4830–4834 *Nucleic Acids Research, 1997, Vol. 25, No. 23*

Extensive RNA editing and possible double-stranded structures determining editing sites in the *atpB* transcripts of hornwort chloroplasts

Koichi Yoshinaga*, Takayuki Kakehi, Yoshihito Shima, Hiroe Iinuma, Takehiro Masuzawa and Masaru Ueno

Faculty of Science, Shizuoka University, Shizuoka 422, Japan

Received July 14, 1997; Revised and Accepted October 7, 1997

DDBJ/EMBL/GenBank accession nos D43695, D86545

ABSTRACT

Three nonsense codons and an unusual initiation codon were located within the putative coding region of the *atpB* gene of chloroplast DNA of the hornwort *Anthoceros formosae*. Nucleotide sequencing of cDNA prepared from transcripts revealed extensive RNA editing. The unusual initiation codon ACG was changed to AUG and three nonsense codons were converted into sense codons. In total 15 C residues of the genomic DNA were replaced by U residues in the mRNA sequences, while 14 U residues were replaced by C residues. This is the highest number of editing events for a chloroplast mRNA reported so far. Partial editing was also shown in a cDNA clone where 23 sites were edited but six sites remained unedited, representing the existence of premature mRNA. The expected two-dimensional structure of the mRNA shows the existence of a sequence complementary to every editing site, which can produce continuous base pairing longer than 5 bp, suggesting that mispairing in the double strand is the site determinant for RNA editing in *Anthoceros* chloroplasts. Comparison of the cDNA sequence with other chloroplast genes suggests that the mechanism arose in the first land plants and has been reduced during evolution.

INTRODUCTION

Editing is a post-transcriptional process which changes the primary sequence of RNAs, as compared with that of the corresponding DNA templates, and can therefore be regarded as a step in the regulation of gene expression. RNA editing was originally detected in the kinetoplast genetic system of trypanosomes (1) and was subsequently observed in nucleus-encoded mRNA (2) and mRNA encoded by mitochondrial genes from higher plants (3–5). More recently, several chloroplast transcripts of higher plants have been shown to be subjected to RNA editing (6,7, for a review see 8). Editing has been found in virtually all mRNAs of plant mitochondria (9). However, observations of editing have been restricted in the chloroplasts of seed plants, with the exception of *rbcL* transcripts of *Anthoceros* (10). All of these editing events in chloroplasts involve C→U conversions, except in *Anthoceros*. We report broad RNA editing of *atpB* transcripts of *Anthoceros formosae*; this is the first evidence for the existence of this gene product and the most extensive editing of chloroplast transcripts observed to date.

MATERIALS AND METHODS

The following oligonucleotide primers designed from the genomic DNA sequence of *A.formosae* were obtained from Biologica (Nagoya, Japan): P1, 5′-ATGCGAAATTTAGCTAAGCAAT (–55 to –34); P2, 5′-CAACTCCATAATGAGTACCGTTTTT (549 to 525); P3, 5′-TTAGTAGATGGAACAATCTCTTCAA (1464 to 1440). Numbering is from the translation start site of *atpB*.

Five clones containing *atpB* of *A.formosae* chloroplasts were prepared and identified as described previously (10). All contained an identical 7556 bp *Kpn*I fragment, the structure of which is shown in Figure 1.

Total RNA was isolated and cDNA was synthesized as described previously (10). The cDNA amplified using the primer pairs P1/P2 and P1/P3 was ligated to pUC18 as described (10) and introduced into *Escherichia coli* DH5α by means of E. coli Pulsar (BioRad).

The nucleotide sequences of genomic DNA and cDNA were determined by dideoxy chain termination methods (11) using 7-deaza Sequenase v.2.0 (US Biochemical) as described (10). The resulting sequences were analyzed with Genetyx-MAC 8.5 software (SDC, Tokyo).

RESULTS

Unusual nucleotide sequence of chloroplast DNA from *A.formosae*

Five clones assumed to contain *rbcL* were selected from a DNA library of *A.formosae* by colony and Southern hybridization as described previously (10). Restriction enzyme mapping and partial DNA sequencing revealed that all clones contained a common 7556 bp *Kpn*I fragment, with sequences homologous to *accD*, *trnR*, *rbcL*, *atpB*, *atpE*, *trnM*, *trnV*, *ndhC* and *ndhK*, in that order (Fig. 1). The gene arrangement is the same as that of

*To whom correspondence should be addressed. Tel: +81 54 238 4760; Fax: +81 54 237 3384; Email: sckyosi@sci.shizuoka.ac.jp

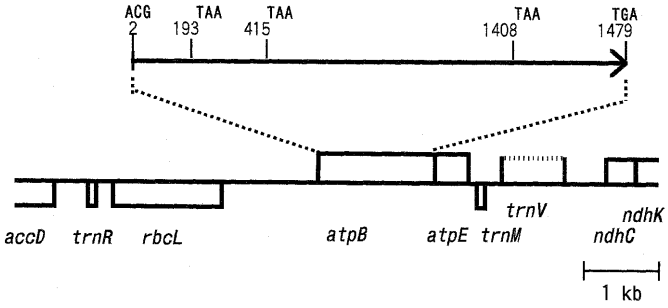

Figure 1. Chloroplast gene arrangement of *A.formosae*. Coding regions are represented by boxes. The intron in *trnV* is indicated by a broken line. The unusual initiation codon, three nonsense codons and the termination codon of the *atpB* gene are shown above.

Angiopteris (12) and *Marchantia* (13), suggesting that the sequence is that of chloroplast DNA. The complete nucleotide sequence of plasmid pK79 containing the 7556 bp *Kpn*I fragment was determined (accession no. D43695). The sequence contains three nonsense codons of TAA in the putative coding region of *atpB*, though TAA is used as the stop codon in *rbcL* (10). In addition to these, ACG instead of ATG was identified at the putative initiation codon. These findings indicate that the RNA transcripts are edited.

Evidence of RNA editing

To examine the possibility of RNA editing we determined the mRNA sequence by analyzing the cDNA. The nucleotide sequences of the resulting five clones containing the entire coding region of *atpB* were determined and compared with that of genomic DNA. Three nonsense codons in the genomic sequence were changed into sense codons by a U→C conversion at positions 193, 415 and 1408. The unusual ACG sequence at the initiation codon was changed to AUG by a C→U conversion. This AUG initiation codon appears in several other chloroplast transcripts (6,7). In total, 15 C residues were converted into U residues and 14 U residues into C residues (accession no. D86545). As shown in Figure 2, only the first and second positions of codons are edited, as in mitochondria, although the second codon position is the primary target of editing in higher plant chloroplasts (14). As a result of RNA editing, 28 amino acids (5.7% of total amino acids) deduced from the mRNA sequence differed from those predicted from the DNA sequence. This is the most extensive RNA editing within a single chloroplast gene product observed to date. An unprecedented amount of reverse editing, U→C conversion, was also observed, although only a few examples of reverse editing have been reported in plant mitochondria and none in other chloroplasts (14). However, we have previously reported 20 sites of RNA editing containing seven reverse events in *rbcL* transcripts of the same *Anthoceros* and generally expected that the lower land plants would contain a higher frequency of editing in chloroplasts (10,15).

Three of the five cDNA clones containing the entire coding region of *atpB* reflected fully edited sequences at 29 sites (Fig. 2). However, one of them, pAB202, reflected unedited sequence at nucleotide positions 415, 1085, 1214, 1232, 1415 and 1433, although it reflected completely edited sequence at every other site. Another clone, pAB105, reflected completely edited sequence except at 484. These results reveal the existence of premature

Site	Genomic	cDNA	Unedited
2	ACG(Thr)	.T.(Met)	0/15
193	TAA(stop)	C..(Gln)	0/10
415	TAA(stop)	C..(Gln)	1/10
475	TCT(Ser)	C..(Pro)	0/10
484	TGT(Cys)	C..(Arg)	1/10
503	CCA(Pro)	.T.(Leu)	0/10
746	TCA(Ser)	.T.(Leu)	0/10
752	GTT(Val)	.C.(Ala)	0/6
799	TCT(Ser)	CT.(Leu)	0/5
800			0/5
820	CTT(Leu)	T..(Phe)	0/5
823	TGT(Cys)	C..(Arg)	0/5
857	TCA(Ser)	.T.(Leu)	0/5
860	TCA(Ser)	.T.(Leu)	0/5
937	CCA(Pro)	T..(Ser)	0/5
986	GTT(Val)	.C.(Ala)	0/5
1003	TCT(Ser)	C..(Pro)	0/5
1031	TCA(Ser)	.T.(Leu)	0/5
1052	TTT(Phe)	.C.(Ser)	0/5
1082	CTT(Leu)	.C.(Pro)	0/6
1085	GTC(Val)	.C.(Ala)	1/5
1109	TTT(Phe)	.C.(Ser)	0/7
1214	ACT(Thr)	.T.(Ile)	1/7
1223	TCA(Ser)	.T.(Leu)	0/7
1232	TCA(Ser)	.T.(Leu)	1/7
1316	TCT(Ser)	.T.(Phe)	0/7
1408	TAA(stop)	C..(Gln)	0/6
1415	TCT(Ser)	.T.(Phe)	1/5
1433	ACT(Thr)	.T.(Ile)	1/7

Figure 2. Codons differing between genomic DNA and cDNA sequences. Nucleotide positions of editing sites are shown on the left. Dots show nucleotides identical to those of genomic DNA and deduced amino acids are in parentheses. Number of unedited observations against total number of clones examined is shown on the right.

mRNA. Another collection of cDNA clones was prepared by amplification of the 5′ portion of the *atpB* transcript (–55 to 549) and the nucleotide sequences were determined. Within them, pAB6 showed unedited sequence at sites 2, 193, 415, 475, 484 and 503, suggesting that it was derived from a growing transcript which had not yet undergone any editing. However, the possibility still remains that the clone was derived from contaminated genomic DNA, therefore it is not included in Figure 2. The other five clones of this collection were completely edited at these six sites. In the case of the trypanosome kinetoplastid the mRNA editing proceeds with a strict 3′→5′ polarity (16). However, our analysis of partially edited transcripts shows no polarity during the editing process, suggesting that the editing mechanism of chloroplasts differs from the trypanosome system.

DISCUSSION

Many processes requiring interaction with specific RNAs recognize not only primary sequence but also the secondary and higher order structures of the target RNAs. Guide RNA complementary to the segments of the edited mRNA has been detected in the trypanosome mitochondrial system (17) and mispairing in RNA was corrected by RNA editing in the stem of mitochondrial tRNA (18) and in the introns of mitochondrial transcripts (19). Therefore, we tried to find *cis*-acting determinants for the transcripts of *atpB*. Partial two-dimensional structures of the mRNA were estimated using Genetyx software, which showed stem structures containing editing sites at 2, 799, 820, 1223 and 1415 (Fig. 3). They were longer than 7 bp, although three of them contain G-U base pairs. The other editing sites were located in loop

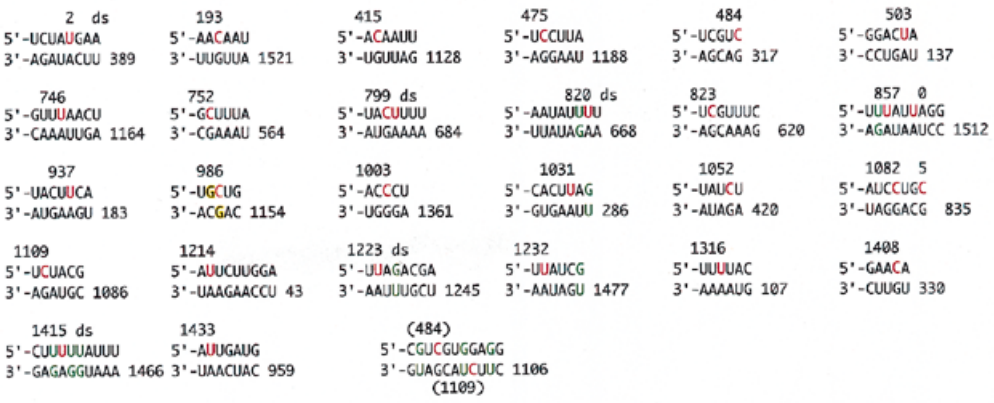

Figure 3. Possible double-stranded structures within the *atpB* mRNA of *Anthoceros*. All of the 29 edited nucleotides (red) and nucleotide positions are shown. Double-stranded structures predicted by the software 'RNA secondary structure pred' are indicated by ds. Numbers behind the complementary strands show their 5′ nucleotide positions. G-U base pairs are shown as green letters. An alternative possible double-stranded structure is shown at the end with nucleotide positions 484 and 1109.

or relatively short stem structures. Sequences complementary to them were found in a similar structure of mRNA using homology search software. They could form Watson–Crick and G-U base pairs which were longer than 5 bp. The G-U pairs usually seen in tRNA species and in the editing system of trypanosome (20) were located outside the edited nucleotides in the possible double-stranded structures.

One sequence of 9 bp, two of 8 bp, six of 7 bp, eight of 6 bp and six of 5 bp formed in the transcript of *atpB* without G-U pairs. A sequence complementary to the pentamer which contains editing sites could be found within an ~1 kb sequence; the possibility of finding a specific sequence of 5 nt within a random sequence is $(1/4)^5$. However, it would require 16 and 65 kb to find a specific sequence of 7 and 8 nt respectively. It would be difficult to find such a long specific sequence within the 1572 bases of *atpB* mRNA examined if the mRNA did not possess a functional sequence to interact with the editing site. This mRNA is highly biased towards A and U content (63%), therefore the probability of finding a specific AU biased sequence in the mRNA is increased. For example, to find a heptamer sequence the probability within a random non-biased sequence is $0.25^7 = 6.10 \times 10^{-5}$. However, the probability of finding a biased heptamer within the biased sequence is given by $P = a^m \times b^n$, where a is the probability that A or U is present and b is the probability that C or G is present and m is the frequency with which A or U appears within 7 nt and n is that for C or G. When the AU content is 63% then $a = 0.63/2$, $b = 0.37/2$, $m = 0.63 \times 7$ and $n = 0.37 \times 7$. Therefore $P = 7.75 \times 10^{-5}$. This value is not high and thus it is not easy to find a heptamer sequence within the mRNA. In addition, the sequences observed in the possible double-stranded structures do not overlap, suggesting that they interact with only one partner, though they contain 350 nt in total, which corresponds to 22% of the analyzed sequence. Similar double-stranded structures containing all the editing sites could form by Watson–Crick base pairing in the *rbcL* mRNA of *Anthoceros* (10), in which three sequences of 8 bp, four of 7 bp, five of 6 bp and five of 5 bp were detected.

Therefore, we believe the complementary sequences within the same mRNA to be factors in selective recognition of the nucleotide to be edited in *Anthoceros*, though such factors have been reported to be of extraplastidic origin in higher plants (21). Mispairing can be corrected by an editing enzyme such as cytidine deaminase (22) or transaminase. This hypothesis can explain several phenomena of RNA editing which have been observed in chloroplasts and mitochondria. No general consensus sequence has been deduced from aligning heterologous sites, thus implying the existence of *trans*-acting specificity factors for individual editing sites (23). Our hypothesis can provide sequences complementary to individual editing sites, although such a functional complementary sequence has not been found in higher plant chloroplast genomes (24). That complementary sequences exist in the chloroplasts of the hornwort but not in those of higher plants could be explained as follows: copies of the chloroplast genes containing the complementary sequence were transferred to the nucleus and the remaining genes lost the sequence during evolution. Several cases of gene transfer from chloroplast to nucleus have been documented; the *tufA* gene was transferred within the green algal lineage giving rise to land plants (25) and spinach nucleus contains integrated sequences that are homologous to chloroplast DNA sequences (26).

The amino acid sequences deduced from the genomic DNA and cDNA of *Anthoceros* were compared with those of other plants. As shown in Figure 4, the homology of the amino acid sequence deduced from cDNA was always higher than that deduced from genomic DNA, suggesting that the former sequence reflects the functional protein sequence. Amino acids at 286, 287 and 344 were changed from Ser to Leu by RNA editing in *Anthoceros*. In these sites TCA codons deduced from genomic DNA were changed to TTA. The same codons were found at 286, 287 and 344 in *Angiopteris* and at 344 in *Pinus*, suggesting that at least three sites in *Angiopteris* and one in *Pinus* are edited. RNA editing occurs at 29 sites in the hornwort, three in the fern, one in the gymnosperm and at no sites in the angiosperm, indicating a tendency towards a reduction in the number of editing sites of *atpB* transcripts during the evolution of plants, with the exception of the liverwort. This suggests that either RNA editing arose in chloroplasts of the first land plants and disappeared from the liverwort or, alternatively, that it arose in the common ancestor of *Anthoceros* and vascular plants (10). The study of RNA editing in chloroplasts has been rather restricted in angiosperms. However, that more editing sites were observed in the gymnosperm than in the angiosperm (27), with the most extensive editing

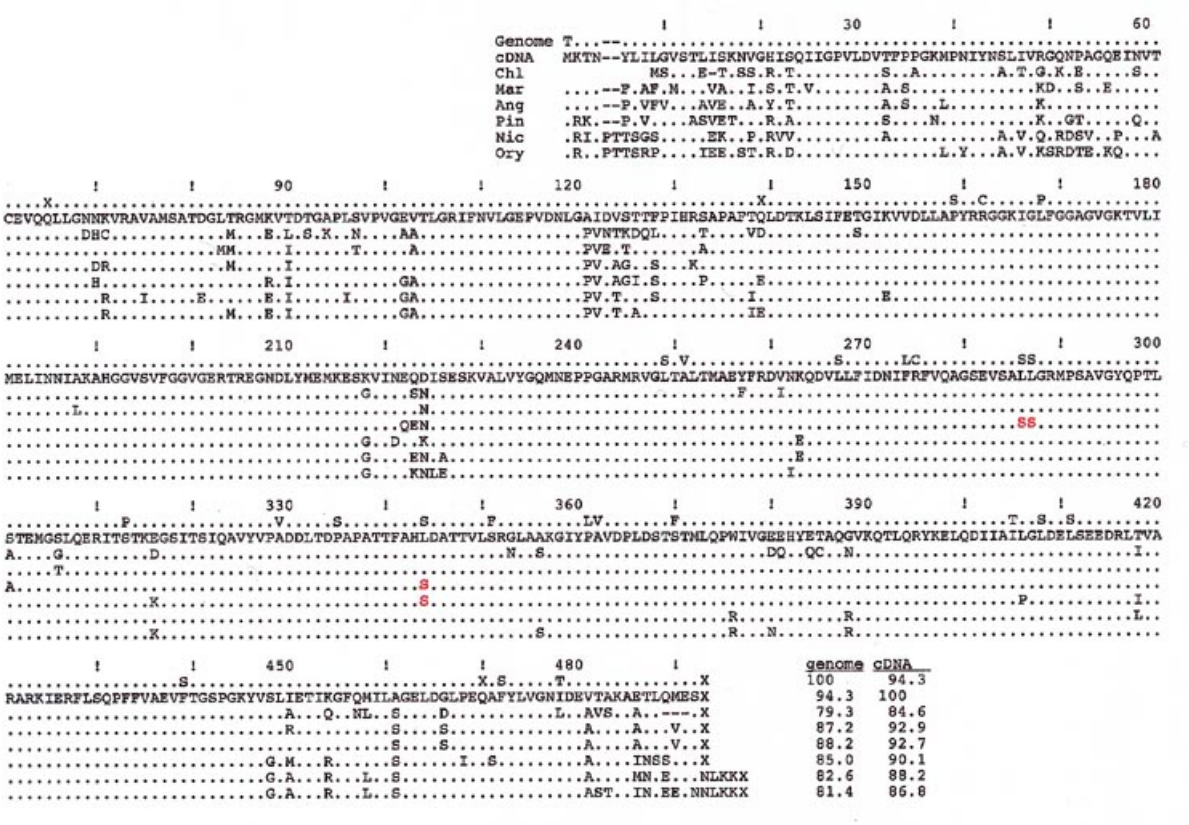

Figure 4. Comparison of amino acid sequences of *atpB*. The amino acid sequence deduced from the cDNA sequence of *A.formosae* is compared with those deduced from genomic sequences of *A.formosae* (Genome), *Chlorella vulgaris* (Chl; accession no. D10997), *Marchantia polymorpha* (Mar; 13), *Angiopteris ligodiiforia* (Ang; 12), *Pinus thunbergii* (Pin; 30), *Nicotiana tabacum* (Nic; 31) and *Oryza sativa* (Ory; 32). Only those amino acids that differed from the cDNA sequence are shown; amino acids identical to those of the cDNA are represented by dots. Amino acids expected to be changed by RNA editing are shown as red letters. Spacers (–) are included to adjust the sequence length. Nonsense codons within coding regions and stop codons are indicated by an X. The amino acids are numbered from the N-terminus deduced from the *A.formosae* cDNA sequence. Homology (%) is shown at the end of the sequence.

observed in the hornwort, suggests that RNA editing is distributed widely in the transcripts of many genes of primitive land plants.

Hiesel *et al*. (28) showed RNA editing in mitochondria of all major groups of land plants and Malek *et al*. (29) showed the highest frequency in a lycoposid *Isoetes* and the highest frequency of reverse events in the hornwort *Anthoceros*, which is among the first land plants in their phylogenetic tree. They showed the existence of RNA editing in some bryophytes but not in the green alga *Chara*. We also analyzed editing events in *rbcL* transcripts of the green alga *Nitella* spp., but could not detect any. These results support a scenario in which the mechanism of RNA editing arose in chloroplasts and mitochondria of the first land plants. RNA editing would affect the evolution of land plants, though further studies of this event in primitive green plants are required.

As already described in the *rbcL* transcripts of hornwort chloroplasts (10), *atpB* transcripts show extensive U→C conversion. This was likely to have been prevalent among primitive land plants and to have disappeared more quickly than C→U conversion, because several events of reverse conversion have been observed in the mitochondria of bryophytes and pteridophytes (29).

ACKNOWLEDGEMENTS

We are grateful to Prof. Kiyomi Wada for helpful advice and for providing *A.formosae* and Tatsuya Ishii for sequencing of genomic DNA. This work was supported in part by a grant-in-aid for scientific research (no. 08640886) from the Ministry of Education of Japan.

REFERENCES

1 Benne,R., van den Burg,J., Brakenhoff,J.P., Sloof,P., Van Boom,J.H. and Tromp,M.C. (1986) *Cell*, **46**, 819–826.
2 Powell,L.M., Wallis,S.C., Pease,R.J., Edwards,Y.H., Knott,T.J. and Scott,J. (1987) *Cell*, **50**, 831–840.
3 Covello,P.S. and Gray,M.W. (1989) *Nature*, **341**, 662–666.
4 Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.H. and Grienenberger,J.M. (1989) *Nature*, **341**, 660–662.
5 Hiesel,R., Wissinger,B., Schuster,W. and Brennicke,A. (1989) *Science*, **246**, 1632–1634.
6 Hoch,B., Maier,R.M., Appel,K., Igloi,G.L. and Kössel,H. (1991) *Nature*, **353**, 178–180.
7 Kudla,J., Igloi,G.L., Metzlaff,M., Hagemann,R. and Kössel,H. (1992) *EMBO J.*, **11**, 1099–1103.

8 Kössel,H., Hoch,B., Maier,R.M., Igloi,G.L., Kudla,J., Zeltz,P., Freyer,R., Neckermann,K. and Ruf,S. (1993) In Brennicke,A. and Kück,U. (eds), *Plant Mitochondria*. VCH Publishers, Weinheim, Germany, pp. 93–102.
9 Bonnard,G., Gualberto,J.M., Lamattina,L. and Grienenberger,J.M. (1992) *Crit. Rev. Plant. Sci*., **10**, 503–524.
10 Yoshinaga,K., Iinuma,H., Masuzawa,T. and Ueda,K. (1996) *Nucleic Acids Res*., **24**, 1008–1014.
11 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
12 Yoshinaga,K., Kubota,Y., Ishii,T. and Wada,K. (1992) *Plant Mol. Biol*., **18**, 79–82.
13 Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S.-I., Inokuchi,H. and Ozeki,H. (1986) *Nature*, **322**, 572–574.
14 Hanson,M.R., Sutton,C.A. and Lu,B. (1996) *Trends Plant Sci*., **1**, 57–64.
15 Yoshinaga,K. (1997) In Iwatsuki,K. and Raven,P.H. (eds), *Evolution and Diversification of Land Plants*. Springer-Verlag, Tokyo, Japan, pp. 45–60.
16 Sloof,P. and Benne,R. (1993) *FEBS Lett*., **325**, 146–151.
17 Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
18 Maréchal-Drouard,L., Ramamonjisoa,D., Cosset,A., Weil,J.H. and Dietrich,A. (1993) *Nucleic Acids Res*., **21**, 4909–4914.
19 Börner,G.V., Mörl,M., Wissinger,B., Brennicke,A. and Schmelzer,C. (1995) *Mol. Gen. Genet*., **246**, 739–744.
20 Stuart,K. (1991) *Trends Biochem. Sci*., **16**, 68–72.
21 Bock,R. and Koop,H.-U. (1997) *EMBO J*., **16**, 3282–3288.
22 Yu,W. and Schuster,W. (1995) *J. Biol. Chem*., **270**, 18227–18233.
23 Chaudhuri,S., Carrer,H. and Maliga,P. (1995) *EMBO J*., **14**, 2951–2957.
24 Bock,R. and Maliga,P. (1995) *Mol. Gen. Genet*., **247**, 439–443.
25 Baldauf,S.L., Manhart,J.R. and Palmer,J.D. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5317–5321.
26 Timmis,J.N. and Scott,N.S. (1983) *Nature*, **305**, 65–67.
27 Wakasugi,T., Hirose,T., Horihata,M., Tsudzuki,T., Kössel,H. and Sugiura,M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 8766–8770.
28 Hiesel,R., Combettes,B. and Brennicke,A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 629–633.
29 Malek,O., Lätting,K., Hiesel,R., Brennicke,A. and Knoop,V. (1996) *EMBO J*., **15**, 1403–1411.
30 Wakasugi,T., Tsudzuki,J., Ito,S., Nakashima,K., Tsudzuki,T. and Sugiura,M. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 9794–9798.
31 Shinozaki,K., Ohme,M., Tanaka,M., Wakasugi,T., Hayashida,N., Matsubayashi,T., Zaita,N., Chunwongse,J., Obokata,J., Yamaguchi-Shinozaki,K., *et al.* (1986) *EMBO J*., **5**, 2043–2049.
32 Hiratsuka,H., Shimada,H., Whittier,R.F., Ishibashi,T., Sakamoto,M., Mori,M., Kondo,C., Honji,Y., Sun,C.R., Meng,B.Y., *et al.* (1989) *Mol. Gen. Genet*., **217**, 185–194.