Theor Appl Genet (1993) 86:579–584

Variations in chloroplast DNA from rice (*Oryza sativa*): differences between deletions mediated by short direct-repeat sequences within a single species

A. Kanno [1,*], N. Watanabe [1], I. Nakamura [2], A. Hirai [3]

[1] Graduate Division of Biochemical Regulation, School of Agriculture, Nagoya University, Nagoya 464-01, Japan
[2] Iwate Biotechnology Research Center, Narita, Kitakami 204, Japan
[3] Laboratory of Radiation Genetics, Faculty of Agriculture, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan

Received: 5 October 1992 / Accepted: 19 November 1992

Abstract. In a previous study, we compared chloroplast DNAs (ctDNAs) from four species of *Oryza* and detected two independent deletions of DNA fragments in the ctDNAs (Kanno and Hirai 1992a). These deletions were genotype-specific variations. Since short direct-repeat sequences were detected at the borders of both deletions, the deletions were apparently the result of intramolecular recombination mediated by these direct-repeat sequences. In the present study, we examined whether or not this type of variation exists within a single species. Ishii et al. demonstrated three types of ctDNA in *O. sativa* (1988), and the source of the variations that they identified seemed to be deletions. We determined the precise locations of the deletions and the sequences around them. As expected, our results showed that these variations were the results of deletions that were mediated by short direct-repeat sequences. While the deletions that had been found previously were located on spacer regions, those found in this study were located within open reading frames (ORFs). Northern hybridization analysis showed that one of the ORFs was transcribed. In the case of this deletion, the amino acid sequence encoded by the C-terminal region of the ORF was altered and the short inverted-repeat sequences downstream of the ORF were deleted. In addition, there were other short inverted-repeat sequences downstream of the altered ORF.

Key words: *Oryza sativa* – Indica – Chloroplast DNA – Deletion – Direct repeat

Communicated by F. Salamini

* *Present address:* Institute of Genetic Ecology, Tohoku University, 2-1-1 Katahira, Aoba, Sendai 980, Japan

Correspondence to: A. Hirai

Introduction

The structure and organization of chloroplast DNA (ctDNA) have been studied in many land plants and algae (Whitfeld and Bottomley 1983, Palmer 1985). Phylogenetic relationships between closely related species or genera have been examined by analyses of restriction fragment length polymorphisms (RFLP) in many groups of plants (Ogihara and Tsunewaki 1982; Kung et al. 1982; Palmer et al. 1983). These analyses have revealed several types of variation in chloroplast genomes, variations that can be accounted for by DNA rearrangements; for example, insertion/deletion, inversion, recombination and transposition (Yamada 1989).

Rice species (genus *Oryza*) have seven types of genomes: AA, BB, BBCC, CC, CCDD, EE, and FF. Ishii et al. carried out phylogenetic studies of the AA genome in *Oryza* species (1986, 1988) and Dally and Second analyzed six genomes, with the exception of the FF genome, in *Oryza* species by endonucleolytic cleavage of ctDNAs (1990). In order to compare ctDNAs among *Oryza* species in detail, we constructed overlapping clone banks of the entire chloroplast genomes of *O. punctata* (W1515), *O. officinalis* (W0002), and *O. australiensis* (W0008) (Kanno and Hirai 1992a). Using these clones, we constructed physical maps of the ctDNAs and compared the ctDNAs from these three species and that from *O. sativa* (cv 'Nipponbare'). The complete nucleotide sequence of the ctDNA from *O. sativa* is available (Hiratsuka et al. 1989). In our analysis, we identified two independent deletions in the ctDNA from *O. punctata* and *O. officinalis*. These deletions represent genotype-specific variations. Short direct-repeat sequences were detected at the borders of both deletions, indicating that these deletions were the result of intramolecular recombination mediated by the adjacent direct-repeat sequences. In wheat,

Ogihara et al. detected similar deletions in a comparison of ctDNAs from *Aegilops crassa*, *Ae. squarrosa*, and *Triticum aestivum* (1988). At the border of these deletions, they also found short direct-repeat sequences. The direct repeats found in these rice and wheat species consisted of AT-rich sequences. Moreover, the characteristics of the deletions found in rice and wheat were genotype-specific and species-specific variations, respectively. These results raise the question of whether or not this type of deletion occurs frequently; in other words, can it be found within a single species?

Ishii et al. (1988) reported that on the basis of RFLP analysis there are three types of ctDNA in *O. sativa*: two ecospecies of *O. sativa*, Japonica and Javanica, contain one type of ctDNA and another ecospecies, Indica, contains three types of ctDNA. Further analysis has shown that these variations are caused by deletions (Ishii et al. 1991). To examine whether these variations in deletions are caused by short direct-repeat sequences, we identified the precise locations of these deletions and determined the nucleotide sequences around these deletions. From our results, we can now report that variations in deletions mediated by short direct-repeat sequences exist even within a single species. Moreover, part of the deletions extends into a transcribed region (ORF), and there is a possibility that the amino acid sequence of a protein is altered by one of the deletions.

Materials and methods

Plant materials

Two strains of *Oryza sativa* were used in this study, ecospecies Japonica (cv 'Nipponbare') and ecospecies Indica (C8005). Seeds of C8005 were generously provided by Dr. Y. Sano of the National Institute of Genetics, Japan.

Southern hybridization

Total DNAs were extracted from 1 g of mature green leaves as described elsewhere (Honda and Hirai 1990). After digestion of DNAs with the appropriate restriction endonucleases, Southern hybridization analysis was carried out using a Nonradioactive DNA Labeling and Detection Kit (Boehringer Mannheim, Germany). The *Pst*I-12 and *Pst*I-10 fragments of ctDNA from *O. sativa* (cv 'Nipponbare') were used as probes (Hirai et al. 1985).

Synthesis of DNA and the polymerase chain reaction (PCR)

DNA synthesis and the PCR were carried out using a DNA synthesizer (Applied Biosystems, USA) and a DNA-Amplification System (Perkin-Elmer Cetus Corp, USA), respectively. The sequences of the primers were as follows: primer 1, 5′-GCCTGCAGTAAGTGGACCTGACTCC-3′; primer 2, 5′-GCGGATCCGAGGTCGTGGTAAATCC-3′; primer 3, 5′-GCCTGCAGTTCCAAAAAAACGTAC-3′; and primer 4, 5′-GCGGATCCGTTTGGTAGTATCGTTGCG-3′.

Sequencing of DNA

Fragments of DNA that contained deletions were subcloned into the appropriate sites of pUC119. Sequencing was carried out on an automated DNA sequencer (Applied Biosystems).

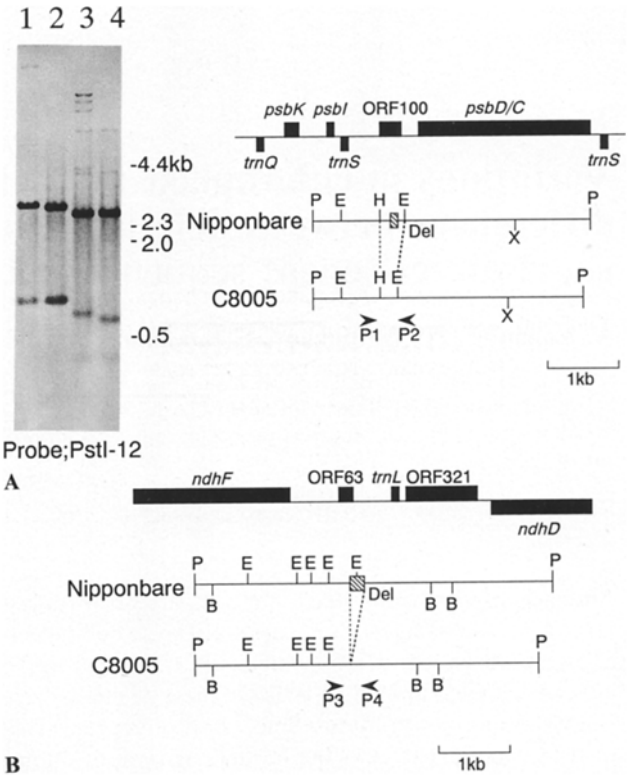


Fig. 1. **A** Southern hybridization and the location of the deletion on the *Pst*I-12 fragment. Total DNAs from *O. sativa* ecosp. Japonica (cv 'Nipponbare') and *O. sativa* ecosp. Indica (C8005) were digested with *Bam*HI/*Pst*I (*lanes 1, 2*) and *Eco*RI/*Pst*I (*lanes 3, 4*). The samples were subjected to electrophoresis on a 0.7% agarose gel, and Southern hybridization was carried out using the *Pst*I-12 fragment of rice ctDNA as probe. *Lanes 1* and *3* and *lanes 2* and *4* show results for 'Nipponbare' and C8005, respectively. The primers used for PCR are indicated by *arrowheads* (*P1* and *P2*), and the genes and open reading frames are indicated by *bold lines*. Restriction sites are designated as follows: *P*, *Pst*I; *E*, *Eco*RI; *X*, *Xba*I; *H*, *Hae*III. **B** The location of the deletion on the *Pst*I-10 fragment. The primers used in PCR are indicated as above (*P3* and *P4*). Restriction sites are designated as above. In addition, *B* represents *Bam*HI. This figure is reproduced with the permission of Ishii et al. (1991)

Northern hybridization

RNA extraction and Northern hybridization were carried out as described by Kanno and Hirai (1992b). As shown in Fig. 4, the fragment of ctDNA from *O. sativa* (cv 'Nipponbare') was used as a probe.

Results

The locations of the deletions found in the ctDNA from O. sativa ecospecies Indica

RFLP analysis of ctDNA from species of *Oryza* with the AA genome was carried out by Ishii et al. (1988). They, described that relative to the ctDNA of Japonica there are two deletions in the ctDNA from C8005 of Indica. To determine the precise location of these deletions, total

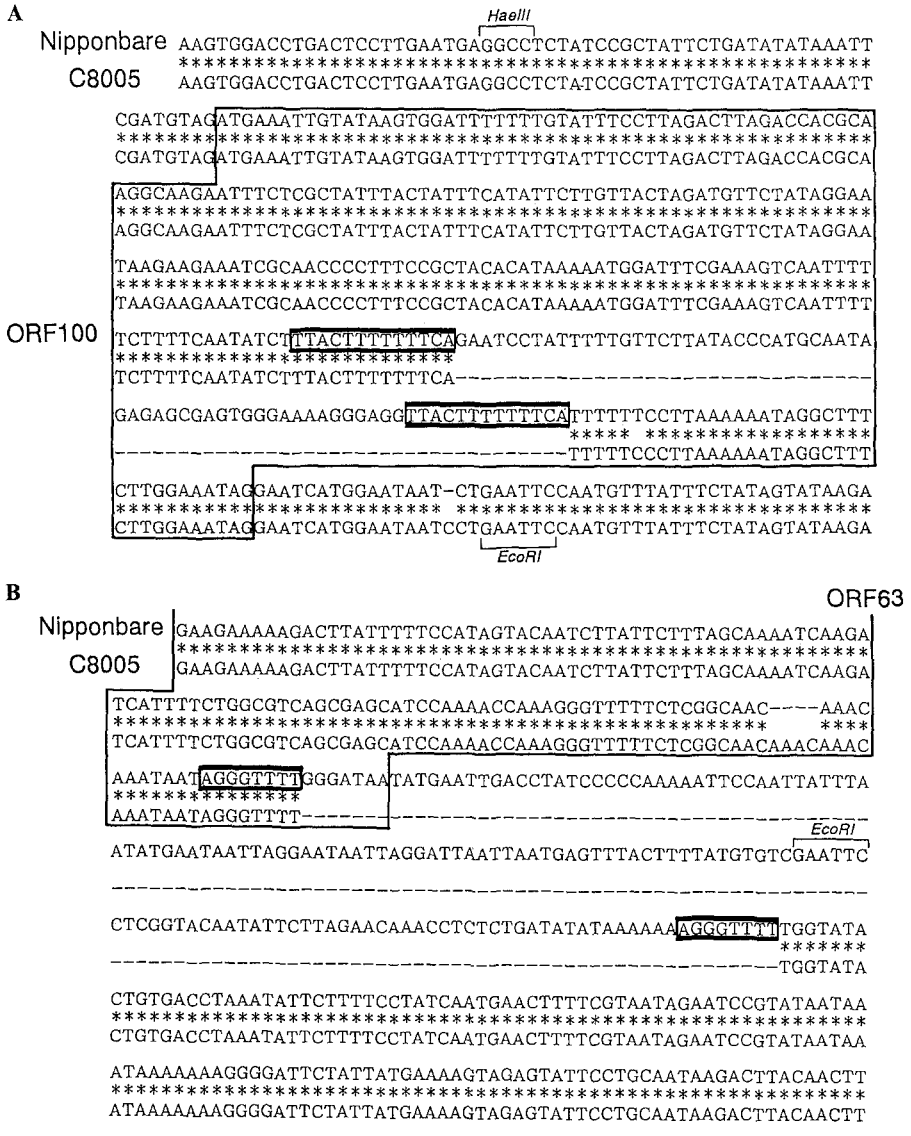


Fig. 2 A, B. Comparison of nucleotide sequences around the deletions on the *Pst*I-12 (**A**) and *Pst*I-10 (**B**) fragments of ctDNA between 'Nipponbare' and C8005. *Stars* denote matching nucleotides and *dashes* indicate gaps (deleted regions). Open reading frames and direct-repeat sequences that are likely to be sites of recombination are *boxed*

DNAs were extracted from green leaves of Japonica ('Nipponbare') and Indica (C8005), and the DNAs were digested with appropriate restriction enzymes. After agarose gel electrophoresis, the DNAs were transferred to nylon membranes and Southern hybridization was carried out using the *Pst*I-12 and *Pst*I-10 fragments of rice ctDNA as probes. The results of the Southern hybridization and the locations of the deletions are shown in Fig. 1. The deletion on the *Pst*I-12 fragment was located on an *Eco*RI fragment of about 0.8 kb, and the length of the deletion was about 60 bp. Further experiments showed that the deletion was located on the fragment between the *Hae*III and *Eco*RI sites and that ORF100 was also located on the same fragment (data not shown). The deletion on the *Pst*I-10 fragment has already been described by Ishii (1991). As shown in Fig. 1 B, the deleted sequence contained an *Eco*RI site. The deletion was about 200 bp long and located around ORF63.

Nucleotide sequences around the deletions

For the cloning of the DNA fragments that contained the deletions, PCR was carried out with total DNA prepared from C8005 as a template and the appropriate primers (primers 1–4) as shown in Fig. 1. The products of PCR were cloned into pUC119, and the nucleotide sequences were determined as described in the Materials and methods.

DNA sequences around the deletions were aligned for 'Nipponbare' and C8005 (Fig. 2). The sequences around the respective deletion-containing regions exhibited a high degree of homology (>98%). The length of the deletion on the *Pst*I-12 fragment was 69 bp and that on the *Pst*I-10 fragment was 158 bp. The short direct-repeat sequences at the borders of the deleted regions were detected in corresponding regions of 'Nipponbare'. The sequences of the direct repeats that were found on the

*Pst*I-12 and *Pst*I-10 fragments were TTACTTTTTTCA and AGGGTTTT, respectively.

The two deletions were located within two independent open reading frames

The deletions found on the ctDNAs of *O. punctata* and *O. officinalis* were located in the spacer region between the *trnQ* and *rps16* genes, and between the *atpI* and *atpH* genes, respectively (Kanno and Hirai 1992a). In contrast, the two deletions found on the ctDNA of Indica (C8005) were located in open reading frames: the deletions detected on the *Pst*I-12 and *Pst*I-10 fragments were located in ORF100 and ORF63, respectively (Fig. 2).

The length of ORF100 was 303 bp, and the precise location of the deletion on the *Pst*I-12 fragment was within the open reading frame, 200–268 bp downstream from the ATG codon. The other deletion, located on the *Pst*I-10 fragment, contained the C-terminal and downstream regions of ORF63. The lengths of the deletions in the C-terminal and downstream regions were 7 bp and 151 bp, respectively. The effects of these deletions are discussed below.

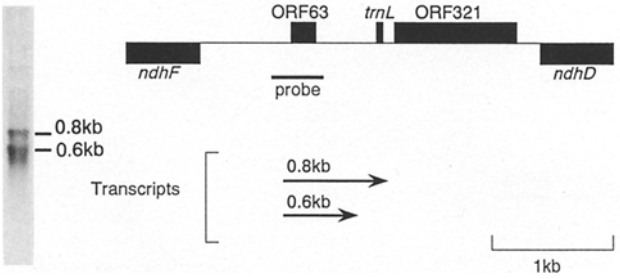


Fig. 3. Northern hybridization and the pattern of transcription of the putative gene, ORF63. The location of the probe and transcripts are shown by a *bold line* and *arrows*, respectively

Transcription of ORF63

To examine the transcription of ORF63, Northern hybridization was carried out using a specific probe for ORF63. As shown in Fig. 3, two transcripts of 0.8 kb and 0.6 kb were detected. In a previous study, we detected a 0.1-kb transcript that corresponded to the *trnL* gene (Kanno and Hirai 1992b). Thus, the 0.8-kb transcript encodes ORF63 and the *trnL* gene, and the transcripts of 0.6 kb and 0.1 kb are the cleaved products of the 0.8-kb transcript.

Discussion

Effects of the deletions found in the ctDNA of Oryza species

In a previous study, we detected two deletions in the ctDNAs from *O. punctata* and *O. officinalis* (Kanno and Hirai 1992a). The lengths of these deletions were 453 bp and 322 bp and at the borders of these deletions there were direct-repeat sequences of AGAAAAAAAT and AATAGAA(T)AATGAG, respectively. Since these deletions were located in the spacer region, they are not likely to have critical effects on gene expression. In contrast, the deletions found in the C8005 type of Indica were different from the deletions mentioned above as they were located within open reading frames.

The deletion on the *Pst*I-12 fragment was located within ORF100. Since the length of the deletion was 69 bp, 23 out of 100 amino acids had been deleted, even though the reading frame of the putative ORF100

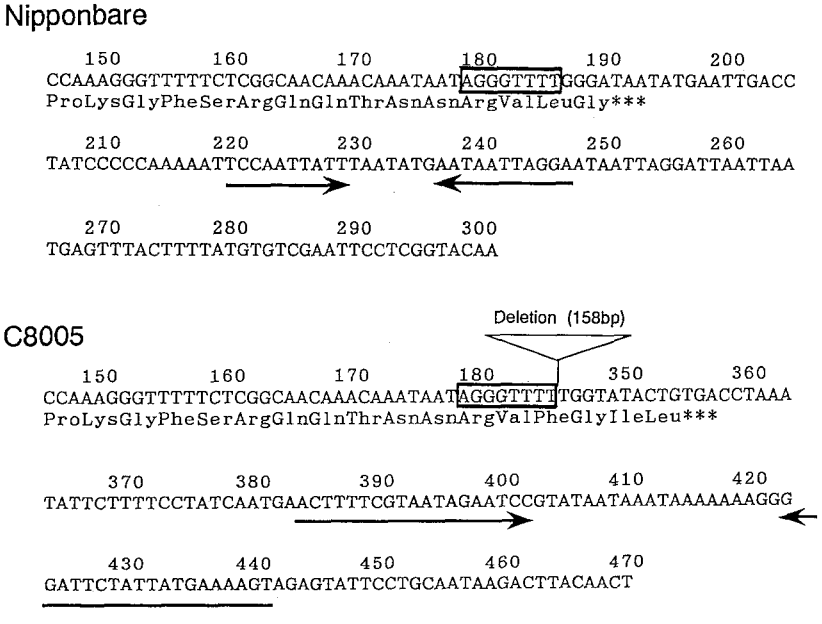
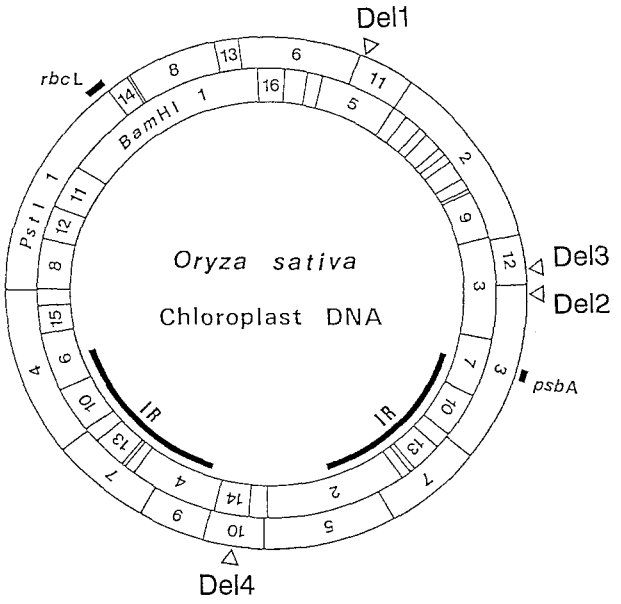


Fig. 4. Amino acid sequences and nucleotide sequences of the C-terminal region and the region downstream from the termination codon of ORF63 in 'Nipponbare' and C8005. *Arrows* indicate inverted-repeat sequences. Direct-repeat sequences that are likely to be sites of recombination are *boxed*

protein remained unchanged. This open reading frame is co-transcribed with *psbK*, *psbI*, *psbD*, *psbC*, ORF62, and *trnG* as a 6.1-kb transcript, and no transcript that encodes only ORF100 alone has been detected (Kanno and Hirai 1992b). From these results, it appears that ORF100 is not likely to be a functional gene.

The length of the other deletion, located on the *Pst*I-10 fragment, was 158 bp. This deleted region contained the sequence of the C-terminal end of the putative ORF63 protein and a short inverted-repeat sequence that might act either as a terminator or might block degradation during processing (Fig. 2B). From the results of Northern hybridization, this open reading frame appears to be co-transcribed with the *trnL* gene as a 0.8-kb transcript, and this transcript seems to be processed into 0.6-kb and 0.1-kb transcripts. Consequently, the 0.6-kb transcript contains only ORF63 (Fig. 3). Thus, ORF63 should be a functional gene and should be expressed in the chloroplast. However, we must still determine whether or not the ORF63 of C8005 is functional.

We compared 'Nipponbare' and C8005 for the C-terminal region of ORF63 and the region downstream from the termination codon (Fig. 4). To our surprise, there were only slight differences with respect to the amino acid sequence encoded by ORF63, and a short inverted-repeat sequence was present in the region downstream of ORF63 in C8005. The amino acid sequences encoded by ORF63 were -Arg-Val-Leu-Gly in 'Nipponbare' and -Arg-Val-Phe-Gly-Ile-Leu in C8005. This variation seems not to have any critical effect on the function of the protein since C-terminal regions of proteins are non-functional in many cases. The inverted-repeat sequence of ORF63 of 'Nipponbare' is located 27 bp downstream from the termination codon, while that of C8005 is probably located 27 bp downstream from the termination codon, even though the sequences and the lengths of the inverted repeats are different in the two ecospecies. From these results, ORF63 seems to be functional in both 'Nipponbare' and C8005, however the function of this protein is not known.



	Sequence of direct repeats	Length of deletions (bp)
Del1	—AATAGAA(T)AATGAG—	322
Del2	—AGAAAAAAAT—	453
Del3	—TTACTTTTTTCA—	69
Del4	—AGGGTTTT—	158

Fig. 5. The locations of the deletions found in the ctDNA from species of *Oryza* (*triangles*). *Del1* and *Del2* are the deletions found in the ctDNAs of *O. officinalis* and *O. punctata*, respectively (Kanno and Hirai 1992a); *Del3* and *Del4* are the deletions identified in the present study

The frequency of the occurrence of deletions that are mediated by short direct-repeat sequences in ctDNA during evolution

Variations in the deletions mediated by short direct-repeat sequences which were found in the ctDNA of the various species of *Oryza* are shown in Fig. 5. The lengths of the direct-repeat sequences varied from 8 to 13 bp, and the lengths of the deletions varied from 69 bp to 453 bp. In the case of similar deletions in wheat species, the lengths of the direct-repeat sequences were 9 bp and 5 bp, and the lengths of the deletions were 791 bp and 1062 bp. Common features of direct repeats found in both *Oryza* and wheat species are the AT-rich sequences. On the ctDNA, there are many candidates for regions in which changes due to deletions mediated by short direct repeats may occur. Therefore, if direct-repeat sequences are located at distances of 100–1000 bp from each other and if the changes due to deletions are not critical, changes in the deletions can occur frequently during evolution. However, there are many copies of ctDNA in one chloroplast, and many chloroplasts in one cell. It will be of interest to investigate why specific deletions are fixed in one plant and fixed in one species.

The deletions found in the ctDNA of *O. punctata* (W1515) and *O. officinalis* (W0002) were genotype-specific variations, while similar deletions found in wheat species were species-specific variations. However, the deletions found in the ctDNA of C8005 represented variations within a single species, *O. sativa*, and even within a single ecospecies. Consequently, it seems that deletion events mediated by short direct-repeat sequences can oc-

cur comparatively easily during evolution, and even the ctDNA within a single species can be altered by such deletions.

Acknowledgements. We wish to express our appreciation to Dr. Y. Sano (National Institute of Genetics, Japan) for providing the rice seeds. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture, Japan.

References

Dally AM, Second G (1990) Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section Oryza). Cladistic-mutation and genetic-distance analysis. Theor Appl Genet 80:209–222

Hirai A, Ishibashi T, Morikami A, Iwatsuki N, Shinozaki K, Sugiura M (1985) Rice chloroplast DNA: a physical map of ribulose 1,5-bisphosphate carboxylase and the 32-kDa photosystem II reaction center protein. Theor Appl Genet 70:117–122

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Honda H, Hirai A (1990) A simple and efficient method for the identification of hybrids using nonradioactive rDNA as probe. Jpn J Breed 40:339–348

Ishii T (1991) Cytoplasmic and nuclear genome differentiation in A-genome diploid species of rice as revealed by the restriction fragment length polymerase analysis of DNAs. PhD thesis, Kyoto University, Japan

Ishii T, Terachi T, Tsunewaki K (1986) Restriction endonuclease analysis of chloroplast DNA from cultivated rice species, *Oryza sativa* and *O. glaberrima*. Jpn J Genet 61:537–541

Ishii T, Terachi T, Tsunewaki K (1988) Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. Jpn J Genet 63:523–536

Kanno A, Hirai A (1992a) Comparative studies of the structure of chloroplast DNA from four species of *Oryza*: cloning and physical maps. Theor Appl Genet 83:791–798

Kanno A, Hirai A (1992b) A transcription map of the chloroplast genome from rice (*Oryza sativa*). Curr Genet 23:166–174

Kung SD, Zhu SY, Shen GF (1982) *Nicotiana* chloroplast genome. III. Chloroplast DNA evolution. Theor Appl Genet 61:73–79

Ogihara Y, Tsunewaki K (1982) Molecular basis of the genetic diversity of the cytoplasm in *Triticum* and *Aegilops*. I. Diversity of the chloroplast genome and its lineage revealed by the restriction pattern of ct-DNAs. Jpn J Genet 57:371–396

Ogihara Y, Terachi T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc Natl Acad Sci USA 85:8573–8577

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Palmer JD, Shields CR, Cohen EB, Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor Appl Genet 65:181–189

Whitfeld PR, Bottomley W (1983) Organization and structure of chloroplast genes. Annu Rev Plant Physiol 34:279–310

Yamada T (1989) Molecular evolution of chloroplast genomes: Important roles of short repeated sequences. Viva Origino 17:1–22