The EMBO Journal vol.15 no.23 pp.6652–6661, 1996

Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals

Koh-ichi Kadowaki[1], Nakao Kubo, Kazuhiro Ozawa and Atsushi Hirai[2]

National Institute of Agrobiological Resources, Department of Molecular Biology, Tsukuba, Ibaraki 305 and [2]University of Tokyo, Faculty of Agriculture, Yayoi, Bunkyo, Tokyo 113, Japan

[1]Corresponding author

We have cloned a gene for mitochondrial ribosomal protein S11 (RPS11), which is encoded in lower plants by the mitochondrial genome, in higher plants by the nuclear genome, demonstrating genetic information transfer from the mitochondrial genome to the nucleus during flowering plant evolution. The sequence s11-1 encodes an N-terminal extension as well as an organelle-derived RPS11 region. Surprisingly, the N-terminal region has high amino acid sequence similarity with the presequence of the β-subunit of ATP synthase from plant mitochondria, suggesting a common lineage of the presequences. The deduced N-terminal region of s11-2, a second nuclear-encoded homolog of *rps11*, shows high sequence similarity with the putative presequence of cytochrome oxidase subunit Vb. The sharing of the N-terminal region together with its 5′ flanking untranslated nucleotide sequence in different proteins strongly suggests an involvement of duplication/recombination for targeting signal acquisition after gene migration. A remnant of ancestral *rps11* sequence, transcribed and subjected to RNA editing, is found in the mitochondrial genome, indicating that inactivation of mitochondrial *rps11* gene expression was initiated at the translational level prior to termination of transcription.

Keywords: *coxVb*/gene transfer/mitochondria/mitochondrial targeting signal/*rps11*

Introduction

It is generally accepted that mitochondria are a descendant of a procaryote that entered into symbiosis with other cell types (reviewed in Gray, 1989). Sequence analyses of mitochondrial DNAs, including entire mitochondrial genomes, indicate that only a small percentage of total mitochondrial proteins are synthesized within the organelle and most mitochondrial proteins depend on nuclear genes (Hartl *et al.*, 1989).

Comparison of entire mitochondrial genome sequences shows relatively conserved gene content among mammals, *Xenopus*, *Drosophila* and yeast (reviewed in Gray, 1992). Recently, however, many additional genes have been identified in lower plant (liverwort and chlorophyte alga) mitochondrial genomes (Oda *et al.*, 1992; Wolff *et al.*, 1994). A striking example is that none of the mitochondrial ribosomal protein genes are encoded by the mitochondrial genome of mammals, *Xenopus*, *Drosophila* and fungi (reviewed in Gray, 1992), except for the yeast *varl* gene (Terpstra *et al.*, 1979). In contrast, numerous ribosomal protein subunits are encoded by the lower plant mitochondrial genomes, with as many as 16 ribosomal protein genes identified in the liverwort mitochondrial genome. In addition, the gene order is similar to the organization of *Escherichia coli* ribosomal protein genes (Takemura *et al.*, 1992; Wolff *et al.*, 1994). This difference in mitochondrial gene content among these organisms results from translocation of genes from the endosymbiont to the nucleus. Thus, plant mitochondrial genomes retain a more primitive genetic content and gene transfer may be an active process in evolution (Brennicke *et al.*, 1993).

In some cases, gene transfer to the nucleus has occurred relatively recently. They are cytochrome oxidase subunit II (*coxII*) from cowpea and soybean (Nugent and Palmer, 1991; Covello and Gray, 1992), *Oenothera* ribosomal protein subunit S12 (Grohmann *et al.*, 1992) and *Arabidopsis* ribosomal protein subunit S10 (Wischmann and Schuster, 1995). By summarizing the recently identified variation of the mitochondrial gene content of different plant species, Brennicke *et al.* (1993) divide transfer of genetic information from mitochondria to the nucleus into several steps: gene translocation; adaptation and activation of the integrated nuclear sequence, such as acquisition of a presequence, promoter, poly(A) signal, etc.; mitochondrial gene inactivation and mitochondrial gene elimination. However, very little is known about the mechanisms for any of the above steps.

In this study, we report interesting evidence for the origin of targeting signals and for the machinery of mitochondrial gene inactivation, by analyzing a gene encoding mitochondrial ribosomal protein subunit S11 (RPS11). In lower plants, the *rps11* gene is present in the mitochondrial genome (Takemura *et al.*, 1992; Wolff *et al.*, 1994), but no information for *rps11* has been available for higher plants. During identification of mitochondrial ribosomal protein genes in a higher plant, we found a mitochondrial pseudogene for *rps11* and a functional nuclear *rps11* gene (s11-1) from rice. A second nuclear copy (s11-2) having 92% nucleotide sequence homology with s11-1 was also identified from rice. The deduced amino acid sequence of s11-1 showed an extended N-terminal region that was very similar to the presequence of the plant ATPase β-subunit peptide (ATPB), strongly suggesting that the RPS11 and ATPB presequences have a common origin. The deduced N-terminal extension in s11-2 was very similar to the putative presequence for cytochrome oxidase subunit Vb peptide (COXVb).

Mitochondrial presequences diverge greatly in amino acid sequence in general. They are rich in basic residues and have few acidic residues and frequently are capable

© Oxford University Press

A

```
CTATTGCGAAATACGGCTTCCTTAATATCAAAATTTCAACAAGGGGGCTGGGCTGCTCGCCTTCATAGAAAGGCGACCGGGCCTAGATTA  -111

GATGTTCGCCTTTTTATAGAATAGAAAGAGAAGGTAAAGGGGGGAAGGACATAGGAAAGAGGGATGATGCCTACAAAAATCAATTGATTC  -21
                       1                                                                p1
GTCAACACTATTCACATAGGATGCCcCAGGAAAAAACAACGGCCGAGACTGATAGGCCTGGGCTTGCGGGTGCAGGCAGTTTCAAATTCA   70
                    M  P  Q  E  K  T  T  A  E  T  D  R  P  G  L  A  G  A  G  S  F  K  F  K

AGAACTTcGTAcAGAGCATCTCGGAGGAATATAAACAATGTGAACGGAAAAAGAAGAGAGATTTTATCCATGTCTTACTGATGAAAAATA  160
 N  F  V  Q<*S  I  S  E  E  Y  K  Q  C  E  R  K  K  K  R  D  F  I  H  V  L  L  M  K  N  K

AGACCTTTGTGACCGTGACAGATGCTAGGGGGAATAAAAAGACTGGGGCCTCCGCAGGTTGTTTGGGGTAAATAAAAGGGCGGTCTCGTC  250
 T  F  V  T  V  T  D  A  R  G  N  K  K  T  G  A  S  A  G  C  L  G  *  I  K  G  R  S  R  L

TTTCTCGATATGCTGCTGAAGCAACTGCGGAACATGTCGGGCGATCCGCCAGAAAGATGGGTTTAAAGTCAGTTGTCATGAAAGTGAAAG  340
 S  R  Y  A  A  E  A  T  A  E  H  V  G  R  S  A  R  K  M  G  L  K  S  V  V  M  K  V  K  G

GATCTACTTATTTcCAGAAAAAGAAGAAAGTGATCTTGAGCTGGGGAGAAGGTTTTCGGGGTGAAAGAGTAAGAGATCAATCTCCTATTA  430
 S  T  Y  F  Q  K  K  K  K  V  I  L  S  W  G  E  G  F  R  G  E  R  V  R  D  Q  S  P  I  M

TGTACATCCACGATGTGACCCAACTTCCACATAATGTCCCCCGACTCCTCCCTCTACCGGGATCGTGATTTCCTCTTTTTTCCATTAATG  520
 Y  I  H  D  V  T  Q  L  P  H  N  V  P  R  L  L  P  L  P  G  S  *

CGCTCTTACTCTTTTCGTAAGGAAGCCATCCGCTTTCGAGCATCGAATCGTTGCTGCTTTTGCTATTGGTCTACAAATCGCTTTGACCCA  610
                 p2
TGTCCCGAAAACTCATTGAATGGGCTTATGAAAGGTGAGATAGATAAAAATGAAGTTGGAATGGAAAGGGATGGCAGGAACAAGGTAG    698
```

B

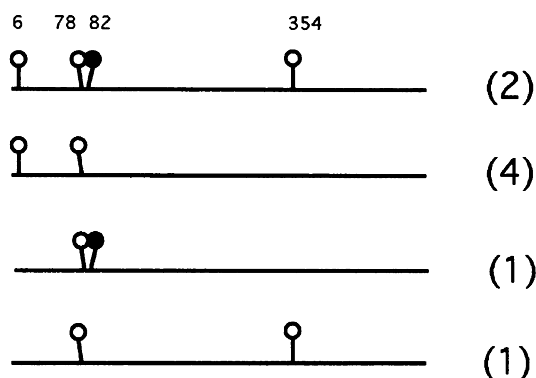

Fig. 1. (**A**) Rice mitochondrial *rps11* sequence similar to the liverwort mitochondrial *rps11* gene. The amino acid sequence deduced from the DNA is denoted by the single letter abbreviations underneath the DNA sequence. The conceptual original ORF is truncated by an internal TAA stop codon (underlined). Positions and orientations of oligonucleotide primers 1 and 2 are indicated by arrows. RNA editing sites (C of genomic DNA to T of cDNA) are indicated by lower case and underlined. A termination codon resulting from the RNA editing event at nucleotide position 82 is indicated by an asterisk. The A of the initiation codon ATG is numbered 1. (**B**) Overview of RNA editing events in the rice mitochondrial *rps11* transcripts. RNA editing sites are schematically indicated in the figure. White and black circles show silent alterations in amino acids and creation of a stop codon respectively. The number of clones examined is indicated in parentheses. Nucleotide positions of RNA editing events are shown above the circles.

of forming an amphiphilic helix (Hartl *et al.*, 1989). Interestingly, however, common presequences were found to be used for different proteins. A scenario of *rps11* gene transfer, inactivation of mitochondrial gene expression, acquisition of targeting presequences and evolutionary relationships of DNA sequences coding for mitochondrial presequences are discussed.

Results

Mitochondrially encoded rps11 is not functional in rice

The mitochondrial *rps11* gene is located in the gene cluster *rps8 rpl6 rps13 rps11* and *rpl5 rpl6 rps13 rps11* in liverwort (Takemura *et al.*, 1992) and chlorophyte alga mitochondrial genomes (Wolff *et al.*, 1994) respectively. In the higher plant rice, most of the ribosomal protein genes were found to be scattered in the mitochondrial genome. This has been characterized by physical mapping using a contiguous rice mitochondrial library and liverwort ribosomal protein genes as probes (data not shown). During the study, a nucleotide sequence homologous with the liverwort *rps11* gene was identified, and a 3.3 kb *Sal*I DNA fragment was cloned and sequenced. The rice mitochondrial *rps11* copy had 57% nucleotide sequence identity and the deduced amino acid sequence showed 34% identity with liverwort mitochondrial *rps11*. However, a TAA stop codon existed at nucleotide position 229, which is located in the middle of the original open reading frame (ORF), suggesting that the rice mitochondrial *rps11* is a pseudogene (Figure 1A). Exon d of the *nad1* and the *rps4* genes were identified 758 bp upstream and 461 bp downstream of the *rps11* sequence on the same strand and a sequence homologous with the liverwort *rps13* gene was identified >5 kb downstream of the *rps11* sequence.

Transcription of the mitochondrial *rps11* sequence was analyzed by RNA gel blot analysis. A 3.3 kb transcript was observed in the mitochondrial RNA (Figure 4C), showing that the *rps11* sequence was transcribed in mitochondria. The 3.3 kb transcript is much larger than the size of the mitochondrial *rps11* sequence. RT–PCR and subsequent DNA sequence analyses showed the presence

```
CTCCCCATCCTCCTCCTCCTCCGCCGCCGCCGCCGTCGGGAAGTACACCTGTGCTCTGGTTCCGGCGACACCGCGAGCGAGGAGAGGAGA  -112
GGTGGGGGTGAGGGAGTGAGGGTGAGGCCGGGGTTGGGCACCACCAGCAAGCCAGCCAGCCCCCCCGTCCCCTTAGATCCGGGCCGGGCG   -22

TTCTCCGGCGATCTCACGGCCATGGCGTCCCGCCGCCTCTTCGCCTCCCTCCTCAGGTCCACCGCGGCGCCGCGCACCCCGTCCACGCCC    69
                     [M] A  S  R  R  L  F  A  S  L  L  R  S  T  A  A  P  R  T  P  S  T  P
                                                                     intron-1
GGGTGCCTCTTCAACCGCGCCGCCGCCGCCGCCTACTCCTCGTCCGCCCCCTACAACGGCCAGGGTTTTCCACTTCCTCAGTCTGAAACT   159
G  C  L  F  N  R  A  A  A  A  A  Y  S  S  S  A  P  Y  N  G  Q  G  F  P  L  P  Q  S  E  T

GCTTCACGTCTGGGTTTGTTCTCCAGTCCCGGCGACACACGCCAACCTTCATACGGAGACCGTCTAATGCAGTCACAACAGCTGAGTCAA   249
A  S  R  L  G  L  F  S  S  P  G  D  T  R  Q  P  S  Y  G  D  R  L [M] Q  S  Q  Q  L  S  Q
                                                      intron-2
GACTACCGTGCTAGGACACAGGCCAATAATGCGCCACGTTTTGGTGACACAATGTCCAGGATAGCTGGTGGCGAGAACTCTTCCTACTTT   339
D  Y  R  A  R  T  Q  A  N  N  A  P  R  F  G  D  T [M] S  R  I  A  G  G  E  N  S  S  Y  F

GGGACCCCATCACGCATCTTTGACGAACACAAACAATCTTTGGTGAAAGGAAAGAGAGATTTTGTCCATGTCTTGCTGAAGAGAAATAAG   429
G  T  P  S  R  I  F  D  E  H  K  Q  S  L  V  K  G  K  R  D  F  V  H  V  L  L  K  R  N  K

ACCTTTGTGACCGTGACAGACGTTAGGGGGAACAAAAAGACCGGGGCATCCGCTGGTTGTTTGGAGGACAGGAAAGGGCGCTCTCGTCTT   519
T  F  V  T  V  T  D  V  R  G  N  K  K  T  G  A  S  A  G  C  L  E  D  R  K  G  R  S  R  L

TCCAAATATGCTGCTGAAGCAACTGCAGAACATGTCGGGCGTGCTGCCAGGAAGATGGGTTTAAAATCTGTGGTCATGAAAGTGAAGGGA   609
S  K  Y  A  A  E  A  T  A  E  H  V  G  R  A  A  R  K  M  G  L  K  S  V  V  M  K  V  K  G

ACTACATTTTTCAATAAGAAGAAGAAAGTGATCCTGAGCTTTAGAGAAGGCTTTCGGGGTGAAAGAGTAAGGGAGCAATCTCCTGTGGTG   699
T  T  F  F  N  K  K  K  K  V  I  L  S  F  R  E  G  F  R  G  E  R  V  R  E  Q  S  P  V  V

CTCATCCACGATGTGACCCAACTTCCACACAACGGATGCCGACTCCCCAAACAACGCCGGGTTTAGGTCTCAGCGAGCTGGAATCGAACC   789
L  I  H  D  V  T  Q  L  P  H  N  G  C  R  L  P  K  Q  R  R  V  *

TGAGTTGAGAGGAGTCGAAGCATTCAAGGTGATCTGGAAGCTCTATTTAAGCTGTGTGATGTTGGTCTAGTTGACCTAAGACAGATATGT   879
ATTGTTTCTCTTGGCTCTGCTACTTGAATTGGTTTTGGCATGACTTATGTTACTCAGTTAGTTAAGTTATCATCAAAAAAAAAAAAAAAA   969
AA                                                                                           971
```

intron-1

```
GTACTCTTGTTTGTATGCGCCATGGATAACGTATATCTCGAAGCTAATGATGAATTGATGATATCCCCAAGCGCACTGGATGTTTTCACC    90
TTGGTTAGCAGTAATTGCAGAGATTATTTTCTTTCCTTTTCTCTTTCAACAATATATTTCAGTAATAATAAGAATGGTTAATGTGAGCCA   180
CCCAGCTGAATCATCGGTACATACCAAAGTACTCTTTAGAGATTTTGCTCAATTCAGTTTAATTTCTCTGCCTCATTTTTTTTCATTCAC   270
CTTGTGGAATATCGATATTACGCCTCGTTATATTATCATAGATGCTGAGATCAGACTGCATCAATCACTGAATCTCACAATTTGACAAGA   360
TCGATTTTGTCTGTCCGTCCAG                                                                       382
```

intron-2

```
GTGAGTAGGCAAATGGTATTTCCATGTACGCTCATATGATTTATGTATTCCATGTTCTGCTGATTGTTCATTCCTGCCTCATTCTGTATT    90
AAG                                                                                           93
```

Fig. 2. Nucleotide sequence of the rice nuclear cDNA encoding mitochondrial RPS11. The cDNA and deduced amino acid sequences are shown. Three Met residues located upstream of the conserved *rps11* region are framed. Two arrowheads indicate the positions of 382 and 93 bp intron sequences identified in a genomic DNA counterpart. The intron sequences are presented below the cDNA sequence.

of a transcript from *nad1* exon d to *rps4* (data not shown). Thus the mitochondrial *rps11* sequence was co-transcribed with the upstream and downstream sequences.

RNA editing of *rps11* transcripts was analyzed to determine whether reverse RNA editing (U to C conversion) may restore the stop codon to a sense codon. cDNAs were synthesized using random hexamers and reverse transcriptase on mitochondrial RNA treated with DNase I. *rps11* cDNAs were amplified by the polymerase chain reaction using primers 1 and 2 and cloned into a plasmid vector. Eight cDNA clones were randomly picked and their DNA sequences were determined. Four RNA editing sites were identified at nucleotide positions 6, 78, 82 and 354 (Figure 1B). All of the cDNAs examined were edited, although the extent of RNA editing was different among the cDNAs. This result, together with RNA gel blot analysis, revealed that mitochondrial *rps11* was transcribed and subjected to RNA editing. Editing events at nucleotide positions 6, 78 and 354 were silent changes (no change in amino acid) and the editing event at position 82 resulted in the creation of a termination codon. None of the RNA editing changed the internal stop codon to a sense codon. On the contrary, RNA editing introduced another stop codon at nucleotide position 82 in the transcripts. The above results indicate that mitochondrial *rps11* is indeed a pseudogene.

The functional copy encoding mitochondrial RPS11 is encoded by the rice nuclear genome

To address the possibility that an *rps11* gene had been transferred from the mitochondrion to the nucleus and expressed, cDNA clones were screened from a rice nuclear cDNA library using the isolated rice mitochondrial *rps11* sequence as probe. Seven clones were isolated, and sequence analysis revealed that the largest clone contained an ORF of 254 amino acid residues with 73% nucleotide sequence identity to the rice mitochondrial *rps11* pseudogene (Figure 2). At the 3′-end of the cDNA, an 18 nt poly(A) tract was identified. Deduced amino acid sequence comparison of the rice cDNA showed that 37 and 33% similarities were observed with liverwort mitochondrial RPS11 and *E.coli* RPS11 respectively. The C-terminal sequence, Arg-Arg-Val, of the nuclear-encoded RPS11 is the same as that of the liverwort mitochondrial and *E.coli* counterparts and the last 20 amino acids of the C-terminal end are well conserved (Figure 3A). On the other hand, an extension of >80 amino acids was observed at the N-terminus of the rice nuclear gene for RPS11, suggesting that the extended region is necessary for mitochondrial targeting. There are three Met codons in the N-terminal extension and the first Met codon was inferred as the probable start codon (see below).

Amino acid sequence deduction around the internal

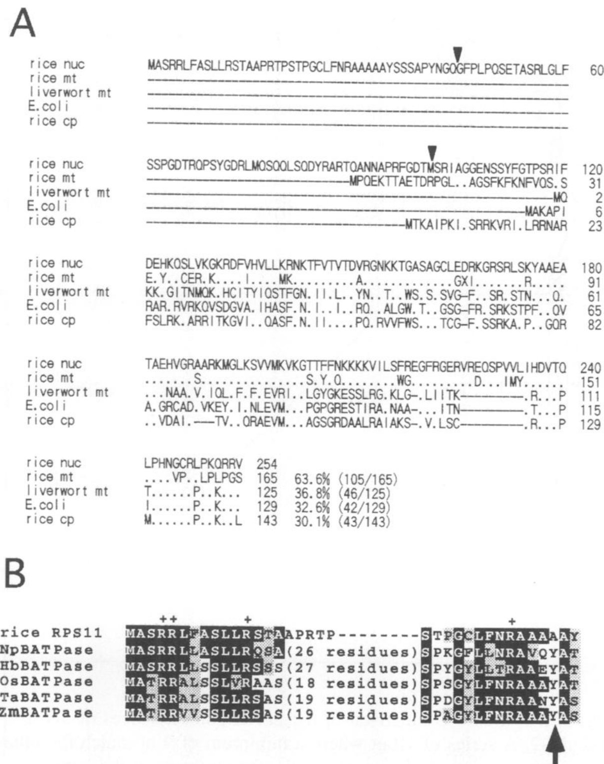

Fig. 3. Comparison of amino acid sequences. (**A**) Amino acid sequences of RPS11 encoded in rice nuclei (this study), rice mitochondria (this study), liverwort mitochondria (Takemura *et al.*, 1992), *E.coli* (Bedwell *et al.*, 1985) and rice chloroplast (Hiratsuka *et al.*, 1989) are aligned. Positions of two introns in the rice nuclear gene are indicated by arrowheads. The extents of amino acid identities are shown as percentages. The numbers of identical residues in total residues are shown in parentheses. (**B**) Amino acid sequence alignment of the N-terminal portion of the nuclear-encoded mitochondrial RPS11 from rice and ATPB from *N.plumbaginifolia* (Boutry and Chua, 1985), rubber plant (Chye and Tan, 1992), rice (Sakamoto *et al.*, 1992), wheat (Abulafia and Breiman, 1993) and maize (Winning *et al.*, 1990) from top to bottom. Residues identical to the rice RPS11 amino acid sequence are shown by reverse contrast. Similar residues, those with non-polar or uncharged polar groups, are indicated by halftone color. An arrow indicates the cleavage site of *N.plumbaginifolia* ATPB (Chaumont and Boutry, 1995). Basic amino acid residues of rice RPS11 are indicated by + above the alignments.

stop codon TAA in the mitochondrial *rps11* sequence shows Gly-Cys-Leu-Gly-X-Ile-Lys-Gly-Arg (X indicates that no amino acid is encoded by the TAA codon), while Gly-Cys-Leu-Glu-Asp-Arg-Lys-Gly-Arg was encoded by the corresponding nuclear gene (Figure 3A). This result suggests that the internal termination codon of the mitochondrial *rps11* copy resulted from nucleotide substitution, rather than a deletion or insertion.

A 1.3 kb transcript was observed by RNA gel blot analysis of poly(A)$^+$ RNA and is different in size from the mitochondrial *rps11* transcript (Figure 4C). The 1.3 kb transcript is similar to the size of the rice nuclear *rps11* cDNA (1172 nt). Thus, both the mitochondrial and nuclear *rps11* sequences were transcribed.

The 5′-terminal extension of rps11 encodes a peptide similar to the plant ATPB presequence

The deduced N-terminal amino acid sequence of RPS11 was used to search for homology in the sequence databases. High similarity was found with presequences for the

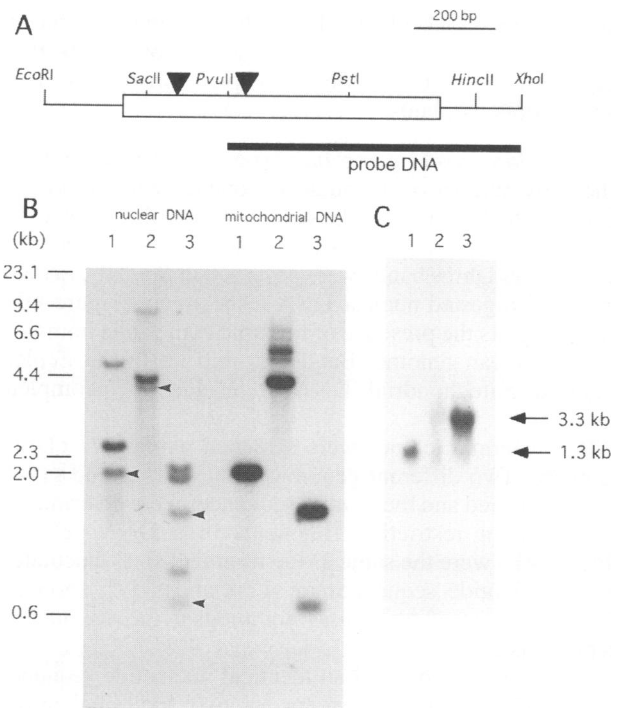

Fig. 4. DNA and RNA gel blot analyses of the rice nuclear gene for mitochondrial RPS11. (**A**) Schematic representation of the rice nuclear *rps11* cDNA and its physical map. An ORF is shown by an open box. Positions of two intron sequences in the genomic DNA are marked by filled triangles. Probe DNA indicated by a thick bar was prepared from the cDNA, hence intron sequences are not included. The 0.7 kb *Pvu*II–*Xho*I fragment from the nuclear *rps11* cDNA was used as probe. (**B**) Autoradiogram of nuclear and mitochondrial DNAs from rice. Lanes 1–3 show DNA digested with *Bam*HI, *Eco*RI and *Dra*I respectively. Arrowheads in the nuclear DNA indicate signals derived from the mitochondrial DNA contamination. Size markers in kb are indicated on the left. (**C**) Poly(A)$^+$ RNA (3.9 μg) (lane 1), 9.6 μg total RNA (lane 2) and 5.0 μg mitochondrial RNA (lane 3) were separated on a 1.5% agarose–formaldehyde gel. Sizes of transcripts are indicated in the figure.

β-subunit of the ATP synthase (ATPB) from plants (Figure 3B). The N-terminal region of mitochondrial RPS11 encoded by the rice nuclear genome had 84% sequence identity (11 out of 13 amino acid residues) with ATPB presequences from *Nicotiana plumbaginifolia* and rubber plant (Boutry and Chua, 1985; Chye and Tan, 1992), provided that the extended region was translated from the first ATG codon (Figure 2A). All of the differences in the first 15 residues were conservative changes in *N.plumbaginifolia*. A second region of amino acid sequence similarity corresponded to the C-terminal portion of the ATPB presequence. The numbers of amino acids between the two blocks were different among the plants. These two blocks of amino acid sequence similarity with plant ATPB presequences may suggest that the N-terminal extension of rice RPS11 is a targeting presequence and that the first Met codon in the *rps11* reading frame is a start codon. No negatively charged amino acid residue was found in the region homologous with the ATPB presequences.

The N-terminal portion of ATPB presequences from rubber plant and *N.plumbaginifolia*, dicotyledonous plants, had a slightly higher extent of amino acid sequence similarity to the rice RPS11 presequence than those from

monocotyledonous plants. In contrast, higher similarity was observed near the processing site of ATPB presequences from monocotyledonous plants than those from dicotyledonous plants.

The nuclear rps11 gene has two intron sequences

The copy number of the nuclear sequence with homology to the mitochondrial *rps11* sequence was determined by DNA gel blot analysis. An autoradiogram showed that two, two and three bands were observed in *Bam*HI-, *Eco*RI- and *Dra*I-digested nuclear DNA respectively (Figure 4B). This suggests the presence of multiple copies of a homolog in the nuclear genome. Bands of ~6.0 kb in the *Eco*RI-digested mitochondrial DNA were due to incomplete digestion.

Rice genomic clones were screened with *rps11* cDNA as probe. Two different genomic clones, s11-1 and s11-2, were identified and their entire sequences were determined. The sizes of restriction fragments in a DNA gel blot (Figure 4B) were the same as the fragment sizes calculated from nucleotide sequences of s11-1 and s11-2, showing that the copy number of *rps11* homologs in the rice nuclear genome is two.

Clone s11-1 contained an identical nucleotide sequence to the cDNA sequence except for two additional intron sequences of 382 and 93 bp (Figure 2). An intron (intron 2) was identified around the junction of the DNA sequence encoding the extended peptide sequence and the RPS11 conservative sequence. The first intron (intron 1) was observed upstream of intron 2 and in the middle of the DNA sequence encoding the extended peptide. The two intron sequences have typical characteristics of nuclear exon–intron junctures (5′-GT/AG-3′), i.e. the first dinucleotide of the intron is GT and the last is AG (Breathnach and Chambon, 1981). The region homologous with the ATPB presequence was encoded only by exon 1 of s11-1.

Amino acid and nucleotide sequence similarity searches with exon 2 of s11-1 have failed to detect any homologous sequence in the database. The origin of the exon 2 sequence is unknown.

Two independent copies similar to mitochondrial rps11 are present in the nuclear genome

Sequence analysis of clone s11-2 showed 92% nucleotide and 94% amino acid homology to s11-1 in the conserved *rps11* region (the sequence of s11-2 is not shown, but is available in DDBJ/EMBL/GenBank under accession No. D85195). s11-2 had nucleotide sequence homology with the region downstream of the junction of intron 1 and exon 2 of s11-1 (Figure 5). The upstream sequence from intron 1 of s11-1, however, is totally different from that of s11-2. An intron-like sequence was present in s11-2 at the same position as intron 2 of s11-1 with 79% nucleotide sequence similarity. However, the junction of the exon–intron sequence of s11-2 was 5′-GG/AG-3′, a deviation from the consensus splicing motif 5′-GT/AG-3′. Nucleotide sequence similarity extends to 615 bp downstream of the TAG stop codon of s11-1 (Figure 5).

The 5′ portion of s11-2 encodes a peptide sequence similar to the putative presequence of COXVb

To examine whether the 5′ portion of s11-2, upstream of the region homologous with s11-1, is present elsewhere and transcribed in the rice nuclear genome, plaque hybridization was performed against a rice cDNA library using the 0.8 kb *Xho*I–*Not*I fragment of s11-2 as probe. Three clones were isolated and their nucleotide sequences were determined. Nucleotide sequence comparison showed that the 5′ portion of the cDNA clone had 81% nucleotide sequence identity with the 5′ portion of s11-2 over 249 nt, but the downstream sequence had no similarity with the *rps11* region of s11-2 (Figure 6). No cDNA clones with an *rps11*-related sequence have been isolated using this probe.

The isolated cDNA clone encoded an ORF of 507 bp and had a 24 bp poly(A) tract (Figure 7). Amino acid sequence comparison showed that the 3′ portion of cDNA encoded a polypeptide with 49 and 35% amino acid sequence identity with the mature portion of human cytochrome c oxidase subunit Vb (COXVb) (Zeviani *et al.*, 1988) and yeast cytochrome oxidase subunit IV (COXIV) (Maarse *et al.*, 1984), respectively (Figure 8). The human *coxVb* is an analogous subunit to yeast *coxIV* (Zeviani *et al.*, 1988). The corresponding gene structure in plant species has not been unraveled to date. Thus, we assigned this cDNA clone as a rice *coxVb* gene based on the amino acid sequence similarity.

The deduced rice COXVb amino acid sequence had an N-terminal extension (>50 amino acids) as well as a conserved COXVb portion, suggesting that the N-terminal region is a presequence for mitochondrial targeting. Human COXVb has a 31 amino acid extension, but the sequence was dissimilar to the presequence of rice COXVb (Figure

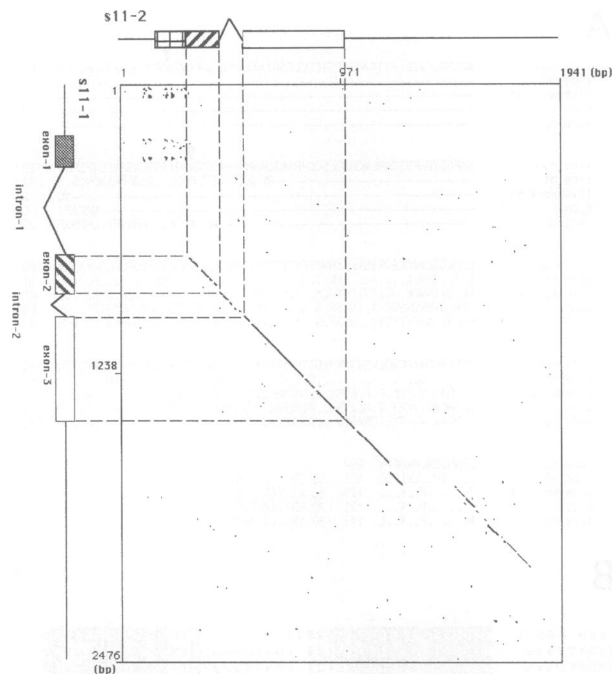

Fig. 5. Dot matrix comparison of the nucleotide sequences of s11-1 and s11-2. A series of 10 nt where a minimum of 9 nt match the other sequences are plotted. Along each axis, the structure of the gene is schematically illustrated. Boxes with different tones indicate exon regions. Broken thin lines indicate intron and intron-related regions. Solid thin lines indicate the 5′ and 3′ flanking regions. The entire nucleotide sequences of s11-1 and s11-2 are registered in DDBJ, EMBL and GenBank under the accession Nos D85382 and D85195 respectively.

```
s11-2 genomic DNA   TATTTCCCCTCCCCACCCTACCCCGTCGTCTCCGTTCGCGCGCATCCAGGTCAAATCAAATCGAGAGAGAGCCCCCATCTCGCCTCGCCTCGCCGCCTCCTCCCTCC-CC   109
                                                                     ***************** ******************     ************ **
coxVb cDNA                                                           ATCAAATCGAGAGAGAG-CCCCATCTCGCCTCGCCT-----CCTCCTCCCTCCTCC   50

CCATTCCCCCGCCGCCGCCGCTGCGAGCGCCGCGAC-----------------------ATGAGAGCACGCCTCCAAACCCTGGCCCCGGCCCTGCGCCGCGCCGCCTCCGCCTCCGCCA   206
******************** * * **********                        *** *  **************************************************
CCATTCCCCCGCCGCCGCCGCCGAGGGCGCCGCGACCTGAGAGCACACGCACGCCCGCCATGTGGCGCCGCCTCCAAACCCTGGCCCCGGCCCTGCGCCGCGCCGCCTCCGCCTCCGCCT   170
                                                          M  W  R  R  L  Q  T  L  A  P  A  L  R  R  A  A  S  A  S  A

CGGGCCCCGCCTCCGCCTCCGCCGCGGGCCCCGCCTCCCTCGCCAGCGCCGCGCCGCTCTCCACGGCGGCCGCCGCCGCCTT-----CCCACCTCCTCAGTCTAAAACCGCTTCACGTATG   322
* *   * ** ** * *    *********************** ************************************      * ****  * *  **     * *   ** *  *
CCGCGGCGGCATCGGGC---GCCGCGGGCCCCGCCTCCCTCGCCCGCGCCGCGCCGCTCTCCACGGCGGCCGCCGCCGCCTTCCGCCGCACCAGCCCTCTCCTGTCCGGGGACAAGCCGG   287
S  A  A  A  S  G     A  A  G  P  A  S  L  A  R  A  A  P  L  S  T  A  A  A  A  A  F  R  R  T  S  P  L  L  S  G  D  K  P
```

Fig. 6. The 5′ portion of s11-2 has high nucleotide sequence similarity with a cDNA encoding the putative signal peptide of rice COXVb protein and its 5′ flanking region. Nucleotide sequence from s11-2 is aligned with the 5′ portion of rice *coxVb* cDNA. Nucleotide sequence identity is indicated by asterisks. Nucleotide sequence numberings are in relation to the 5′-end of cloned fragments at position 1. The region in s11-2 corresponding to exon 2 of s11-1 is shaded.

```
ATCAAATCGA GAGAGAGCCC CATCTCGCCT CGCCTCCTCC TCCCTCCTCC CCATTCCCCC    60

GCCGCCGCCG CCGAGGGCGC CGCGACCTGA GAGCACACGC ACGCCCGCCA TGTGGCGCCG   120
                                                      M  W  R  R
CCTCCAAACC CTGGCCCCGG CCCTGCGCCG CGCCGCCTCC GCCTCCGCCT CCGCGGCGGC   180
 L  Q  T   L  A  P  A   L  R  R   A  A  S   A  S  A  S   A  A  A
ATCGGGCGCC GCGGGCCCCG CCTCCCTCGC CCGCGCCGCG CCGCTCTCCA CGGCGGCCGC   240
 S  G  A   A  G  P  A   S  L  A   R  A  A   P  L  S  T   A  A  A
                                     intron-1
CGCCGCCTTC CGCCGCACCA GCCCTCTCCT GTCCGGGGAC AAGCCGGCGA CGGTGGAGGA   300
 A  A  F   R  R  T  S   P  L  L   S  G  D   K  P  A  T   V  E  D
                                                        intron-2
CGTCATGCCC ATCGCCACGG GGCTGGAGCG CGAGGAGCTC GCGGCGGAGC TCAAGGGGGA   360
 V  M  P   I  A  T  G   L  E  R   E  E  L   A  A  E  L   K  G  E
                                                  intron-3
GAAGCGGTTC GACATGGATC CTCCCGTAGG CCCATTTGGT ACCAAGGAGG CCCCAGCTGT   420
 K  R  F   D  M  D  P   P  V  G   P  F  G   T  K  E  A   P  A  V
                                                        intron-4
CATTGAATCC TACTATAACA AGAGGATAGT TGGTTGTCCT GGTGGTGAAG GAGAGGATGA   480
 I  E  S   Y  Y  N  K   R  I  V   G  C  P   G  G  E  G   E  D  E
GCATGATGTT GTATGGTTCT GGTTGAAAAA AGATGAACCG CATGAGTGCC CAGTCTGTTC   540
 H  D  V   V  W  F  W   L  K  K   D  E  P   H  E  C  P   V  C  S
             intron-5
ACAATATTTT GTGCTTAAGG TCATTGGTGA TGGTGGTGAT CCAGATGGAC ATGATGATGA   600
 Q  Y  F   V  L  K  V   I  G  D   G  G  D   P  D  G  H   D  D  D
CGACGAGCAT CACCACTAAG GACGCCCATG CCTCGTGGTT CTGATAATAA GAATTTGATG   660
 D  E  H   H  H  *
AGGCAGATGA TTTCATTTAA ACCTTTTTAG CACCAGACTG CTCTACGGCG TCCTTGTTTA   720
TGAATGCACA CATTTGGCCA TCAAGCAGGT GGTAGTTTTT TCCTTAGGGA TTTTGTCCAA   780
ACAAAAAACT CATCGAGGCT TGACCTGTGC ATTCCTTCGG AGCATCATCG AATAAAAGAT   840
GACTGCATCA GCTGCGTTTT TCTTGTGTAT CCTAATGGCC ATCTGGTGTG ATGGACCTGT   900
TCTATTCAGA TCATGCAACA GTTTATTTTC AAAAAAAAAA AAAAAAAAAA AAAA         954
```

Fig. 7. Nucleotide sequence of the rice *coxVb* cDNA. Positions of intron sequences identified in the corresponding genomic DNA sequence are marked by open triangles. Intron sequences of intron 1 to intron 5 are 98, 124, 3160, 94 and 147 bp respectively. A possible poly(A) addition signal AATAAA is underlined. The entire genomic *coxVb* gene sequence is registered in DDBJ, EMBL and GenBank under the accession No. D85381.

8). The deduced amino acid sequence of the N-terminal portion of s11-2 had 83% amino acid sequence similarity over 46 amino acids with the putative presequence of rice COXVb. These high nucleotide and amino acid similarities strongly suggest that the 5′ portion sequence of s11-2 shares an evolutionary origin with the rice COXVb presequence.

Structure of the rice coxVb gene and transcript analysis

Complete nucleotide sequence analyses for liverwort and chlorophyte alga mitochondrial genomes have revealed that the *coxVb* gene is absent from their mitochondrial genomes (Oda *et al.*, 1992; Wolff *et al.*, 1994). These results suggest that migration of the *coxVb* gene into the nuclear genome and subsequent sequence loss from the mitochondrial genome took place long before the evolution of flowering plants.

Unexpectedly, the plant *coxVb* gene was picked up from the rice nuclear genome. This occurred because the 5′ portion of the *coxVb* reading frame has a very similar nucleotide sequence to the 5′ portion of s11-2. The nucleotide and deduced amino acid sequences of rice *coxVb* are shown in Figure 7. A consensus polyadenylation signal (AATAAA) is present 84 nt upstream of the poly(A) tail. A genomic *coxVb* DNA was isolated from the rice library using the *coxVb* cDNA as probe and the genomic DNA structure was determined. Five intron sequences of 98, 124, 3160, 94 and 147 bp were identified in the genomic sequence and their respective positions are indicated in Figure 7 (the entire genomic DNA sequence is not shown, but is available in the DDBJ/EMBL/GenBank sequence databank under accession No. D85381). A putative presequence is encoded by the first exon of *coxVb* genomic DNA and the conserved COXVb peptide is encoded by the other five exon sequences. The 5′ flanking sequence of the *coxVb* gene had 86% nucleotide sequence homology with s11-2 over 145 nt.

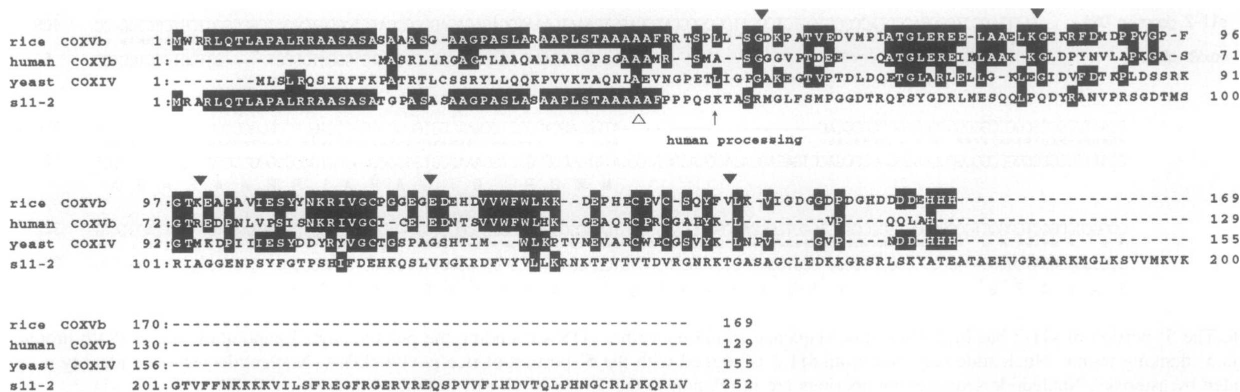

Fig. 8. Alignment of the amino acid sequences deduced from rice *coxVb* cDNA (this study), human *coxVb* (Zeviani *et al.*, 1988), yeast *coxIV* (Maarse *et al.*, 1984) and rice s11-2 (this study). The intron 2-like sequence of s11-2 is not considered for translation. Amino acids identical to rice COXVb are highlighted by inverse contrast. Intron positions of the rice *coxVb* genomic sequence are marked by filled triangles. The processing site of human COXVb is indicated by an arrow. The position corresponding to the intron 1–exon 2 juncture of s11-1 is indicated by an open triangle.

DNA gel blot analysis was carried out to determine the copy number of the *coxVb* gene. Two, two and three major bands together with several minor bands were observed in *Bam*HI-, *Eco*RI- and *Dra*I-digested rice nuclear DNA respectively. No hybridization to rice mitochondrial DNA was detected (Figure 9B). DNA sequence analysis together with physical mapping of the genomic *coxVb* gene indicated a single copy of the *coxVb* gene in the rice nuclear genome. RNA gel blot analysis showed that *coxVb* was transcribed as a single transcript of 1.1 kb (Figure 9C).

Computer analysis showed that the putative presequence of *coxVb* has the potential to fold into an amphiphilic α-helix with charged and hydrophobic amino acid residues distributed on opposite sides, which may be important for receptor recognition (Tamm, 1991). None of amino acid residues in the N-terminal extension are acidic, in either the COXVb or RPS11 presequence.

Discussion

RNA editing in the mitochondrial rps11 transcript

Four RNA editing sites were detected in the transcripts of the mitochondrial *rps11* pseudogene; one site creates a stop codon and the other three are silent changes. None of the four editing events change the predicted amino acid to a more conserved amino acid. This observation is unusual, because editing frequently changes specified amino acids to a more highly conserved residue (reviewed in Walbot, 1991). Because rice has an *rps11* pseudogene in mitochondria and a functional RPS11 protein can be supplied by the nuclear genome, a functional RNA editing machinery for the mitochondrial *rps11* sequence is not necessary and the specificity of RNA editing may be lost. Alternatively, the mitochondrial pseudo-*rps11* may have sequence matches with some guide RNAs used for other mitochondrial genes.

Pathway of rps11 gene migration from mitochondrion to nucleus

The process of organelle to nucleus gene transfer is thought to be ongoing in plant mitochondria (Brennicke *et al.*, 1992). We have used the *rps11* gene to investigate a detailed pathway of the gene transfer event and to study the mechanism of presequence acquisition. A schematic representation for a scenario of mitochondrial *rps11* gene migration to the nuclear genome is shown in Figure 10 and is explained as follows.

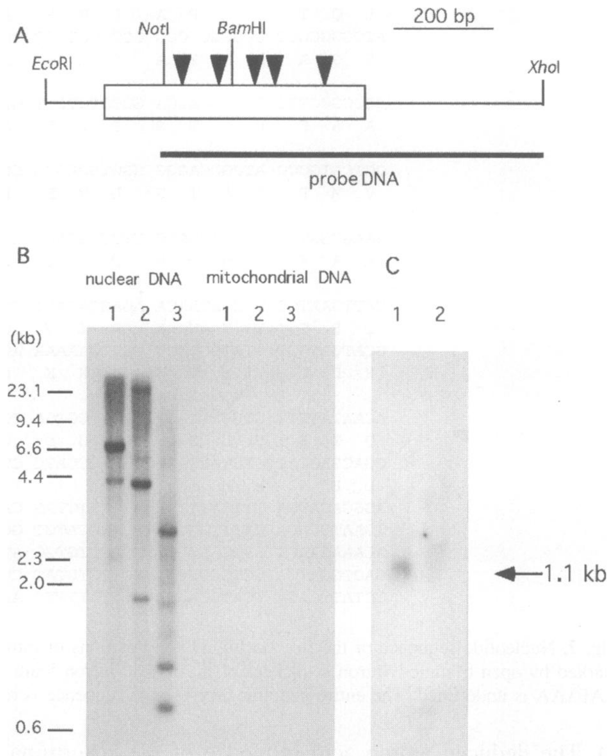

Fig. 9. DNA and RNA gel blot analyses of the rice *coxVb* gene. **(A)** Schematic representation of the *coxVb* cDNA sequence and the region used as probe. Positions of five intron sequences in genomic *coxVb* are shown by filled triangles. **(B)** A DNA gel blot of rice nuclear and mitochondrial DNAs was hybridized with ^{32}P-labeled probe. Restriction enzymes *Bam*HI (lane 1), *Eco*RI (lane 2) and *Dra*I (lane 3) were used for digestion of nuclear and mitochondrial DNA. Molecular size markers are shown on the left. **(C)** An RNA gel blot of poly(A)$^+$ RNA (lane 1) and total RNA (lane 2) was hybridized with the labeled probe. The size of the rice *coxVb* transcript is given on the right.

1. The *rps13 rps11* gene order in *E.coli* (Bedwell *et al.*, 1985) is also conserved in the liverwort mitochondrial genome (Takemura *et al.*, 1992). In the higher plant rice,

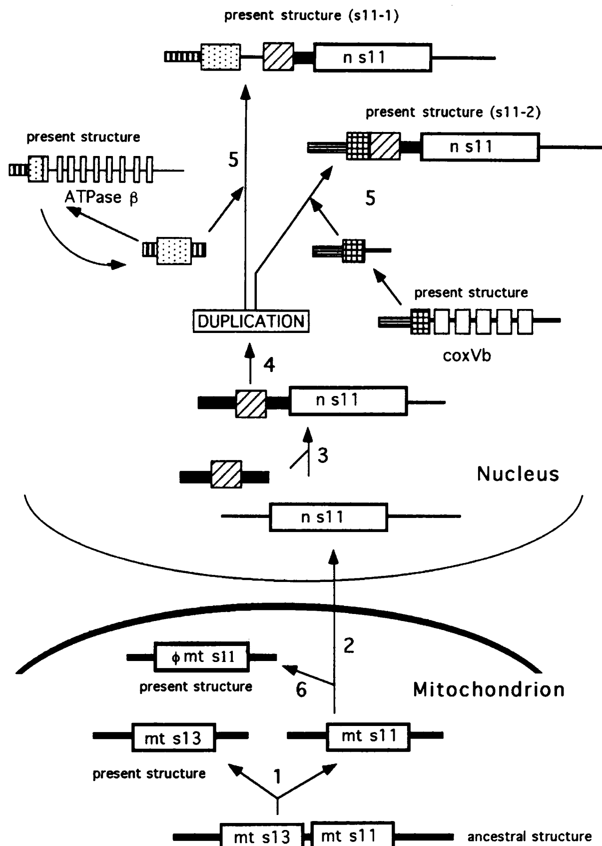

Fig. 10. A proposed model for the gene migration pathway of the *rps11* gene from the mitochondrion to the nucleus and the relationship among conceptual ancestral counterparts. Numbers in the figure correspond to the numbers in the text.

an *rps11* pseudogene was identified downstream of *nad1* exon d and upstream of *rps4*. The mitochondrial *rps13* gene was located >5 kb downstream of the *rps11* sequence in the rice mitochondrial genome, suggesting a gene order rearrangement between *rps13* and *rps11* and a subsequent gain of the necessary transcription signals.

2. The *rps11* sequence migrated from the mitochondrion to the nucleus. Whether the transfer was DNA mediated or RNA (edited RNA) mediated is not clear in the case of *rps11*, because of large nucleotide substitutions between the mitochondrial and the nuclear *rps11* copies, the absence of intron sequences in the mitochondrial *rps11* copy and a lack of information about gene structure (including editing) of *rps11* from other higher plants.

3. The *rps11* sequence that originated in the mitochondrion acquired exon 2 and intron 2 sequences from the nuclear genome. The mechanism for exon 2 and intron 2 sequence integration upstream of the mitochondrion-derived *rps11* sequence and the role of exon 2 are unknown.

4. The exon 2, intron 2 and *rps11* sequence was duplicated in the rice nuclear genome. The fact that both s11-1 and s11-2 sequences contain exon 2 and intron 2 in the same position strongly suggests the occurrence of a duplication in the nuclear genome, rather than independent transfer events from mitochondrion to nucleus. s11-2 has a high extent of nucleotide sequence homology with downstream sequences from the poly(A) addition site of s11-1. In addition, both sequences retain intron 2. The evidence suggests the duplication has taken place via a DNA molecule rather than an RNA molecule as an intermediate.

5. One of the duplicated copies acquired the exon 1 and intron 1 sequence, forming s11-1. The region homologous with the ATPB presequence is encoded only by exon 1 of s11-1. In addition, the 5′ flanking sequences to exon 1 of s11-1 and rice *atpB* (Sakamoto *et al.*, 1992) have 93% nucleotide sequence identity (13 out of 14 nt). The evidence suggests that the RPS11 presequence has a genetical relationship with the ATPB presequences, and the DNA sequence coding presequence and its flanking sequence seems to result from duplication and recombination.

The other duplicated copy obtained a sequence coding for a putative presequence of *coxVb*, forming s11-2. High nucleotide sequence similarity (86% over 145 nt) was observed between 5′ flanking regions of the *coxVb* gene and s11-2. This also strongly suggests involvement of duplication and subsequent recombination of the *coxVb* presequence region in forming s11-2.

Gene migrations of *atpB* and *coxVb* from mitochondrion to nucleus took place long before the evolution of flowering plants. Rice still retains an *rps11* pseudogene in its mitochondrial genome, suggesting that translocation of the *rps11* sequence to the nucleus was a relatively recent event. The above results suggest that the *atpB* and *coxVb* sequences were a template for the 5′ portion of s11-1 and s11-2 respectively. After gain of the presequence, DNA encoding the internal part of the s11-1 presequence may have been deleted or there may be a common ancestral sequence for RPS11 and ATPB presequences. The mechanism of transfer by duplication and recombination of already existing targeting sequences to new genes is, as yet, unexplained.

Other explanations could be that s11-1 was the primitive structure and s11-2 was produced by duplication of the s11-1 sequence. It is less likely that s11-2 was a primitive form and s11-1 was produced by duplication of s11-2 sequence because s11-1 is apparently more homologous to the liverwort *rps11* gene than s11-2, in nucleotide and deduced amino acid sequences.

In any case, it is highly likely that duplication of the *rps11* sequence occurred in the nuclear genome and acquisition of the COXVb and the ATPB-related presequences were independent. In addition, both s11-1 and s11-2 have sequences highly homologous with the 5′ flanking sequences of the *atpB* and the *coxVb* genes respectively. This suggests sequences necessary for transcription and regulation signals may have originated from an ancestral counterpart sequence. We could not detect transcription of s11-2 by two kinds of independent RT–PCR analyses using two sets of specific primer pairs (data not shown). Nucleotide sequence homology of s11-2 extends to ~200 bp upstream of the *coxVb* gene, however, the sequence further upstream is dissimilar. This sequence divergence in the upstream region, as well as sequence differences in the 200 bp region in s11-2, may lack signals necessary for transcription. However, it is possible that s11-2 was functional in the past and became non-functional or that this copy is strictly regulated in a tissue-specific manner.

6. Both the mitochondrial *rps11* and the nuclear *rps11*

sequences are transcribed in rice. The nuclear copy, s11-1, is functional, while the mitochondrial copy is a pseudogene. These results suggest that RPS11 production by the mitochondrial copy was terminated by introduction of the stop codon in the original coding region. This would have to occur after the nuclear *rps11* gene was activated and could supply appropriate protein product to the mitochondrion. The results also show that inactivation of rice mitochondrial *rps11* gene expression was at the translational level rather than by elimination of transcription. This is in contrast to the case of soybean *coxII*, which has an intact ORF in the mitochondrial genome, but is not transcribed in mitochondria (Covello and Gray, 1992). In addition, most legumes have *coxII*-related sequences in both the nucleus and the mitochondrion, although there is no evidence for simultaneous *coxII* transcription in both compartments (Nugent and Palmer, 1993). There might be many switches for the inactivation of mitochondrial gene expression and for the transition of gene expression from the mitochondrion to the nucleus.

Origin of the targeting signal for the gene recently migrated from mitochondria

A number of amino acid sequences for the N-terminal extensions of nuclear-encoded mitochondrial proteins have been analyzed so far. A notable result is that no significant sequence homology has been detected among the primary sequences (Hartl *et al.*, 1989). One example is the bovine mitochondrial ATP synthase proteolipid subunit gene, which has two isoforms that encode identical mature proteins with highly diverged signal sequences (Gay and Walker, 1985). However, the primary structures of mitochondrial presequences do exhibit several common features; they are rich in positively charged residues, lack acidic amino acid residues and fold in amphiphilic α-helices (Hartl *et al.*, 1989).

Baker and Schatz (1987) have reported that mitochondrial precursor proteins could arise as a result of DNA rearrangements that place potential mitochondrial presequences at the N-terminus of existing ORFs. Vassarotti *et al.* (1987) have reported that point mutations in the N-terminus of mature ATPB can generate a functional mitochondrial import sequence.

These cases could represent possible mechanisms for creation of a presequence for a gene newly introduced from mitochondria to the nucleus in plants. However, what we have found in this study is a mechanism different from the previous two. One of the *rps11* sequences in the nucleus acquired a presequence that is related to the presequence of ATPB. The other *rps11* sequence obtained a presequence that is related to the presequence of COXVb. Both presequences are in-frame with the evolutionarily conserved RPS11 sequence. Random recombination or random exon shuffling would be less likely to produce s11-1 and s11-2. How the integrated mitochondrial copy recruited a DNA sequence encoding a mitochondrial targeting presequence remains an important question.

Our most notable findings are that highly conserved presequences are present in different mitochondrial proteins and that a presequence-information was used to isolate a new mitochondrial protein gene. Such common use of one targeting presequence for different proteins may accelerate functional reconstruction of a gene that has migrated from mitochondria. Thus, our results provide a plausible explanation for the origin of the targeting signal of a gene recently transferred from the mitochondrion to the nucleus.

Materials and methods

Plant material and nucleic acid extraction

Etiolated seedlings of rice (*Oryza sativa* L, cv Nipponbare) were used as the plant material. Mitochondrial DNA, mitochondrial RNA and nuclear DNA were prepared as described by Umbeck and Gengenbach (1983), Stern and Newton (1986) and Rivin *et al.* (1982) respectively. Total RNAs were extracted by the method of Logemann *et al.* (1987) and poly(A)$^+$ RNAs were subsequently enriched by Oligotex-dT30<super> according to the manufacturer's instructions (Takara Shuzo, Kyoto, Japan).

Construction of mitochondrial DNA, genomic DNA and cDNA libraries

*Sal*I-digested rice mitochondrial DNA was ligated into the pBluescript vector (Stratagene, CA) and a mitochondrial DNA library was made. A rice cDNA library was constructed in the Uni-ZAP XR cloning vector (Stratagene) with oligo(dT) primer for first strand synthesis. Rice nuclear DNA was digested partially with *Sau*3AI and resultant DNA fragments were size fractionated. The DNA fragments were ligated into the λ DASH II cloning vector (Stratagene) and a rice genomic DNA library was prepared.

Screening of the recombinant library

DNA fragments used as probes were recovered from low melting point agarose after electrophoresis (Sambrook *et al.*, 1989) and probe DNAs were labeled with [α-^{32}P]dCTP using a multilabeling system according to the supplier's instructions (Amersham International, Little Chalfont, UK). Approximately 600 clones of a rice mitochondrial *Sal*I DNA library were screened by colony hybridization with the liverwort *rps11* gene as probe (Takemura *et al.*, 1992). Isolated mitochondrial *rps11* sequence was used as probe for screening for a nuclear *rps11* cDNA on ~5×10^5 clones from the cDNA library. Then the isolated *rps11* cDNA was used as probe for screening against 1×10^6 clones of the genomic library. An *Xho*I–*Not*I fragment of 0.8 kb from s11-2 was used for screening the cDNA library (5×10^5 clones). The *Not*I–*Xho*I fragment from *coxVb* cDNA was used as a probe for screening the genomic library (5×10^5 clones). Colony and plaque hybridizations were performed as described (Sambrook *et al.*, 1989). The final washing condition was 2× SSC/0.1% SDS at 42°C.

DNA sequencing

DNA was sequenced using single-stranded templates and the fluorescent T7 primer according to the manufacturer's instruction (Applied Biosystems, CA) and analyzed with the Applied Biosystems model 373A sequencer.

Computer analysis

Nucleotide and deduced peptide sequences were analyzed by the computer software GENETYX (Software Development Co. Ltd, Tokyo, Japan). The survey of nucleotide and amino acid sequence similarities was performed by the FASTA program (Pearson and Lipman, 1988) and BLAST (Altschul *et al.*, 1990) through the DDBJ/EMBL/GenBank database.

Oligonucleotides

The primers synthesized were: primer 1, 5′-ATTCGTCAACAGAATT-CACATAGG-3′; primer 2, 5′-AAAAGAGTAAGAGCGAATTCATGG-3′. The primers were designed according to the rice mitochondrial *rps11* sequence (this study) to harbor an *Eco*RI recognition site (underlined) for convenient cloning. Positions and orientations of the oligonucleotides are indicated in Figure 1A.

RNA editing analysis

cDNAs were prepared as described by Kubo *et al.* (1996). Subsequent PCR amplification was performed using primers 1 and 2 by denaturation at 92°C for 2 min and 30 cycles of 92°C for 1 min, 58°C for 2 min and 72°C for 2 min, with a final extension at 72°C for 5 min. The resultant fragment was subsequently subcloned and sequenced.

RNA and DNA gel blot analyses

Total RNA (9.6 μg), poly(A)$^+$ RNA (3.9 μg) and mitochondrial RNA (5.0 μg) were denatured at 65°C for 5 min and electrophoresed through a 1.5% agarose gel containing formaldehyde. Transfer of RNA onto a Hybond-N$^+$ membrane (Amersham) and subsequent hybridization was carried out by the conventional method (Sambrook *et al.*, 1989).

Total DNA (2.5 μg) and mitochondrial DNA (0.5 μg) were digested with *Bam*HI, *Eco*RI or *Dra*I and separated on a 0.7% agarose gel and blotted onto Hybond-N$^+$ membrane by the conventional capillary method (Sambrook *et al.*, 1989).

A *Pvu*II–*Xho*I fragment of 0.7 kb from the rice *rps11* cDNA and a *Not*I–*Xho*I fragment of 0.7 kb from the rice *coxVb* cDNA were used as probes for hybridization analyses. The final washing condition for the membrane was 0.1× SSC/0.1% SDS at 42°C.

Nucleotide accession number

The nucleotide sequences reported in this paper will appear in the DDBJ, EMBL and GenBank nucleotide sequence database under accession Nos D85128 (mitochondrial *rps11* pseudogene), D85382 (s11-1 genomic sequence), D85195 (s11-2 genomic sequence) and D85381 (*coxVb* genomic sequence).

Acknowledgements

We thank Dr K.Ohyama for the generous gift of liverwort mitochondrial clones. We are grateful to Drs K.Korth, R.M.Mulligan, M.Nakazono and M.Takeda for helpful discussions. We also thank Drs T.Hino, T.Akihama and M.Nishiguchi for their continuous encouragement. This work was supported in part by grants from the Human Frontier Science Program (RG-437/94) and the Ministry of Agriculture, Forestry and Fisheries to K.K.

References

Abulafia,S. and Breiman,A. (1993) EMBL Data Library, S47350.

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.

Baker,A. and Schatz,G. (1987) Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc. Natl Acad. Sci. USA*, **84**, 3117–3121.

Bedwell,D., Davis,G., Gosink,M., Post,L., Nomura,M., Kestler,H., Zengel,J.M. and Lindahl,L. (1985) Nucleotide sequence of the alpha ribosomal protein operon of *Escherichia coli*. *Nucleic Acids Res.*, **13**, 3891–3904.

Boutry,M. and Chua,N.-H. (1985) A nuclear gene encoding the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia*. *EMBO J.*, **4**, 2159–2165.

Breathnach,R. and Chambon,P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, **50**, 349–383.

Brennicke,A., Grohmann,L., Hiesel,R., Knoop,V. and Schuster,W. (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett.*, **325**, 140–145.

Chaumont,F. and Boutry,M. (1995) Protein import into plant mitochondria. In Levings,C.S. and Vasil,I.K. (eds), *The Molecular Biology of Plant Mitochondria*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 207–235.

Chye,M.-L. and Tan,C.-T. (1992) Isolation and nucleotide sequence of a cDNA clone encoding the beta subunit of mitochondrial ATP synthase from *Hevea brasiliensis*. *Plant Mol. Biol.*, **18**, 611–612.

Covello,P.S. and Gray,M.W. (1992) Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. *EMBO J.*, **11**, 3815–3820.

Gray,M.W. (1989) Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.*, **5**, 25–50.

Gray,M.W. (1992) The endosymbiont hypothesis revisited. *Int. Rev. Cytol.*, **141**, 233–357.

Gay,N.J. and Walker,J.E. (1985) Two genes encoding the bovine mitochondrial ATP synthase proteolipid specify precursors with different import sequences and are expressed in a tissue-specific manner. *EMBO J.*, **4**, 3519–3524.

Grohmann,L., Brennicke,A. and Schuster,W. (1992) The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. *Nucleic Acids Res.*, **20**, 5641–5646.

Hartl,F.-U., Pfanner,N., Nicholson,D.W. and Neupert,W. (1989) Mitochondrial protein import. *Biochim. Biophys. Acta*, **988**, 1–45.

Hiratsuka,J. *et al.* (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.*, **217**, 185–194.

Kubo,N., Ozawa,K., Hino,T. and Kadowaki,K. (1996) A ribosomal protein L2 gene is transcribed, spliced and edited at one site in rice mitochondria. *Plant Mol. Biol.*, **31**, 853–862.

Logemann,J., Schell,J. and Willmitzer,L. (1987) Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.*, **163**, 16–20.

Maarse,A.C., Van Loon,A.P.G.M., Riezman,H., Gregor,I., Schatz,G. and Grivell,L.A. (1984) Subunit IV of yeast cytochrome c oxidase: cloning and nucleotide sequencing of the gene and partial amino acid sequencing of the mature protein. *EMBO J.*, **3**, 2831–2837.

Nugent,J.M. and Palmer,J.D. (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell*, **66**, 473–481.

Nugent,J.M. and Palmer,J.D. (1993) Evolution of gene content and gene organization in flowering plant mitochondrial DNA: a general survey and further studies on *coxII* gene transfer to the nucleus. In Brennicke,A. and Kück,U. (eds), *Plant Mitochondria*. VCH Publishers, Weinheim, Germany, pp.163–170.

Oda,K. *et al.* (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. *J. Mol. Biol.*, **223**, 1–7.

Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

Rivin,C.J., Zimmer,E.A. and Walbot,V. (1982) Isolation of DNA and DNA recombinants from maize. In Seridan,W.F. (ed.), *Maize for Biological Research*. University of North Dakota Press, Grand Forks, ND, pp.161–164.

Sakamoto,M., Shimada,H. and Fujimura,T. (1992) Nucleotide sequence of a cDNA encoding a β subunit of the mitochondrial ATPase from rice (*Oryza sativa*). *Plant Mol. Biol.*, **20**, 171–174.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Stern,D.B. and Newton,K.J. (1986) Isolation of plant mitochondrial RNA. *Methods Enzymol.*, **118**, 488–496.

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.

Tamm,L.K. (1991) Membrane insertion and lateral mobility of synthetic amphiphilic signal peptides in lipid model membranes. *Biochim. Biophys. Acta*, **1071**, 123–148.

Terpstra,P., Zanders,E. and Butow,R.A. (1979) The association of Var1 with the 38 S mitochondrial ribosomal subunit in yeast. *J. Biol. Chem.*, **254**, 12653–12661.

Umbeck,P.F. and Gengenbach,B.G. (1983) Reversion of male-sterile T-cytoplasm maize to male fertility in tissue culture. *Crop Sci.*, **23**, 584–588.

Vassarotti,A., Stroud,R. and Douglas,M. (1987) Independent mutations at the amino terminus of a protein act as surrogate signals for mitochondrial import. *EMBO J.*, **6**, 705–711.

Walbot,V. (1991) RNA editing fixes problems in plant mitochondrial transcripts. *Trends Genet.*, **7**, 37–39.

Winning,B.M., Bathgate,B., Purdue,P.E. and Leaver,C.J. (1990) Nucleotide sequence of a cDNA encoding the beta subunit of the mitochondrial ATP synthase from *Zea mays*. *Nucleic Acids Res.*, **18**, 5885.

Wischmann,C. and Schuster,W. (1995) Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. *FEBS Lett.*, **374**, 152–156.

Wolff,G., Plante,I., Lang,B.F., Kück,U. and Burger,G. (1994) Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. *J. Mol. Biol.*, **237**, 75–86.

Zeviani,M., Sakoda,S., Sherbany,A.A., Nakase,H., Rizzuto,R., Samitt,C.E., DiMauro,S. and Schon,E.A. (1988) Sequence of cDNAs encoding subunit Vb of human and bovine cytochrome *c* oxidase. *Gene*, **65**, 1–11.

Received on June 10, 1996; revised on August 13, 1996