.

An additional notable finding is that the transcripts of the *ccmC* gene are truncated before the stop codon. Previously we reported the unusual structure and expression of the *ccmFN* gene in rapeseed mitochondria. Although there are no precise data about the expression of CcmC protein in rapeseed mitochondria so far, the present finding raises another question about expression of Ccm proteins in rapeseed.

ACKNOWLEDGEMENTS

The author wishes to thank Ms Kazuko Hashimoto and Kyoko Minagawa for their skilful technical assistance. The author is greatly indebted to Dr Géraldine Bonnard for helpful advice and for help in the revision of the manuscript. Appreciation is expressed to Drs Stefan Binder and Philippe Giége for providing unpublished data.

REFERENCES

1. Palmer,J.D. and Herbon,L.A. (1987) Unicircular structure of the *Brassica hirta* mitochondrial genome. *Curr. Genet.*, **11**, 565–570.
2. Ward,B.L., Anderson,R.S. and Bendich,A.J. (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell*, **25**, 793–803.
3. Unseld,M., Marienfeld,J.R., Brandt,P. and Brennicke,A. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genet.*, **15**, 57–61.
4. Kubo,T., Nishizawa,S., Sugawara,A., Itchoda,N., Estiati,A. and 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.
5. Notsu,Y., Masood,S., Nishikawa,T., Kubo,N., Akiduki,G., Nakazono,M., Hirai,A. and Kadowaki,K. (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics*, **268**, 434–445.
6. Palmer,J.D. and Herbon,L.A. (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.*, **28**, 87–97.
7. Itani,K. and Handa,H. (1998) Rapeseed mitochondrial *ccb206*, a gene involved in cytochrome *c* biogenesis is cotranscribed with the *nad3* and *rps12* genes: organization, transcription and RNA editing of the *nad3/rps12/ccb206* locus. *Curr. Genet.*, **34**, 318–325.
8. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
9. Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
10. Brandt,P., Unseld,M., Eckert-Ossenkopp,U. and Brennicke,A. (1993) An *rps14* pseudogene is transcribed and edited in *Arabidopsis* mitochondria. *Curr. Genet.*, **24**, 330–336.
11. Handa,H., Bonnard,G. and Grienenberger,J.M. (1996) The rapeseed mitochondrial gene encoding a homologue of the bacterial protein Ccl1 is divided into two independently transcribed reading frames. *Mol. Gen. Genet.*, **252**, 292–302.
12. Wolstenholme,D.R. and Fauron,C.M.-R. (1995) Mitochondrial genome organization. In Levings,C.S.,III and Vasil,I.K. (eds), *The Molecular Biology of Plant Mitochondria.* Kluwer Academic Publishers, Dordrecht, pp. 1–59.
13. Sato,S., Nakamura,Y., Kaneko,T., Asamizu,E. and Tabata,S. (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.*, **29**, 283–290.
14. Handa,H., Itani,K. and Sato,H. (2002) Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (*Brassica napus* L.). *Mol. Genet. Genomics*, **267**, 797–805.
15. Binder,S., Marchfelder,A. and Brennicke,A. (1996) Regulation of gene expression in plant mitochondria. *Plant Mol. Biol.*, **32**, 303–314.
16. Dombrowski,S., Hoffmann,M., Kuhn,J., Brennicke,A. and Binder,S. (1998) On mitochondrial promoters in *Arabidopsis thaliana* and other flowering plants. In Møller,I.M., Gardeström,P., Glimelius,K. and Glaser,E. (eds), *Plant Mitochondria: From Gene to Function*. Backhuys Publishers, Leiden, Netherlands, pp. 165–170.
17. Giese,A., Thalheim,C., Brennicke,A. and Binder,S. (1996) Correlation of nonanucleotide motifs with transcript initiation of 18S rRNA genes in mitochondria of pea, potato and *Arabidopsis*. *Mol. Gen. Genet.*, **252**, 429–436.
18. Sabar,M., Gagliardi,D., Balk,J. and Leaver,C.J. (2003) ORFB is a subunit of $F_1F_{(O)}$-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. *EMBO Rep.*, **4**, 381–386.
19. Giegé,P. and Brennicke,A. (1999) RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl Acad. Sci. USA*, **96**, 15324–15329.
20. L'Homme,Y., Stahl,R.J., Li,X.Q., Hameed,A. and Brown,G.G. (1997) *Brassica* nap cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the pol CMS-associated *orf224* gene. *Curr. Genet.*, **31**, 325–335.
21. Sutton,C.A., Conklin,P.L., Pruitt,K.D., Calfee,A.J., Cobb,A.G. and Hanson,M.R. (1993) Editing of *rps3/rpl16* transcripts creates a premature truncation of the *rpl16* open reading frame. *Curr. Genet.*, **23**, 472–476.
22. Sakamoto,W., Tan,S.-H., Murata,M. and Motoyoshi,F. (1997) An unusual mitochondrial *atp9-rpl16* cotranscript found in the maternal distorted leaf mutant of *Arabidopsis thaliana*: Implication of GUG as an initiation codon in plant mitochondria. *Plant Cell Physiol.*, **38**, 975–979.
23. Bock,H., Brennicke,A. and Schuster,W. (1994) *Rps3* and *rpl16* genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria? *Plant Mol. Biol.*, **24**, 811–818.
24. Takemura,M., Oda,K., Yamato,K., Ohta,E., Nakamura,Y., Nozato,N., Akashi,K. and Ohyama,K. (1992) Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*. *Nucleic Acids Res.*, **20**, 3199–3205.
25. Wahleithner,J.A. and Wolstenholme,D.R. (1988) Ribosomal protein S14 genes in broad bean mitochondria. *Nucleic Acids Res.*, **16**, 6897–6913.

26. Schuster,W., Unseld,M., Wissinger,B. and Brennicke,A. (1990) Ribosomal protein S14 transcripts are edited in *Oenothera* mitochondria. *Nucleic Acids Res.*, **18**, 229–233.
27. Hoffmann,M., Dombrowski,S., Guha,C. and Binder,S. (1999) Cotranscription of the *rpl5-rps14-cob* gene cluster in pea mitochondria. *Mol. Gen. Genet.*, **261**, 537–545.
28. Figueroa,P., Gomez,I., Carmona,R., Holuigue,L., Araya,A. and Jordana,X. (1999) The gene for mitochondrial ribosomal protein S14 has been transferred to the nucleus in *Arabidopsis thaliana*. *Mol. Gen. Genet.*, **262**, 139–144.
29. Figueroa,P., Gomez,I., Carmona,R., Holuigue,L., Araya,A. and Jordana,X. (1999) Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J.*, **18**, 601–609.
30. Kubo,N., Harada,K., Hirai,A. and Kadowaki,K. (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: Common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl Acad. Sci. USA*, **96**, 9207–9211.
31. Bonnard,G. and Grienenberger,J.M. (1995) A gene proposed to encode a transmembrane domain of an ABC transporter is expressed in wheat mitochondria. *Mol. Gen. Genet.*, **246**, 91–99.
32. Heazlewood,J.L., Whelan,J. and Millar,A.H. (2003) The products of the mitochondrial *orf25* and *orfB* genes are F_O components in the plant F_1F_O ATP synthase. *FEBS Lett.*, **540**, 201–205.
33. 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.
34. Handa,H., Mizobuchi-Fukuoka,R. and Pinyarat,W. (1997) The rapeseed mitochondrial gene for subunit 2 of the NADH dehydrogenase complex: A trans-spliced structure is conserved on one of the smallest plant mitochondrial genome. *Curr. Genet.*, **31**, 336–342.
35. Handa,H. and Nakajima,K. (1992) RNA editing of *atp6* transcripts from male-sterile and normal cytoplasms of rapeseed (*Brassica napus* L.). *FEBS Lett.*, **310**, 111–114.
36. Handa,H., Gualberto,J.M. and Grienenberger,J.M. (1995) Characterization of the mitochondrial *orfB* gene and its derivative, *orf224*, a chimeric open reading frame specific to the mitochondrial genome of 'Polima' male-sterile cytoplasm in rapeseed (*Brassica napus* L.). *Curr. Genet.*, **28**, 546–552.
37. Handa,H. (1993) RNA editing of rapeseed *atp9* transcripts: RNA editing changes four amino acids, but termination codon is already encoded by genomic sequence. *Jpn. J. Genet.*, **68**, 47–54.
38. Handa,H., Kubo,N. and Kadowaki,K. (1998) Genes for the mitochondrial ribosomal S4 protein encoded by higher plant mitochondria are transcribed, edited and translated. *Mol. Gen. Genet.*, **258**, 199–207.