Organization of DNA sequences highly repeated in tandem in rice genomes*

Hisako OHTSUBO, Masaaki UMEDA and Eiichi OHTSUBO

*Institute of Applied Microbiology, University of Tokyo,
Bunkyo-ku, Yayoi 1-1-1, Tokyo 113, Japan*

(Received 21 February 1991)

ABSTRACT

Digestion of the total genomic DNA from rice *Oryza sativa* L. cv. C5924 with *Eco*RI generated an intense band of a DNA fragment of about 0.36 kb long. The DNA fragment cloned into pUC19 was used to hybridize with the total rice genomic DNA partially digested with *Eco*RI. A ladder of bands of DNA fragments with multiplied length of 0.36 kb was observed, demonstrating that this sequence occurs in tandem in the genome. The copy number of the sequence estimated by dot blot hybridization analysis was 2000–3000 copies per haploid genome from callus or seedling of C5924. This sequence was present in other *O. sativa* cultivars, such as Sasanishiki in 700–900 copies, Koshihikari in 3400–4300, and Nipponbare in 4600–6000 copies. Another rice species, *O. glaberrima*, also had this sequence in 540–680 copies, but four lines of foxtail millet had none. The DNA fragments containing the repeated sequences in Nipponbare were then cloned into λ EMBL3, and sequences of nine units consecutively repeated and an AT-rich sequence connected with them in a phage clone could be determined. Each repeating unit showed sequence divergency mostly by substitution of bases in a range from 3% to 7%, when compared with a 355-bp consensus sequence. Analyses of the substituted bases indicate that these are due to spontaneous mutations which occurred at random, after reiteration of a unit sequence by unequal crossing over events. Gene conversion within the repeated sequences might have further diversified their sequences.

1. INTRODUCTION

The genomes of higher plants, like those of other eukaryotes, contain families of repeated DNA sequences. Some are interspersed in a genome, whereas others are arranged in tandem and can be detected as satellite DNA (Flavell, 1982; Grellet et al., 1986). True or cryptic satellite DNAs have been identified in a number of plant species and some of these sequences have been determined: Examples are 185-bp repeat sequence from maize (Peacock et al., 1981), 172-bp sequence from mustard (Capesius, 1983), 59-bp sequence from *Vicia faba* (Kato et

* DDBJ/EMBL/GenBank accession number: The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number X55418, X55419, X55420, X55421 and X55422.

al., 1984), 180-bp sequence from *Arabidopsis thaliana* (Martinez-Zapater et al., 1986), 351-bp sequence from *Cucurbita pepo* (Leclerc and Siegel, 1987), 300-bp sequence from mung bean (Roy et al., 1988) and 162-bp sequence from tomato (Ganal et al., 1988). In most of these cases, one or two units of the repeats have been cloned and sequenced; thus little is known about the sequence divergence in the consecutively repeating units as well as the DNA sequences flanking the repeated sequences.

The rice genome, which we have been interested in, also has repetitive DNA sequences like other plants. Cot analysis has indicated that repetitive sequences consist of 50% of the rice nuclear genome, and 30% of them are the repetitive sequences with a higher GC content (63.9%) than that of the whole rice nuclear genomic DNA (38.5%) (Dhar et al., 1988). These include the sequences with very long stretches (4–20 kb), occupying more than half of the fraction of the repetitive sequences. Kikuchi et al. (1987) have identified two kinds of rice repetitive sequences which change their copy numbers during callus formation, plant regeneration and the growth process. One of these sequences was assumed to be derived from a segment of the chloroplast DNA which had been integrated into the rice nuclear genome. Zheng et al. (1987) and Cuzzoni et al. (1990) have also reported the amplification of another kind of repetitive sequences in cultured rice cell. Wu and Wu (1987) have identified a moderately repeated sequence in Indica type of *O. sativa*, which is transcribed in rice, showing some homology with wheat phenylalanine tRNA as well as with wheat 5S RNA.

In this paper, we report isolation, characterization and genomic organization of highly repetitive DNA sequences in Japonica type of *O. sativa* and *O. glaberrima*. The repeated sequences, 0.36 kb in one unit length, appear in tandem array and show sequence divergence. An AT-rich sequence showing no homology with the repeated sequences connected with them at a boundary. The repeating unit sequence is homologous with that reported by Wu and Wu (1987). We discuss how these repeated sequences have diverged in the rice genomes based on the organization of these sequences.

2. MATERIALS AND METHODS

Plants

The rice strains used were *Oryza sativa* L. cv. C5924 (obtained from Dr. K. Shono), *O. sativa* L. cv. Sasanishiki, *O. glaberrima* (from Dr. K. Okuno), *O. sativa* L. cv. Nipponbare, and *O. sativa* L. cv. Koshihikari (from Dr. H. Uchi-miya). C5924 was grown in a suspension culture using Murashige and Skoog media (Flow Laboratories). The foxtail millet strains used were *Setaria italica* Beauv. cv. Aso 1, *S. italica* Beauv. strain Taiwan 7, *S. italica* Beauv. strain Taiwan 7-9-1-7-2 and *S. italica* Beauv. strain Taiwan 15.

Bacterial plasmids

Plasmid vectors used for cloning were pUC19 (Yanisch-Perron et al., 1985) and pUC118 (Vieira and Messing, 1987) which were obtained from Takara Shuzo Co. Plasmid pREE61, carrying rice glutelin cDNA, was kindly supplied by Dr. F. Takaiwa. Plasmids pHOC154 and pHOC193 contained the *Eco*RI fragments of the repeated sequence in C5924 which were cloned into the *Eco*RI site of pUC19. Plasmid pH081X19 and its derivatives carrying the repeated sequences will be described below.

Isolation of the total rice genomic DNA

The total genomic DNA was isolated from leaves of 2 weeks-old seedlings or calli of a week-old suspension culture. Thirty grams of the tissue were frozen in liquid nitrogen, ground into powder with motor and pestle, suspended and homogenized in 100 ml of buffer A (15% sucrose, 50 mM Tris-HCl (pH 8.0), 50 mM EDTA, 250 mM NaCl and 10 μ l/ml mercaptoethanol), and then passed through two layers of Miracloth (Calbiochem Co.). The crude extract obtained was centrifuged at 20,000*g* for 5 min at 4°C, and the pellet was resuspended in 20 ml of ice-cold buffer B (15% sucrose, 50 mM Tris-HCl (pH 8.0), 50 mM EDTA and 0.1% diethyl pyrocarbonate). The extract was incubated at 70°C for 10 min, further incubated for 60 min on ice after addition of 2 ml of 5 M potassium acetate, and centrifuged at 20,000*g* for 15 min at 4°C. DNA in the supernatant was precipitated with 3 volumes of cold ethanol, and DNA fibers were wound up to a siliconized glass rod. The DNA was redissolved in 10 ml of buffer C (10 mM Tris-HCl (pH 7.5), 5 mM EDTA) and incubated with protease K (final concentration: 1.25 mg/ml) at 37°C for 60 min. DNA was extracted with phenol/chloroform twice and chloroform once, and precipitated by adding an equal volume of isopropanol. DNA fibers were again wound up onto a glass rod, washed by passing through 80% and 90% ethanol sequentially, dried, and resuspended in 1 ml of buffer C. DNA concentration was measured by absorbance at 260 nm or by agarose gel electrophoresis.

Southern blot hybridization

DNA was digested with *Eco*RI, *Kpn*I, *Bam*HI, *Hind*III or *Sal*I (Takara Shuzo Co.), and electrophoresed in a 1.2% agarose gel containing Tris-acetate/EDTA buffer (40 mM Tris-acetate (pH 8.0), 1 mM EDTA). Blotting and hybridization of the DNA were carried out under the stringent condition as described (Umeda and Ohtsubo, 1989). The probe DNA used was the insert C154 in pHOC154 labeled at its 5'-ends with [γ -³²P]ATP (Amersham, 185 TBq/mmol) using phage T4 polynucleotide kinase (Takara Shuzo Co.).

Construction of a genomic library

A genomic library was constructed as described by Maniatis et al. (1982). The total rice DNA from seedling leaves of *O. sativa* L. cv. Nipponbare was partially digested with *Sau*3AI (Takara Shuzo Co.) and ligated with the λ EMBL3 DNA (STRATAGENE) which had been digested with *Bam*HI. An *in vitro* packaging kit, Gigapack Gold (STRATAGENE), was used to obtain phage particles.

Nucleotide sequencing

DNA sequencing was carried out by the dideoxynucleotide chain termination method (Messing, 1983; Sanger et al., 1977). A 7-deaza sequencing kit (Takara Shuzo Co.) and Sequenase Ver. 2.0 (United States Biochemicals) were used.

Estimation of copy numbers of repeated sequences

A series of fifteen 2-fold serial dilutions of the rice genomic DNA was prepared. Each solution (20 μ l) was mixed with 60 μ l of 0.4 M NaOH, incubated at room temperature for 10 min, and chilled on ice for 10 min. Eighty μ l of 2 M ammonium acetate (pH 7.0) were added, and 50 μ l of the resulting solution were filtered through a nitrocellulose filter (Schleicher & Schuell) equilibrated in 1 M ammonium acetate (pH 7.0) using Minifold II (Schleicher & Schuell). The filter was then baked at 80°C under vacuum and hybridized with the 32 P-labeled probe DNA as described (Umeda and Ohtsubo, 1989). The individual slots were separated into squares of nitrocellulose, and radioactivities were counted using a liquid scintillation spectrometer (Packard TRI-CARB 300C). The radioactivity was plotted *versus* the quantity of DNA on the filter, and the slope of the line fitted by linear regression was evaluated as radioactivity hybridized with 1 μ g of genomic DNA. The probe DNAs used were *Eco*RI fragments of C154 and rice glutelin cDNA, generated from plasmids pHOC154 and pREE61, respectively. Probes with the same specific activity were prepared by labeling at their 5'-ends with [γ - 32 P]ATP (Amersham, 185 TBq/mmol) using phage T4 polynucleotide kinase (Takara Shuzo Co.).

3. RESULTS AND DISCUSSION

Isolation and characterization of a sequence highly repeated in tandem in rice

The total rice genomic DNA prepared from *O. sativa* L. cv. C5924 was digested with *Eco*RI and electrophoresed in a 5% polyacrylamide gel. A band of a fragment of about 0.36 kb in length was detected in the gel stained with ethidium bromide (Fig. 1), indicating that the fragment is derived from the sequence highly repeated in the rice genome. To characterize the sequence, the DNA fragment was isolated from the gel and cloned into the *Eco*RI site of pUC19. Two independent clones, pHOC154 and pHOC193, were obtained, and DNA inserts,

Fig. 1. An ethidium bromide-stained 5% polyacrylamide gel, showing the *Eco*RI digest of the total genomic DNA from *O. sativa* L. cv. C5924. A closed triangle indicates the band of the fragment of about 0.36 kb in length.

C154 and C193, in these clones were sequenced. The inserts were found to be 352 and 353 bp in length, respectively, with 17 base substitutions and 2 deletions/insertions when compared with each other (Fig. 2). The average GC content is 48.6% which is significantly higher than that of the total rice chromosomal DNA (38.5%) reported by Dhar et al. (1988). There are two short inverted repeats, but are no other notable characteristics in the sequence (Fig. 2). Recently, Wu and Wu (1987) have described moderately repeated sequences of rice *O. sativa* L. var. Labelle (Indica type), which were cloned from a rice cDNA library as well as

Fig. 2. Nucleotide sequences of inserts, C154 and C193. Bases altered are only shown in the sequence of C193. *d* indicates bases deleted. Arrows indicate stem-loop structures. Relevant restriction sites are also indicated.

from a genomic library. Our sequences showed about 90% homology with the sequences of Wu and Wu (1987).

To know the genomic organization of the repeated sequence, the total rice genomic DNA isolated from callus or seedling of C5924 was partially digested with *EcoRI*, electrophoresed in an agarose gel, and blotted onto a filter for Southern hybridization with the ^{32}P -labeled C154 as probe. As shown in Fig. 3A and B, a ladder of at least 20 bands of fragments was observed with the interval of the

Fig. 3. Autoradiograms, showing the results of Southern blot hybridization analysis using the ^{32}P -labeled C154 sequence as probe. (A) The total genomic DNA from callus of *O. sativa* L. cv. C5924. The DNA (8 μg) was digested with *EcoRI* (42, 30, 18, 6, or 3 units/ μg DNA) (lanes a~e lanes), *KpnI* (4.7 units/ μg DNA) (lane f), *BamHI* (4.7 units/ μg DNA) (lane g) or *HindIII* (4.7 units/ μg DNA) (lane h), respectively. (B) The *EcoRI*-digested total DNA from seedling of C5924 (lanes a~c), from callus of C5924 (lanes d~f), from seedling of Sasanishiki (lanes g~i), and from seedling of Nipponbare (lanes j~l). These DNAs were digested with 5 units (lanes a, d, g and j), 1 unit (lanes b, e, h and k), or 0.3 units (lanes c, f, i and l) of *EcoRI* per μg DNA, respectively. Numbers indicate positions of monomers (0.36 kb) (1), dimers (0.72 kb) (2), etc. which could be identified from the *HindIII* digest of the λ DNA used as length standards.

length 0.36 kb. This indicates that the C154 sequence is repeated as a unit in tandem more than twenty times in C5924. A similar ladder of bands was observed also in the total DNA isolated from other cultivars of *O. sativa*, Sasanishiki and Nipponbare (Fig. 3B), indicating that the C154 sequence is also repeated in tandem in each cultivar of *O. sativa*.

Note here that even upon extensive digestion of the total DNA with *Eco*RI, bands of fragments with length of 3 to 6 kb which differed in intensity were observed (Fig. 3A and B). These probably correspond to the fragments consisting of more than 9 copies of the unit sequence and are generated due to the absence of the *Eco*RI site in some of the repeating units (Fig. 3A and B). The band patterns of the *Eco*RI-resistant fragments are different among C5924, Nipponbare and Sasanishiki (Fig. 3B), indicating that the genome of each cultivar has the repeated sequences with unique sequence divergency and reiteration frequency.

As shown in Fig. 2, the C154 sequence has neither the *Kpn*I, *Bam*HI nor *Hind*III site. However, digestion of C5924 with *Bam*HI or *Hind*III generated several bands of DNA fragments hybridized with the probe, although most of them were larger than 5 kb in size (Fig. 3A, lanes g and h). This suggests that a small fraction of the repeated sequences contain several units with the cleavage sites for these restriction enzymes. The *Hind*III digest contained the DNA fragments with the interval of two unit size (Fig. 3A, lane h). This suggests that *Hind*III sites occur every two units in certain area of the repeated sequences.

Copy numbers of the repeated sequence

To determine approximate copy numbers of the repeated sequence in genomes of several cultivars or lines of rice, each total genomic DNA was serially diluted, denatured with alkaline, blotted onto a nitrocellulose filter, and hybridized with the ³²P-labeled C154 fragment as one probe, and with the ³²P-labeled rice glutelin cDNA in plasmid pREE61 as the other probe. The radioactivity measured was plotted *versus* the quantity of DNA on a filter to calculate a value per 1 µg of DNA. The rice glutelin gene is present in four to five copies per haploid genome (Takaiwa et al., 1987), and thus the copy number of a repeated sequence could be estimated from the value for the glutelin gene. As summarized in Table 1, copy numbers of the repeated sequence in seedling of C5924 was 1900 to 2400 as many as those in callus. The copy number of the repeated sequence, however, varied in each line of Japonica type of *O. sativa* (Table 1): Sasanishiki showed only 700 to 900 copies per haploid genome; Koshihikari and Nipponbare showed 3400 to 5800 copies per haploid genome. Another rice cultivar *O. glaberrima* contained 540–680 copies per haploid genome (Table 1), which are significantly lower than those in *O. sativa*. Wu et al. (1988) have described that *O. sativa* L. var. Labelle (Indica type) and *O. glaberrima* contain about 2000 and 50 copies of the repetitive sequence per haploid genome, respectively. The discrepancy in the copy num-

Table 1. Copy numbers of the repeated sequence in rice

Strain	Source	Copy number/haploid
<i>O. sativa</i>		
C5924	seedling	1900~2400
C5924	callus	2300~2900
Sasanishiki	seedling	700~ 900
Koshihikari	seedling	3400~4300
Nipponbare	seedling	4600~5800
<i>O. glaberrima</i>	seedling	540~ 680

bers in *O. glaberrima* between ours and Wu et al.'s may be due to the difference of lines used to estimate them.

Besides rice plants, DNAs from four strains of foxtail millet were analyzed by dot blot hybridization as above, but no hybridized sequences were detected. This may suggest that the C154 sequence is specifically present in rice. Probably, the repeated sequences in rice have been derived from a common ancestor sequence. The copy number would have been increased by multiple unequal crossing over events whose frequencies differed in each line or cultivar of rice in the process of evolution.

Cloning and sequencing the segment containing tandem repeat sequences

To further analyze the genomic organization of the repeated sequences, we constructed a λ EMBL3 genomic library of Nipponbare, screened 5×10^4 clones, and selected five independent clones which hybridized with the probe used above. The phage DNA from each clone was then partially digested with *EcoRI* or *SalI* and subjected to Southern blot hybridization analysis. As shown in Fig. 4, three (81, 101 and 1020) out of five clones clearly showed a ladder of many bands of DNA with multiplied length of 0.36 kb. This indicates that these clones carry tandem repeats with various numbers of a unit sequence. Among these clones obtained, we chose clone 81 for the analysis of the tandem repeat sequences.

Digestion of clone 81 DNA with *XmaI* generated one fragment containing a tandem repeat of several unit sequences (data not shown). The fragment was cloned into the *XmaI* site of pUC118 to obtain plasmid pH081X19 (Fig. 5). pH081X19 was then partially digested with *XbaI* or *SalI* and self-ligated to give 17 sets of deletion derivatives (D1–D17 in Fig. 5) carrying various portions of the inserted fragment. Using these derivatives, we determined the nucleotide sequence of the 3005-bp region shown in Fig. 6A. The region was found to contain sequences of about 355 bp which were repeated nine times in tandem (See R1–R9 in Fig. 6B), except for a non-unit sequence, which was connected with the tandem repeat (Figs. 5 and 6A). The region sequenced contained no open reading frames with a significant size, indicating that these sequences do not encode proteins.

Fig. 4. Autoradiograms, showing the results of Southern blot hybridization analysis of five λ phage clones. Panels A and B are DNAs partially digested with *Eco*RI (10 units/ μ g DNA) and *Sal*I (10 units/ μ g DNA), respectively. Lanes a~e are DNAs from λ phage clones, 410, 610, 81, 101 and 1020, respectively. Numbers are sizes in bp of the *Hind*III digest of λ DNA used as standards.

Fig. 5. Structures of plasmid pH081X19 carrying an insert containing tandem repeat sequences and its deletion derivatives. Open arrows indicate the repeating units in the insert. The solid lines with D1~D17 indicate the portions carried by deletion derivatives from pH081X19. Solid arrowheads and circles indicate cleavage sites for *Xba*I and *Sal*I, respectively. Thick solid lines flanking the insert DNA are the sequences of vector plasmid pUC118.

Fig. 6. Nucleotide sequences of the insert in pH081X19. (A) The nucleotide sequence determined. Nine units (R1 to R9) of the repeated sequence and an AT rich sequence connected with them are indicated. Asterisks in the non-unit sequence indicate the bases homologous to the homeobox sequence of *Drosophila*. Hyphens in the sequences of R6 and R8 indicate positions of bases which are possibly present in the insert, but not sequenced (see below). *d* indicates bases possibly deleted (see below). (B) Alignment of the repeating unit sequences together with the consensus sequence (top line). The repeated sequences C154 and C193 isolated from C5924 and the consensus sequence (W & W) of those reported by Wu and Wu (1987) are also aligned for comparison. Nucleotides which are

Sequence relations among the repeating unit sequences

There is heterogeneity in the nucleotide sequences of the nine repeating units. However, their sequences are homologous with one another in a range from 86 to 97.5% (Table 2), which is relatively higher than those of the repeated sequence in the animal system, in which the alphoid sequences, for example, show 75–80%

Table 2. Comparison of the nucleotide sequences between two repeating units

	R2	R3	R4	R5	R6	R7	R8	R9	Consensus Sequence
R1	34 (9.6)	30 (8.5)	32 (9.0)	35 (9.9)	— [4.1]	31 (8.7)	— [8.9]	— [11.8]	20 (5.6)
R2		31 (8.7)	30 (8.5)	36 (10.1)	— [5.0]	35 (9.9)	— [7.4]	— [14.3]	21 (5.9)
R3			24 (6.8)	31 (8.7)	— [3.3]	26 (7.3)	— [6.4]	— [10.0]	14 (3.9)
R4				35 (9.9)	— [2.5]	18 (5.1)	— [5.0]	— [8.5]	17 (4.8)
R5					— [6.6]	34 (9.6)	— [6.4]	— [12.5]	22 (6.2)
R6						— [0.8]	— [2.5]	— [5.5]	— [0.83]
R7							— [4.0]	— [11.4]	13 (3.7)
R8								— [4.9]	— [2.4]
R9									— [7.0]
C154									25 (7.0)
C193									22 (6.2)
W & W									12 (3.4)

R1~R9 are the repeating unit sequences in *O. sativa* L. cv. Nipponbare; C154 and C193 are those from *O. sativa* L. cv. C5924; W & W is the consensus sequence of the repetitive sequences reported by Wu and Wu (1987). Numbers not parenthesized are changes between one complete unit sequence and another one or the consensus sequence, 355 bp in length; the entire sequences of units R6, R8 and R9 were not determined, and thus such numbers are not shown. The percent values of base changes are indicated by numbers in parentheses () and [], in which values in the latter are calculated by comparing the determined sequence of R6, R8 or R9 with that of the corresponding region in each of the other units.

different from the consensus sequence are only shown. Note that the sequences C154 and C193 have insertions between positions 234 and 235 and between 124 and 125, respectively. Dots in the sequences of R6 and R8 indicate positions of bases which are possibly present in the insert, but not sequenced. d indicates bases deleted. Relevant restriction sites are also indicated.

homology (Darling et al., 1982; Jørgensen et al., 1986). A consensus sequence deduced from the nine repeating units, R1–R9, is depicted in Fig. 6B. Most of the units have a similar number of base changes when compared with the consensus sequence (Table 2). Base changes do not seem to be clustered in any restricted regions (see Fig. 6B). Almost all the changes are single-base substitutions in which transition and transversion occur evenly, although some changes occur more frequently than others (A to G, G to A, C to T, and T to A; Table 3). Some substitutions caused differences in restriction sites: R1 and R2 do not have the *SalI* sites, but the other seven units do; seven units have two *HinfI* sites, but the others have a single one; all the units have no *BamHI* and *HindIII* sites, which were present in some units in the repeated sequences in the C5924 DNA, as described in the previous section.

Although substitutions occur at 86 positions on the 355 bp consensus sequence upon comparison of it with each of the repeating unit sequences, the substitution at either one of 65 positions occurs in only one of the nine units, and the substitution at either one of the other positions occurs in two or more of the nine units, such that the nucleotides substituted at a position are usually different from one another (see Fig. 6B). This indicates that these substitutions have occurred at random by point mutations. McCutchan et al. (1982) have reported that the closer the alpha-satellite sequences in the African green monkey DNA to the junctions with the non alpha-satellite DNA, the more abundant the divergent forms of the sequences are, when compared to the average substitution frequency of such forms within total alpha-satellite DNA. However, we did not observe such a tendency in our sequences, since the frequency of divergency is almost the

Table 3. Types and numbers of base substitutions

Bases in consensus sequence	Bases substituted	Number of substitutions	Total number of	
			transitions	transversions
A	G	18		
	C	4	18	9
	T	5		
G	A	16		
	C	6	16	19
	T	13		
C	A	8		
	G	5	26	13
	T	26		
T	A	11		
	G	5	4	16
	C	4	(Total) 64	57

All the sequences in R1 to R9 (see Fig. 6) were subjected to analysis.

same among the nine consecutive units (see Table 2).

The repeated sequences, C154 and C193, isolated from C5924 showed more than 93% homology with the consensus sequence (Table 3 and Fig. 6B), but interestingly, they have four deletions occurring at the same positions (Fig. 6B). This indicates that the repeated sequences in C5924 have their own consensus sequence which is slightly different from that in Nipponbare. This implies that the repeated sequences have been generated from its own consensus type of sequence in each cultivar by unequal crossing over events. The sequence divergence in each unit may have occurred by spontaneous mutation after reiteration of the sequences. Gene conversion events might have occurred also to diversify the repeating unit sequences.

Notes on characteristics of the sequence joined with the repeated sequence at a boundary

The non-unit sequence joined with our repetitive sequences has high AT content (75%) (Fig. 6A). Whether this is a unique sequence or another repetitive sequence is under investigation. We observed partial homology to the consensus sequence of homeoboxes (5'-TCAATTAAAT-3') of *Drosophila* (see sequences with asterisks in Fig. 6A). The significance of such sequence homology remains unknown.

We wish to thank Drs. K. Shono, H. Uchimiya, and K. Okuno for providing rice strains; Dr. T. Kokubu for providing foxtail millet strains; Dr. F. Takaiwa for providing glutelin cDNA clone.

This work was supported by a Grant-in-Aid for Scientific Research from Ministry of Education, Science and Culture of Japan, and by a grant Pioneering Research Project in Biotechnology from Ministry of Agriculture, Forestry and Fisheries of Japan. M.U. is the recipient of a JSPS fellowship for Japanese Junior Scientists.

REFERENCES

- Capesius, I. (1983). Sequence of the cryptic satellite DNA from the plant *Sinapis alba*. *Biochim. Biophys. Acta* **739**, 276-280.
- Cuzzoni, E., Ferretti, L., Giordani, C., Castiglione, S. and Sala, F. (1990). A repeated chromosomal DNA sequence is amplified as a circular extrachromosomal molecule in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* **222**, 58-64.
- Darling, S. M., Crampton, J. M. and Williamson, R. (1982). Organization of a family of highly repetitive sequences within the human genome. *J. Mol. Biol.* **154**, 51-63.
- Dhar, M. S., Dabak, M., Gupta, V. S. and Ranjekar, P. K. (1988). Organization and properties of repeated DNA sequences in rice genome. *Plant Science* **55**, 43-52.
- Flavell, R. (1982). Amplification, deletion and rearrangement: Major sources of variation during species divergence. In: *Genome Evolution*, (eds.: G. A. Dover and R. B. Flavell), pp. 301-324. Academic Press, New York.
- Galli, G., Hofstetter, H. and Birnstiel, M. L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* **294**, 626-631.
- Ganal, M. W., Lapitan, N. L. V. and Tanksley, S. D. (1988). A molecular and cytogenetic survey of

- major repeated DNA sequences in tomato (*Lycopersicon esculentum*). *Mol. Gen. Genet.* **213**, 262–268.
- Grellet, F., Delcasso, D., Panabieres, F. and Delseny, M. (1986). Organization and evolution of a higher plant aliphoid-like satellite DNA sequence. *J. Mol. Biol.* **187**, 495–507.
- Jørgensen, A. L., Bostock, C. J. and Bak, A. L. (1986). Chromosome-specific subfamilies within human aliphoid repetitive DNA. *J. Mol. Biol.* **187**, 185–196.
- Kato, A., Yakura, K. and Tanifuji, S. (1984). Sequence analysis of *Vicia faba* repeated DNA, the *FokI* repeat element. *Nucleic Acids Res.* **12**, 6415–6426.
- Kikuchi, S., Takaiwa, F. and Oono, K. (1987). Variable copy number DNA sequences in rice. *Mol. Gen. Genet.* **210**, 373–380.
- Leclerc, R. F. and Siegel, A. (1987). Characterization of repetitive elements in several *Cucurbita* species. *Plant Mol. Biol.* **8**, 497–507.
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Martinez-Zapater, J. M., Estelle, M. A. and Somerville, C. R. (1986). A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **204**, 417–423.
- McCutchan, T., Hsu, H., Thayer, R. E. and Singer, M. F. (1982). Organization of African green monkey DNA at junctions between α -satellite and other DNA sequence. *J. Mol. Biol.* **157**, 195–211.
- Messing, J. (1983). New M13 vectors for cloning. *Methods Enzymol.* **101**, 20–78.
- Peacock, W. J., Dennis, E. S., Rhoades, M. M. and Pryor, A. J. (1981). Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proc. Natl. Acad. Sci. USA* **78**, 4490–4494.
- Roy, P., Bhattacharyya, N. and Biswas, B. B. (1988). Isolation, characterization and sequencing of a novel repetitive DNA from the mung bean *Vigna radiata*. *Gene* **73**, 57–66.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Takaiwa, F., Kikuchi, S. and Oono, K. (1987). A rice glutelin gene family—A major type of glutelin mRNAs can be divided into two classes. *Mol. Gen. Genet.* **208**, 15–22.
- Umeda, M. and Ohtsubo, E. (1989). Mapping of insertion elements IS1, IS2 and IS3 on the *Escherichia coli* K-12 chromosome: Role of the insertion elements in formation of Hfrs and F' factors and in rearrangement of bacterial chromosomes. *J. Mol. Biol.* **208**, 601–614.
- Vieira, J. and Messing, J. (1987). Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**, 3–11.
- Wu, T. and Wu, R. (1987). A new rice repetitive DNA shows sequence homology to both 5S RNA and tRNA. *Nucleic Acids Res.* **15**, 5913–5923.
- Wu, R., Zhao, R., Wu, T. and Xie, Y. (1988). Genome-type specific repetitive DNA sequence in rice. *Rice Genetics News Letter* **5**, 136–137.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.
- Zheng, K. L., Castiglione, S., Biasini, M. G., Biroli, A., Morandi, C. and Sala, F. (1987). Nuclear DNA amplification in cultured cells of *Oryza sativa* L. *Theor. Appl. Genet.* **74**, 65–70.