Theor Appl Genet (1989) 78: 201–209

Genome-specific repetitive sequences in the genus *Oryza*

X. Zhao, T. Wu, Y. Xie and R. Wu*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received August 8, 1988; Accepted January 24, 1989
Communicated by A.L. Kahler

Summary. Repetitive DNA sequences are useful molecular markers for studying plant genome evolution and species divergence. In this paper, we report the isolation and characterization of four genome-type specific repetitive DNA sequences in the genus *Oryza*. Sequences specific to the AA, CC, EE or FF genome types are described. These genome-type specific repetitive sequences will be useful in classifying unknown species of wild or domestic rice, and in studying genome evolution at the molecular level. Using an AA genome-specific repetitive DNA sequence (pOs48) as a hybridization probe, considerable differences in its copy number were found among different varieties of Asian-cultivated rice (*O. sativa*) and other related species within the AA genome type. Thus, the relationship among some of the members of AA genome type can be deduced based on the degree of DNA sequence similarity of this repetitive sequence.

Key words: Rice – Repetitive sequences – *Oryza*

Introduction

One distinguishing feature of the genome of most higher eukaryotes is the presence of large amounts of repetitive DNA. In higher animals, many repetitive sequences are well-characterized in terms of their length, abundance, chromosomal distribution, and even nucleotide sequence (Singer 1982; Waye and Willard 1985). Recently, a number of studies on repetitive sequences have been reported in higher plants, such as rye (Bedbrook et al. 1980), wheat and barley (Dennis et al. 1980), *Scilla* (Deumling 1981), maize (Peacock et al. 1981; Viotti et al. 1985), mustard (Capesius 1983), broad bean (Kato et al. 1984), radish (Grellet et al. 1986), *Arabidopsis thaliana* (Martinez-Zapater et al. 1986), flax (Cullis and Cleary 1986), and rice (Wu and Wu 1987). One important conclusion drawn from these studies is that repetitive sequences appear to change rapidly during evolution. Although the function and origin of these sequences remain to be elucidated, highly repetitive DNA sequences are useful in studying genome evolution at the molecular level.

Rice belongs to the genus *Oryza,* which includes 20 wild species and two cultigens (*Oryza sativa* and *Oryza glaberrima* steud) (Chang 1984). Rice has been classified into six genome types: AA, BB, CC, DD, EE, and FF (Hiroka 1984). The rice nuclear genome contains approximately 50% repetitive DNA as determined by C_0t analysis (Deshphane and Ranjekar 1980; Zhou 1986). With a view to analyzing the divergence of rice species and the evolutionary relationships among related genomes, we report here the isolation and characterization of four genome-type specific repetitive sequences. In addition, the AA genome-specific repetitive sequence, pOs48, was analyzed in 31 different rice entries by hybridization analysis. Considerable differences in copy number were observed.

Materials and methods

Plant materials and growth conditions

Thirty-seven rice entries were used in these experiments, and their origins and genome types are tabulated in Table 1. All wild rice and IR-derived varieties were obtained from T. T. Chang of the International Rice Research Institute (IRRI) through R. Coffman. *O. sativa* var. Labelle was obtained from C. N. Bollich, and *O. sativa* var. Calrose 76 was obtained from N. Rutger. All rice plants were grown in pots in greenhouses (26° ± 3 °C;

* To whom reprint requests should be sent

Table 1. Rice entries, genome type and origin

Rice entries	Type	Genome	Origin	IRRI accession No.
O. sativa				
Taipei 309	*Sinica* (*Japonica*)	AA	China	42576
Tainung 67	*Sinica*	AA	China	47743
Fujiminori	*Sinica*	AA	Japan	10901
Fujisaka 5	*Sinica*	AA	Japan	244
Nipponbare	*Sinica*	AA	Japan	12731
Calrose 76	*Sinica*	AA	U.S.A.	
Belle Patna	*Indica*	AA	U.S.A.	BL-PT
Dawn, CI9534	*Indica*	AA	U.S.A.	109
M202	*Indica*	AA	U.S.A.	
MR365	*Indica*	AA	U.S.A.	
CP231	*Indica*	AA	U.S.A.	
Lemont	*Indica*	AA	U.S.A.	
Labelle	*Indica*	AA	U.S.A.	
DGWG	*Indica*	AA	China	
Peta	*Indica*	AA	Philippines	
TKM6	*Indica*	AA	India	237
Tetep	*Indica*	AA	Vietnam	PI 280682
CR94-13	*Indica*	AA	India	15791
IR26	*Indica*	AA	IRRI	RYT 1806
IR36	*Indica*	AA	IRRI	
IR54	*Indica*	AA	IRRI	RYT 3348
IR1561-228-3-3	*Indica*	AA	IRRI	4884
IR1737-19-7-8-3	*Indica*	AA	IRRI	32645
IR34583-19-3-3	*Indica*	AA	IRRI	HB 469
Bulu Dalam	*Javanica*	AA	Indonesia	HB 1203
O. glaberrima	African cultigen	AA	Africa	103971
O. meridionalis	Wild	AA	Australia	101147
O. rufipogon	Wild	AA	Asia, U.S.A.	103823
O. glumaepatula	Wild	AA	S. America	100894
O. nivara	Wild	AA	Asia	102165
O. longistaminata	Wild	AA	Africa	100930
O. punctata	Wild	BB, BBCC[a]	Africa	103897
O. officinalis	Wild	CC	Asia	103286
O. alta	Wild	CCDD[a]	U.S.A.	100161
O. latifolia	Wild	CCDD[a]	U.S.A.	100165
O. australiensis	Wild	EE	Australia	101467
O. brachyantha	Wild	FF	Africa	101236

[a] Tetraploid ($4n=48$). All other entries are diploid ($2n=24$)

50%–65% relative humidity; 14 h light, 10 h dark, light intensity 160 μmol M^{-2} sec^{-1} maximum at the pot level).

Isolation of rice total DNA

Large-scale preparation of total rice DNA was carried out according to the following procedure: ten grams of finely chopped leaf tissue from plants (usually around 2 months old) were frozen in liquid nitrogen, powdered, and then quickly homogenized using a cold mortar and pestle in 30 ml of homogenization buffer (1 × SSC, 2.5% SDS, 0.25% sarkosyl, 100 μg/ml proteinase K). The homogenate was incubated at 65 °C for 10 min. An equal volume of buffered phenol (pH 8.0) was added, mixed by gentle inversion and centrifuged at 3,000 × g for 10 min. DNA in the supernatant was precipitated by adding 2 vol. of ethanol and centrifuging for 10 min at 12,000 × g at 4 °C. The DNA pellet was resuspended in 1 ml of TE (10 m*M* TRIS-HCl, 1 m*M* EDTA, pH 8.0) and digested with 50 μg/ml RNase A at 37 °C for 1 h. After extraction twice with equal volume of buffered phenol, once with equal volume of phenol : chloroform (1 : 1), and then once with chloroform, the DNA was precipitated with 2 vol. of ethanol. The DNA was immediately spooled with a glass hook, resuspended in TE, and stored at 4 °C. On the average, 1 mg of DNA was obtained from 10 g of fresh leaf tissue. DNA concentration was estimated from the absorbance at 260 nm and calibrated by agarose gel electrophoresis, using a defined amount of lambda DNA as a standard.

Cloning of repetitive DNA

The isolation and cloning of a repetitive sequence from *O. australiensis* was carried out as follows. Total rice DNA was completely digested with EcoRV, fractionated electrophoretically on a 0.8% agarose gel, and stained with ethidium bromide. A prominent band corresponding to DNA fragments of approximately 500 bp was eluted from the gel and cloned into the HincII site of pUC13 by blunt-end ligation (Maniatis et al. 1982). The ligation mixture was used to transform *E. coli* JM101 cells, and a plasmid pOa4 containing a 511-bp repetitive sequence was obtained. The cloning of repetitive sequences from *O. brachyantha* and *O. officinalis* were carried out as follows. Total DNA from these two wild species was digested with HincII and EcoRV, respectively. The bands corresponding to DNA fragments of approximately 340 bp were eluted from the gel and cloned into the HincII site of pUC13. The other steps were the same as described above.

Slot-blot hybridization

In order to quantify the copy numbers of repetitive DNA sequences in different rice genomes, we applied defined amounts of total rice DNA and recombinant plasmid DNA to a nitrocellulose filter through a slot-blot template. The DNA samples were denatured in 0.3 *M* NaOH and 6 × SSC at 80 °C for 15 min, transferred to 0 °C, and then neutralized with an equal volume of 2 *M* ammonium acetate. Filters were preincubated in 6 × SSC. Hybridization solution contained 50% formamide, 5 × SSC, 5 × Denhardt's solution, 0.1% SDS, 100 μg/ml denatured salmon sperm DNA, and ^{32}P-labeled cloned repetitive DNA (0.4 μg, 10^8 cpm/μg). The molar ratio of the probe over the filter-bound DNA was at least 100. After overnight hybridization at 42 °C, the filter was washed sequentially at 55 °C, 60 °C, 65 °C, and 70 °C in 0.2 × SSC, 0.1% SDS. After washing (3 times) at each temperature, the filters were exposed to X-ray films for at least two different lengths of time, and the films were scanned with a Quick Scan (Flur-vis) densitometer. Several known quantities of DNA were also spotted on the same filter for constructing a standard curve. Variation between duplicate experiments were approximately ±4% (data not shown). Copy numbers of repetitive sequences were calculated according to Rivin et al. (1986).

Nick translation and genomic blot hybridization

Repetitive sequence fragments used as probes were separated from recombinant plasmids after digestion with restriction endonuclease and fractionated by gel electrophoresis. ^{32}P-labeled repetitive DNA fragments were prepared according to the methods of Maniatis et al. (1982). Digested genomic DNA samples were fractionated by electrophoresis in 0.8% agarose gels. DNA fragments were transferred to a NYTRAN filter which was then hybridized to ^{32}P-labeled probes at 42 °C in the same hybridization solution as that used in the slot-blot hybridization.

A

```
        10         20         30         40         50         60         70         80         90        100
ATCTCTCCAA AGGAGGGCAA ATTCCATCTT GATCACTCAC ATCCCACTCC ATGTTTCATA GCAAACCCGA AAACTACCTT TATAACTACC CAGTTACGGA
       110        120        130        140        150        160        170        180        190        200
ATAGCTGTTT GGTAGTCCCA AAGTAGGCTA CTACACATGT TGGAATGCAT GGTGATCTCA GGTCTAAGGG CTTTGTACCA ACACTACCTG AGACCAACTA
       210        220        230        240        250        260        270        280        290        300
ATGGCATATT AATTCCTCTA GTGTCTCATG GTGGGTCTAT CCAACAATGT GTTCTCTAAC ACATAAGTCC ATGAAAATTG ATTTGGTATC TCCATACTCA
       310        320        330        340        350        360        370        380        390        400
TGATCTATGA GACATGATCA TCAATCAATA AACATGCTGA TCTTAGAATC ATATTAGTTC CAATAATATC ATATATGATC AGTGATCATT TAGAAATAGA
       410        420        430        440        450        460        470        480        490        500
TTCATACATA TATAGTCTCA TAAATAAGTC ACATACTATT GATCAATACA AGATGTCTAT TGATGGAACT GAATAACACT TATTCATAAG AACATAAACA
       510
TTGACCATGA T
```

B

```
        10         20         30         40         50         60         70         80         90        100
ATCTAGCAAC GGAGACATTC TTCAAGGCCC TGAGTGGCAC TCATGCGTGC TACGTCGTGG AACATGCAAC CTTTTCGGGG GGAATGTTTA GAAACCTGGT
       110        120        130        140        150        160        170        180        190        200
GAATAAACAC ATTCCTCACC ATGGTTGGCA CATTCCTTTG GATATGCGAT GCGTTTTAGG GCAATGTCCT TAATGTTTCG ATGGAAATAA CCCCACAAGA
       210        220        230        240        250        260        270        280        290        300
AGTTAATCTG GTCAGTTGAG GGCCCTTCTA CACTGAGCAC GTCAGGTTTA GGAAATAATT TATGGTAGCG AGGAAGGAAA AGAACGACAT TAGACGAGCT
       310        320        330        340        350        360
AAAAAACTCT CTTCAAGTTC CATGTCTTCA TGCATTTCCA TCATAACGGA GTTCGTCTCG ACTGAT
```

C

```
        10         20         30         40         50         60         70         80         90        100
AACCAAGATT CAATTATTAA CTTCGAAATA ACTACACAAA TTGTGTATCC ACTTGGTTTA CACTTCGTGC TCGAACACAA GTTCGTTTGC GAAAGACCCG
       110        120        130        140        150        160        170        180        190        200
ACCATGAAGG GTACTCTTGG ATTCATTCCG TCATACTCTT CACCATTCAT ACAAAGTTGA ATAAGATTCG ACTATTAACT TCTTGATAAC TACAGAAATT
       210        220        230        240        250        260        270        280        290        300
GTATATCAAC TTGGTTTACA CTTTGGTCTC GAACACAACT TCGGTTGCGA AACTCTCGGG CTGGAAGGGT GACCTTGGAT TCATTCCATC ACCAACTTCA
       310
CCATTCATAC AAAGTC
```

Fig. 1A–C. DNA sequence of three repetitive DNAs from wild rice. **A** Sequence of a 511-bp pOa4 isloated from *O. australiensis*. **B** Sequence of a 366-bp pOo2 isolated from *O. officinalis*. **C** Sequence of a 316 bp-pOb1 isolated from *O. brachyantha*, which is a dimer of 159-bp repeats with 80% homology between monomer units

Results

Identification and organization of genome-type specific repetitive DNA sequences in rice

We were interested in studying repetitive DNA sequences which are unique to the different genome types of rice. To begin this analysis, three repetitive DNAs, designated pOa4, pOo2, and pOb1, were cloned from three wild rice species. The sequence of these repetitive DNAs are shown in Fig. 1. A fourth repetitive DNA, pOs48 (previously referred to as RC48) was isolated from *Oryza sativa* L. var. Labelle (Wu and Wu 1987).

The four cloned repetitive DNAs (pOs48, pOa4, pOo2 and pOb1) were used as probes for genomic blot hybridization on 37 rice entries, including domestic varieties and wild species of the repesentative genome types. DNAs from four types of rice (*Oryza sativa* var. Labelle, *O. australiensis*, *O. officinalis*, and *O. brachyantha*) were separately digested with several different restriction enzymes. Using a given cloned repetitive DNA as a probe, the hybridization patterns of these four rice species were found to be different. When probed with labeled pOs48, two bands were observed on EcoRI-digested Labelle DNA (Fig. 2A, lane I), however, a ladder pattern was obtained from Pst 1-digested DNA (Wu and Wu 1987). With pOa4, only one strong band in the 511 bp position was found in DNA from *O. australiensis* (Fig. 2A, lane II). A ladder pattern was found with EcoRV-digested *O. officinalis* DNA probed with pOo2, and HincII-digested *O. brachyantha* DNA probed with pOb1 (Fig. 2A, lanes III and IV). The ladder pattern, with observed band representing monomer, dimer, trimer, etc., is that expected for a tandemly repeated sequence. Since the genomic DNA was completely digested with restriction enzymes, the observed ladder of repeats was probably the result of either mutations or methylation, which altered the frequency of restriction sites in the genomic copies of the corresponding repetitive DNA sequences. To substantiate this conclusion and to determine to what extent these sequences are present as tandem repeats, genomic DNA from Labelle (AA genome) was digested with restriction enzymes HpaII and MspI, which recognize the same sequence GGCC. HpaII is sensitive to methylation, but MspI is insensitive to methylation. Southern blots were probed with pOs48 plasmid. Little or no detectable digestion was observed with HpaII, indicating that methylation occurred at most of the HpaII/MspI sites, while MspI digestion gave a ladder pattern (data not shown). The ladder pattern from MspI digestion is due to

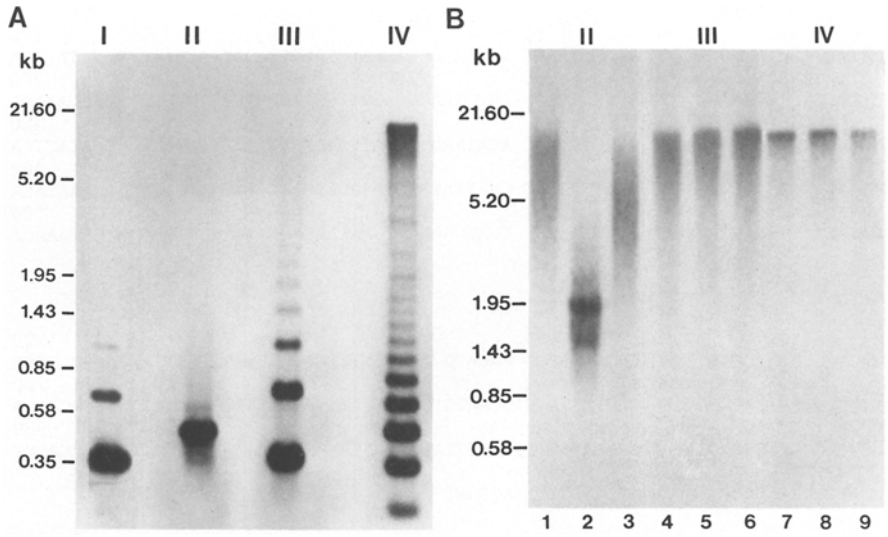

Fig. 2 A and B. Genomic blot analysis (Southern blot) of different rice species with repetitive DNA probes. **A** Total rice DNA (10 μg/lane) was digested with different restriction enzymes and fractionated electrophoretically on a 0.8% agarose gel. All DNA samples showed approximately the same intensity after staining the gel with ethidium bromide and visualizing under UV light (data not shown). *Hin*dIII and *Eco*RI digested λDNA were run as size markers. The DNA samples were transferred to a nitrocellulose filter for hybridization using different ^{32}P-labeled rice repetitive DNA sequences as probes. *Lane I:* DNA from *O. sativa* var. Labelle was digested with *Eco*RI and probed with pOs48. *Lane II:* DNA from *O. australiensis* was digested with *Eco*RV and probed with pOa4. *Lane III:* DNA from *O. officinalis* was digested with *Eco*RV and probed with pOo2. *Lane IV:* DNA from *O. brachyantha* was digested with *Hin*cII and probed with pOb1. Exposure times of the hybridized filters were 120 min, 30 min, 5 min, and 30 min, which correspond to *lanes I, II, III* and *IV*, respectively. **B** Total rice DNA was digested with different restriction enzymes and fractionated electrophoretically on a 0.8% agarose gel. *Hin*dIII and *Eco*RI digested λDNA were run as markers. *Lanes 1, 2* and *3* (*II*): DNA samples (5 μg/lane) from *O. australiensis* were digested with *Bgl*II, *Bst*NI and *Hin*dIII, respectively, and probed with pOa4. *Lanes 4, 5,* and *6* (*III*): DNA samples (9 μg/lane) from *O. officinalis* were digested with *Bam*HI, *Bgl*II and *Hind*III, respectively, and probed with pOo2. *Lanes 7, 8* and *9* (*IV*): DNA samples (10 μg/lane) from *O. brachyantha* were digested with *Bam*HI, *Bgl*II and *Hin*dIII, respectively, and probed with pOb1. Exposure times of the X-ray film to different filters were 6 h, 30 min, and 60 min for samples II, III and IV, respectively

random point mutations as shown by DNA sequence analysis of nine Os48 related clones. We found that one out of every three HpaII/MspI sites has a base substitution which changes the recognition sequence GGCC to AGCC (Wu and Wu 1987).

Genome DNA from *O. australiensis, O. officinalis,* and *O. brachyantha* was digested with several restriction enzymes which are cytosine methylation insensitive and do not cut within the repeats. The genomic blots were then probed with pOa4, pOo2, and pOb1, respectively. Results in Fig. 2 B indicate that the CC and FF genomic-specific DNA are largely present as tandem repeats of up to 16 kb in length. In this experiment, the largest DNA found in the ethidium bromide stained gel is about 16 kb (data not shown). The EE genome-specific DNA is also present, at least in part, as tandem repeats from 5 kb to 16 kb in total DNA digested with BglII and HindIII (Fig. 2 B, lanes 1 and 3). However, when the DNA was digested with BstNI (Fig. 2 B, lane 2), two lower molecular weight bands (2 kb and 1.3 kb) were observed. The largest DNA found in the BstNI-digested DNA was about 5 kb in the ethidium bromide stained gel (data not shown).

The abundance of repetitive sequences in the rice genome

The copy number of these four repetitive sequences, pOs48, pOa4, pOo2, and pOb1, in the rice genome was determined by quantitative slot-blot hybridization. A dilution series of each cloned repetitive DNA was used as a copy number standard (Fig. 3, columns A and C). Three different amounts of genomic DNA were loaded on the same filter (Fig. 3, column B). The extent of hybridization was quantitated by tracing the autoradiogram with a densitometer as described under "Materials and methods". For the copy number calculation, the size of the rice haploid genome was taken as 1.2×10^9 bp (Oono 1984) for all four genome types, even though only the AA genome type was analyzed. The copy number of repetitive sequences corresponding to pOs48, pOa4, pOo2, and pOb1 was estimated to be 2,000, 80,000, 170,000, and 184,000, respectively (Fig. 3 and Table 2). If the size of the rice haploid genome is taken as 0.6×10^9 bp (Bennett and Smith 1976), the copy number of these repetitive sequences would be twice as high. No cross-hybridization was found among these four repetitive sequences even when hybridization was carried out

Table 2. Characterization of four genome-type specific repetitive DNAs

Rice entry	Genome	Name of clone	Size of repeat unit	G+C (%)	Expected[a] T_m	Copy No.	Genome (%)
O. sativa var. Labelle	AA	pOs48	355 bp	49.1	76 °C	2,000	0.064
O. australiensis	EE	pOa4	511 bp	35.6	71 °C	80,000	3.3
O. officinalis	CC	pOo2	366 bp	44.5	74 °C	170,000	5.7
O. brachyantha	FF	pOb1	159 bp	39.2	72 °C	184,000	2.6

[a] Expected T_m value at 0.03 *M* Na^+ in the washing buffer

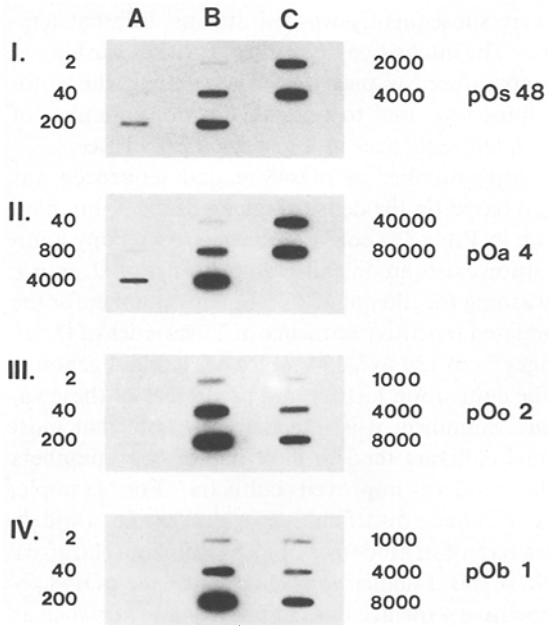

Fig. 3. Copy number determination of rice repetitive DNA sequences of the AA, CC, EE, and FF genome types. Different amounts of DNA were loaded on nitrocellulose filters using a slot-blot apparatus. Copy number standards (the same as the probes used for a specific genome) were used in *columns A* and *C*, and the numbers to the left of *column A* and to the right of *column C* refer to the copy numbers. Different amounts of total rice DNA were spotted in *column B: top lane* 10 ng; *middle lane* 100 ng; *bottom lane* 1,000 ng (except *sample II*, in which, *top lane* 0.5 ng, *middle lane* 5 ng, *bottom lane* 50 ng were spotted in *column B*). Quantitation of the hybridization signal was determined by tracing the X-ray film with a densitometer. *Column B, sample I:* DNA from *O. sativa* var. IR36 was probed with ^{32}P-labeled pOs48 and exposed for 120 min. The hybridization signal of the *bottom lane* in *column B* is equal to that of the standard with 2,000 copies. Thus, the copy number of the repetitive sequence, pOs48, in the IR36 genome is 2,000. *Column B, sample II:* DNA from *O. australiensis* was probed with ^{32}P-labeled pOa4 and exposed for 15 min. The hybridization signal in the *middle lane* in *column B* is twice as high as that of the standard with 4,000 copies, and the *bottom lane* was equal to the standard with 80,000 copies. Thus, the measured copy number of pOa4 is 80,000 in the *O. australiensis* genome. *Column B, sample III:* DNA from *O. officinalis* was probed with ^{32}P-labeled pOo2 and exposed for 5 min. Calculations showed that there are 170,000 copies of the pOo2 sequence in the *O. officinalis* genome. *Column B, sample IV:* DNA from *O. brachyantha* was probed with ^{32}P-labeled pOb1 and exposed for 15 min. Calculation showed that there are 184,000 copies of the 159-bp pOb1 repetitive sequence in the *O. brachyantha* genome

at 42 °C in 5 × SSC (without formamide), which is a relatively low stringency hybridization condition. According to the copy number, the length of the cloned repetitive sequence, and the rice haploid genome size, we estimated that the repetitive sequence in pOa4, pOo2, or pOb1 comprises 3.3%, 5.7% and 2.6%, respectively, of each particular rice genome (Table 2).

Genome-type specificity of the repetitive DNA sequences among different rice entries

Because the repetitive sequences were isolated from different rice genomes, it seemed likely that some of them might be genome-type specific. Using DNA from 37 rice entries, which cover all the known genome types (AA, BB, BBCC, CC, CCDD, EE, FF) and two cultivated species of the genus *Oryza*, enabled us to examine this possibility. The four repetitive sequences were used as probes to screen all the rice DNA samples from the 37 entries by slot-blot hybridization. Figure 4 shows the results on some of the rice entries. When pOs48 was used as the probe (Fig. 4, column I), only those DNA samples from the AA genome showed hybridization. This indicates that the repetitive sequence of pOs48 is AA genome-specific, although the copy number in various AA genome rice varieties is different. The hybridization signal to *O. glaberrima* is very weak but is visible on the original X-ray film.

When pOa4 was used as a probe, strong hybridization (Fig. 4, column II) was observed mainly with *O. australiensis* DNA (EE genome). Much weaker hybridization was barely visible with *O. alta* and *O. latifolia* (CCDD genome), but no hybridization was found with DNA samples of other genome types. These results show that the repeated sequence of pOa4 is enriched in the EE genome of the genus *Oryza*, and can be considered EE genome-specific. Weak hybridization to the CCDD genome suggests that the CCDD and EE genomes are more closely related to each other than to other rice genome types. Figure 4, column III shows that pOo2 only hybridizes to DNA from *O. officinalis* (CC genome), indicating that this repetitive sequence is CC genome-specific. It is interesting to note that the highly repetitive sequence (pOo2) present in *O. officinalis* (CC genome) is absent in *O. alta* and *O. latifolia* (CCDD genome)

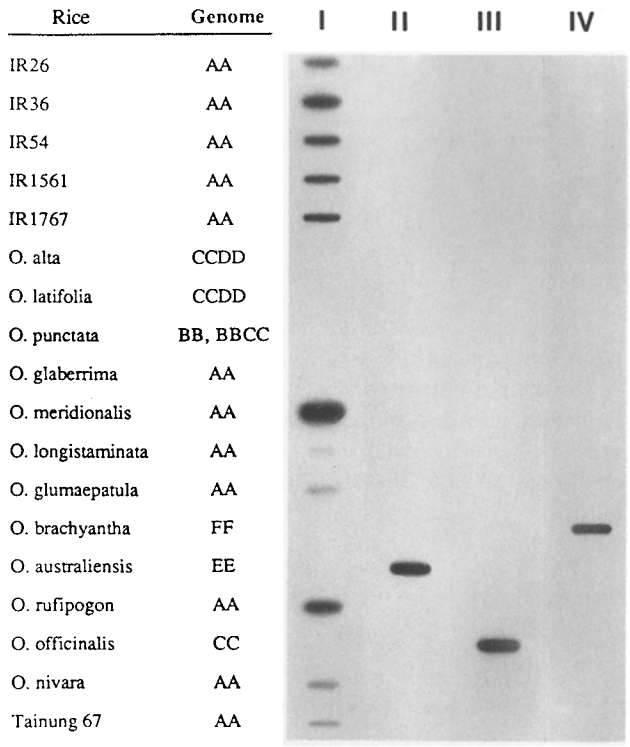

Fig. 4. Screening of different rice genome types using cloned repetitive DNA sequences as probes. Total rice DNA (40 ng) from different genome types was loaded onto a filter, using a slot-blot template, and probed with pOs48 (*column I*), pOa4 (*column II*) pOo2 (*column III*) and pOb1 (*column IV*). The filter was hybridized at 42 °C in 5 × SSC and 50% formamide (corresponding to T_m plus 1 °C) (see 'Materials and methods'), washed at 45 °C in 0.2 × SSC, and exposed for 8 h, 2 h, 70 min and 90 min for *columns I–IV*, respectively. In a parallel experiment, the same results were obtained (data not shown) when hybridization was carried out at 45 °C in 5 × SSC without formamide and the filter was washed at 45 °C in 0.2 × SSC

(Fig. 4, column III). The absence of this sequence in certain species with a CC complement suggests that this CC genome-specific sequence may be lost in the CCDD genome of *O. alta* and *O. latifolia*. Alternatively, there may be two subtypes of the CC genome which contain different repetitive DNAs. The repetitive sequence of pOb1 is FF genome-specific since it only shows hybridization to FF genome DNA from *O. brachyantha* (Fig. 4, column IV).

Variation in the copy number of Os48-related repetitive DNA sequences among rice entries of the AA genome type

Quantitative variation between the amounts of DNA in closely related cereal species frequently involves repetitive sequences (Flavell et al. 1977). Much of the variation in the total DNA content between *S. cereale* and *S. silvestre* can be accounted for by differences in the amounts of the same highly repetitive sequences (Bedbrook et al. 1980). In maize, the majority of repetitive sequences vary markedly in copy number among ten maize species examined (Rivin et al. 1986). However, information on the intraspecific variation of a given repetitive sequence in rice is scarce. To determine whether there is copy number variation of a given repetitive DNA sequence among several related rice species, we carried out the following experiments. DNA from 25 varieties of Asian-cultivated rice species (*O. sativa* of AA genome type) and six other rice species of the AA genome type was probed with the AA genome-specific clone, pOs48, on slot-blots. The filters were subsequently washed at four different temperatures. The intensity of hybridization after washing at each temperature, as measured by scanning the autoradiographs, was used to estimate the copy number of pOs48-related sequences in a given variety of rice.

The copy number of pOs48-related sequences was estimated based on the density tracing of the X-ray film. As shown in Table 3, a considerable degree of copy number variation exists among different varieties of *O. sativa*. When washing the filter at 55 °C, the copy number for the pOs48-related repetitive sequence in 25 varieties of *O. sativa* ranges from 130 to 3,350 copies per haploid genome. When the cultivation history and predigrees of these varieties are examined, it is interesting to note that most traditional cultivars tend to have higher copy numbers than the modern improved cultivars. For example, DGWG, a Chinese dwarf cultivar which has been widely used as a parent in modern rice improvement (Hargrove et al. 1980), has a higher copy number of the pOs48-related repetitive sequence than improved cultivars such as IR26 and IR54. Similarly, Peta also contains a high copy number of this repetitive sequence. However, a higher degree of copy number variation was found among the more distantly related members of the AA genome. For example, an African cultivated species, *O. glaberrima*, shows very low copy numbers of the pOs48-related repetitive sequence (only 50 copies), while *O. meridionalis* shows very high copy numbers (7,000 copies).

We can make the following general conclusions based on the data presented in Table 3: (1) All varieties so far examined under the subspecies *sinica* (also known as *japonica* or *keng*) contain relatively low copy numbers (between 130 and 650 under the 55 °C washing conditions) of pOs48-related repetitive DNA sequences. The variations in copy number between washing at 55 °C and 70 °C are only fivefold, with the exception of Calrose 76. In the latter, there were no repetitive sequences homologous to pOs48 found under the 65 °C washing conditions. The data also indicate that Fujiminori, Fujisaka 5, and Taipei 309 are closer to one another than to other sinica varieties. (2) The varieties of *indica* (also known as *hsien*) subspecies contain relatively high and more variable copy numbers (280–3,350, spanning a 12-fold range) of pOs48-related repetitive sequences. Some varieties, such as M202, MR365, DGWG, Peta, and

Table 3. Copy number determination for pOs48-related repetitive sequences in rice entries containing the AA genome

Name	Type	Copy no.[a] at washing temperature			
		55 °C	60 °C	65 °C	70 °C
O. sativa					
Taipei 309	*Sinica*	230	230	220	150
Tainung 67	*Sinica*	650	630	170	130
Fujiminori	*Sinica*	280	280	160	130
Fujisaka 5	*Sinica*	310	270	220	170
Nipponbare	*Sinica*	190	190	190	130
Calrose 76	*Sinica*	130	80	0	0
Belle Patna	*Indica*	650	630	190	110
Dawn CI	*Indica*	3,350	2,700	80	80
M202	*Indica*	370	330	240	140
MR365	*Indica*	1,130	1,050	830	670
CP231	*Indica*	880	850	30	0
Lemont	*Indica*	350	330	120	70
Labelle	*Indica*	280	280	70	50
DGWG	*Indica*	3,050	3,050	2,440	2,020
Peta	*Indica*	3,250	3,250	2,500	1,970
TKM6	*Indica*	1,920	1,870	50	30
CR94-13	*Indica*	1,550	1,390	50	30
Tetep	*Indica*	1,000	920	30	0
IR26	*Indica*	750	730	510	260
IR36	*Indica*	2,000	1,700	1,220	1,100
IR54	*Indica*	840	750	660	470
IR1561	*Indica*	730	730	550	420
IR1737	*Indica*	640	640	400	260
IR34583	*Indica*	1,210	1,170	90	90
Bulu Dalam	*Javanica*	810	300	0	0
O. glaberrima	African *cultigen*	50	0	0	0
O. meridionalis	Wild	7,000	5,890	3,900	3,700
O. rufipogon	Wild	1,290	1,290	440	380
O. glumaepatula	Wild	650	610	520	470
O. nivara	Wild	450	340	290	220
O. longistaminata	Wild	190	180	110	90

[a] Copy number is defined as the number of copies of pOS48-related sequences with a certain extent of sequence identity with the probe as detected by hybridization at a specified condition. For example, at 70 °C (T_m −6 °C), any pOS48-related sequence with at least 90% sequence identity will hybridize to the probe. Thus, at 70 °C, the copy number represents those pOs48-related sequences with at least 90% sequence identity to the probe. The copy number was calculated using 1.2×10^9 bp as the size of the rice haploid genome (Oono 1984)

IR36, show relatively little difference in copy numbers under the 55 °C and 70 °C washing conditions. However, CP231, TKM6, and Tetep show much larger differences between the 55 °C and 70 °C washes.

Relationship between copy number, percent of sequence mismatch and hybridization temperature of Os48-related repetitive sequences

According to the formula $T_m = 81.5 + 16.6 (\log M) + 0.41$ (% G+C) (Hayes et al. 1970; Bender et al. 1978), the melting temperature (T_m) of a DNA duplex is a function of its G+C content and the salt concentration (M) during filter hybridization or washing. This formula was derived for solution hybridization, and is only an approximation for filter hybridization (Beltz et al. 1983). However, for filter hybridization, the calculated T_m value can be verified by experiments for a particular family of related sequences. pOs48 DNA has a 49% G+C content; when used at a salt concentration of 0.03 M Na^+ in the washing buffer (0.2 × SSC), the T_m for pOs48 is calculated to be 76 °C. The T_m for pOs48 was determined experimentally as follows: pOs48 DNA was spotted on a filter, probed with ^{32}P-labeled pOs48, and washed at 55 °C, 60 °C, 65 °C, 70 °C, 73 °C, 76 °C, and 79 °C. We found the same number of counts (between 390 and 400 cpm) were bound to the filter between 55 °C and 70 °C. Thus, up to 70 °C the pOs48 homoduplex was stable. However, washing the filter at 76 °C resulted in the loss of 50% of the counts (data not shown). Thus, the T_m is 76 °C for pOPs48 under the specified conditions. Using these values as the baseline for comparison, any decrease in copy number in different varieties of rice at washing temperatures of 70 °C or lower can be assumed to be due to mismatches between pOs48 and Os48-related sequences in the rice genomes examined.

The relationship between the decrease of T_m and the percent of mismatch of nucleotides in DNA-DNA heteroduplexes has been estimated for different DNA sequences. Values between 0.3 °C and 1.3 °C decrease in T_m for 1% of mismatch have been reported (Laird et al. 1969; Britten et al. 1974; Yang et al. 1980; Beltz et al. 1983). For pOs48-related sequences, we found a decrease of approximately 0.5 °C for 1% of mismatch by comparing DNA sequences and hybridization results. By DNA sequence analysis, nine cloned pOs48-related genomic clones showed 90%–91% (average 90.5%) of nucleotide identity to pOs48 (Wu and Wu 1987). After IR36 genomic DNA was hybridized to pOs48 and the filters were washed at 65 °C and 70 °C, the copy number of pOs48-related sequence was estimated to be 1,220 and 1,100, respectively (Table 2). Thus, a 10% decrease of the copy number and a 9.5% sequence mismatch can be correlated with a 5 °C difference in the washing temperature after hybridization. The above estimation is valid if one assumes that the thermal stability differences result from single base changes occurring at random in different copies of the pOs48-related sequences. This assumption is basically valid because we found that the differences between the nine clones of pOs48-related sequences are mainly attributable to randomly distributed point mutations (Wu and Wu 1987).

IR36 gave 2,000 copies of pOs48-related sequences when the filter was washed at 55 °C as compared to 1,100 copies at 70 °C (Table 3). Based on the value of 0.5 °C for 1% of mismatch, we estimate that 900 copies of the repetitive sequences of this family contained up to ap-

proximately 30% of mismatch when compared to pOs48 and thus lowered the T_m by 15 °C. Again, this interpretation is valid only if most of these 900 copies of repetitive sequences differ from pOs48 by randomly distributed point mutations, and the percent of G+C remains unchanged. It is likely that some of the 900 copies resulted from deletions or transposition.

Discussion

We have presented evidence for the isolation of repetitive DNA sequences in rice specific to the AA, CC, EE, and FF genomes. These genome-specific repetitive sequences can be used as molecular markers in simple hybridization experiments to distinguish a particular rice genome type from others. This method is simpler and more accurate in classifying unknown rice entries than other currently employed procedures such as comparison of morphological traits, determination of fertility after crosses, or cytological analyses. Moreover, one can follow the transfer of a genome-specific repetitive DNA sequence when monitoring the success of crosses between different rice genome types during plant breeding experiments.

Recently, efforts have been made to study plant species and genome evolution using repetitive DNA sequences (Dover and Flavell 1982; Evans et al. 1983). Our results provide information which augments our understanding of species divergence and genome evolution in rice. The result that all rice species containing the AA genome show sequence identity to the repetitive sequence pOs48 (Fig. 4, column I) confirms the close relationship between these rice species (Chang 1976; Hiroka 1984; Pental and Barnes 1985). The fact that the other genomes show no hybridization to an AA genome-specific repetitive sequence indicates that there is a clear distinction at the DNA sequence level between the AA genome and the other genomes.

Wild rice species have received more and more attention because of their potential utility in rice improvement. Traits such as cold tolerance and insect and disease resistance are present in different wild species (Chang 1976; Hiroka 1984; Hargrove et al. 1980; Shao et al. 1986). Although several wild species of the AA genome are close to the cultivated rice, there are still limitations preventing them from crossing with cultivated varieties. Transformation of rice using cloned sequences is a promising alternative for introducing useful genes. If the integration of the input gene into the host genome involves reciprocal recombination, the inclusion of repetitive sequence(s) that are similar to those in the recipient genomes could facilitate integration. This hypothesis is being tested.

Acknowledgements. We thank T. T. Chang, R. Coffman, C. N. Bollich, and N. Rutger for rice seeds. We are grateful to M. Rothenberg, D. McElroy, and S. Goldman for critically reading this manuscript. X. P. Zhao thanks the National Education Commission of the People's Republic of China for the one-year fellowship. This research was supported by grants RF84066, Allocation No. 3, (to R. Wu) and RF86059 (to S. Tanksley) from the Rockefeller Foundation.

References

Bedbrook JR, Jones J, O'Dell M, Thompson RD, Flavell RB (1980) A molecular description of telomeric heterochromatin in *Secale* species. Cell 19:545–560

Beltz GA, Jacobs KA, Eickbush TH, Cherbas PT, Kafatos FC (1983) Isolation of multigene families and determination of homologies by filter hybridization methods. Meth Enzymol 100:266–285

Bender W, Davidson N, Kindle KL, Taylor WC, Silverman M, Firtel RA (1978) The structure of M6, a recombinant plasmid containing dictyostelium DNA homologous to actin messenger RNA. Cell 15:779–788

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc London 274:227–274

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. Meth Enzymol 29:363–418

Capesius I (1983) Sequence of the cryptic satellite DNA from the plant *Sinapis alba*. Biochem Biophys Acta 739:276–280

Chang TT (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25:425–441

Chang TT (1984) Conservation of rice genetic resources: Luxury or necessity? Science 224:251–256

Cullis CA, Cleary W (1986) Rapidly varying DNA sequences in flax. Can J Genet Cytol 28:252–259

Dennis ES, Gerlach WL, Peacock WJ (1980) Identical polypyrimidine-polypurine satellite DNAs in wheat and barley. Heredity 44:349–366

Deshphane VG, Ranjekar PK (1980) Repetitive DNA in three *Gramineae* species with low DNA content. Hoppe-Seyler's Z Physiol Chem 361:1223–1233

Deumling B (1981) Sequence arrangement of a highly methylated satellite DNA of a plant, *Scilla:* A tandemly repeated inverted repeat. Proc Natl Acad Sci USA 78:338–342

Dover GA, Flavell RB (1982) Genome evolution. Academic Press, London

Evans IJ, James AM, Barnes SR (1983) Organization and evolution of repeated DNA sequences in closely related plant genomes. J Mol Biol 170:803–826

Flavell RB, Rimpav J, Smith DB (1977) Repeated sequence DNA relationships in four cereal genomes. Chromosoma 63:205–222

Grellet F, Delcasso D, Panabieres F, Delseny M (1986) Organization and evolution of a higher plant alphoid-like satellite DNA sequence. J Mol Biol 187:495–507

Hargrove TR, Coffman WR, Cabanilla VL (1980) Ancestry of improved cultivars of Asian rice. Crop Sci 20:721–727

Hayes FN, Lilly EH, Tarliff RL, Smith DA, Williams DL (1970) Thermal transitions in mixtures of polydeoxyribonucleotides. Biopolymers 9:1105–1117

Hiroka M (1984) Wild plants and domestication. In: Tsunoda S, Takahashi N (eds) Biology of rice. Jap Sci Press, Tokyo, pp 3–30

Kato A, Yakura K, Tanifuji S (1984) Sequence analysis of *Vicia faba* repeated DNA, the FokI repeat element. Nucleic Acids Res 16:6415–6426

Laird CD, McConaughy BL, McCarthy BJ (1969) Rate of fixation of nucleotide substitutions in evolution. Nature 224:149–154

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

Martinez-Zapater JM, Estelle MA, Somerville CR (1986) A high repeated DNA sequence in *Arabidopsis thaliana*. Mol Gen Genet 204:417–423

Oono K (1984) Tissue culture and genetic engineering in rice. In: Tsunoda S, Takahashi N (eds) Biology of rice. Jap Sci Press, Tokyo, pp 339–358

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ (1981) Highly repeated DNA sequences limit to knob heterochromatin in maize. Proc Natl Acad Sci USA 78:4490–4494

Pental D, Barnes SR (1985) Interrelationship of cultivated rices *Oryza sativa* and *O. glaberrima* with wild *O. perennis* complex. Theor Appl Genet 70:185–191

Rivin CJ, Cullis CA, Walbot V (1986) Evaluating quantitative variation in the genome of *Zea mays*. Genetics 113:1009–1019

Shao Q, Yi H, Chen Z (1986) New findings concerning the origin of rice. In: Rice genetics, Proc. Int. Rice Genet. Symp. Island, Manila, pp 53–58

Singer MR (1982) Highly repeated sequences in mammalian genomes. Int Rev Cytol 76:67–112

Viotti A, Privitera E, Sala E, Pogna N (1985) Distribution and clustering of two highly repeated sequences in the A and B chromosomes of maize. Theor Appl Genet 70:234–239

Waye JS, Sillard HF (1985) Chromosome-specific alpha satellite DNA: Nucleotide sequence analysis of the 2.0 kilobase pair repeat from the human X chromosome. Nucleic Acids Res 13:2731–2743

Wu TY, Wu R (1987) A new rice repetitive DNA shows sequence homology to both 5S RNA and tRNA. Nucleic Acids Res 15:5913–5923

Yang RCA, Young A, Wu R (1980) BK virus DNA sequence coding for the t and T antigens and evaluation of methods for determining sequence homology. J Virol 34:416–430

Zhou GY (1986) Distantly related hybridization and genetic engineering of crops. In: Rice genetics, Proc. Int. Rice Genet. Symp. Island, Manila, pp 867–876