The EMBO Journal vol.12 no.4 pp.1437–1446, 1993

Processing followed by complete editing of an altered mitochondrial *atp6* RNA restores fertility of cytoplasmic male sterile rice

Mari Iwabuchi, Junko Kyozuka and Ko Shimamoto

Plantech Research Institute, 1000 Kamoshida, Midori-ku, Yokohama 227, Japan

Communicated by J.H.Weil

Two *atp6* genes were found in the mitochondrial genome of cytoplasmic male sterile (CMS) rice carrying the [*cms-bo*] cytoplasm. One (N-*atp6*) was identical to the normal cytoplasmic gene, while the second (B-*atp6*) was identified as a candidate CMS gene by Southern analysis of the mitochondrial genome of CMS cybrid rice. The coding sequence of B-*atp6* was identical to the normal N-*atp6* gene but its 3′-flanking sequence was different starting at 49 bases downstream from the stop codon. Northern analysis showed that B-*atp6* is transcribed into a 2.0 kb RNA in the absence of the *Rf-1* gene, whereas two discontinuous RNAs, of ~1.5 and 0.45 kb, were detected in the presence of the *Rf-1* gene. Determination of the 3′ and 5′ ends of these RNAs suggested that the two discontinuous RNAs were generated from the 2.0 kb RNA by RNA processing at sites within the B-*atp6*-specific sequences by the action of the *Rf-1* gene. Sequence analysis of cDNA clones derived from the N-*atp6* RNA and the processed and unprocessed RNAs of B-*atp6* indicated that the processed B-*atp6* RNAs were edited as efficiently as the N-*atp6*, whereas unedited and partially edited RNAs were detected among unprocessed RNAs. RNA processing by *Rf-1* thus influences the sequential post-transcriptional editing of the B-*atp6* RNAs. Because the unprocessed RNAs of B-*atp6* are possibly translated into altered polypeptides, our results suggest that interaction of RNA processing and editing plays a role in controlling CMS expression and the restoration of fertility in rice.

Key words: *atp6*/cytoplasmic male sterility/rice/RNA editing/RNA processing

Introduction

Cytoplasmic male sterility (CMS) is a widespread phenomenon in the plant kingdom (reviewed in Hanson and Conde, 1985; Newton, 1988; Levings and Brown, 1989). CMS plants have normal female fertility but fail to shed functional pollen, and the male sterility is maternally inherited. CMS is caused by an incompatibility between nucleus and cytoplasm and has been used to produce hybrid seed from crop plants, because it eliminates the need for hand emasculation. Nuclear-encoded genes which restore fertility (*Rf* genes) are necessary to obtain seeds by self-pollination from F_1 plants.

Mitochondrial genes have been found to be associated with CMS cytoplasms in maize, sunflower, *Petunia*, sorghum, and radish (Boeshoer *et al.*, 1985; Bailey-Serres *et al.*, 1986; Dewey *et al.*, 1986; Crouzillat *et al.*, 1987; Young and Hanson, 1987; Siculella and Palmer, 1988; Makaroff *et al.*, 1989; Rasmussen and Hanson, 1989; Levings, 1990; Bonhomme *et al.*, 1991). In maize, *Petunia* and sorghum a chimeric mitochondrial gene which produces a novel protein has been found in CMS plant mitochondria (Bailey-Serres *et al.*, 1986; Dewey *et al.*, 1987; Nivison and Hanson, 1989). In the mitochondria of CMS sunflower a new open reading frame (ORF) is co-transcribed with the *atp1* gene (Köhler *et al.*, 1991) and the protein supposed to be encoded by this ORF is uniquely present (Laver *et al.*, 1991). Among these examples only the T-*urf13* associated with maize *cms-T* cytoplasm has been strongly correlated with the CMS phenotype based on the analysis of a number of mutants which are toxin resistant and fertile (Levings and Brown, 1989). Although novel proteins uniquely present in the CMS mitochondria or altered gene expression caused by mitochondrial genome rearrangements are thought to be responsible for CMS, the molecular mechanism underlying CMS expression remains to be elucidated.

Nuclear-encoded factors required for normal mitochondrial gene expression have been described for several mitochondrial genes in yeast (reviewed in Grivell, 1989). They regulate the expression of mitochondrial genes at both the transcriptional and post-transcriptional levels. It has been shown that the size and the number of plant mitochondrial transcripts are also influenced by the nuclear background (Cooper *et al.*, 1990; Gupta and Abbott, 1991), including the nuclear *Rf* genes (Bailey-Serres *et al.*, 1986; Dewey *et al.*, 1986; Kennell *et al.*, 1987; Kennell and Pring, 1989; Makaroff *et al.*, 1989; Köhler *et al.*, 1991; Pruitt and Hanson, 1991) suggesting that *Rf* genes may be involved in transcriptional or post-transcriptional regulation of CMS gene expression. It is not, however, understood how *Rf* genes change the mitochondrial gene expression and restore fertility.

The analysis of plant mitochondrial genomes is usually complex because of the multipartite genome organization and frequent intramolecular recombination (Pring and Lonsdale, 1985). Mitochondrial genome organization and transcripts are known to be altered by nuclear backgrounds (Escote-Carlson *et al.*, 1990; Gupta and Abbott, 1991). Furthermore, the pattern of plant mitochondrial transcripts is complex due to multiple sites of transcription initiation and post-transcriptional processing (Mulligan *et al.*, 1988). This makes it difficult to identify genes responsible for the CMS phenotype by the simple comparison of normal plants and CMS plants. Many structural and/or transcriptional variations that are not directly related to CMS expression can be found in CMS plants. Thus, somatic hybrid plants, which often contain recombinant mitochondrial genomes, are valuable tools for the analysis of genes involved in CMS expression (Boeshore *et al.*, 1985; Bonhomme *et al.*, 1991).

A number of CMS systems have been genetically defined

© Oxford University Press

by their distinct responses to different restorer genes in rice (Virmani and Shinjyo, 1988). The [*cms-bo*] cytoplasm, which is derived from an indica rice, Chinsurah boro II, is one of these and its fertility is gametophytically restored by a single, nuclear-encoded *Rf-1* gene (Shinjo, 1984). In [*cms-bo*]*Rf-1*/*rf-1* plants 50% of the pollen is aborted. We previously generated a CMS cybrid by asymmetric cell fusion between male fertile japonica rice, Nipponbare and Chinsurah boro II (Kyozuka *et al.*, 1989). The CMS trait of the cybrid was shown to derive from the [*cms-bo*] of Chinsurah boro II and the mitochondrial genome of the cybrid was a recombinant between the parents. Furthermore, our study indicated that a small portion of the [*cms-bo*] mitochondrial genome was transferred to the fertile Nipponbare by cell fusion resulting in the CMS cybrid (Kyozuka *et al.*, 1989). Here we report the analysis of the mitochondrial genome of this CMS cybrid. We found that a second *atp6* gene, B-*atp6*, which originated from Chinsurah boro II, was uniquely associated with the CMS cytoplasm. The predicted amino acid sequence of this additional *atp6* gene is identical to that of the normal type. We found that the presence of the *Rf-1* gene influences the processing and editing of the B-*atp6* RNA. Based on this observation we propose a novel model for the mechanisms of CMS and for the restoration of fertility by the *Rf-1* gene in rice.

Results

B-atp6 is uniquely associated with the CMS phenotype

The nuclear and cytoplasmic genotypes and phenotypes of the plant materials used in this study are listed in Table I. To search for a mtDNA region that correlates with CMS we performed Southern analysis of the CMS cybrid [referred to as C(r) in this paper] which carries a recombinant mitochondrial genome of the fusion parents, Nipponbare carrying normal cytoplasm [referred to as N(r)] and Chinsurah boro II carrying [*cms-bo*] cytoplasm [referred to as B(R)], using 10 cloned mtDNA fragments containing 13 genes as probes (Figure 1). Hybridization patterns of the parental lines were virtually identical with seven probes examined (Figure 1A). For the *atp1* probe an additional faint band was observed in N(r); however, because neither C(r) nor B(R) carried this fragment (Figure 1A) this could not be related to the CMS phenotype. When the *cox2, cob, rrn18* and *atp6* genes were used as probes, N(r) and B(R) exhibited different hybridization patterns (Figure 1B and C). The hybridization pattern of C(r) with three of these four probes, namely *cox2, cob* and *rrn18*, was the same as that of N(r) suggesting that they are not related to the CMS trait, but *atp6* was found to be uniquely associated with the CMS cytoplasm (Figure 1C). MtDNA isolated from both C(r) and B(R) carried the 0.9 kb band in addition to the 1.3 kb band found in N(r) which carries the normal *atp6* gene (N-*atp6*). Based on these observations, we identified the second *atp6* gene (B-*atp6*), present in mitochondria of B(R) and C(r), as a candidate for the CMS gene.

The nucleotide sequence organization of N-atp6 and B-atp6

We isolated N-*atp6* and B-*atp6* genes from the three lines N(r), B(R) and C(r). The physical maps of N-*atp6* genes isolated from the three lines were identical, and those of B-*atp6* genes isolated from B(R) and C(r) were also identical (Figure 2). We then sequenced the two *atp6* genes isolated from C(r) and compared the nucleotide sequences. The nucleotide sequence from −713 to 1062 was completely identical in the two *atp6* genes; this region includes the *atp6* coding region deduced from the genomic sequence (Figure 3A and B). Divergence between N-*atp6* and B-*atp6* was, however, found in the 3′ region that starts at the nucleotide 49 bp downstream from the stop codon (Figure 3A and B).

Sequence analysis also revealed that the 3′ region of the B-*atp6* gene could be generated by multiple recombinations involving at least three mitochondrial gene regions, *cox2*, $tRNA^{fMet}$ and *cox1*. As shown in Figure 3B, the proximal region of the B-*atp6*-specific sequence contains ~120 bp that are homologous to the 5′ region of the *cox2* gene (Figure 3B, underline 1) that partially overlaps with the sequence which has a homology with the downstream region of the $tRNA^{fMet}$ gene (Figure 3B, underline 2, unpublished data). Southern analysis with the $tRNA^{fMet}$ gene probe showed that the hybridization pattern of C(r) was similar to that of N(r) (data not shown), indicating that the $tRNA^{fMet}$ region was not involved in CMS expression. Moreover, a stretch of 80 bp having homology with the *cox1* coding region is present ~20 bp downstream from the $tRNA^{fMet}$ region (Figure 3B, underline 3). Thus, the 3′ region of B-*atp6* appeared to be generated by complex intermolecular recombinations of the mitochondrial genome. In addition, two sets of inverted repeats (Figure 3B, arrows above the sequence) were also found in this region, and were probably generated by these recombination events.

We performed Southern analysis using the cloned *atp6* genes as probes with the materials used in this study (Table I). All the cybrid lines with different nuclear backgrounds [C(r), CT(R) and CT(r)] contained the two *atp6* genes (Figure 4). The intensity of the 2.1 kb band relative to the 1.6 kb band was different between the plants with different

Table I. The plant materials used in this study

	Genotype	Nuclear background	*atp6* gene	Phenotype
N(r)	[+]*rf-1*/*rf-1*	Nipponbare (N)	N-*atp6*	fertile
B(R)	[*cms-bo*]*Rf-1*/*Rf-1*	Chinsurah boroII	N-*atp6*, B-*atp6*	fertile
C(r)	[*cms-bo*]*rf-1*/*rf-1*	Nipponbare	N-*atp6*, B-*atp6*	sterile
CT(r)	[*cms-bo*]*rf-1*/*rf-1*	N/Taichung 65	N-*atp6*, B-*atp6*	sterile
CT(R)	[*cms-bo*]*Rf-1*/*rf-1*	N/Taichung 65	N-*atp6*, B-*atp6*	fertile
Bm(R)	[*cms-bo*]*Rf-1*/*Rf-1*	Chinsurah boroII	B-*atp6*	–

MtDNA and mtRNA used in this study were isolated from suspension culture cells of the plants listed. CT(R) and CT(r) are isogenic lines except the *Rf-1* gene. Bm(R) ([*cms-bo*]*Rf-1*/*Rf-1*) is a mutant derivative of suspension culture line originated from B(R). The symbols used for male-sterile cytoplasm and fertility-restoring gene conform to the suggestion by the committee on gene symbolization in rice (Virmani and Shinjyo, 1988).

nuclear backgrounds. However, because CT(r) and CT(R), which carry the same nuclear background except for the *Rf-1* gene, exhibited identical hybridization patterns, it was concluded that the structure of the *atp6* genes was not changed by the presence or absence of the *Rf-1* gene. The observed stoichiometric difference of the two bands between different plants (panels 1, 2 and 4 in Figure 4) could possibly have arisen due to the multipartite nature of the plant mitochondrial genome (Palmer and Shields, 1984; Iwahashi *et al.*, 1992; Yamato *et al.*, 1992). A mutant cell line, Bm(R), carrying only B-*atp6* arose in cultured B(R) lines in which N-*atp6* was deleted during the culture period of 2 years (Figure 4).

The B-atp6 RNA is regulated by the Rf-1 gene

Mitochondrial RNAs (mtRNAs) were hybridized with single-stranded probes covering various regions of the *atp6* genes

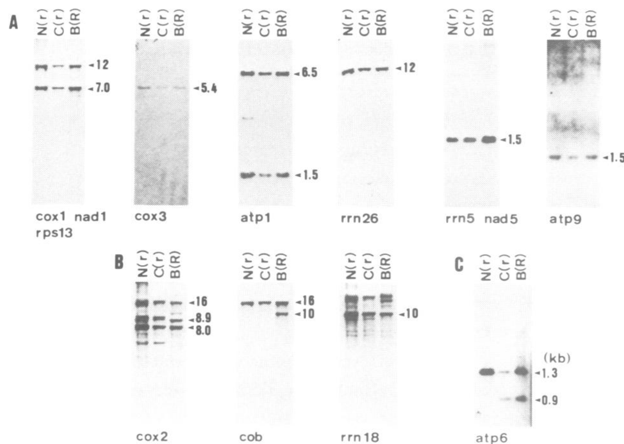

Fig. 1. Analysis of mitochondrial genome organization. Southern blots were hybridized with 10 cloned mtDNA fragments containing 13 genes. The mtDNAs were digested with *Taq*I for the *atp6* probe, with *Eco*RI for the *atp9* probe and with *Bam*HI for the others. Plant genotypes are as follows: N(r), [+]*rf-1*/*rf-1*; B(R), [*cms-bo*]*Rf-1*/*Rf-1*; C(r), [*cms-bo*]*rf-1*/*rf-1*. (**A**) The hybridization patterns are identical in the three lines. (**B**) The hybridization pattern of C(r) is identical to those of N(r). (**C**) The hybridization pattern of C(r) is identical to that of B(R).

(Figure 5). No transcript was detected with the sense-strand probes of the *atp6* gene (data not shown). Analysis of the Bm(R) line carrying the B-*atp6* gene alone (Figure 4) indicated that the B-*atp6* gene was transcribed into an ~1.5 kb RNA similar in size to the N-*atp6* RNA in the presence of the *Rf-1* gene (Figure 5, probe 1). A 2.0 kb RNA was detected in the CT(r) male-sterile line in addition to the 1.5 kb RNA (Figure 5, probe 1). The 2.0 kb RNA was shown to derive from the B-*atp6* by hybridization with probe 2 and 3 (Figure 5). On the other hand, a 0.45 kb RNA derived from the 3′ region of the B-*atp6* was detected in the Bm(R) line and CT(R) line carrying the *Rf-1* gene (Figure 5, probe 2 and 3). No transcripts were found with probes 4 and 5 (data not shown). Taken together, the B-*atp6* generates a continuous 2.0 kb RNA in the absence of the *Rf-1* gene, but two discontinuous RNAs (1.5 and 0.45 kb) in the presence of the *Rf-1* gene. Although the origin of the 1.5 kb RNA detected in CT(r) line was not precisely identified, it is most likely to be derived from N-*atp6* since there is little possibility that it is not transcribed in this nuclear background. Based on Southern and Northern analyses, it is most likely that the 2.0 kb RNA of B-*atp6* is processed into the 1.5 and 0.45 kb RNAs at a site within the B-*atp6*-specific region by the action of the *Rf-1* gene product (Figure 2).

The B-atp6 RNA is processed by the Rf-1 gene

To study the processing phenomenon, we first determined the 3′ termini of the 1.5 kb RNA and the 5′ termini of the 0.45 kb RNA derived from B-*atp6*. For the 3′ mapping of the RNAs, the mtRNAs were polyadenylated, then the cDNAs were synthesized using an oligo(dT) primer. The cDNAs including the 3′ termini were amplified by the polymerase chain reaction (PCR, see Materials and methods). The amplified cDNAs had the expected sizes of ~0.6, 0.3, 0.2 and 0.15 kb (Figure 6A) using primers a, b, c and d (Figure 2), respectively. Although two kinds of cDNAs were expected to be amplified by PCR using primer a in the CT(r) line (Figure 2); one with the 3′ terminus of the 1.5 kb N-*atp6* RNA, and the other with the 2.0 kb B-*atp6* RNA, the size of the cDNA species derived from the 2.0 kb RNA was not clear on the gel (Figure 6A). We were,

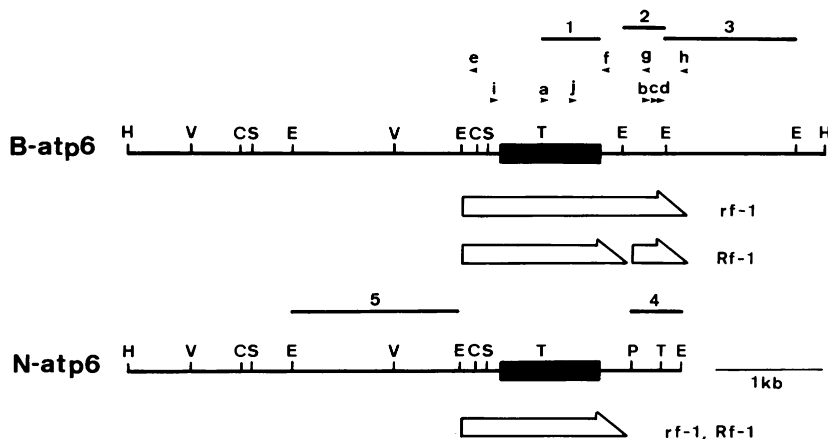

Fig. 2. Schematic diagram of *atp6* region. The N-*atp6* and B-*atp6* genes were isolated from N(r), B(R) and C(r). The coding region is shown as a black box. The probes used for nucleic acid hybridization are indicated by bars (1–5). The sizes of probes 1–5 are 0.55, 0.4, 1.3, 0.45 and 1.7 kb, respectively. The primers used for sequencing, PCR amplification and primer extension experiment are shown as arrowheads (a–j). The *atp6* RNAs are represented by arrows. The transcript of B-*atp6* is altered, whereas that of N-*atp6* is not changed by the presence or absence of the *Rf-1* gene (see the results of Northern analysis in Figure 5 and the 5′- and 3′-mapping of the *atp6* RNAs in Figures 6 and 7). The restriction sites are shown as E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; T, *Eco*T22I; S, *Sal*I; C, *Cla*I; and V, *Eco*RV.

A

```
-713 GGTAGGAAATTCATCCCAACCAATAGCAGTGAGTTAGCTATAACTACTAGGGATCCTAGAGTAAAGTAGTAGGGCTTTCTGAGGAGTAAGCCTAATTCCGTTAATGCAGAAAGATCTTTC
-593 CCAGTCCCAGATAGAAGAGTACCAACCAAAGAAGCAAAAGCAAGACCTCACTAGTAAGGAAGGCACTTGCTGCCGGAGTTCAACAGGCAAATATAAGAAAAGAAGTCCTGTTCACTTCAT
-473 CATCTGTGGGTTGTACTGCTTGAAGGTTCTTCTGAGGGGTAGAATTTGAATTCCTTCTTTGCTTGTGAGATAACCATTTCCAGAAACTCATATATAGAGAGCGGGTATCGGTGAAAATGG
-353 ATCTTACCAGGAGTGGCATTGAATAGGCAGGCTCTGGGATGTAATCTCACTCAAGAGGTCATTTGTTGGCCCCGCCTTCACTAGACTAGAGTTTTAGGATAGGTTGGGGAACCTATACGT
-233 CAAGCCCCTACGAAGATTGAGAAAAATCGATGCACATAAGCCATCCGAAACCAGTATTGGAAAGTGTTCAGTTTCGTTTTCCATTCTGAAATGTTCATAGTAGTATAGTATGTTTTCCGT
-113 TGGGTCGACGCCATGTGATCGCTACTAAAGATAGAGTTTCCTTGGAAAAACCGAGGCCAGTTGAGATCAGTCTCCCTTTCTAGGAGCAGAGCTTAAAAAGATGGGAAATTCCAATGAATT
                                                                                                                         M  N  F
   8 TCGATCACAATCATGTGGTAATAATGGGTTTGAATCAGAGAGACTCGATCTGGAAACTCCTCAATGATTATAACGTGAACTCGTTGAAGAGAAGGAGACAAGCAGAAATAGACGCTTTTT
      D  H  N  H  V  V  I  M  G  L  N  Q  R  D  S  I  W  K  L  L  N  D  Y  N  V  N  S  L  K  R  R  R  Q  A  E  I  D  A  F  F
 128 TTGAACCATTTGAGAGGGCGCAGCGTATCCGTTTCAATAACTGGCAGAACGGAATAGAGTTGTTAGATGGGGCTGAATGGAGGAACGGCGATATAGTTATCCCTGGAGGCGGCGGACCAG
      E  P  F  E  R  A  Q  R  I  R  F  N  N  W  Q  N  G  I  E  L  L  D  G  A  E  W  R  N  G  D  I  V  I  P  G  G  G  G  P  V
 248 TAATTTCAAGCCCCTTGGATCAATTTTTCATTGATCCATTATTTGGTCTTGATATGGGTAACTTTTATTTATCATTCACAAATGAATCCTTGTCTATGGCGGTAACTGTCGTTTTGGTGC
      I  S  S  P  L  D  Q  F  F  I  D  P  L  F  G  L  D  M  G  N  F  Y  L  S  F  T  N  E  S  L  S  M  A  V  T  V  V  L  V  P
 368 CATCTTTATTTGGAGTTGTTACGAAAAAGGGCGGGGGAAAGTCAGTGCCAAATGCATGGCAATCCTTGGTAGAGCTTATTTATGATTTCGTGCTGAACCTGGTAAACGAACAAATAGGTG
      S  L  F  G  V  V  T  K  K  G  G  G  K  S  V  P  N  A  W  Q  S  L  V  E  L  I  Y  D  F  V  L  N  L  V  N  E  Q  I  G  G
 488 GAAATGTTAAACAAAAGTTTTTCCCTCGCATCTCGGTCACTTTTACTTTTTCGTTATTTCGTAATCCCCAGGGTATGATACCCTTTAGCTTCACAGTGACAAGTCATTTTCTCATTACTT
      N  V  K  Q  K  F  F  P  R  I  S  V  T  F  T  F  S  L  F  R  N  P  Q  G  M  I  P  F  S  F  T  V  T  S  H  F  L  I  T  L
 608 TGGCTCTTTCATTTTCCATTTTTATAGGCATTACGATCGTTGGATTTCAAAGACATGGGCTTCATTTTTTTAGCTTCTTATTACCAGCGGGAGTCCCACTGCCATTAGCACCTTTTTTAG
      A  L  S  F  S  I  F  I  G  I  T  I  V  G  F  Q  R  H  G  L  H  F  F  S  F  L  L  P  A  G  V  P  L  P  L  A  P  F  L  V
 728 TACTCCTTGAGCTAATCTCTCATTGTTTTCGTGCATTAAGCTCAGGAATACGTTTATTTGCTAATATGATGGCCGGTCATAGTTCAGTAAAGATTTTAAGTGGGTTCGCTTGGACTATGC
      L  L  E  L  I  S  H  C  F  R  A  L  S  S  G  I  R  L  F  A  N  M  M  A  G  H  S  S  V  K  I  L  S  G  F  A  W  T  M  L
 848 TATTTCTGAATAATATTTTCTATTTCATAGGAGATCTTGGTCCCTTATTTATAGTTCTAGCATTAACCGGTCTGGAATTAGGTGTAGCTATATTACAAGCTCATGTTTCTACGATCTCAA
      F  L  N  N  I  F  Y  F  I  G  D  L  G  P  L  F  I  V  L  A  L  T  G  L  E  L  G  V  A  I  L  Q  A  H  V  S  T  I  S  I
 968 TTTGTATTTACTTGAATGATGCTATAAATCTCCATCAAAATGAGTAATTTCATAATTGAATAAAAACGAGGAGCCGAAGATTTTAGGGGGCGGGAGAAAAGGTCAAGTCTAAGCGACATA
      C  I  Y  L  N  D  A  I  N  L  H  Q  N  E  *
1088 TGGCACGAAAAGGAAATCCAATTTCGGTAATCCGAAGCCCACCTTACCTTAATCCATTTTCTTGAGGATCTTGCATTCGTTTGTTCTCTGCCGACCTATTGTTTCCATCACTCCTGTTGC
1208 TTCTTCCTTTCTCTATCCAACCCGAGCTCTATCGCCCATGGCTGAACTCTTCTCTGCAGCCTCATCGCCAACAAAAGGTCCACATCCTTCAATCGGCATGTACTATGGAATGGCTTTAAT
1328 GTTTCTATATCCGGAAGTAGGGAATCTCGTGCTTGCATATCTAAATCTAAGTTTTGAGACAGACCTTTCATGGGTTCAAAGAAAAGAAGAGTACGAGTGGGTGATGTGATTGACTAATAG
1448 GAAAGTGATAGGAGTTGTGATCGGGTTTCTTTCTTTTTTTGCCAAGGATATTCTTATTCACTGCTCAGAGAAAATGAAAAAAAACAAACTAGGATTCTTCCGAACTCAACTCCGTGGCCA
1568 AAACCATCTTCTCACTCTGACCCCCACATATCAGATCCCAGATGCATAGGAAAAGCGGTATCAAGAATAGTAGTATAAAGAAAGATAGTACAGTACTCTTGTAAAT
```

B

```
-713 GGTAGGAAATTCATCCCAACCAATAGCAGTGAGTTAGCTATAACTACTAGGGATCCTAGAGTAAAGTAGTAGGGCTTTCTGAGGAGTAAGCCTAATTCCGTTAATGCAGAAAGATCTTTC
-593 CCAGTCCCAGATAGAAGAGTACCAACCAAAGAAGCAAAAGCAAGACCTCACTAGTAAGGAAGGCACTTGCTGCCGGAGTTCAACAGGCAAATATAAGAAAAGAAGTCCTGTTCACTTCAT
-473 CATCTGTGGGTTGTACTGCTTGAAGGTTCTTCTGAGGGGTAGAATTTGAATTCCTTCTTTGCTTGTGAGATAACCATTTCCAGAAACTCATATATAGAGAGCGGGTATCGGTGAAAATGG
-353 ATCTTACCAGGAGTGGCATTGAATAGGCAGGCTCTGGGATGTAATCTCACTCAAGAGGTCATTTGTTGGCCCCGCCTTCACTAGACTAGAGTTTTAGGATAGGTTGGGGAACCTATACGT
-233 CAAGCCCCTACGAAGATTGAGAAAAATCGATGCACATAAGCCATCCGAAACCAGTATTGGAAAGTGTTCAGTTTCGTTTTCCATTCTGAAATGTTCATAGTAGTATAGTATGTTTTCCGT
-113 TGGGTCGACGCCATGTGATCGCTACTAAAGATAGAGTTTCCTTGGAAAAACCGAGGCCAGTTGAGATCAGTCTCCCTTTCTAGGAGCAGAGCTTAAAAAGATGGGAAATTCCAATGAATT
                                                                                                                         M  N  F
   8 TCGATCACAATCATGTGGTAATAATGGGTTTGAATCAGAGAGACTCGATCTGGAAACTCCTCAATGATTATAACGTGAACTCGTTGAAGAGAAGGAGACAAGCAGAAATAGACGCTTTTT
      D  H  N  H  V  V  I  M  G  L  N  Q  R  D  S  I  W  K  L  L  N  D  Y  N  V  N  S  L  K  R  R  R  Q  A  E  I  D  A  F  F
 128 TTGAACCATTTGAGAGGGCGCAGCGTATCCGTTTCAATAACTGGCAGAACGGAATAGAGTTGTTAGATGGGGCTGAATGGAGGAACGGCGATATAGTTATCCCTGGAGGCGGCGGACCAG
      E  P  F  E  R  A  Q  R  I  R  F  N  N  W  Q  N  G  I  E  L  L  D  G  A  E  W  R  N  G  D  I  V  I  P  G  G  G  G  P  V
 248 TAATTTCAAGCCCCTTGGATCAATTTTTCATTGATCCATTATTTGGTCTTGATATGGGTAACTTTTATTTATCATTCACAAATGAATCCTTGTCTATGGCGGTAACTGTCGTTTTGGTGC
      I  S  S  P  L  D  Q  F  F  I  D  P  L  F  G  L  D  M  G  N  F  Y  L  S  F  T  N  E  S  L  S  M  A  V  T  V  V  L  V  P
 368 CATCTTTATTTGGAGTTGTTACGAAAAAGGGCGGGGGAAAGTCAGTGCCAAATGCATGGCAATCCTTGGTAGAGCTTATTTATGATTTCGTGCTGAACCTGGTAAACGAACAAATAGGTG
      S  L  F  G  V  V  T  K  K  G  G  G  K  S  V  P  N  A  W  Q  S  L  V  E  L  I  Y  D  F  V  L  N  L  V  N  E  Q  I  G  G
 488 GAAATGTTAAACAAAAGTTTTTCCCTCGCATCTCGGTCACTTTTACTTTTTCGTTATTTCGTAATCCCCAGGGTATGATACCCTTTAGCTTCACAGTGACAAGTCATTTTCTCATTACTT
      N  V  K  Q  K  F  F  P  R  I  S  V  T  F  T  F  S  L  F  R  N  P  Q  G  M  I  P  F  S  F  T  V  T  S  H  F  L  I  T  L
 608 TGGCTCTTTCATTTTCCATTTTTATAGGCATTACGATCGTTGGATTTCAAAGACATGGGCTTCATTTTTTTAGCTTCTTATTACCAGCGGGAGTCCCACTGCCATTAGCACCTTTTTTAG
      A  L  S  F  S  I  F  I  G  I  T  I  V  G  F  Q  R  H  G  L  H  F  F  S  F  L  L  P  A  G  V  P  L  P  L  A  P  F  L  V
 728 TACTCCTTGAGCTAATCTCTCATTGTTTTCGTGCATTAAGCTCAGGAATACGTTTATTTGCTAATATGATGGCCGGTCATAGTTCAGTAAAGATTTTAAGTGGGTTCGCTTGGACTATGC
      L  L  E  L  I  S  H  C  F  R  A  L  S  S  G  I  R  L  F  A  N  M  M  A  G  H  S  S  V  K  I  L  S  G  F  A  W  T  M  L
 848 TATTTCTGAATAATATTTTCTATTTCATAGGAGATCTTGGTCCCTTATTTATAGTTCTAGCATTAACCGGTCTGGAATTAGGTGTAGCTATATTACAAGCTCATGTTTCTACGATCTCAA
      F  L  N  N  I  F  Y  F  I  G  D  L  G  P  L  F  I  V  L  A  L  T  G  L  E  L  G  V  A  I  L  Q  A  H  V  S  T  I  S  I
 968 TTTGTATTTACTTGAATGATGCTATAAATCTCCATCAAAATGAGTAATTTCATAATTGAATAAAAACGAGGAGCCGAAGATTTTAGGGGGCGGGACAAACGCGGAAGTGTATTGCGTTAC
      C  I  Y  L  N  D  A  I  N  L  H  Q  N  E  *                                                     1
1088 AAAAAATGACAACTAGCATTTGTTTTTTCATTTCATGTTCGAATTCGTTTTTCGTTGGAAAAACCAACGCCGACCCCAAACAAGTCTCTCCAATATAAGGAGAGCGGAGCTTAAAAATAT
                                                                                                                       2
1208 TATTTTATTGTGCTATGGCAAATCTGGTCCGATGGCTCTTCTCCACTACCCGAGGGACTAACGGTCTTCCATATTTCATCTTCGGTGTCGTTGTAGGAGGCGCCTGTTGTTTGCTTTGCT
                       3
1328 AAAGTATCAGGCCCCTCTGTACGACCCGGCTTTAATGGAAAAAATCATAGATCATAATATAAAAGCCGGGCACCCTATAGAGGTTGACTATTCGTGGTGGGGCACCTCTATTCGTGTAGT
1448 CTTTCCTAAGTAAGAAAGACAGGACAGTGGTGGTTTGCTCATACTTTCATTACAAAACCATACTATGGAATTCTGGTGGACCAGGCAAATATCCCACCCCTACAAGGTGGGAGGGCGGCC
1568 GGGAACCCGTGGTGGCGGTTCCTCCCGTGATGATCGACTAATGGATAAGTCGCAATGATCGACTAATGGATAAGTCGCATTGAGATATTGGTTCGAATCCAATTCATTACGAGCTC
```

Fig. 3. The nucleotide and deduced amino acid sequences of N-*atp6* and B-*atp6*. (**A**) N-*atp6*. The 5′ terminus (▼) and the 3′ termini (▲) of N-*atp6* RNA. (**B**) B-*atp6*. The 5′ termini of the 1.5 and 2.0 kb RNAs (▼), the 3′ termini of 1.5 kb RNA obtained from Bm(R) (▲), the 5′ terminus of the 0.45 kb RNA (▽) and the 3′ termini of the 0.45 kb RNA and 2.0 kb RNA (△). The 3′ regions which are different in N-*atp6* (A) and B-*atp6* (B) are boxed. The deduced amino acid sequence is shown under the genomic nucleotide sequence. The underlined nucleotides represent the region homologous to the 5′ region of *cox2* (underline 1, Kao *et al.*, 1984), the downstream sequence of $tRNA^{fMet}$ (underline 2, data not shown) and the N-terminus of *cox1* (underline 3, Kadowaki *et al.*, 1989). The nucleotides which form the inverted repeats are indicated by arrows above the nucleotides.

however, able to detect the cDNAs derived from the 2.0 kb RNA by using sequence-specific primers (see Figure 8), suggesting that the cDNA derived from the 1.5 kb RNA could be amplified more efficiently than that of the 2.0 kb RNA with the non-specific adaptor-$(dT)_{17}$ primer. The nucleotide at the 3′ terminal position (Figure 3) was

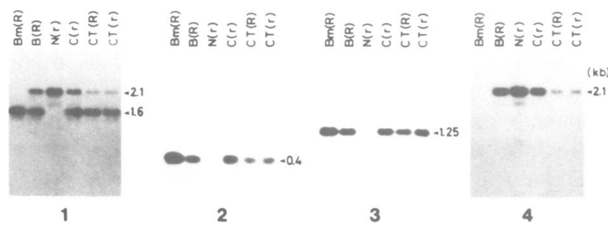

Fig. 4. Hybridization of mtDNAs with N-*atp6* and B-*atp6*. The mtDNAs were digested wtih *Eco*RI and hybridized with probes (1−4, under photos) represented in Figure 2. Plant genotypes are as follows: N(r), [+]*rf-1*/*rf-1*; B(R), [*cms-bo*]*Rf-1*/*Rf-1*; Bm(R), [*cms-bo*]*Rf-1*/*Rf-1*; C(r), [*cms-bo*]*rf-1*/*rf-1*; CT(r), [*cms-bo*]*rf-1*/*rf-1*; CT(R), [*cms-bo*]*Rf-1*/*rf-1*.

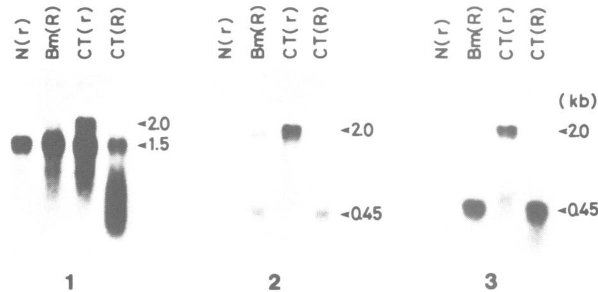

Fig. 5. Hybridization of mtRNAs with N-*atp6* and B-*atp6*. Hybridization probes (1−3, under photos) are shown in Figure 2. Plant genotypes are as follows: N(r), [+]*rf-1*/*rf-1*; Bm(R), [*cms-bo*]*Rf-1*/*Rf-1*; CT(r), [*cms-bo*]*rf-1*/*rf-1*; CT(R), [*cms-bo*]*Rf-1*/*rf-1*.

determined by sequencing five cDNA clones for each RNA; some examples are shown in Figure 6B. The 3′ termini of the 1.5 kb RNAs derived from B-*atp6* in the Bm(R) line were determined by sequencing cDNAs amplified with primer a (Figure 2). Two cDNA clones were mapped at the same position as the N-*atp6* RNA, but three other cDNA clones mapped at three different positions within the 3′ region of the B-*atp6* (Figure 3B). The 5′ terminus of the 0.45 kb RNA as determined by the primer extension experiment (Figure 7B) was found at a site downstream from the determined 3′ termini of the 1.5 kb RNA of the B-*atp6* (Figure 3B).

Next we examined whether the 5′ termini of the 1.5 and the 2.0 kb RNAs, and also the 3′ termini of the 2.0 and the 0.45 kb RNAs are identical. We found that the 5′ termini of the 1.5 kb RNAs are the same by the analysis of N(r) and Bm(R) lines (Figure 7A). In addition, primer extension experiments indicated that the 5′ terminus of the 2.0 kb RNA is the same as that of the 1.5 kb RNAs because a single band was found in CT(r) at the same position as in N(r) and Bm(R) (Figure 7A). The 3′ termini of the 2.0 and 0.45 kb RNAs were determined by sequencing the cDNAs derived from CT(r) and CT(R) which were amplified by PCR using primers c and d, respectively and the adaptor-$(dT)_{17}$ (Figure 6A). All the cDNAs carried the identical 3′ terminus (Figures 3B and 6B). The results of 5′ and 3′ mapping of the transcripts support the notion that the 1.5 and 0.45 kb RNAs are generated from the 2.0 kb RNA by RNA processing within the B-*atp6*-specific region.

To examine this possibility further, we next tried by PCR to detect the unprocessed 2.0 kb RNA of the B-*atp6* which was not detected by Northern analysis in the presence of

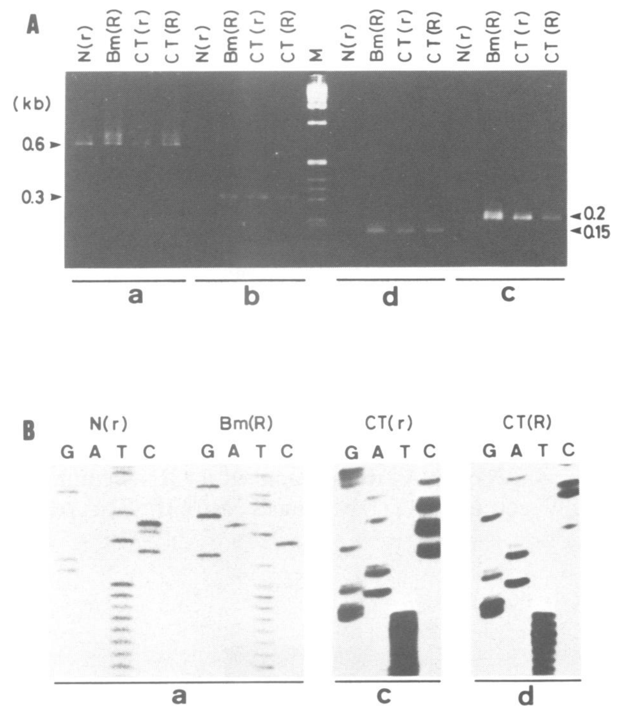

Fig. 6. Determination of 3′ termini of *atp6* RNAs. (**A**) PCR amplification of cDNAs containing 3′ termini. The cDNAs were synthesized using polyadenylated mtRNAs and were amplified with the *atp6*-specific primers (a−d, Figure 2) and the adaptor-$(dT)_{17}$ primer. (**B**) Representative sequencing gels of the cDNAs including the 3′ termini. The cDNAs amplified by using the *atp6*-specific primers (a, c and d, under photos) and the adaptor-$(dT)_{17}$ primer were ligated into the BluescriptII and five cDNA clones for each RNA were sequenced. These representatives are included in the 3′ termini of *atp6* RNAs indicated by upward arrowheads in Figure 3 (the non-coding strands of the *atp6* can be read from the gels).

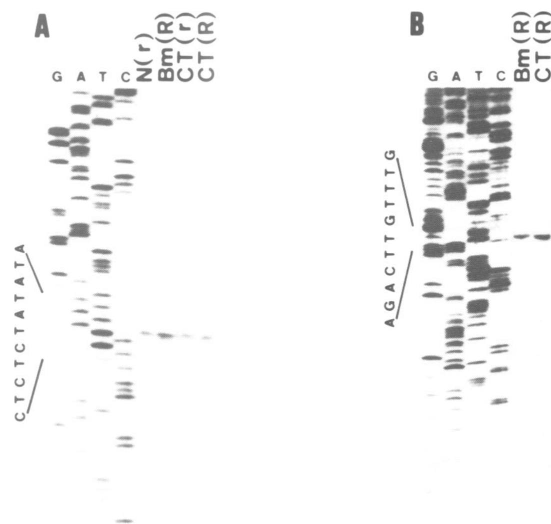

Fig. 7. Mapping of the 5′ termini by primer extension. (**A**) Results with primer e (Figure 2) for mapping the 5′ termini of the 1.5 and the 2.0 kb RNAs. (**B**) Results with primer g (Figure 2) for mapping the 5′ terminus of the 0.45 kb RNA.

the *Rf-1* gene. In this experiment we first amplified cDNAs with a sequence-specific primer (i, see Figure 3) and a non-specific, adaptor-$(dT)_{17}$ primer (data not shown), then

the PCR product was further amplified by using primers i and h (Figure 8). In all lines carrying [*cms-bo*] mitochondria, the 1.6 kb band, the size of which was as expected from the B-*atp6* sequence, was detected, indicating that the unprocessed RNA was indeed present in plants carrying the *Rf-1* gene although it was not detected by Northern analysis. In addition to the 1.6 kb band, a 1.1 kb fragment was visible in all lines examined. This band was derived from N-*atp6* or processed B-*atp6* RNA during the first round of amplification with primer i and the adaptor-$(dT)_{17}$ primer. The result is consistent with our hypothesis that the B-*atp6* is first transcribed into 2.0 kb RNA, then processed into 1.5 and 0.45 kb RNAs by the action of the *Rf-1* gene.

The Rf-1 gene indirectly normalizes RNA editing through processing

The amplified cDNAs were examined for RNA editing by partially sequencing 43 cDNA clones (Table II). RNA editing was found at nine positions; all the changes observed were transitions from C to U (Figure 9B). No RNA editing was observed in the 3′ non-coding region. We found six amino acid changes within the analyzed sequence. The editing at positions 5 and 6 does not result in amino acid exchange. RNA editing at position 1 creates the stop codon instead of glutamine encoded by the genomic sequence (Figure 9B). The editing events at position 1, 2, 3, 4, 7, 8 and 9 are the

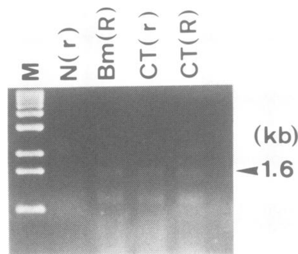

Fig. 8. Detection of the unprocessed B-*atp6* RNA by PCR. The cDNAs were first amplified by PCR with primer i and the adaptor-$(dT)_{17}$ primer, and next with primers i and h, and electrophoresed as described in Materials and methods. Plant genotypes are as follows: N(r), [+]*rf-1*/*rf-1*; Bm(R), [*cms-bo*]*Rf-1*/*Rf-1*; CT(r), [*cms-bo*]*rf-1*/*rf-1*; CT(R), [*cms-bo*]*Rf-1*/*Rf-1*.

Table II. The editing patterns of the *atp6* RNAs

cDNA	Source	Position									No. of cDNAs
		1	2	3	4	5	6	7	8	9	
N-*atp6*	N(r)	+	+	+	+	−	−	+	+	+	7
		+	+	+	+	+	−	+	+	+	1
		+	+	+	+	−	+	+	+	+	2
B-*atp6*											
processed	Bm(R)	+	+	+	+	−	−	+	+	+	9
		+	+	+	+	−	−	+	−	+	1
unprocessed	CT(r)	+	+	+	+	−	−	+	+	+	4
		+	+	+	+	−	+	+	+	+	1
		+	+	−	+	−	−	+	−	+	2
		−	−	−	+	−	−	−	+	+	1
		−	−	−	−	−	−	+	−	+	1
		−	−	−	−	−	−	−	−	−	2
unprocessed	Bm(R)	+	+	+	+	−	−	+	+	+	5
		+	+	+	+	−	+	+	+	+	1
		+	+	+	+	−	−	−	−	−	1
		−	−	−	−	−	−	+	−	+	1
		+	−	−	−	−	−	−	−	−	1
		−	−	−	−	−	−	−	−	−	2

For each *atp6* RNA the cDNAs were obtained by two independent amplifications with different sets of primers (see Figure 2). For the N-*atp6* RNA and the processed B-*atp6* RNA, a sequence-specific primer (a or j) and the adaptor-$(dT)_{17}$ primer were used for amplification. For the unprocessed B-*atp6* RNAs, amplification was performed with two sets of primers (a−g and j−g) by the procedure described in Materials and methods.

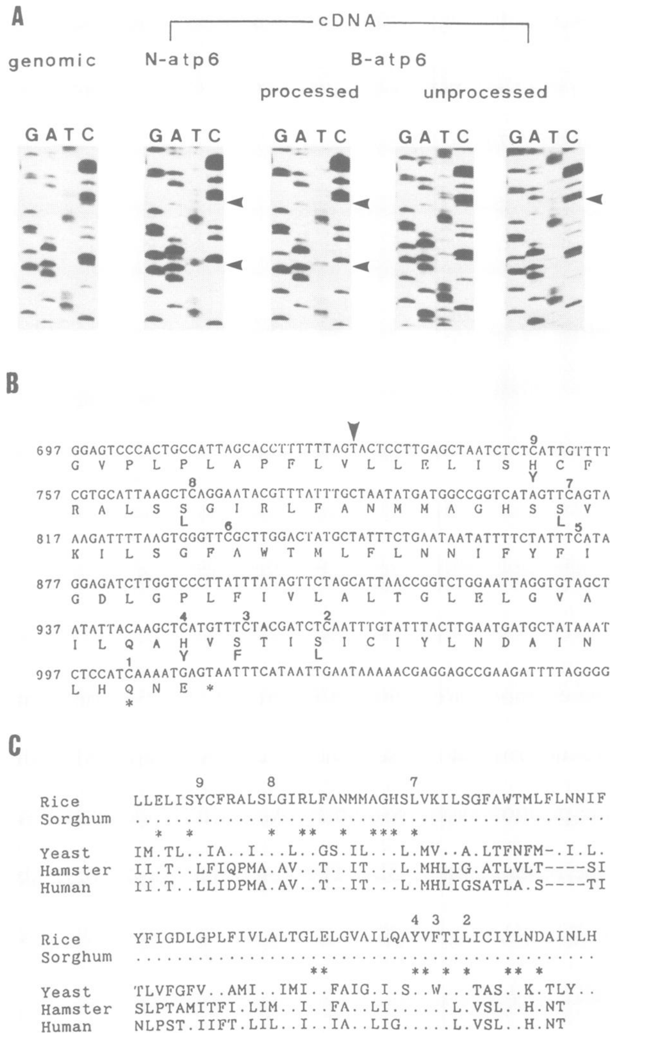

Fig. 9. The RNA editing of the *atp6* RNAs. (**A**) Representative sequencing gels of cDNAs of *atp6* RNAs. The cDNAs derived from the N-*atp6* RNA, processed B-*atp6* RNA and unprocessed B-*atp6* RNAs were amplified by PCR, ligated into BluescriptII and sequenced as described in Materials and methods. The coding strand of *atp6* can be read from the gels. The edited nucleotides (transition from C to T shown in the sequencing gels) are indicated by arrowheads: lower arrowheads, position 8; upper arrowheads, position 9 (see Figure 9B). (**B**) Positions of editing detected in the *atp6* RNA. The cDNA clones for each RNA were partially sequenced within the region downstream from the arrowhead. The edited positions are numbered above the genomic nucleotide sequence. The amino acids deduced from the edited nucleotide sequence are represented below those deduced from genomic sequence. (**C**) Comparison of amino acid sequences of ATP6 proteins from various organisms. The amino acid sequence deduced from the edited nucleotide sequence of rice (downstream of the arrowhead in B) is compared with the corresponding sequences of sorghum (Kempken *et al.*, 1991), yeast (Macino and Tzagoloff, 1980), Chinese hamster (Breen *et al.*, 1986) and human (Anderson *et al.*, 1981). The amino acids conserved in all five organisms are indicated by asterisks.

same as those found in sorghum *atp6* and result in an identical polypeptide in the two species (Figure 9C). RNA editing causes insertion of three leucine residues instead of serine which makes the polypeptide more hydrophobic (Figure 9B). Amino acid changes occur at positions highly conserved in other non-plant species (Figure 9C). To compare the degree of editing between processed and unprocessed B-*atp6* RNAs, a total of 10–12 cDNA clones for each RNA species obtained from two independent amplifications with different sets of primers were sequenced. The cDNAs derived from the processed 1.5 kb RNA of B-*atp6* were as efficiently edited as those of N-*atp6*, except for one position in one of the cDNA clones (Table II). On the other hand, various patterns of editing, from unedited to completely edited, were found among the cDNAs derived from the unprocessed 2.0 kb RNA of B-*atp6* (Figure 9A, Table II). No editing polarity was shown by partially edited species. The processed RNA of B-*atp6* encodes polypeptides with amino acid sequences almost identical to that generated by the normal 1.5 kb RNA of the N-*atp6* (Table II), whereas the unprocessed RNA species should produce altered amino acid sequences. Comparison of degrees of editing in unprocessed RNA in the presence and absence of the *Rf-1* gene indicated that the *Rf-1* gene does not itself change the editing efficiency but indirectly normalizes editing through RNA processing (Table II).

Discussion

The second *atp6* gene, B-*atp6* , which is uniquely present in the CMS mitochondria, was identified as a candidate CMS gene based on Southern analysis of the mitochondrial genome of the CMS cybrid by using 10 cloned mtDNA fragments containing 13 genes as probes. The coding sequence of B-*atp6* deduced from the genomic DNA sequence was identical to that of N-*atp6*. However, its expression was post-transcriptionally regulated by the nuclear *Rf-1* gene. The nucleotide sequence of the N-*atp6* gene present in normal rice has previously been reported by Kadowaki *et al.* (1990); however, part of the sequence was found to be different in this study: the three bases at positions 151, 152 and 156 or 157 (Figure 3) are missing in the sequence of Kadowaki *et al.*, resulting in tyrosine (position 51) instead of arginine (position 51) and isoleucine (position 52). In addition, the transcription initiation sites determined by Kadowaki *et al.* (1990) are also different from ours. However, the initiation site determined in this study is supported by the following observations. First, we determined the 3′ termini and calculated sizes of N-*atp6* RNAs by 5′ and 3′ mapping as 1419–1515 nucleotides; this is consistent with the result of Northern analyses. Second, the 5′ termini determined in this study contain a consensus motif (TCATA/CG/TANAAA) similar to those found in the 5′ ends of *atp9, cob-1* and dicistronic *orf25/cox3* mRNAs of rice (E.Kaleikau, personal communication).

Kadowaki *et al.* (1990) reported that [*cms-bo*] mitochondria contain an extra *atp6* gene, *urf-rmc*, which is a truncated gene of the normal *atp6* (N-*atp6*), and they discussed its relationship to the CMS trait (Kadowaki *et al.*, 1990). Available evidence suggests, however, that *urf-rmc* is only present in some [*cms-bo*] cytoplasms and not related to the CMS phenotype. First, the 4.3 kb *Eco*RI fragment containing the *urf-rmc* was recovered in [*cms-bo*] mitochondrial genome with Taichung 65 nuclear background (Kadowaki *et al.*, 1990); however, the same authors did not detect this fragment in a CMS rice carrying [*cms-bo*] mitochondrial genome wtih Reimei nuclear background in the previous study (Kadowaki and Harada, 1989). Similarly in the present study, we were not able to detect this fragment in the [*cms-bo*] mitochondrial genome we analyzed (Figure 4, probe 1). Because Reimei [*cms-bo*] and our cybrid are both male sterile despite the lack of *urf-rmc*, this gene is not likely to be involved in CMS. In contrast to the study of Kadowaki *et al.* (1990), our analyses indicated that the B-*atp6* present in the [*cms-bo*] mitochondria, which was identified as a normal gene cloned in pOSB1139 (Kadowaki *et al.*, 1990), is related to the CMS trait based on the analysis of the CMS plants generated by cell fusion.

Results of Northern analyses suggested that the 2.0 kb RNA derived from B-*atp6* was processed into 1.5 and 0.45 kb RNAs by the action of the *Rf-1* gene. This hypothesis is supported by the results of 5′ and 3′ mapping of *atp6* RNAs and by detection of the unprocessed RNA by PCR in the presence of the *Rf-1* gene. The result is consistent with the notion that the observed alterations of *atp6* RNAs are due not to creation of novel transcriptional units but to processing. However, we could not rigorously exclude the possibility that alternative initiation and termination of the transcripts create such changes in B-*atp6* RNAs within the processed region. The processed 0.45 kb RNA contains a small ORF (35 amino acids) homologous to the N-terminus of the *coxI* gene. It is not known whether this ORF is functional and it is still possible that the 0.45 kb RNA influences the restoration of fertility.

The 3′ termini of the processed 1.5 kb RNA of B-*atp6* were mapped at various positions within the B-*atp6*-specific 3′ region and at similar positions to those of N-*atp6* RNA (Figure 3A and B). Both the 5′ terminus of the 0.45 kb RNA and the 3′ terminus of the longest transcript among the processed 1.5 kb RNAs of B-*atp6* mapped within the stretch including the two inverted repeat sequences present in the 3′ region of the B-*atp6* gene (Figure 3B). Therefore, it can be envisioned that the B-*atp6* RNA is first endonucleolytically processed at a specific site within the two inverted repeats and then further processed in the 5′ direction by exonucleolytic reaction possibly due to lack of protection of the 3′ terminus. The stretch of the two inverted repeats appears to have been generated by recombination between the 5′ region of *cox2* and the 3′ region of the $tRNA^{fMet}$ gene (Figure 3B). It has been demonstrated that wheat mitochondrial extracts contain 5′- and 3′-endonucleolytic activities able to precisely process tRNA precursors *in vitro* (Hanic-Joyce and Gray, 1990). Similar activities have been detected in mammalian cells (Apirion, 1983) and yeast (Hollingsworth and Martin, 1986; Chen and Martin, 1988).

Recently, a novel post-transcriptional modification, RNA editing, has been reported in several organisms (Benne *et al.*, 1986; Powell *et al.*, 1987; Thomas *et al.*, 1988; Simpson and Shaw, 1989), including plant mitochondria (Covello and Gray, 1989; Gualberto *et al.*, 1989; Hiesel *et al.*, 1989) and chloroplasts (Hoch *et al.*, 1991). It has been shown that in the *cox2* genes of maize and *Petunia*, polysomal RNAs and spliced RNAs were almost completely edited, while various degrees of editing were detected in unspliced RNAs (Sutton *et al.*, 1991; Yang and Mulligan, 1991). Gualberto *et al.* (1991) demonstrated that RNA editing correlates with RNA

maturation of the transcription unit that contains the *nad3* and *rps12* genes and two unidentified ORFs in wheat. Our results also demonstrate that structural changes in RNAs influence the efficiency of editing. The processed B-*atp6* RNA is edited as efficiently as N-*atp6* RNA and encodes a protein with a normal amino acid sequence whereas the unprocessed RNA is not efficiently edited, presumably generating altered proteins if they are translated (Table II). The unprocessed B-*atp6* RNAs from the *rf-1* and *Rf-1* plants are edited to a similar degree (Table II), indicating that the *Rf-1* gene does not directly influence editing. Northern analyses show that the unprocessed RNA is not detectable in the presence of the *Rf-1* gene, while in the absence of the *Rf-1* gene no processing of the B-*atp6* RNA takes place, therefore a large amount of unedited B-*atp6* RNA must be present in the mitochondria of CMS plants.

It has been demonstrated that all or most of the RNAs are fully edited in mitochondria (Bégu *et al.*, 1990; Lamattina and Grienenberger, 1991) and that the major form of edited RNA is translated into proteins in wheat *atp9* (Bégu *et al.*, 1990). It is not known, however, whether partially edited and unedited RNAs, which are present in small amounts and detected by PCR amplification, are translated into polypeptides. A large amount of unedited *atp6* RNA present in the CMS plants could theoretically be translated into proteins because the 5′ termini of both N-*atp6* and B-*atp6* RNAs are identical. The amino acid sequence deduced from edited nucleotide sequence is completely identical to that of sorghum *atp6* (Kempken *et al.*, 1991) and amino acid substitutions were observed at positions highly conserved in eukaryotes (Figure 9C). The analyzed sequence is therefore a functional part of the protein and RNA editing

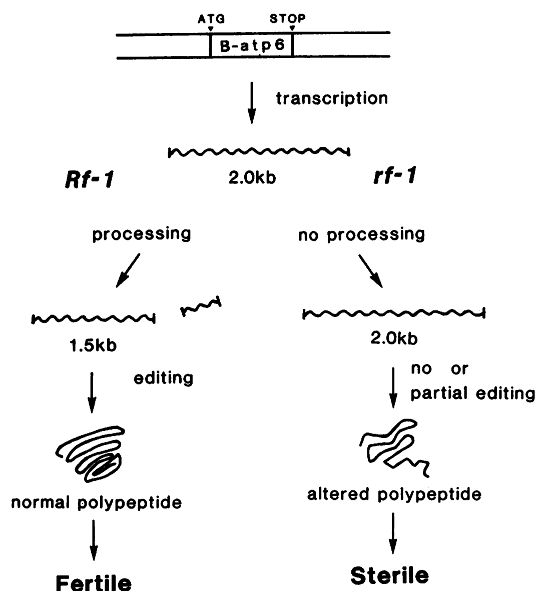

Fig. 10. Proposed model of B-*atp6* gene regulation. The second *atp6* gene, B-*atp6*, which is uniquely present in [*cms-bo*] mitochondria, is first transcribed into a 2.0 kb RNA. In the presence of the *Rf-1* gene which may encode an RNA processing enzyme, the 2.0 kb RNA is processed into 1.5 kb and 0.45 kb RNAs. The processed RNA is efficiently edited and translated into the normal polypeptide. In contrast, in the absence of the *Rf-1* gene, the 2.0 kb RNA is not processed. The unprocessed RNA is not efficiently edited and is translated into an altered polypeptide. A large amount of altered ATP6 polypeptides may cause CMS by competing with normal ATP6 protein to reduce ATPase activity.

appears to be essential for production of the functional ATP6 protein. Three of the six cases of amino acid changes are serine to leucine. As leucine is a hydrophobic residue, the ATP6 protein translated from unedited or partially edited RNAs may be different from the normal ATP6 protein in function and structure. A large amount of altered ATP6 polypeptides may result in CMS expression by either competing with normal ATP6 protein or inhibiting the assembly of functional ATPase proteins. A model to explain CMS and fertility restoration by the *Rf-1* gene is summarized in Figure 10. Chu *et al.* (1972) described that pollen mother cells of [*cms-bo*]*rf-1*/*rf-1* plants are normally produced and that meiosis is normal. The microspores, however, fail to develop and disintegrate after release from tetrads. We propose that the altered ATP6 polypeptides which do not have normal function interfere with sequential maturation of the pollen by decreasing the ATPase activity. It is not known, however, why the altered protein affects only pollen production. Both the 13 kDa protein associated with *cms-T* maize (Dewey *et al.*, 1987) and the *Petunia* 25 kDa protein (Nivison and Hanson, 1989) are known to be expressed in non-reproductive tissues, but only affect pollen production. Our hypothesis that CMS is caused by synthesis of the altered ATP6 protein could be experimentally tested by targeting the altered *atp6* polypeptide generated by the unedited *atp6* RNA and the normal polypeptide produced from the edited *atp6* RNA into mitochondria of normal rice plants by transformation and subsequent examination of their phenotypes, because rice transformation (Shimamoto *et al.*, 1989) and targeting polypeptides into plant mitochondria *in vivo* (Boutry *et al.*, 1987) are possible.

In this study we demonstrated that RNA processing influences sequential post-transcriptional editing. The nuclear *Rf-1* gene, which could code for an RNA processing enzyme, is shown to promote the production of functional protein post-transcriptionally. Furthermore, we also showed a novel post-transcriptional regulation of mitochondrial gene expression in which interaction of RNA processing and editing plays a role in controlling CMS expression and the restoration of fertility in plants. Although this type of regulation was uncovered through the analysis of rice CMS expression, it is possible that alteration of RNA processing in conjunction with editing may be a means to regulate expression of other plant mitochondrial genes.

Materials and methods

Isolation of mitochondrial DNA and RNA

Nucleic acids were extracted from isolated mitochondria of suspension culture cells. MtDNA was extracted essentially by the method of Kyozuka *et al.* (1989). MtRNA was extracted by the method of Chomczynski and Sacchi (1987).

Cloning and sequencing of atp6 genes

MtDNA was digested with restriction endonucleases and ligated into the BluescriptII plasmid vector (Stratagene). Mitochondrial genomic clones were screened with the *Oenothera atp6* gene (Schuster and Brennicke, 1987b). Sequencing was performed by the chain-termination method of Sanger *et al.* (1977). Nucleic and amino acid sequence analyses were performed by the computer programs furnished by GENETYX (Software Development) accessing the National Institute of Health (GenBank) and EMBL DNA sequence libraries.

Gel electrophoresis and nucleic acid hybridization

The DNA fragments were electrophoresed on 0.8% agarose gels in TBE buffer and transferred to nylon membrane (Reed and Mann, 1985). RNA was heat-denatured and electrophoresed on 1.2% agarose gels containing

6% formaldehyde in MOPS buffer [20 mM MOPS/5 mM sodium acetate (pH 7.0)/1 mM EDTA]. RNA was transferred to nylon membrane in 20 × SSPE [3 M NaCl/0.2 M sodium phosphate (pH 7.2)/20 mM EDTA]. Double-stranded DNA was labeled with [α-^{32}P]dCTP (NEN, 3000 Ci/mM) using a random prime DNA labeling kit (Amersham). Single-stranded probe was obtained by transcription of the nucleotide sequence cloned into the BluescriptII by using T7 or T3 RNA polymerase with [α-^{32}P]CTP (NEN, 3000 Ci/mM). Nucleic acid hybridization was peformed under the conditions described by Maniatis *et al.* (1982). DNA and RNA ladder markers (BRL) were used for estimation of the sizes of nucleic acids.

MtDNA genes used as hybridization probes are as follows: *cox1*, *rps13* and *nad1* (*Oenothera*: Schuster and Brennicke, 1987a), *atp6* (*Oenothera*: Schuster and Brennicke, 1987b), *cox3* (*Oenothera*: Hiesel *et al.*, 1987), *atp1* and *atp9* (pea: Morikami and Nakamura, 1987), *rrn26* (pea: Morikami,A. and Nakamura,K., unpublished), *rrn5* and *nad5* (*Oenothera*: Brennicke *et al.*, 1985), *rrn18* (pea: Morikami,A. and Nakamura,K., unpublished), *cox2* (maize: Fox and Leaver, 1981), *cob* (maize: Dawson *et al.*, 1984).

Oligonucleotide primers

The primers used for PCR amplification and primer extension experiment are as follows: adaptor-(dT)$_{17}$, 5′-GACTCGAGTCGACAAGCTTTTTTT-TTTTTTTTTTT-3′; primer a, 5′-GTGCCAAATGCATGGCAATCC-TTG-3′; primer b, 5′-AAAGTATCAGGCCCCTCTGTACGA-3′; primer c, 5′-GACTATTCGTGGTGGGGTACCTCTA-3′; primer d, 5′-TTTCCT-AGGTAAGAAAGACAGGACA-3′; primer e, 5′-CCTCTTGAGTGA-GATTAC-3′; primer f, 5′-CAGAGGGGCCTGATACTT-3′; primer g, 5′-CCTCGTTTTTATTCAATT-3′; primer h, 5′-GGGAGGATCCGCCAC-CACGG-3′; primer i, 5′-TCTCCCTTTCTAGAAGCAGAGC-3′; and primer j, 5′-TCCCACTGCAGCATTAGCAC-3′. Primers a−j are *atp6*-specific primers. The underlined nucleotides were replaced or inserted to create restriction sites for cloning.

Primer extension analysis

The experiment was performed as described by Geliebter (1988) with some modifications. Labeled oligonucleotide (1 pM) was annealed with 10 μg of mtRNA in 12 μl of 250 mM KCl, 10 mM Tris−HCl (pH 8.3) for 3 min at 80°C and 45 min at 50°C. Then 1 μl of H_2O and 3.3 μl of extension mixture [24 mM Tris−HCl (pH 8.3), 16 mM $MgCl_2$, 8 mM dithiothreitol, 0.4 mM dNTPs, 5 U of AMV reverse transcriptase] was added to 2 μl of annealing mixture, and extension was allowed to proceed for 45 min at 50°C. After 3 min at 95°C in 24% formamide, the products were electrophoresed in an 8% acrylamide−8 M urea gel. For the 5′ mapping of the 1.5 and 2.0 kb RNAs, primer e was used; and for the 0.45 kb RNA primer g was used.

Construction of cDNAs and PCR amplification for 3′ mapping of RNAs

MtRNAs were treated with RNase-free DNase I (Pharmacia) and the ATP:RNA adenyltransferase was used under the conditions recommended by the supplier (Takara) to add poly(A) tails to RNAs. cDNA was synthesized with an oligo(dT) primer using a cDNA synthesis system (Amersham). In PCR amplification, 5 ng of cDNA was amplified in 50 mM Tris−HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM of each of the four dNTPs with 2.5 U of *Taq* polymerase (Takara) and 100 pM of the adaptor-(dT)$_{17}$ primer and 100 pM of an *atp6*-specific primer. Each of the first 10 cycles of 35 cycles included 1 min at 94°C, 2 min at 40°C and 3 min at 72°C; the following 25 cycles included 1 min at 94°C, 2 min at 45°C and 3 min at 72°C with a 10 min extension at 72°C of the final cycle. Amplified cDNAs were digested with restriction endonucleases and ligated into the BluescriptII plasmid vector. The cloned cDNAs were sequenced with T3 primer and SK primer (Stratagene).

Detection of unprocessed RNA by PCR

5 ng of cDNA was amplified by PCR under the conditions described above using a sequence-specific primer and the adaptor-(dT)$_{17}$ primer. One-twentieth of the PCR product was further amplified using sets of the sequence-specific primers under the following conditions. Each of the 35 cycles included 1 min at 94°C, 2 min at 55°C and 3 min at 72°C with a 10 min extension at 72°C of the final cycle. Amplified cDNAs were cloned and sequenced with primer f or T3 primer for detecting the editing of the unprocessed RNA.

Acknowledgements

We are grateful to H.Kurihara and Y.Kimura for providing and maintaining the plant materials, to J.Imamura and T.Sakai for providing information on the mitochondrial genes and discussions, and to M.Yamamoto and H.Fujimoto for technical advice. We also thank C.André of Stanford University for helpful suggestions on the manuscript and E.Kaleikau of Stanford University for allowing us to cite their unpublished results.

References

Apirion,D. (1983) *Prog. Nucleic Acid Res. Mol. Biol.*, **30**, 1−40.

Anderson,S. *et al.* (1981) *Nature*, **290**, 457−465.

Bailey-Serres,J., Hanson,D.K., Fox,T.D. and Leaver,C.J. (1986) *Cell*, **47**, 567−576.

Bégu,D., Graves,P.-V., Domec,C., Arselin,G., Litvak,S. and Araya,A. (1990) *Plant Cell*, **2**, 1283−1290.

Benne,R., Van Den Burg,J., Brakenhoff,J.P.J., Sloot,P., Van Boom,J.H. and Tromp,M.C. (1986) *Cell*, **46**, 819−826.

Boeshore,M.L., Hanson,M.R. and Izhar,S. (1985) *Plant Mol. Biol.*, **4**, 125−132.

Bonhomme,S., Budar,F., Férault,M. and Pelletier,G. (1991) *Curr. Genet.*, **19**, 121−127.

Boutry,M., Nagy,F., Poulsen,C., Aoyagi,K. and Chua,N.-H. (1987) *Nature*, **328**, 340−342.

Breen,G.A.M., Miller,D.L., Holmans,P.L. and Welch,G. (1986) *J. Biol. Chem.*, **261**, 11680−11685.

Brennicke,A., Möller,S. and Blanz,P. (1985) *Mol. Gen. Genet.*, **198**, 404−410.

Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156−159.

Chen,J.-Y. and Martin,N.C. (1988) *J. Biol. Chem.*, **263**, 13677−13682.

Chu,Y.E., Li,Y.S., Shinjo,C. and Li,H.W. (1972) *Jap. J. Genet.*, **47**, 179−183.

Cooper,P., Butler,E. and Newton,K.J. (1990) *Genetics*, **126**, 461−467.

Covello,P.S. and Gray,M.W. (1989) *Nature*, **341**, 662−666.

Crouzillat,D., Leroy,P., Perrault,A. and Ledoigt,G. (1987) *Theor. Appl. Genet.*, **74**, 773−780.

Dawson,A.J., Jones,V.P. and Leaver,C.J. (1984) *EMBO J.*, **3**, 2107−2113.

Dewey,R.E., Levings,C.S.,III and Timothy,D.H. (1986) *Cell*, **44**, 439−449.

Dewey,R.E., Timothy,D.H. and Levings,C.S.,III (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5374−5378.

Escote-Carlson,L.J., Gabay-Laughnan,S. and Laughnan,J.R. (1990) *Mol. Gen. Genet.*, **223**, 457−464.

Fox,T.D. and Leaver,C.J. (1981) *Cell*, **26**, 315−323.

Geliebter,J. (1988) *Focus*, **9**, 5−8.

Grivell,L.A. (1989) *Eur. J. Biochem.*, **182**, 477−493.

Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.-H. and Grienenberger,J.-M. (1989) *Nature*, **341**, 660−662.

Gualberto,J.M., Bonnard,G., Lamattina,L. and Grienenberger,J.M. (1991) *Plant Cell*, **3**, 1109−1120.

Gupta,D. and Abbott,A.G. (1991) *Theor. Appl. Genet.*, **82**, 723−728.

Hanic-Joyce,P.J. and Gray,M.W. (1990) *J. Biol. Chem.*, **265**, 13782−13791.

Hanson,M.R. and Conde,M.F. (1985) *Int. Rev. Cytol.*, **94**, 213−267.

Hiesel,R., Schobel,W., Schuster,W. and Brennicke,A. (1987) *EMBO J.*, **6**, 29−34.

Hiesel,R., Wissinger,B., Schuster,W. and Brennicke,A. (1989) *Science*, **246**, 1632−1634.

Hoch,B., Maier,R.M., Appel,K., Igloi,G.L. and Kössel,H. (1991) *Nature*, **353**, 178−180.

Hollingsworth,M.J. and Martin,N.C. (1986) *Mol. Cell. Biol.*, **6**, 1058−1064.

Iwahashi,M., Nakazono,M., Kanno,A., Sugino,K., Ishibashi,T. and Hirai,A. (1992) *Theor. Appl. Genet.*, **84**, 275−279.

Kadowaki,K. and Harada,K. (1989) *Jap. J. Breed.*, **39**, 179−186.

Kadowaki,K., Suzuki,T., Kazama,S., Oh-fuchi,T. and Sakamoto,W. (1989) *Nucleic Acids Res.*, **17**, 7519−7519.

Kadowaki,K., Suzuki,T. and Kazama,S. (1990) *Mol. Gen. Genet.*, **224**, 10−16.

Kao,T.-h., Moon,E. and Wu,R. (1984) *Nucleic Acids Res.*, **12**, 7305−7315.

Kempken,F., Mullen,J.A., Pring,D.R. and Tang,H.V. (1991) *Curr. Genet.*, **20**, 417−422.

Kennell,J.C. and Pring,D.R. (1989) *Mol. Gen. Genet.*, **216**, 16−24.

Kennell,J.C., Wise,R.P. and Pring,D.R. (1987) *Mol. Gen. Genet.*, **210**, 399−406.

Köhler,R.H., Horn,R., Lössl,A. and Zetsche,K. (1991) *Mol. Gen. Genet.*, **227**, 369−376.

Kyozuka,J., Kaneda,T. and Shimamoto,K. (1989) *BioTechnology*, **7**, 1171−1174.

Lamattina,L. and Grienenberger,J.M. (1991) *Nucleic Acids Res.*, **19**, 3275–3282.
Laver,H.K., Reynolds,S.J., Moneger,F. and Leaver,C.J. (1991) *Plant J.*, **1**, 185–193.
Levings,C.S.,III (1990) *Science*, **250**, 942–947.
Levings,C.S.,III and Brown,G.G. (1989) *Cell*, **56**, 171–179.
Macino,G. and Tzagoloff,A. (1980) *Cell*, **20**, 507–517.
Makaroff,C.A., Apel,I.J. and Palmer,J.D. (1989) *J. Biol. Chem.*, **264**, 11706–11713.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Morikami,A. and Nakamura,K. (1987) *Nucleic Acids Res.*, **15**, 4692.
Mulligan,R.M., Maloney,A.P. and Walbot,V. (1988) *Mol. Gen. Genet.*, **211**, 373–380.
Newton,K.J. (1988) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **39**, 503–532.
Nivison,H.T. and Hanson,M.R. (1989) *Plant Cell*, **1**, 1121–1130.
Palmer,J.D. and Shields,C.R. (1984) *Nature*, **307**, 437–440.
Powell,L.M., Wallis,S.C., Pease,R.J., Edwards,Y.H., Knott,T.J. and Scott,J. (1987) *Cell*, **50**, 831–840.
Pring,D.R. and Lonsdale,D.M. (1985) *Int. Rev. Cytol.*, **97**, 1–46.
Pruitt,K.D. and Hanson,M.R . (1991) *Mol. Gen. Genet.*, **227**, 348–355.
Rasmussen,J. and Hanson,M.R . (1989) *Mol. Gen. Genet.*, **215**, 332–336.
Reed,K.C. and Mann,D.A. (1985) *Nucleic Acids Res.*, **13**, 7207–7221.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schuster,W. and Brennicke,A. (1987a) *Mol. Gen. Genet.*, **210**, 44–51.
Schuster,W. and Brennicke,A. (1987b) *Nucleic Acids Res.*, **15**, 9092.
Shimamoto,K., Terada,R., Izawa,T. and Fujimoto,H. (1989) *Nature*, **338**, 274–276.
Shinjo,C. (1984) In Tsunoda,S. and Takahashi,N.(eds), *Biology of Rice*. Japan Sci. Soc. Press, Tokyo, pp. 321–338.
Siculella,L. and Palmer,J.D. (1988) *Nucleic Acids Res.*, **16**, 3787–3799.
Simpson,L. and Shaw,J. (1989) *Cell*, **57**, 355–366.
Sutton,C.A., Conklin,P.L., Pruitt,K.D. and Hanson,M.R. (1991) *Mol. Cell. Biol.*, **11**, 4274–4277.
Thomas,S.M., Lamb,R.A. and Paterson,R.G. (1988) *Cell*, **54**, 891–902.
Virmani,S.S. and Shinjyo,C. (1988) *Rice Genet. Newslet.*, **5**, 9–15.
Yamato,K., Ogura,Y., Kanegae,T., Yamada,Y. and Ohyama,K. (1992) *Theor. Appl. Genet.*, **83**, 279–288.
Yang,A.J. and Mulligan,R.M. (1991) *Mol. Cell. Biol.*, **11**, 4278–4281.
Young,E.G. and Hanson,M.R. (1987) *Cell*, **50**, 41–49.

Received on March 17, 1992; revised on January 4, 1993