Theor Appl Genet (2006) 113:419–428
DOI 10.1007/s00122-006-0307-1

ORIGINAL PAPER

Sequence comparison of distal and proximal ribosomal DNA arrays in rice (*Oryza sativa* L.) chromosome 9S and analysis of their flanking regions

Masaki Fujisawa · Harumi Yamagata · Kozue Kamiya · Mari Nakamura · Shoko Saji · Hiroyuki Kanamori · Jianzhong Wu · Takashi Matsumoto · Takuji Sasaki

Received: 18 October 2005 / Accepted: 6 May 2006 / Published online: 30 May 2006

Abstract Rice (*Oryza sativa* ssp. *japonica* cv. Nipponbare) harbors a ribosomal RNA gene (rDNA) cluster in the nucleolar-organizing region at the telomeric end of the short arm of chromosome 9. We isolated and sequenced two genomic clones carrying rice rDNA fragments from this region. The rice rDNA repeat units could be classified into three types based on length, which ranged from 7,928 to 8,934 bp. This variation was due to polymorphism in the number of 254-bp subrepeats in the intergenic spacer (IGS). Polymerase chain reaction (PCR) analysis suggested that the rDNA units in rice vary widely in length and that the copy number of the subrepeats in the IGS ranges from 1 to 12 in the rice genome. PCR and Southern blot analyses showed that most rDNA units have three intact and one truncated copies of the subrepeats in the IGS, and distal (telomere-side) rDNA units have more subrepeats than do proximal (centromere-side) ones. Both genomic clones we studied contained rDNA-flanking DNA sequences of either telomeric repeats (5′-TTTAGGG-3′) or a chromosome-specific region, suggesting that they were derived from the distal or proximal end, respectively, of the rDNA cluster. A similarity search indicated that retrotransposons appeared more frequently in a 500-kb portion of the proximal rDNA-flanking region than in other subtelomeric regions or sequenced regions of the genome. This study reveals the repetitive nature of the telomeric end of the short arm of chromosome 9, which consists of telomeric repeats, an rDNA array, and a retrotransposon-rich chromosomal region.

Communicated by J. S. Heslop-Harrison

Sequence accession numbers in DDBJ assigned for OSJNOa063K24 and OSJNBb0013K10 are AP009051 and AP008245, respectively.

M. Fujisawa · H. Yamagata · K. Kamiya · M. Nakamura · S. Saji · H. Kanamori · J. Wu · T. Matsumoto · T. Sasaki (✉)
Rice Genome Research Program (RGP), National Institute of Agrobiological Sciences/Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305-8602, Japan
e-mail: tsasaki@nias.affrc.go.jp

Introduction

Ribosomal RNAs are well known as structural components of the ribosome complex, which translates mRNA into protein in both prokaryotes and eukaryotes. In many eukaryotes, including higher plants, nuclear 16(−18)S, 5.8S, and 25(−28)S rRNA genes (rDNAs) typically are present in this order, with two internal transcribed spacers (ITSs) in the transcriptional unit (35S rDNA). An rDNA repeat unit, composed of a transcriptional unit and a nontranscribed intergenic spacer (IGS), is highly repetitive and clustered in tandem array in these genomes. In addition, rDNA units are highly homogeneous as a result of concerted evolution through unequal crossover and genetic exchange (Arnheim et al. 1980; Coen et al. 1982; Flavell 1986; Liao 1999). Sequence comparison of rDNAs from various plant species and subspecies could lead to a better understanding of the evolutionary relationships among them (Zimmer et al. 1988; Cordesse et al. 1993; Hemleben and Zentgraf 1994).

The chromosomal location of rDNA has been analyzed cytologically in cultivated and wild rice species and subspecies (Ohmido et al. 2000; Shishido et al. 2000). One cultivated rice subspecies, *Oryza sativa* ssp. *japonica*, has a nucleolar-organizing region (NOR) at the telomeric end of the short arm of chromosome 9 (Chung et al. 1993; Ohmido et al. 2000; Kato et al. 2003). The rice *O. sativa* ssp. *japonica* has approximately 850 copies of 8-kb rDNA repeat unit per diploid genome (Oono and Sugiura 1980). For several cultivated and wild rice species and subspecies, the nucleotide sequences of rDNAs and their spacers have been registered in public databases. In addition, in rice, as well as in the other organisms, heterogeneity of the rDNA units due to polymorphism of subrepeats in intergenic regions has been identified (Oono and Sugiura 1980). However, it remains unclear how rice rDNA unit variants are ordered on the chromosome. Moreover, characterization of the DNA sequences of the regions flanking rDNA clusters in rice also remains insufficient. We isolated rice genomic clones carrying rDNA sequences to compare the structures of the distal and proximal rDNA arrays and to analyze the sequences of their flanking regions. Here we describe the characteristics of the rice rDNA arrays in relation to their chromosomal position, sequence structure of the boundary region between the rDNA array and telomeric repeats, and sequence composition of the 500 kb flanking the proximal rDNA region.

Materials and methods

Isolation and mapping of genomic clones each carrying an rDNA array and its flanking region

A bacterial artificial chromosome (BAC) genomic library (coded by OSJNBb; Chen et al. 2002) and a fosmid genomic library (coded by OSJNOa; Ammiraju et al. 2005) from *O. sativa* ssp. *japonica* (cv. Nipponbare) provided by the Clemson University Genomics Institute (CUGI) and the Arizona Genomics Institute (AGI), respectively, were used. A BAC clone that carried both a proximal rDNA array and its flanking region was selected on the basis of BAC contigs (http://www.genome.arizona.edu/fpc/WebAGCoL/rice/WebFPC/) generated by using fingerprinting and sequence-tagged connectors (Chen et al. 2002). In contrast, the fosmid library was screened by overgo hybridization (Chen et al. 2002) using high-density fosmid filters. A telomere-specific overgo probe (54 bp long) was labeled with ^{32}P-dATP and -dCTP by using a pair of 32-mer oligonucleotides (5′-AGGGTTAGGGTTTAGGGTTAGGGTTTAAGGGG-3′ and 5′-ACCCTAAACCCTAAACCCTAAACCCCTTAAAC-3′) containing complementary 10-bp overlaps at each 3′ end. DNA labeling and hybridization conditions were the same as described previously (Mizuno et al. 2006).

DNA sequencing and assembly

The nucleotide sequences of the isolated genomic clones were determined by using shotgun sequencing with tenfold overlap, as described (Sasaki et al. 2002). The sequence data were assembled, edited, and verified for accuracy using the PHRED, PHRAP, and CONSED computer programs (Ewing et al. 1998; Ewing and Green 1998; Gordon et al. 1998). For correct assembly of repeat sequences, the entire sequences of some shotgun clones were determined independently using the GPS-1 Genome Priming System (New England Biolabs, Ipswich, MA, USA) and integrated with the whole assembly of genomic clones. The resulting assembly was confirmed by restriction digestion and subjected to pulsed-field agarose gel electrophoresis (CHEF DR-III, Bio-Rad, Hercules, CA, USA). All determined nucleotide sequences have been warranted as highly accurate (PHRAP quality score $\geq$ 30, or by manual confirmation or both), according to International Rice Genome Sequencing Project standards (International Rice Genome Sequencing Project 2005).

Database searches and sequence analysis

Database searches of nucleotide sequences were performed using the BLASTN program (Altschul et al. 1997). We used the sequences of our genomic clones as queries to search the nonredundant (nr) nucleotide databases (http://www.ncbi.nlm.nih.gov/BLAST/). Repeat sequences were detected using RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html) in combination with the TIGR rice repeat sequence database (http://www.tigr.org/tdb/e2k1/osa1/). Sequence analysis, alignment, and comparison were performed using SEQUENCHER 4.1 software (Gene Codes, MI, USA), the CLUSTALW program (Thompson et al. 1994), and the EMBOSS program (Rice et al. 2000; http://www.hgmp.mrc.ac.uk/Software/EMBOSS/).

Southern blot hybridization

A 1.6 kb-DNA fragment of the 18S rDNA was amplified by polymerase chain reaction (PCR) in a 50-μl reaction mixture containing 250 ng of rice total DNA, 250 μM of each dNTP, 2.5 mM $MgCl_2$, 0.5 μM of each primer (5′-CGAACTGTGAAACTGCGAATGGC-3′

and 5′-TAGGAGCGACGGGCGGTGTG-3′), 1× *ExTaq* buffer, and 1 U *ExTaq* DNA polymerase (Takara Bio, Otsu, Japan). The amplification conditions were: 94°C for 1 min; 25 cycles of 98°C for 20 s and 68°C for 5 min: and 72°C for 10 min (GeneAmp PCR System 9700, Applied Biosystems, Foster City, CA, USA). DNA blotting and hybridization procedures were as previously described (Kurata et al. 1994; Nagamura et al. 1995).

PCR analysis

The IGS subrepeat array of the rDNA unit was amplified by PCR in a 50-μl reaction mixture containing 50 ng of rice total DNA, 250 μM of each dNTP, 0.5 μM of each primer (IGS-F: 5′-CTCGCCCCAGCTCCCGAG-3′ and IGS-R: 5′-GGCTACGTGCCCGAACAC-3′), 2× GC buffer I, and 2.5 U *LATaq* DNA polymerase (Takara Bio). The amplification conditions were: 94°C for 1 min, followed by 30 cycles of 94°C for 30 s and 68°C for 10 min (GeneAmp PCR System 9700, Applied Biosystems). PCR products were electrophoresed in a 0.8% agarose gel in 0.5× TBE buffer.

Results

Sequence analysis of genomic clones carrying rice rDNA arrays

To study the relationship between polymorphism of rice rDNA units and chromosomal location, we selected genomic clones derived from the boundaries between the rDNA cluster and chromosome-specific segments. In the case of the distal boundary region, a genomic clone carrying a distal rDNA array also was expected to carry telomeric repeats, as shown in previous cytologic studies (Ohmido et al. 2000; Shishido et al. 2000). Furthermore, the rDNA cluster is located at the distal end of the short arm of chromosome 9, suggesting that it is close to the telomeric region. We therefore first screened the fosmid library using a telomere-repeat-specific probe to isolate the boundary clone from the distal rDNA array. We found that a genomic clone, OSJNOa063K24, which was one of the fosmid clones isolated with an overgo probe for telomeric repeats, carried both the rDNA array as well as telomeric repeats, according to analysis of end-sequences. Therefore this clone was confirmed to represent the boundary between the distal rDNA array and a chromosome-specific segment close to the telomere.

For the proximal boundary, we selected BAC clone OSJNBb0013K10 from a contig (contig no. 60) located at 0 cM of rice chromosome 9 (Chen et al. 2002). This clone was found to carry an rDNA array as well as a chromosome-specific segment that overlapped with a previously sequenced chromosomal region, in light of analysis of the BAC end-sequences provided by AGI (Chen et al. 2002). Therefore OSJNBb0013K10 was confirmed to represent the boundary between the proximal rDNA array and a chromosome-specific segment.

The nucleotide sequences of the two genomic clones were determined by shotgun sequencing and carefully assembled and edited by manual confirmation of restriction profiles. In addition, several shotgun clones were fully sequenced independently and the data applied to the assembly of the whole genomic clones. The resulting assemblies of OSJNOa063K24 and OSJNBb0013K10 showed that the sizes of the cloned DNA fragments were 30,177 and 129,118 bp, respectively. These assemblies were consistent with the diagnostic restriction digestion profiles of these clones (data not shown).

Sequence comparison of rice rDNA units

To determine the nucleotide sequences of the rDNA units in the two clones, we compared them with known sequences of rice rDNA and spacer regions (accession nos. M11585, M16845, X00755, and X54194; Takaiwa et al. 1984, 1985a, b, 1990) using BLASTN. The results showed that OSJNOa063K24 and OSJNBb0013K10 carried two and four copies of the complete rDNA unit, respectively, each of which consisted of 18S, 5.8S, and 25S rDNAs, two internal transcribed spacers (ITS1, ITS2), and a nontranscribed IGS between the 25S and 18S rDNAs (Fig. 1a, c). These six copies of rDNA units could be classified into three types according to size (Fig. 1b). The type 1 and type 2 units were 8,934 and 8,426 bp, respectively, with one copy each in OSJNOa063K24 (Fig. 1c), whereas the type three unit was 7,928 bp, with four copies in OSJNBb0013K10 (Fig. 1c).

The sequences were aligned to reveal polymorphism among them. The three types of rDNA units had the same sequence length with >99.9% identity in the transcribed region from the 18S to the 25S rDNA (data not shown). However, sequence alignment gaps in the IGSs appeared at 254-bp intervals, consistent with the length of the subrepeat array (Fig. 2a). The results also showed that the differences were due to the polymorphism in copy number of subrepeats among the three types of unit: type 1 units carried 7 subrepeats in the IGS, type 2 units had 5, and type 3 units had 3 (Fig. 2b). The number of putative transcription initiation sites

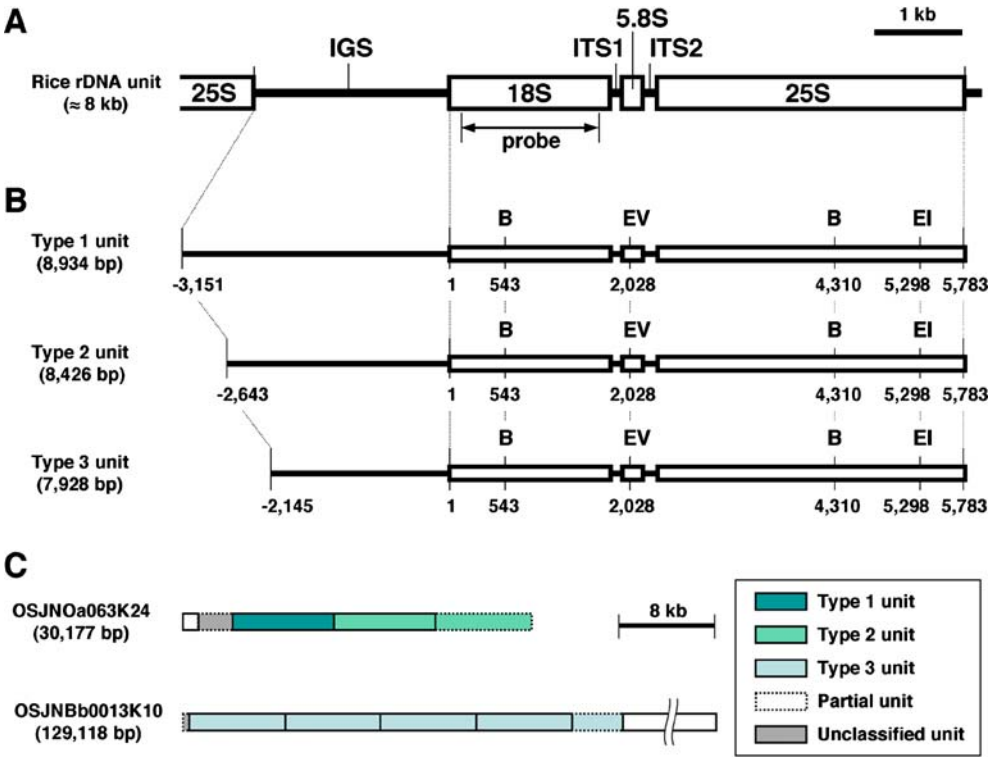

Fig. 1 Structure of rDNA units of rice. *Open boxes* in **a** and **b** indicate the positions of rRNA genes. **a** Structure of a typical rDNA unit of rice. A 1.6-kb DNA fragment of the 18S rDNA was used as a probe for Southern blot analysis. **b** Comparison of the rDNA units identified. The 5′ end of the 18S rDNA is indicated as 1, and the positions of restriction sites for *Bam*HI (*B*), *Eco*RI (*EI*), and *Eco*RV (*EV*) are indicated accordingly. **c** Composition of the rDNA units in the two genomic clones

(5′-TATAGTAGGGG-3′, Takaiwa et al. 1990) in the IGS subrepeats also varied according to changes in the copy number of subrepeats (Fig. 2a). Therefore the subrepeating element may function as transcription enhancer as in other systems (Schiebel et al. 1989; Piller et al. 1990). Interestingly, the subrepeats commonly present in the IGSs had identical sequences among units (Fig. 2a), although variation in size (254 bp, 244 bp, and truncated at 62 bp) occurred among subrepeats. In addition, both upstream and downstream regions of the subrepeat arrays as well as the transcribed regions were highly conserved, with >99.9% sequence identity. As far as possible, we also categorized partial rDNA units in the two clones according to subrepeat types (Fig. 1c).

Size variation among rice rDNA units

To survey variation of rDNA units in rice, Southern blot hybridization was done using the 18S rDNA sequence described in Fig. 1a as a probe. Consequently, a single 7.9-kb band after *Eco*RI or *Eco*RV digestion, and two bands at 4.2 and 3.8 kb after *Bam*HI digestion (Fig. 3a), were detected. The sizes of these bands were equivalent with those of the DNA fragments expected from the sequence of the type 3 unit (Fig. 3a). Therefore, most of the rDNA units in rice are represented by the type 3 unit. In contrast, Southern analysis failed to reveal bands characteristic of type 1 and type 2 units. In addition, a weak band at 7.9 kb appeared after *Bam*HI digestion, implying the presence of a minor class of rDNA units in rice that was not sequenced in this study. However, this minor rDNA unit was considered to be a variant of type 3 in light of the size similarity between the units. There is another possibility that this minor band might arose because of the partial interference of restriction digestion due to DNA methylation (Torres-Ruiz and Hemleben 1994).

For a more detailed survey of the variation in rice rDNA units, we performed PCR analysis of the rice genome by using a pair of primers that amplified the IGS subrepeat array (Fig. 2a). A laddered collection of amplification products resulted (Fig. 3b). Part of the ladder occurred at intervals of about 250 bp, reflecting the copy number of the subrepeats. An approximately 860-bp fragment was amplified predominantly (Fig. 3b), and the size of this PCR product was consistent with that expected from the type 3 unit sequence—that is, 3 254-bp subrepeats and a single 62-bp truncated subrepeat in the IGS

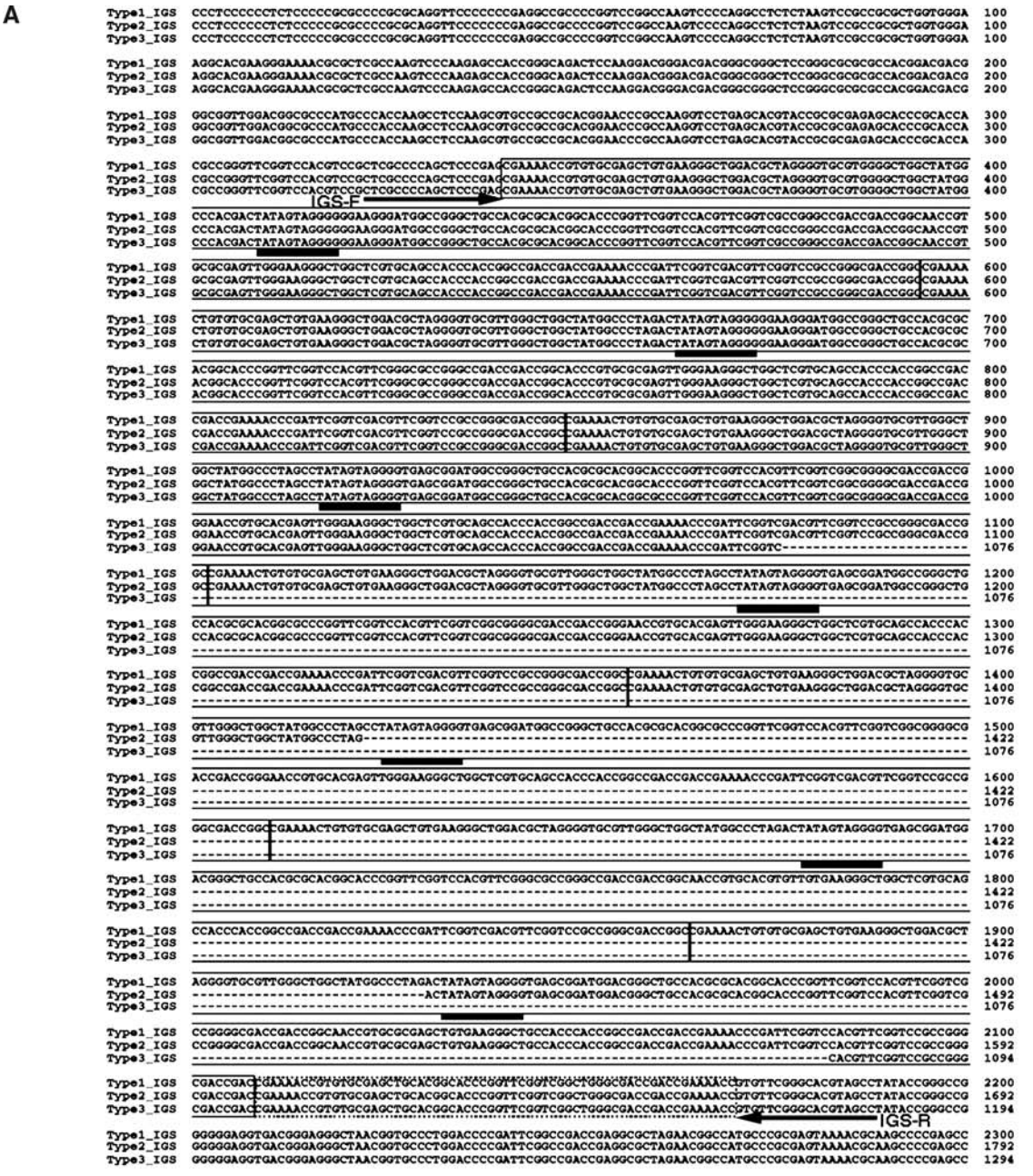
A
Type1_IGS
Type2_IGS
Type3_IGS
IGS-F
IGS-R
B
250 bp
Type 1 unit IGS
Type 2 unit IGS
Type 3 unit IGS
254
244
62
Subrepeat
Truncated subrepeat

Fig. 2 Comparison of the IGS sequences in the three types of rDNA units of rice. **a** Sequence alignment of a 5′ region of the IGS. *Hyphens* indicate sequence gaps. The *numbers in the right column* indicate the position of the rightmost base in each line. Subrepeats (254 or 244 bp) or truncated subrepeats (62 bp) are enclosed by *lines* or *broken lines*, respectively. Putative transcriptional initiation sites for RNA polymerase I are indicated by *bold underlines*. The positions of the primers for PCR analysis (IGS-F and IGS-R) are indicated by *arrows* below the sequence alignment. **b** Composition of subrepeats of the IGSs in three types of rDNA units. Sizes of the full-length (*solid lines*) or truncated (*broken lines*) subrepeats are shown by the *numbers above the arrows*

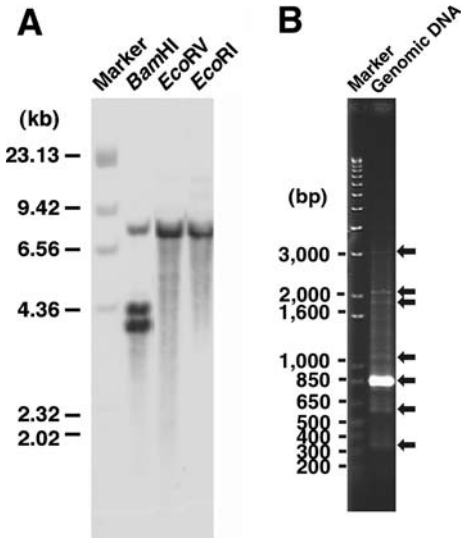

Fig. 3 Variation in rDNA units of rice. **a** Genomic Southern blot analysis of rice rDNA. Genomic DNA of rice was digested with *Bam*HI, *Eco*RI, or *Eco*RV. The position of probed 18S rDNA is shown in Fig. 1. **b** PCR analysis of the IGS subrepeats of rice rDNA. The primer pair (IGS-F and IGS-R) indicated in Fig. 2a was used for PCR amplification. *Arrows* indicate the sizes of amplified DNA fragments detected

(Fig. 2a). This result supports the idea that the type 3 unit represents the majority of rDNA units in the rice genome. In addition, several other fragments were amplified weakly, suggesting the existence in rice of low-copy minor rDNA units that vary according to the number of IGS subrepeats present (Fig. 3b). The smallest PCR product was estimated to 350 bp, consistent with a fragment carrying a single 254-bp subrepeat and one 62-bp truncated subrepeat. The largest product was approximately 3.1 kb, consistent with a fragment carrying 12 full-length (254-bp) subrepeats and a single 62-bp truncated subrepeat as well as the two oligonucleotide primers. These results suggests that rice rDNA units range from 7.4 to 10.2 kb.

Sequence analysis of the boundary between the distal rDNA array and telomeric repeats on rice chromosome 9

Sequence analysis of OSJNOa063K24 showed that, in addition to the distal rDNA array, this clone carried the telomeric repeat sequence (5′-TTTAGGG-3′) arrayed in tandem (Fig. 4a). A DNA fragment corresponding to bases 1 through 1,065 in OSJNOa063K24 consisted of 135 copies of the telomeric repeat interspersed with its variants (5′-TTAGGG-3′ or 5′-TTTTAGGG-3′) (Fig. 4a). Therefore, we concluded that this clone originated from the boundary between the distal rDNA array and the telomeric region on the short arm of chromosome 9. We also detected a 53-bp DNA fragment between the rDNA sequence and the telomeric repeats (bases 1,066 through 1,118 of OSJNOa063K24). This fragment contained six inverted telomeric repeats of (5′-CCCTAAA-3′) (Fig. 4a).

Sequence analysis of the boundary between the proximal rDNA array and a chromosome-specific segment

Sequence analysis of OSJNBb0013K10 showed that this clone included the boundary between the proximal rDNA array and a chromosome-specific segment already mapped and sequenced (Wu et al. 2003; International Rice Genome Sequencing Project 2005). In contrast to the distal boundary and its association with telomeric repeats, the proximal boundary showed no specific sequence characteristics upstream of the rDNA array.

Further, to characterize the chromosome-specific segment flanking the proximal rDNA array, we extended the sequence analysis of the region to 500 kb adjacent to the rDNA array (Fig. 4b). The G+C content of the 500-kb region was 46%, lower than that of the rDNA units (62–63%). A total of 121 genes were identified from the manual annotation of the genomic clones covering this region (http://www.rgp.dna.affrc.go.jp/cgi-bin/statusdb/). Among them, 66 were potential protein-coding genes, and the remainder were pseudogenes, noncoding transcripts, or single-program predicted genes. Moreover, the prediction of seven of the protein-coding genes was supported by the full-length cDNA sequences (Kikuchi et al. 2003) from this 500-kb region (Fig. 4b). Most of the remaining 59 predicted protein-coding genes showed similarity to transposable elements, suggesting that this 500-kb region is not gene-dense.

Detection of rice repeat sequences by using RepeatMasker showed that 49.6% of the 500-kb region was occupied by several classes of the repeat sequences (Table 1). Moreover, 79.7% of the repeat sequences in

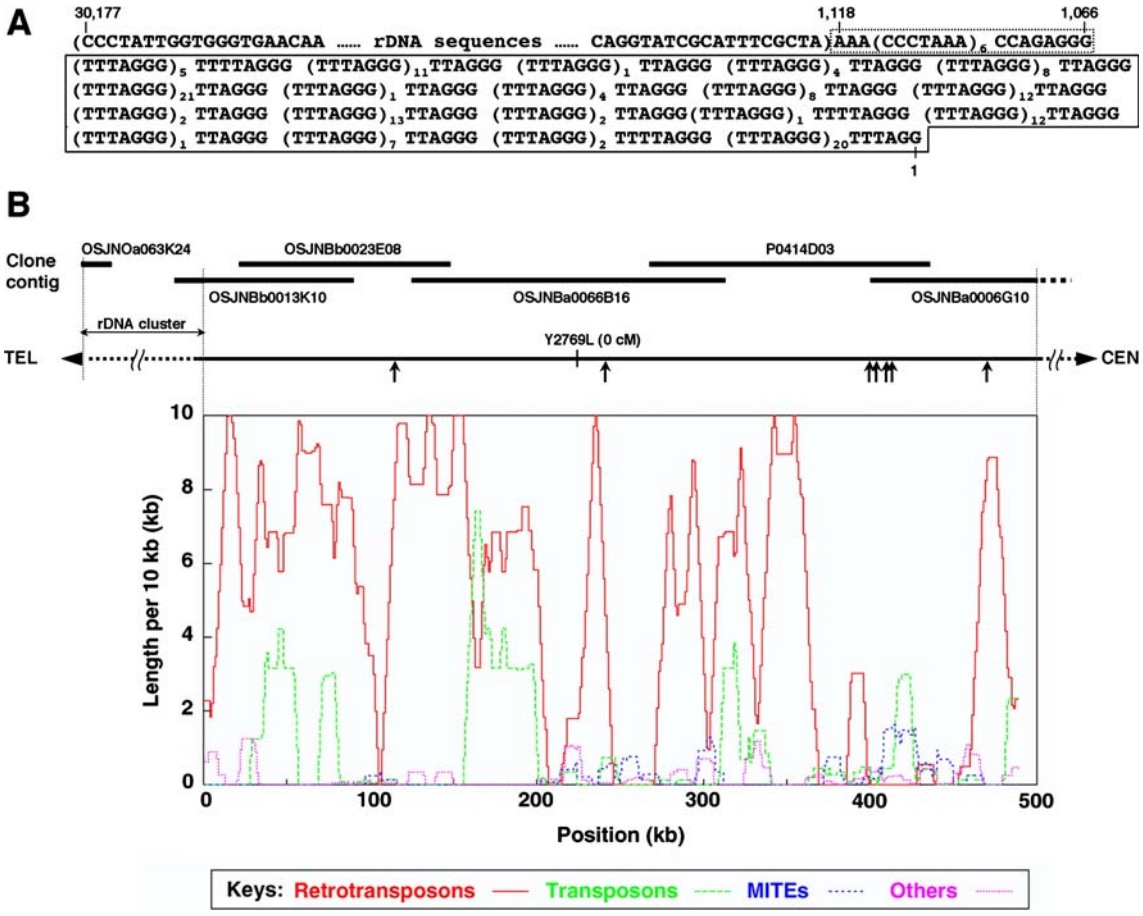

Fig. 4 Sequence of DNA fragments flanking the rDNA array. **a** Nucleotide sequence of the telomere region and adjacent region to the rDNA array in OSJNOa063K24. A complementary strand of the registered sequence of the clone is described here. Tandem telomere repeats are as with $(TTTAGGG)_n$, where *n* is copy number. A 1,065-bp of the telomeric region is enclosed by *solid lines*. A 53-bp section of the region adjacent to the rDNA array is enclosed by *broken lines*. Inverted telomere repeats are shown as $(CCCTAAA)_n$, where *n* is copy number. Base positions are indicated by *numbers below the vertical lines*. **b** Structure of a 500-kb section of the region flanking the proximal rDNA array at the short arm (0 cM) of chromosome 9. The physical map with the clone contig is shown in the upper figure. Positions of predicted genes supported by the full-length cDNA sequences of rice are indicated by *arrows*. Distribution of repeat sequences is shown in the lower figure. Repeat length was calculated for every 10-kb window with 1-kb sliding

this region (corresponding to 39.6% of the 500-kb region) were classified as retrotransposons, especially Ty3/*gypsy*-class retrotransposons (corresponding to 74.6% of the total number of retrotransposon sequences and 29.5% of the 500-kb region). The distribution of the repeat sequences showed that a 150-kb region next to the proximal rDNA-flanking region was particularly enriched in Ty3/*gypsy*-class retrotransposons (Fig. 4b).

Discussion

Variation in the length of rice rDNA units is related to chromosomal location

Through sequence analysis of two genomic clones, OSJNOa063K24 and OSJNBb0013K10, which originated from the distal and proximal regions, respectively, of the short arm of chromosome 9, this study revealed that the size variation of rDNA units in *O. sativa* ssp. *japonica* (cv. Nipponbare) is related to their chromosomal location. OSJNOa063K24 is a fosmid clone that was generated from a randomly sheared DNA (Ammiraju et al. 2005), thus enabling cloning of a DNA fragment of the telomeric region. The two rDNA arrays that we sequenced are located in the rDNA cluster, as summarized in Fig. 5. In the case of the sequenced rDNA units, the rDNA units at the distal end are larger (because of the presence of more copies of subrepeats) than are those at the proximal end (Fig. 5). The only known example similar to this finding for rice is that of *Arabidopsis thaliana NOR4*, where rDNA units were divided into at least three classes according to the length (long, intermediate, or

Table 1 Repeat sequences in the proximal flanking region of the rice rDNA array

Class[a]	Code[b]	No. of sites	Total length (bp)	Repeat sequence length/500 kb (%)	Class repeat sequence length/total repe at length (%)
Retrotransposon	TERT	94	197,726	39.5	79.7
Transposon	TETN	43	36,698	7.3	14.8
MITE	TEMT	41	8,495	1.7	3.4
Others[c]	Not applicable	39	5,070	1.0	2.0
Total		217	247,989	49.6	100

[a] Classifications of repeat sequences provided by the TIGR_Oryza_Repeats v 2.1 database were used

[b] Classifications of codes of repeat sequences provided by the TIGR_Oryza_Repeats v2.1 database were used

[c] Repeat classes other than retrotransposons, transposons, and MITE were included

short) of the spacer region, and each class appeared in this order (from telomere to centromere) in the NOR (Copenhaver and Pikaard 1996b). These researchers also suggested that the intermediate spacer-length class may have originated by recombination between short- and long-spacer variants in *A. thaliana NOR4*.

However, unlike *A. thaliana NOR4*, in which each of the three classes is clustered in a region spanning several hundred kilobases to ≈1.7 Mb , the rice NOR was occupied predominantly by type 3 rDNA units, each of which has three 254-bp subrepeats and one 62-bp truncated subrepeat in the IGS. Results of PCR analysis indicated the existence of other rDNA units in which the number of subrepeats varied, but Southern blot analysis failed to detect them. These observations support our conclusion that the majority of rice rDNA units are of type 3. Although this result is apparently inconsistent with those of the study by Oono and Sugiura (1980), in which they detected two kinds of rDNA units in regard to size (7.6 and 7.9 kb), we consider that the inconsistency might arise because of the use of different rice cultivars in the two studies. In fact, the nucleotide sequences of the rDNA units showed discrepancies between the two studies (data not shown). Therefore, we conclude that type 3 rDNA units represent the majority of rDNA units in *O. sativa* ssp. *japonica* cv. Nipponbare. We believe that it is important to characterize the rDNA unit sequences of the Nipponbare in detail because it is considered to be a standard rice cultivar for molecular biology and genetics. Regardless, rice rDNA units are more homogeneous than those of *A. thaliana*, as established through this and previous (Oono and Sugiura 1980) studies.

Structural features of the regions flanking rice rDNA arrays

In this study we also revealed various features of regions flanking the distal and proximal rDNA arrays. A typical plant telomeric repeat (5′-TTTAGGG-3′) and its variants were repeated in the distal rDNA-flanking region. As in the ends of other rice chromosomes (Yang et al. 2005; Mizuno et al. 2006), the telomeric repeat arrays in the distal region were oriented as 5′-TTTAGGG-3′ in the direction of centromere to telomere. The telomeric region of rice has been estimated to be 3.5 kb long on average (Yang et al. 2005), suggesting that the telomeric region of chromosome 9 most likely extends beyond the 1,069 bp of telomeric repeats we sequenced. In addition, we identified a 53-bp

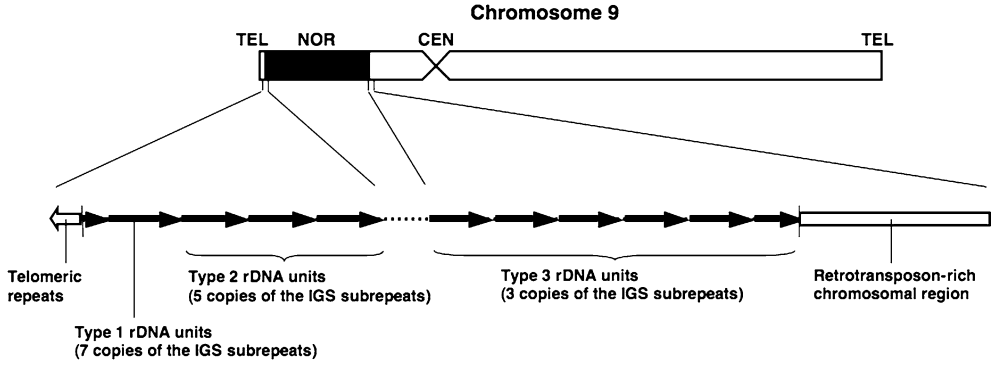

Fig. 5 Schematic illustration of the structure of the distal end of the short arm of chromosome 9. Each rDNA unit is represented by a *horizontal arrow*

DNA fragment that intervened between the telomeric region and the rDNA array. This fragment carries six copies of inverted telomeric repeats. However, this sequence may not have functional or other importance for chromosome maintenance, because this structure is not common to all the telomere-adjacent regions of rice chromosomes sequenced thus far (Yang et al. 2005). Unlike in rice, in *A. thaliana* a 13-bp sequence has been duplicated directly in rDNA–telomere junctions, likely because of incomplete recombination, (Copenhaver and Pikaard 1996a).

In contrast to the case with the distal array, we found that almost half of the 500-kb region flanking the proximal rDNA array was occupied with repeat sequences, especially Ty3/*gypsy*-class retrotransposons. The proportion of retrotransposons in this region is twice as high as the average for the entire rice genome (International Rice Genome Sequencing Project 2005). Further, the gene density in this region is 7.6 kb/gene (66 genes per 500 kb), which appears higher than the average gene density for the entire rice genome. However, many of the predicted genes presumably are derived from transposable elements, and only seven of the 66 genes predicted are supported by full-length rice cDNA sequences in this region. Therefore, we conclude that the actual gene density of this region is lower than the average for the entire rice genome.

The structural features of an increased proportion of Ty3/*gypsy*-class retrotransposons and low gene density are rather similar to the structural features of centromeric regions (Presting et al. 1998; Nagaki et al. 2004; Wu et al. 2004) than to those in telomeric and subtelomeric regions (Mizuno et al. 2006). A kind of Ty3/*gypsy*-class retrotransposon is found specifically in the centromere or pericentromeric regions of maize and rice (Zhong et al. 2002; Nagaki et al. 2004; Nagaki and Murata 2005) and may bind centromere-specific histone H3 protein, suggesting that the retrotransposon sequence plays a role in the maintenance of chromatin structure, chromosome pairing, or chromosome segregation. Regardless, the retrotransposons in the 500-kb region that we analyzed differ from centromere-specific retrotransposons (CRRs), and we cannot presently suggest a rational hypothesis regarding the functional meaning of the Ty3/*gypsy*-class retrotransposons amassed in the flanking region of the proximal rDNA array. Apart from any functional meaning, the presence of abundant retrotransposon sequences and the rDNA cluster suggests that much of the short arm of rice chromosome 9 is potentially heterochromatic in nature.

Acknowledgements We thank Dr. Rod A. Wing of the Arizona Genomics Institute for kindly providing the rice Nipponbare BAC and fosmid genomic libraries. We also thank Dr. Baltazar A. Antonio for critical reading of the manuscript and Drs. Masahiro Nakagahra and Kyozo Eguchi for their advice and encouragement. This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (GS1101, GS1201, and GD2007).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Ammiraju JS, Yu Y, Luo M, Kudrna D, Kim H, Goicoechea JL, Katayose Y, Matsumoto T, Wu J, Sasaki T, Wing RA (2005) Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. Nipponbare) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome. Theor Appl Genet 111:1596–1607

Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E (1980) Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. Proc Natl Acad Sci USA 77:7323–7327

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Chung MC, Ning CN, Wu HK (1993) Localization of ribosomal RNA genes o rice chromosomes. Bot Bull Acad Sin 34:43–55

Coen E, Strachan T, Dover G (1982) Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. J Mol Biol 158:17–35

Copenhaver GP, Pikaard CS (1996a) RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. Plant J 9:259–272

Copenhaver GP, Pikaard CS (1996b) Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J 9:273–282

Cordesse F, Cooke R, Tremousaygue D, Grellet F, Delseny M (1993) Fine structure and evolution of the rDNA intergenic spacer in rice and other cereals. J Mol Evol 36:369–379

Ewing B, Green P (1998) Base-calling of automated sequencer traces using *Phred.* II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res 8:175–185

Flavell R (1986) Repetitive DNA and chromosome evolution in plants. Philos Trans R Soc Lond B 312:227–242

Gordon D, Abajian C, Green P (1998) *Consed*: a graphical tool for sequence finishing. Genome Res 8:195–202

Hemleben V, Zentgraf U (1994) Structural organization and regulation of transcription by RNA polymerase I of plant nuclear ribosomal RNA genes. Plant promoters and transcription

factors. In: Nover L (ed) Results and problems in cell differentiation, vol 20. Springer, Berlin Heidelberg New York, pp 3–24

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Kato S, Ohmido N, Fukui K (2003) Development of a quantitative pachytene chromosome map in *Oryza sativa* by imaging methods. Genes Genet Syst 78:155–161

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie QH, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashidume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376–379

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300-kilobase interval genetic map of rice including 883 expressed sequences. Nat Genet 8:365–372

Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64:24–30

Mizuno H, Wu J, Kanamori H, Fujisawa M, Namiki N, Saji S, Katagiri S, Katayose Y, Sasaki T, Matsumoto T (2006) Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. Plant J 46:206–217

Nagaki K, Murata M (2005) Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. Chromosome Res 13:195–203

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Nagamura Y, Inoue T, Antonio BA, Shimano T, Kajiya H, Shomura A, Lin SY, Kubiki Y, Harushima Y, Kurata N, Minobe Y, Yano M, Sasaki T (1995) Conservation of duplicated segments between rice chromosomes 11 and 12. Breed Sci 45:373–376

Ohmido N, Kijima K, Akiyama Y, de Jong JH, Fukui K (2000) Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in *indica* and *japonica* rice. Mol Gen Genet 263:388–394

Oono K, Sugiura M (1980) Heterogeneity of the ribosomal RNA gene clusters in rice. Chromosoma 76:85–89

Piller KJ et al (1990) Structural analysis of the short length ribosomal DNA variant from *Pisum sativum* L. cv. Alaska. Nucleic Acids Res 18:3135–3145

Presting G, Malysheva L, Fuchs J, Schubert I (1998) A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16:721–728

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16:276–277

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y, Antonio BA, Kanamori H, Hosokawa S, Masukawa M, Arikawa K, Chiden Y, Hayashi M, Okamoto M, Ando T, Aoki H, Arita K, Hamada M, Harada C, Hijishita S, Honda M, Ichikawa Y, Idonuma A, Iijima M, Ikeda M, Ikeno M, Ito S, Ito T, Ito Y, Ito Y, Iwabuchi A, Kamiya K, Karasawa W, Katagiri S, Kikuta A, Kobayashi N, Kono I, Machita K, Maehara T, Mizuno H, Mizubayashi T, Mukai Y, Nagasaki H, Nakashima M, Nakama Y, Nakamichi Y, Nakamura M, Namiki N, Negishi M, Ohta I, Ono N, Saji S, Sakai K, Shibata M, Shimokawa T, Shomura A, Song J, Takazaki Y, Terasawa K, Tsuji K, Waki K, Yamagata H, Yamane H, Yoshiki S, Yoshihara R, Yukawa K, Zhong H, Iwama H, Endo T, Ito H, Hahn JH, Kim HI, Eun MY, Yano M, Jiang J, Gojobori T (2002) The genome sequence and structure of rice chromosome 1. Nature 420:312–316

Schiebel Kv, Waldburg G, Gerstner J, Hemleben V (1989) Termination of transcription of ribosomal RNA genes of mung bean occurs within a 175-bp repetitive element of the spacer region. Mol Gen Genet 218:302–307

Shishido R, Sano Y, Fukui K (2000) Ribosomal DNAs: an exception to the conservation of gene order in rice genomes. Mol Gen Genet 263:586–591

Takaiwa F, Oono K, Sugiura M (1984) The complete nucleotide sequence of a rice 17S rRNA gene. Nucleic Acids Res 12:5441–5448

Takaiwa F, Oono K, Iida Y, Sugiura M (1985a) The complete nucleotide sequence of a rice 25S rRNA gene. Gene 37:255–259

Takaiwa F, Oono K, Sugiura M (1985b) Nucleotide sequence of the 17S–25S spacer region from rice rDNA. Plant Mol Biol 4:355–364

Takaiwa F, Kikuchi S, Oono K (1990) The complete nucleotide sequence of the intergenic spacer between 25S and 17S rDNAs in rice. Plant Mol Biol 15:933–935

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res 22:4673–4680

Torres-Ruiz RA Hemleben V (1994) Pattern and degree of methylation in ribosomal RNA genes of *Cucurbita pepo* L. Plant Mol Biol 26:1167–1179

Wu J, Mizuno H, Hayashi-Tsugane M, Ito Y, Chiden Y, Fujisawa M, Katagiri S, Saji S, Yoshiki S, Karasawa W, Yoshihara R, Hayashi A, Kobayashi H, Ito K, Hamada M, Okamoto M, Ikeno M, Ichikawa Y, Katayose Y, Yano M, Matsumoto T, Sasaki T (2003) Physical maps and recombination frequency of six rice chromosomes. Plant J 36:720–730

Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi M, Yoshihara R, Kobayashi H, Ito K, Karasawa W, Yamamoto M, Saji S, Katagiri S, Kanamori H, Namiki N, Katayose Y, Matsumoto T, Sasaki T (2004) Composition and structure of the centromeric region of rice chromosome 8. Plant Cell 16:967–976

Yang TJ, Yu Y, Chang SB, de Jong H, Oh CS, Ahn SN, Fang E, Wing RA (2005) Toward closing rice telomere gaps: mapping and sequence characterization of rice subtelomere regions. Theor Appl Genet 111:467–478

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Bichler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Zimmer EA, Jupe ER, Walbot V (1988) Ribosomal gene structure, variation and inheritance in maize and its ancestors. Genetics 120:1125–1136