Palindromic repetitive elements in the mitochondrial genome of *Volvox*[1]

Naoki Aono[a], Toshinobu Shimizu[a,2], Tan Inoue[a,b], Hideaki Shiraishi[a,b,*]

[a] *Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*
[b] *Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan*

Received 30 January 2002; revised 10 May 2002; accepted 10 May 2002

First published online 23 May 2002

Edited by Takashi Gojobori

Abstract Group I introns were found in the *cob* and *cox I* genes of *Volvox carteri*. These introns contain tandem arrays of short palindromic sequences that are related to each other. Inspection of other regions in the mtDNA revealed that similar palindromic repetitive sequences are dispersed in the non-protein coding regions of the mitochondrial genome. Analysis of the group I intron in the *cob* gene of another member of Volvocaceae, *Volvox aureus*, has shown that its sequence is highly homologous to its counterpart in *V. carteri* with the exception of a cluster of palindromic sequences not found in *V. carteri*. This indicates that the palindromic clusters were inserted into the introns after divergence of the two species, presumably due to frequent insertions of the palindromic elements during evolution of the Volvocaceae. Possible involvement of the palindromic repetitive elements in the molecular evolution of functional RNAs is discussed. © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Mitochondrial genome; Group I intron; Palindromic repetitive sequence; Molecular evolution of functional RNA; *Volvox*

1. Introduction

Palindromic repetitive sequences have been described in the mitochondrial genomes of various organisms; e.g. GC-rich clusters in *Saccharomyces cerevisiae* [1], PRSs (palindromic repeated sequences) in *Oryza sativa* [2] and the short dispersed repeats in Chlamydomonadalean algae [3,4]. Of these sequences, PRSs in *O. sativa* and GC-rich clusters in *S. cerevisiae* have been reported to be mobile elements. Such repetitive sequences are thought to be one of the causes of mtDNA rearrangements during evolution. For example, it has been suggested, from the comparison of the mtDNA of Chlamydomonadalean algae, that fragmented and scrambled rRNA coding regions of *Chlamydomonas reinhardtii* are results of transposition [3] or recombination [5] between short palindromic repetitive sequences [6]. Recently, palindromic repetitive sequences, RPEs (*Rickettsia* palindromic elements), were found to be inserted in-frame within many unrelated protein coding sequences in *Rickettsiae*, the closest extant relatives of mitochondria, suggesting the potential role of these elements in the creation of new protein sequences [7]. Thus, this class of short repetitive sequences seems to have potential to cause both the rearrangement of gene organization and the molecular evolution of protein sequences.

In the course of the structural analysis of group I introns in the mtDNA of *Volvox carteri*, a multicellular organism closely related to Chlamydomonadalean algae, we found that clusters of short palindromic repetitive sequences are inserted in the mitochondrial introns. Analysis of the palindromic repetitive sequences in the mtDNA of *V. carteri* and another member of Volvocaceae, *Volvox aureus*, suggested that they are mobile invasive elements. Finally, we will discuss a potential role of the palindromic mobile elements in the molecular evolution of functional RNAs.

2. Materials and methods

2.1. Algal materials

Strain 'EVE' isolated from *V. carteri* f. *nagariensis* 'HK10' was obtained from D.L. Kirk [8]. *V. aureus* '31202-2-9' was obtained from H. Nozaki (Graduate School of Science, University of Tokyo, Japan).

2.2. Preparation of DNA and RNA

Total DNAs of *V. carteri* and *V. aureus* were isolated as described [9]. To obtain the fraction of DNA enriched with mtDNA, total DNA of *V. carteri* was mixed with CsCl to make a final concentration of 0.95 g/ml and with bisbenzimide (Hoechst 33258) (50 μg/ml) [10]. Buoyant density equilibrium gradients were run at 36 000 rpm for 36 h at 20°C in an SW55Ti rotor (Beckman). The upper UV-fluorescent band was collected using a 26-gauge needle, extracted with isopropanol saturated with CsCl and precipitated with 3 volumes of 70% ethanol. The mtDNA-enriched fraction was digested with *Hin*dIII or *Eco*RI, ligated into pBluescript II SK, and used to transform *Escherichia coli* strain, SURE (Stratagene).

RNA was prepared as described by Kirk and Kirk [11].

2.3. RT-PCR and PCR

Complementary DNA of apocytochrome *b* (*cob*) transcripts were synthesized with AMV Reverse Transcriptase XL (TaKaRa). PCR was carried out with LA Taq DNA polymerase (TaKaRa). Oligonucleotides for the synthesis and amplification of cDNAs are COB1, 5′-GAA CGA AGA ATT GCA TAA ACC CAC AG, and COB2, 5′-ATG RTN RTN CAN GCN TTY AT (Y: T or C, R: A or G, N: A or C or G or T). Oligonucleotides for the amplification of Vc.cob-1 and Va.cob-1 are COB3, 5′-ACC AAC AAC TGG AAT AGC TGT TGC TAA GC, and COB4, 5′-TTC ATT GGC TAT GTA CTA CCA TGG G. Oligonucleotides for the amplification of Vc.cox-1

*Corresponding author. Fax: (81)-75-753 3996.
E-mail address: siraisi@kuchem.kyoto-u.ac.jp (H. Shiraishi).

[1] Nucleotide sequence data reported in this paper are available in the DDBJ/EMBL/GenBank databases under the accession numbers AB075042, AB075043, AB075044, AB075195, AB075196 and AB075197.

[2] Present address: Mitsubishi Pharma Corporation, 5-100, Yana, Kisarazu, Chiba 292-0812, Japan.

Abbreviations: cob, apocytochrome *b*; *cox I*, cytochrome *c* oxidase subunit I; ORF, open reading frame

0014-5793/02/$22.00
PII: S0014-5793(02)02832-6

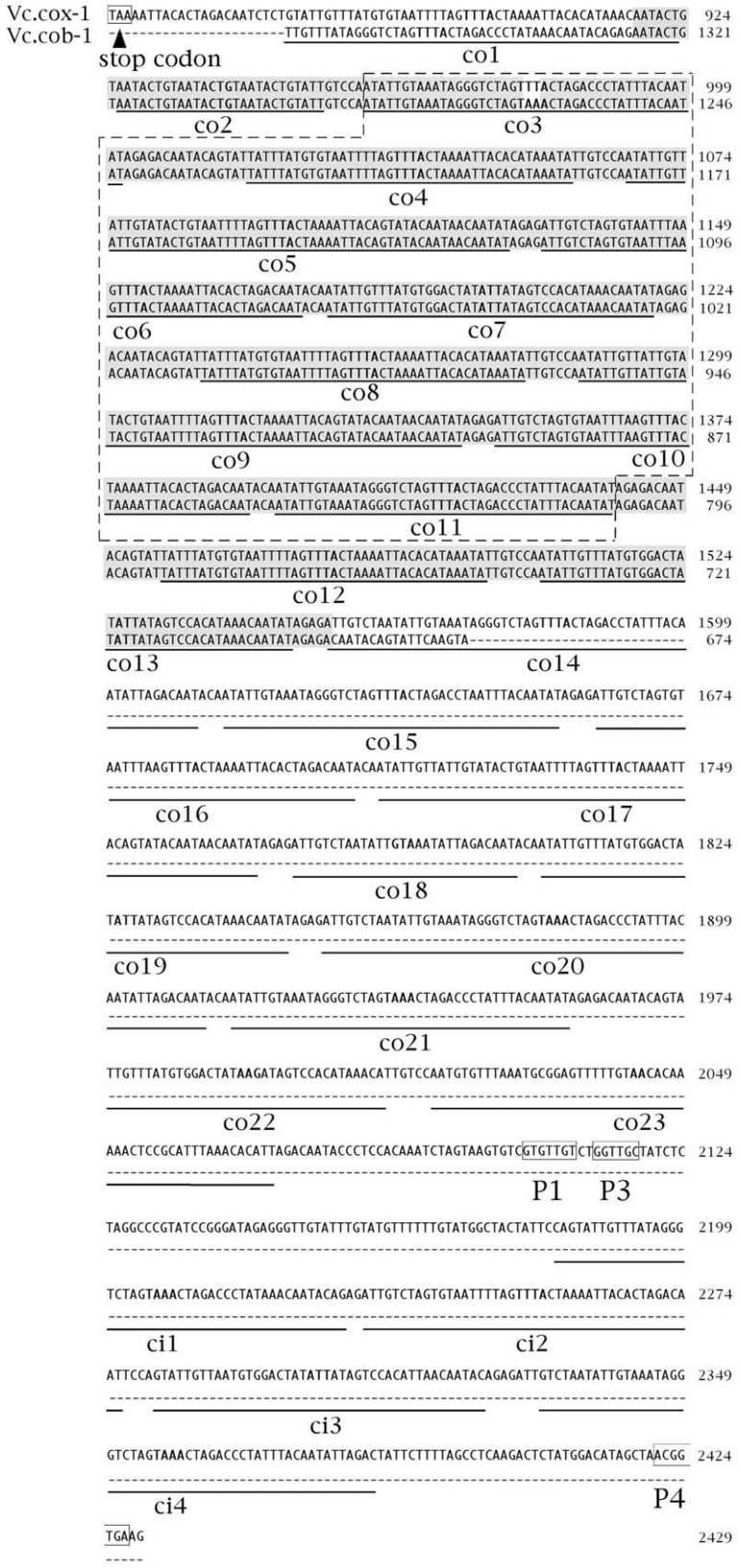

Fig. 1. Alignment of the palindromic repeated sequences of Vc.cob-1 and Vc.cox-1. The sequence of Vc.cob-1 shown in this figure is the complementary sequence of Vc.cob-1 transcript. Numbering is from the first nucleotide of each intron. Shaded areas show identical sequences between Vc.cox-1 and Vc.cob-1. The repeat units of Vc.cox-1 (co1 to co23) are underlined. Nucleotides that constitute loops in the potential stem-and-loop structures are shown in bold. Sequences in dashed boxes are able to form a large stem-and-loop structure (483 bp in length). Stop codon of the ORF in Vc.cox-1 and the regions that constitute P1, P2 and P3 stems of Vc.cox-1 are shown by open boxes.

are COXI1, 5′-TAA TGC AGG TGT AGA TAT GCT TGT CCA CG, and COXI2, 5′-CCC ATA CTC AGC ACG TAA TGA AAG TGT GC.

3. Results and discussion

3.1. Cloning of mitochondrial introns

In many organisms, the introns in the *cob* gene of their mitochondria are present in the conserved region of the gene [12–14]. To determine the nucleotide sequence in the *cob* gene region of *V. carteri*, we obtained partial cDNAs, and carried out PCR to amplify the conserved region of the *cob* gene from the total DNA of *V. carteri*. The amplified genomic DNA fragment was longer than the fragment amplified from cDNA, suggesting that the amplified genomic region contained intron(s). Cloning and sequence analysis of the amplified fragment revealed that it contained an intron at the site corresponding to the position where an intron is inserted in *Chlamydomonas smithii* [12]. We named the intron Vc.cob-1. The nucleotide sequence and predicted secondary structure of

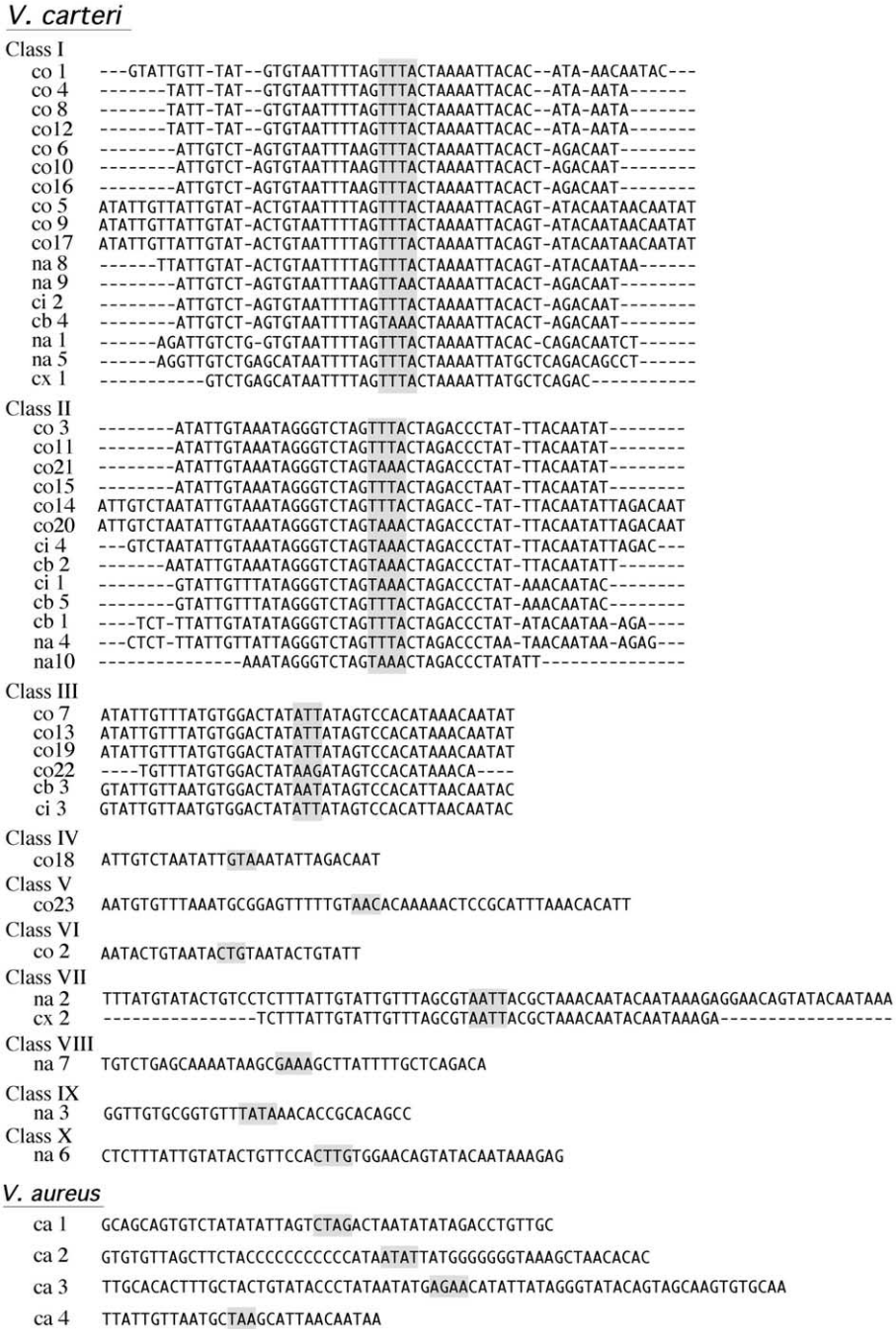

Fig. 2. Classification of the palindromic units. Shaded nucleotides constitute loops in the predicted stem-and-loop structures in their transcripts. Origin of each unit is shown by the following prefixes: co: the cluster downstream of the ORF in Vc.cox-1; ci: the cluster between P3 and P4 of Vc.cox-1; cx: the cluster downstream of the *cox I* gene; cb: the cluster in the *cob* intron other than Vc.cob-1; na: the cluster in the intergenic region between *nad2* and *nad6* genes; ca: the cluster downstream of the ORF in Va.cob-1.

Vc.cob-1 (1682 bp) indicate that it belongs to subgroup ID [15]. Vc.cob-1 contains an open reading frame (ORF) of 227 amino acids in-frame with the 5′ exon. The amino acid sequence of the polypeptide encoded by the ORF has similarity with maturases encoded by other group I introns.

In parallel with the isolation of the *cob* gene intron, we attempted to characterize other regions of the mitochondrial genome. In the course of the experiments, two clones containing the central portion and C-terminus of the cytochrome *c* oxidase subunit I (*cox I*) gene were isolated from a mtDNA plasmid library. When the region between the fragments was amplified by PCR, the size of the amplified DNA fragment was larger than that of the expected coding sequence. Sequence analysis of the fragment showed that it contained an intron at the site corresponding to the position where the *cox I* gene intron of *Chlamydomonas eugametos* and the intron 15 of *Podospora anserina* are inserted [16]. The nucleotide sequence and the predicted secondary structure of the intron, Vc.cox-1 (2702 bp), indicate that it belongs to subgroup IB3 [15]. It contains an ORF of 283 amino acids in-frame with the 5′ exon, and the polypeptide encoded by the ORF is similar to LAGLIDADG type endonucleases that have been found in other group I introns such as intron 15 of *P. anserina* [16].

3.2. Clusters of palindromic sequences in the introns

Downstream of the ORFs, unusually long clusters of palindromic sequences were found. The clusters in Vc.cob-1 and Vc.cox1 are inserted between P1 and P3 and in the L1 region of the introns, respectively (Fig. 1). The nucleotide sequences of the clusters are closely related; the cluster in Vc.cob-1 is complementary to a portion of the cluster in Vc.cox-1 (Fig. 1). In Vc.cox-1, another cluster consisting of four palindromic sequences is inserted between P3 and P4 in addition to the cluster in the L1 region (Fig. 1).

The cluster in the L1 region of Vc.cox-1 consists of 23 palindromic sequences (from 23 to 60 bp) separated by 3–16-bp spacers. Based on their sequence similarities, the palindromic units in the L1 region of Vc.cox-1 can be classified into six classes (see co1 to co23 in Fig. 2). Classes I, II and III contain 10, 6 and 5 types of palindromic sequences, respectively, whereas classes IV, V and VI consist of only one. The sequence from the nucleotide positions 958 to 1440 of Vc.cox-1 (from co3 to co11 in Fig. 1) predicts the potential to form a large stem-and-loop structure in which co3 to co6 form base pairings with co11 to co8, respectively; co7 is located at the top of the stem-and-loop (Fig. 1). The four palindromic sequences (ci1 to ci4) inserted between P3 and P4 of Vc.cox-1 are related to the palindromic sequences in the L1 region of the same intron. They belong to class I, II or III (Fig. 2).

In the palindromic sequences, compensatory mutations which are often observed in the members of each class allow the retention of the stem-and-loop structures. For example, of the 26 positions in the central parts of the stems of class I, II and III where base substitutions, deletions or insertions are observed, 22 positions have compensatory base changes which retain the Watson–Crick base pairings (Fig. 3). Although the reason these compensatory mutations occur is unclear, their occurrence suggests that the stem regions have some role in the mechanism of the expansion of the palindromic repetitive sequences.

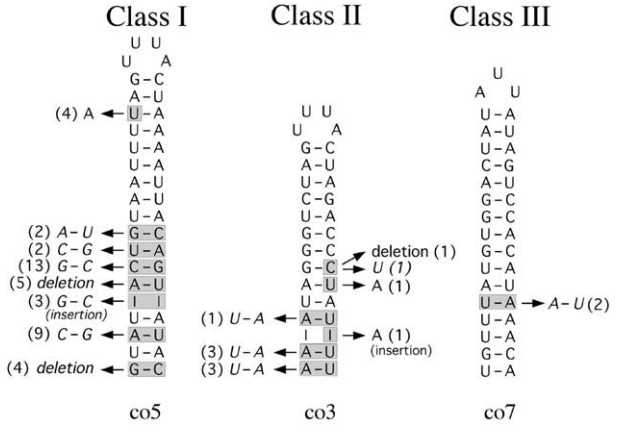

Fig. 3. Base conversions in the stem regions of the palindromic sequences. Shaded boxes are the regions where base substitution, deletion or insertion occurred from co5 (class I), co3 (class II) and co7 (class III). Numbers in parentheses show the same base changes found in different palindromic units. Base changes that retain base pairings are shown in italics.

3.3. Palindromic repetitive sequences in other regions of the mtDNA of V. carteri

The clusters of palindromic sequences found in the mitochondrial introns of *V. carteri* share similar sequences. The fact that such sequences have not been found in group I introns from other organisms and that the related sequences are inserted at two different positions in Vc.cox-1 suggest that the palindromic sequences are selfish DNAs.

To see whether similar clusters of palindromic sequences are dispersed in the mitochondrial genome of *V. carteri*, we next investigated other regions of the mtDNA. By sequencing the DNA clones from the mtDNA-enriched plasmid library, we found a cluster of 10 palindromic sequences in the intergenic region between *nad2* and *nad6* and a couple sequences downstream of the *cox I* gene. Another intron was also found in the *cob I* gene. Although the complete sequence of this intron has not been determined, an array of five palindromic sequences was found near its 3′ end. Some of the newly found palindromic sequences belong to classes I–III (Fig. 2). Thus, the clusters of palindromic sequences are dispersed in the non-protein coding regions of the mtDNA of *V. carteri*.

3.4. Palindromic repeats in the cob intron of another member of Volvocaceae

To investigate whether the clusters of palindromic sequences are present in the mtDNA of other Volvocaceae, we analyzed *V. aureus*, a green alga closely related to *V. carteri* [17,18]. By using the total DNA of *V. aureus* as a template for PCR, we amplified portions of the *cob* and *cox I* genes that correspond to the regions where Vc.cob-1 and Vc.cox-1 are inserted in *V. carteri*. The size of the fragment amplified from the *cob* gene was longer than that predicted from the length of the coding region. The nucleotide sequence of the fragment showed the insertion of a group I intron at the position corresponding to the insertion site of Vc.cob-1 in *V. carteri*. In contrast, the length of the fragment amplified from the *cox I* gene was the same as that expected from the length of the coding region, thus it was not analyzed further.

The intron in the *cob* gene of *V. aureus* (Va.cob-1) contained an ORF and, downstream of the ORF, a cluster of palindromic sequences as in the case of Vc.cob-1. However,

the palindromic sequences in Va.cob-1 have no sequence similarity with those in Vc.cob-1 nor with those found in *V. carteri* (Fig. 2). In contrast, other regions of the introns are quite similar; the sequence identity of Vc.cob-1 and Va.cob-1 is 96% except for the palindromic repeats. These fact suggest that the clusters were inserted into the preexisting introns after the divergence of the two species. Thus, the insertion and expansion of the palindromic repetitive sequences seem to have occurred frequently in the mitochondrial genomes of these species over evolution.

Palindromic repetitive sequences in the mtDNA of *C. reinhardtii* share similarity with those in *V. carteri* and *V. aureus* in that they are present in tandem array [3]. However, their primary sequences are quite divergent; although the repeats in *C. reinhardtii* are reported to have in common the palindromic motif TRCTCGG(N_{4-14})CCGAGYA or a slight variant thereof, those of *V. carteri* and *V. aureus* have no such sequences. Thus it is unclear whether the palindromic repeats in the Volvocales are evolutionarily related to those in *C. reinhardtii*.

3.5. Possible involvement of the palindromic repetitive sequences in the molecular evolution of functional RNAs

In the genome of Rickettsiae, short palindromic mobile elements, RPEs, have been identified [7]. The observation that these RPEs are inserted into many unrelated protein coding regions suggests their potential role in creating new protein sequences during evolution [7].

In this study, we have found insertion of palindromic repetitive sequences in the group I introns from two species of *Volvox*. The frequent insertion of palindromic elements in group I introns of *Volvox* mitochondria prompts us to propose their potential role in the evolution of functional RNAs as follows.

In general, functional RNAs consist of stem-and-loop structures with bulges connected by single-stranded linkers, and tertiary interactions among them stabilize their active structures. Recently, we have reported that the insertions of various stem-and-loop structures with bulges can re-activate defective deletion mutants of the *Tetrahymena* group I intron ribozyme [19–21]. Also, deletion analyses of the group I introns have shown that a simple pseudoknot structure, P3-P7 domain, constitutes a catalytic core; other domains, mostly stem-and-loop structures with bulges, help the P3–P7 domain to form an active conformation [22]. From these observations, it has been proposed that an ancestral group I intron consists of a simple pseudoknot structure resembling the P3–P7 domain, and that other regions were subsequently added to accomplish more reliable and efficient splicing [22]. For such a mode of evolution, the insertion of the palindromic sequences could be a far more efficient way than gradual nucleotide changes, since the insertion of a palindromic sequence would create a new structural domain in the parental RNA instantaneously. The inserted palindromic mobile elements themselves or their evolved forms might contribute to improving or modifying the activity of the parental RNA.

Acknowledgements: We thank Drs. Ruth T. Yu and Y. Ikawa for critical reading of the manuscript and Drs. T. Miyata, K. Kuma and M. Shishikura for valuable discussions. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] Weiller, G., Schueller, C.M. and Schweyen, R.J. (1989) Mol. Gen. Genet. 218, 272–283.
[2] Nakazono, M., Kanno, A., Tsutsumi, N. and Hirai, A. (1994) Plant Mol. Biol. 24, 273–281.
[3] Boer, P.H. and Gray, M.W. (1991) Curr. Genet. 19, 309–312.
[4] Nedelcu, A.M., Lee, R.W., Lemieux, C., Gray, M.W. and Burger, G. (2000) Genome Res. 10, 819–831.
[5] Denovan-Wright, E.M. and Lee, R.W. (1994) J. Mol. Biol. 241, 298–311.
[6] Nedelcu, A.M. (1997) Mol. Biol. Evol. 14, 506–517.
[7] Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P.E., Raoult, D. and Claverie, J.M. (2000) Science 290, 347–350.
[8] Harper, J.F., Huson, K.S. and Kirk, D.L. (1987) Genes Dev. 1, 573–584.
[9] Zhu, H., Qu, F. and Zhu, L.H. (1993) Nucleic Acids Res. 21, 5279–5280.
[10] Moore, L.J. and Coleman, A.W. (1989) Plant Mol. Biol. 13, 459–465.
[11] Kirk, M.M. and Kirk, D.L. (1985) Cell 41, 419–428.
[12] Colleaux, L., Michel-Wolwertz, M.R., Matagne, R.F. and Dujon, B. (1990) Mol. Gen. Genet. 223, 288–296.
[13] Dujon, B. (1989) Gene 82, 91–114.
[14] Kroymann, J. and Zetsche, K. (1998) Mol. Evol. 47, 431–440.
[15] Michel, F. and Westhof, E. (1990) J. Mol. Biol. 216, 585–610.
[16] Cummings, D.J., Michel, F. and McNally, K.L. (1989) Curr. Genet. 16, 381–406.
[17] Larson, A., Kirk, M.M. and Kirk, D.L. (1992) Mol. Biol. Evol. 9, 85–105.
[18] Nozaki, H., Misawa, K., Kajita, T., Kato, M., Nohara, S. and Watanabe, M.M. (2000) Mol. Phylogenet. Evol. 17, 256–268.
[19] Ikawa, Y., Shiraishi, H. and Inoue, T. (1995) in: Tracing Biological Evolution in Proteins and Gene Structures (Go, M. and Schimmel, P., Eds.), pp. 115–123, Elsevier, Amsterdam.
[20] Ikawa, Y., Shiraishi, H. and Inoue, T. (1996) FEBS Lett. 394, 5–8.
[21] Naito, Y., Shiraishi, H. and Inoue, T. (2000) FEBS Lett. 466, 273–278.
[22] Ikawa, Y., Shiraishi, H. and Inoue, T. (2000) Nature Struct. Biol. 7, 1032–1035.