Theor Appl Genet (2005) 111: 467–478
DOI 10.1007/s00122-005-2034-4

ORIGINAL PAPER

Tae-Jin Yang · Yeisoo Yu · Song-Bin Chang
Hans de Jong · Chang-Sik Oh · Sang-Nag Ahn
Eric Fang · Rod A. Wing

Toward closing rice telomere gaps: mapping and sequence characterization of rice subtelomere regions

Received: 12 January 2005 / Accepted: 11 April 2005 / Published online: 18 June 2005

Abstract Despite the collective efforts of the international community to sequence the complete rice genome, telomeric regions of most chromosome arms remain uncharacterized. In this report we present sequence data from subtelomere regions obtained by analyzing telomeric clones from two 8.8 × genome equivalent 10-kb libraries derived from partial restriction digestion with *Hae*III or *Sau*3AI (OSJNPb *Hae*III and OSJNPc *Sau*3AI). Seven telomere clones were identified and contain 25–100 copies of the telomere repeat $(CCCTAAA)_n$ on one end and unique sequences on the opposite end. Polymorphic sequence-tagged site markers from five clones and one additional PCR product were genetically mapped on the ends of chromosome arms 2S, 5L, 10S, 10L, 7L, and 7S. We found distinct chromosome-specific telomere-associated tandem repeats (TATR) on chromosome 7 (TATR7) and on the short arm of chromosome 10 (TATR10s) that showed no significant homology to any International Rice Genome Sequencing Project (IRGSP) genomic sequence. The TATR7, a degenerate tandem repeat which is interrupted by transposable elements, appeared on both ends of chromosome 7. The TATR10s was found to contain an inverted array of three tandem repeats displaying an interesting secondary folding pattern that resembles a telomere loop (t-loop) and which may be involved in a protective function against chromosomal end degradation.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-005-2034-4

Communicated by Q. Zhang

T.-J. Yang
Brassica Genomics Team, National Institute of Agricultural Biotechnology, RDA, Suwon, 441-707, Korea

T.-J. Yang · Y. Yu · R. A. Wing (✉)
Arizona Genomics Institute, University of Arizona, 303 Forbes building, Tucson, AZ 85721, USA
E-mail: rwing@ag.arizona.edu
URL: http://www.genome.arizona.edu

S.-B. Chang · H. de Jong
Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands

C.-S. Oh · S.-N. Ahn
Department of Agronomy, ChungNam Nat'l University, Daejeon, Korea

E. Fang
Clemson University, Clemson, SC 29760, USA

Introduction

Telomeres are specialized chromosomal end structures which play an essential role in maintaining chromatin structure by preventing both end degradation and end-to-end fusion during recombination and in promoting chromosomal end replication (Gasser 2000; Knight and Flint 2000; Riha et al. 2001). Telomeric DNA mediates many biological activities associated with cell-cycle regulation, cellular aging, and the movement and localization of chromosomes within the nucleus and with the transcriptional regulation of subtelomeric genes (de Bruin et al. 2001; Riethman et al. 2001).

Subtelomeric repeats belong to the most rapidly evolving chromosomal sequences and, consequently, vary considerably between chromosomes in a cell and between genotypes or related species. In the human genome, subtelomeres vary in size from 8 kb up to 300 kb (Riethman et al. 2001; Der-Sarkissian et al. 2002), whereas in plant genomes, such as tomato (Broun et al. 1992; Zhong et al. 1998), rice (Ohmido and Fukui 1997; Ohmido et al. 2000, 2001), and tomato (+) potato hybrids (de Jong et al. 2000), such regions can measure up to 1,000 kb. The highly variable distribution of large duplicated subtelomeric segments are caused by homology-based, non-allelic (ectopic) recombination events between nonhomologous chromosomes (Knight

and Flint 2000; Scherf et al. 2001; Der-Sarkissian et al. 2002). Subtelomeric regions have also been shown to be gene-rich (Bishop et al. 2000; Riethman et al. 2001; Scherf et al. 2001; Bringaud et al. 2002; reviewed by Barry et al. 2003). Intense efforts to close telomere gaps and integrate telomere repeat stretches $(TTAGGG)_n$ into the human genome sequence have been successful using genetic and physical mapping of the human telomere regions (Lese et al. 1999; Knight et al. 2000) and large telomere-terminal fragments cloned in specialized yeast artificial chromosome (YAC) cloning vehicles called half-YACs. The result has been the integration of 32 of the 96 telomere regions into the human genome draft sequence (Riethman 1997; Riethman et al. 2001; Xiang et al. 2001). In tomato, species-specific subtelomeric repeats (162-bp unit, TGR1, 500–10,000 copies per locus) have been identified in 20 of the 24 telomeres near the telomere repeat stretch (Ganal et al. 1991; Broun et al. 1992; Zhong et al. 1998). Some subtelomere sequences have also been identified and mapped using a degenerate telomere primer and the Vectorette PCR approach in wheat (Mao et al. 1997), barley (Kilian and Kleinhofs 1992; Röder et al. 1993), and rice (Ashikawa et al. 1994). Following completion of the *Arabidopsis* genome sequence, all ten chromosomal ends, including the telomere repeat regions, were integrated into the genome sequence with the exception of the short arm ends of two chromosomes, 2S and 4S, which contain nucleolar organizer regions (NORs) (http://mips.gsf.de/proj/thal/db/gv/gv_frame.html, December 2004) (Kotani et al. 1999).

Rice (*Oryza sativa* L.) is the most important human food crop in the world and a model system for monocot genomic and evolutionary studies. Rice centromeres and flanking pericentromeric heterochromatin (Singh et al. 1996a, b; Cheng et al. 2001, 2002; Feng et al. 2002) and telomeres (Ohmido and Fukui 1997) have been well studied at the cytogenetic level. Fluorescence in situ hybridization (FISH) revealed that the *Arabidopsis* type $(CCCTAAA)_n$ telomere repeat sequence hybridized to all 24 chromosomal ends (Ohmido et al. 2000, 2001). In that same study, the average size of a rice telomere was measured at 3.5 kb.

At the molecular genetic level, three short subtelomeric sequences have been mapped on the distal ends of chromosome arms 5L, 11S, and 12S (Ashikawa et al. 1994), whereas another three pulsed field gel electrophoresis (PFGE) markers have been mapped on the ends of chromosomes 8, 9, and 11 (Wu and Tanksley 1993). One rice genome-specific tandem repeat (TrsA or Os48: 355-bp tandem repeat) was found on the distal ends of eight chromosome pairs of *indica* rice (*Oryza sativa* ssp. *indica*) and two chromosomal ends of *japonica* rice (*Oryza sativa* ssp. *japonica*) (Ohtsubo et al. 1994; Ohmido and Fukui 1997; Cheng et al. 2001). FISH analysis of extended DNA fibers of *japonica* rice revealed that TrsA was organized in two arrays of 82 kb and 241 kb each that were located adjacent to the telomere tandem arrays, on 6L and 12L, respectively (Ohmido et al. 2000, 2001).

Genome sequencing of *O. sativa* ssp. *japonica* var. Nipponbare has been completed recently under the auspices of the International Rice Genome Sequencing Project (IRGSP), a consortium of research institutions from ten countries. The sequence included the complete sequence of the centromeres from chromosomes 4 (Zhang et al. 2004) and 8 (Nagaki et al. 2004; Wu et al. 2004). Unfortunately, all of the rice telomeric regions still remain as physical gaps despite exhaustive analyses of the almost complete bacterial artificial chromosome (BAC)-, P1 artificial chromosome (PAC)-, and YAC-based physical maps (Chen et al. 2002) and the IRGSP genome sequence (Feng et al. 2002; Sasaki et al. 2002; The Rice Chromosome 10 Sequencing Consortium 2003; The Rice Chromosome 3 Sequencing Consortium 2005; IRGSP 2005). The final frontier in achieving a completed rice genome sequence is to fill the approximately 50 remaining gaps that include ten centromeres, 24 telomeres, as well as several highly repetitive heterochromatin regions. This paper reports on the sequence characterization of seven unique subtelomeric clones containing 25–100 copies of the telomeric array sequence at one end and unique sequence on the opposite end. In-depth sequence annotation provides two chromosome-specific telomere-associated tandem repeats (TATR7 and TATR10s), the occurrence of various transposon insertions, and interesting features of chromosomal ends.

Materials and methods

Library screening

We screened the OSJNPb and OSJNPc rice genomic libraries, representing 8.8 × genome equivalents, with 166,752, and 138,960 clones, respectively; the average size of each insert was 10.8 kb (10-kb libraries) for clones containing rice telomeric repeats (http://www.genome.arizona.edu; Yang et al. 2003). Eleven high-density colony filters of the two libraries were gridded in a 5×5-array pattern on 22.5×22.5-cm Hybond N+ filters (Amersham, UK) and hybridized with a telomere-specific overgo probe (OVG-A: CCCTAAACCCTAAACCCTAAACCC; OVG-B: TAGGGTTTAGGGTTTAGGGTTTAG). Radioactive labeling and hybridization was performed as described by Budiman et al. (2000) and Chen et al. (2002).

DNA sequencing

A total of 96 clones showing strong hybridization signals with the telomere repeat were picked into a 96-well plate, end sequenced with the T3 and T7 vector primers of pCUGIblu21 (Yang et al. 2004) using BigDye terminator chemistry v3.0 [Applied Biosystems (ABI), Foster City Calif.] and electrophoresed on a ABI 3730 x1 automated DNA sequencer. Base-calling was performed

automatically using PHRED, and vector sequences were removed by CROSS_MATCH (Ewing and Green 1998; Ewing et al. 1998). We applied the TGS system F-700 transposon method (Finnzymes, Espoo, Finland) for complete sequencing of the selected 10-kb clones (Yang et al. 2003). High-quality, vector-trimmed sequences were then used for the sequence assembly of the 10-kb clones using PHRAP and CONSED (Gordon et al. 1998).

During the sequencing of the telomere clones, we found insert size variation for two clones—pb005D12 and pb273O07. Clone pb005D12 consisted of only telomere repeat sequences, and the insert size was unstable during growth in *Escherichia coli*, varying between 1000 bp and 40 bp. Clone pb273O07 showed size variation derived from the deletion of 7-bp unit(s) of the telomere repeat (CCCTAAA) (up to six units: 42 bp) and four occurrences of the identical nucleotide substitution in the telomere stretch (TTTAGGG → T *A*TAGGG) [see figure in Electronic Supplmentary Material (ESM)].

Sequence analysis

We analyzed the DNA sequences of the telomeric clones by pairwise comparison using PIPMAKER (Schwartz et al. 2000) and MIROPEAT software (Parsons 1995). Further BLAST and repeat survey analyses were conducted using BLAST-NR (http://www.ncbi.nlm.nih.gov/BLAST/) and REPEATMASKER (http://ftp.genome.washington.edu/RM/webrepeatmaskerhelp.html). The GC composition of every 50-bp window was calculated using GENOMATIX (http://www.genomatix.de/cgi-bin/tools/tools.pl), while the detection of putative genes was analyzed using several web-based gene prediction programs including: FGENE-SH MONOCOT (http://www.softberry.com/berry.phtml), GENESCAN RICE (http://genes.mit.edu/GENSCAN.html), RICE GAAS (http://ricegaas.dna.affrc.go.jp/), and GENEMARK (http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi). The GenBank accession numbers of the sequences described in this paper are given in Table 1.

Genetic mapping

Sequence-tagged site primers were created from non-redundant sequence regions of telomeric clones using the software package PRIMER3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html; Rozen and Skaletsky 2000). Genetic mapping was performed using 96 backcross inbred lines (BILs; BCF_5 of the F_1 between the rice cultivars *Nipponbare* and *Kasalath*) and their genotype scores of reference restriction fragment length polymorphism (RFLP) markers (http://rgp.dna.affrc.go.jp/publicdata/genotypedataBILs/genotypedata.html). An additional population of 80 BILs (BC_1F_6) derived from a cross between the rice lines Milyang 23 and Hapcheonaengmi 3 was also employed to confirm their map position (Oh et al. 2004). Approximately 10 ng of genomic DNA, 200 μ*M* of dNTP, 200 μ*M* of STS primers designed from MWG (Philadelphia, Pa.; http://www.mwg-biotech.com/html/i_custom/i_primer.shtml), and 1 U *Taq* polymerase (Promega, Madison, Wis.) were used for STS PCR in a total volume of 12.5 μl. After 35 cycles of 20 s at 94°C, 30 s at 56°C, and 1 min at 72°C, the PCR products were separated on a 2.5–3% agarose gel [mixture of 1.5–2% Metaphor (BMA, Rockland, Me.) and 1% normal agarose (Fisher Scientific)] in 1× TAE buffer at 4 V/cm for 2 h. When increased resolution was necessary, we separated the PCR products on a 4% acrylamide sequencing gel and visualized the DNA bands by silver staining (Cho et al. 1996).

Results

Identification of telomere clones

In an effort to identify and characterize telomere regions in *japonica* rice (*O. sativa* cv. Nipponbare), we screened the two 10-kb libraries with the $(CCCTAAA)_5$ telomere repeat (Yang et al. 2003) and identified around 200 positive clones, of which 96 were end-sequenced. The ten clones showing telomere repeats at one end were fully sequenced. Three out of these ten, pc174K02 (18 kb), pc201O09 (15 kb), and pb375C08 (6.5 kb), were determined not to be telomeric based on genetic mapping or BLAST analysis with the rice genome sequence and were mapped to chromosomes 5, 1, and 3, respectively.

Table 1 shows the seven remaining candidate clones with arrays of 25–100 copies of the telomere repeat. Their inserts begin with any one nucleotide of the CCCTAAA sequence (italics with underscoring in the table) and end with the expected restriction enzyme cloning site (lowercase letters with underlining in the table) used to construct the libraries. Genomic DNA inserts from the *Hae*III 10-kb library (start with CC and end with GG: cc-----gg) are supposed to be cloned into the *Eco*RV end site (end with -GAT and start with ATC-). However, for two *Hae*III clones, pb106I21 and pb273O03, the inserts begin with the telomere repeat (ctaaacc and aaaccct, respectively) instead of the predicted *Hae*III digestion site (cc) and end with correct *Hae*III digestion (gg) site ligated to the *Eco*RV cloning site of the vector (Fig. 1a, Table 2). The pb005D12 insert, which contains only telomere motifs, starts with ac- and ends with -aa, thereby showing incorrect cloning sites at both ends. For the *Sau*3AI library, genomic inserts (*gatc*—gatc) are supposed to be cloned into a *Bam*HI digested cohesive end site (ending with -gatc). The two *Sau*3AI clones, pc311K23 and pc198E15, begin with the telomere motif (ccctaaa and aaaccct, respectively) instead of a Sau3AI end (gatc). For both of these clones, the cloning vector also showed truncation of the cohesive end (gatc) that resulted in a blunt end. Two *Hae*III clones, pb083I20 and pb027O22, fortuitously

Table 1 Nucleotide sequences of the telomere, degenerate telomere stretch, and cloning site of the telomeric clones

Clone.	Enzyme	Nucleotide sequence and repeat numbers {The position of unique nucleotide}
pb106I21 (Tel10S) <AC134380>	*Hae*III	***CT***AAA(CCCTAAA)$_{13}$*CCCTAA*(CCCTAAA)$_2$*CCCTAACCCCCAAA*CCCTAAA*TCCTAAA* (CCCTAAA)$_2$*(CCCTAAT)*$_2$CCCTAAA*CCGTAATCCATAATG*{atattatcatage···gg} {180···6207}
pb083I20 (Tel7L) <AY367130>	*Hae*III	CCCTAAA(CCCTAAA)$_{15}$*CCCTAA*(CCCTAAA)$_2$*(CCTATA)*$_2$CCCTAAA*(CCTATA)*$_4$(CCCTAAA)$_2$ *(CCTATA)*$_4$(CCCTAAA)$_2$*CCCAAATTCCTAAAACCCTATACCCAATA*CCCTAAAA*CCGTAAACCC* *TAATTGCAACCCTAAAGCCCTATGCCCTAAACCC*{tttacaaatcttac···gg} {255···6116}
pc311K23 (Tel3L) <AY367132>	*Sau3*AI	*CCCTAAA*(CCCTAAA)$_{12}$*CCCTCAACCCTCAACCCTAAACCCCGAACCCAAAACCCTGAA* *CACTGAA(CCCTGAA)*$_4$*CCCTAAGCCCTAAGCCCTAAGCCCTAAATCCTAAA*(CCCTAAA)$_2$ *CCTTTTA*{aatctcaagatgct···gatc} {224···435}
pb027O22 (Tel10L) <AY367131>	*Hae*III	CCTAAA(CCCTAAA)$_{24}$*CCCTAA*(CCCTAAA)$_{17}$*ACCCTAA*(CCCTAAA)$_{14}$*CCCTAA* (CCCTAAA)$_{20}$*CCCTAA*(CCCTAAA)$_{14}$*CCCTATT*{tgagtgggttgtga···gg} {661···2258}
pc198E15 (Tel2S) <AY367133>	*Sau3*AI	***AAA(C***CCTAAA)$_4$*CCCTAA*CCCTAAA*CCCTAC*(CCCTAAA)$_6$*CCCTAA*(CCCTAAA)$_5$*CCCTAA* (CCCTAAA)$_8$*CCCTAAC*{cccaaccttaatga···gatc} {223···897}
pb273O07 <AY367134>	*Hae*III	***AA***A(CCCTAAA)$_{24}$*CCCTAA*(CCCTAAA)$_4$*CCCTCAATCCTAACCCCTCAATCCTAAG* {ecgtctgcagtgg} {233··260}
pb005D12	*Hae*III	***A(C***CCTAAA)$_{6-100}$CCCTA*AA*

[a]GenBank Accession numbers are noted in <>

[b]Capital and italic capital letter indicate telomere and degenerate telomere stretch, respectively. Unique degenerate telomere repeats are represented as gray boxes. Fourteen unique sequence following telomere stretch represented as small letters. The putative restriction enzyme sites were designated with under lines. Left ends showing difference with the expected restriction enzyme sites are designated with bold italics letters with underlining

showed the correct *Hae*III site (cc) by starting with cc of the telomere motif. The resulting analyses indicate that most of the distal telomere sequences were illegitimately ligated into the *Eco*RV-digested blunt end vector for the *Hae*III library, whereas, it is likely that a fraction of the cohesive ends of the *Bam*HI-digested cloning vector were damaged during preparation of dephophorylated linear vector (Yang et al. 2003; Kim et al. 2004) (e.g., mechanical breakage or exonuclease contamination of the restriction enzyme or CIP), resulting in blunt ends that would be suitable for telomere cloning. Distal telomere ends are known to contain a single G overhang. We assume that mechanical breakage inside the telomere stretch or removal of distal single strands of the telomere repeat stretch during the preparation of insert DNA resulted in blunt ends and, consequently, the insert DNAs were competent to be cloned into blunt end vectors.

The telomere repeat array in the seven clones ranges from 180 bp (pb106I21) to 661 bp (pb027O22) and includes a degenerate telomere repeat unit which is relatively unique in each clone [e.g., CCCTAA *T* in pb106I21 (Tel10S), CCCTA *T*A in pb083I20 (Tel7L), CCCT *G*AA in pc311K23 (Tel3L), CCCT-AA in pc098E15 (Tel2S) and pb027O22)] (Table 1).

Mapping and sequence characterization of candidate telomeric clones

Five telomere clones and an additional PCR band derived from a set of primers designed from a telomere-associated tandem repeat (TATR7) were genetically mapped by STS mapping at the termini of six chromosomes. All STS band polymorphisms were dominant and amplified only in *japonica* (cv. Nipponbare) DNA (Fig. 1). To eliminate mapping errors, we developed two separate STS markers from each clone and (or) used two independent mapping populations. The polymorphic STS markers were named based on rice map positions (Table 2).

Tel7L (pb083I20)

Clone pb083I20 contained a 6,116-bp insert, 255 bp of which was the telomere repeat sequence. This clone mapped to the end of the long arm of chromosome 7 (named: Tel**7L**) using two STS markers, 7L-A and 7L-B (Table 2, Fig. 1). The 7L-A marker produced one distinct Nipponbare-specific band (the lower band in Fig. 1), while, interestingly, primer pair 7L-B amplified two independent PCR bands—7-B-u and 7L-B (top two

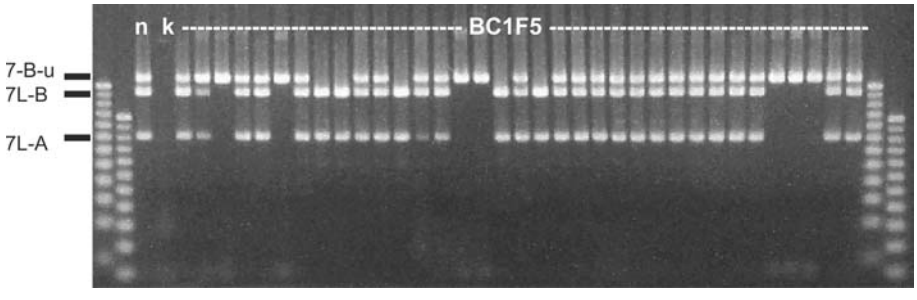


Fig. 1 Duplex agarose gel electrophoresis of two STS PCR products. Two PCRs, 7L-A and 7L-B, were carried out against 98 backcross inbred lines (BIL, BC_1F_5 of rice cultivars *Nipponbare* and *Kasalath*). Both PCR products were separated on a mixture of 1.5% metaphor agarose and 1% regular agarose gel with 4 V/cm for 3 h with 30 min of loading time interval. Two bands, 7L-A and 7L-B, cosegregated and mapped on the telomeric region of the chromosome 7 long arm. An additional unexpected band, 7-B-u, segregated independently and mapped on the other end of chromosome 7. The 25-bp DNA ladders were loaded at 30-min time intervals. *Lanes n* and *k* represent the parental rice cultivars, Nipponbare and Kasalath, respectively

Table 2 Telomere-specific STS markers, their map positions, and primer sequence (*n.d.* not determined)

Original clones	STS markers[a]	Nucleotide (5′ → 3′)	Product (bp)[a]	Position[b]
pc098E15	2S	5′-cctaaaccctaaccccaacc-3′ 5′-gatttcgaccccaacgacta-3′	215	185
pc311K23	3L	5′-tcaccattcttcgttgcatt-3′ 5′-accctgaacactgaaccctg-3′	199	111
pb083I20	7L-A	5′-gcattggagtcattgtgcttt-3′ 5′-tagtgaaattttgggccgac-3′	248	559
	7L-B/7-B-u	5′-ggggttttagccaaagggta-3′ 5′-tctccagccccaaaaattcac-3′	276/317	5,991/n.d.
pb106I21	10S-A	5′-tggattaaaatggagctcgg-3′ 5′-ccgatctgaaccatcgatct-3′	237	4,180
	10S-B	5′-ggcgatgtacgagaacctgt-3′ 5′-cccccaaaccctaaatccta-3′	487	132
pb027O22	10L	5′-ccctaaaccctaaccctaaacc-3′ 5′-acccaaaaactgtccagtcg-3′	360	495

[a]The size (in basepairs) of the PCR product from *Nipponbare* (*O. sativa* ssp. *japonica*). All bands were null in *Kasalath* (*O. sativa* ssp. *indica*)
[b]Represents the nucleotide position of STS primers from the telomeric end
[c]The 7-B-u is an additional band amplified with the same STS primer set and mapped at the termini of the short arm of chromosome 7

bands in Fig. 1). The 7L-B band co-segregated with the 7L-A band (Fig. 1) as expected; however, the unexpected upper band, 7-B-u, segregated independently and mapped to the opposite terminus of the same chromosome, i.e., on the short arm of chromosome 7 (Tel7L). The 7-B-u fragment was cloned into pCUGIblu31 (Yang et al. 2004) and sequenced using the T3 and T7 primers. Two PCR products, 7LB and 7-B-u, showed significant sequence similarity each other except for two insertions of 17 bp and 20 bp at 7-B-u (Fig. 2). Sequence comparison showed that the 7-B-u sequence is a part of a group of telomere-associated tandem repeats (TATR) which appear in more than ten copies in the 6,116-bp insert of the Tel7L clone (Figs. 2a, 3a, b), suggesting that the TATR is present on both arms of chromosome 7 (named as TATR7). The TATR7 has no significant homology with rice sequences in GenBank (http://www.ncbi.nim.nih.gov, December 2004), indicating that TATR7 is likely present in the confined subtelomere region of chromosome 7 of *japonica* rice.

Sequence annotation revealed that the contiguous TATR7 array is interrupted by at least two repetitive elements: first, by an unknown middle repetitive element (speckled box in Fig. 3b, c) and then by a subsequent nested insertion of a highly repetitive element into the unknown element (gray box in Fig. 3b). The highly repetitive element showed identical sequence similarity with part of the RIRE9 solo-long terminal repeat (LTR; GenBank accession no. AB033547; Han et al. 2000). Our study revealed that the complete structure of RIRE9 was not predicted correctly by Han et al. (2000). The existence of the related empty sites (GenBank accession no. AE017070) and flanking 5-bp TSD sequence revealed that the 2,283-bp sequence (between 668 bp and 3,520 bp

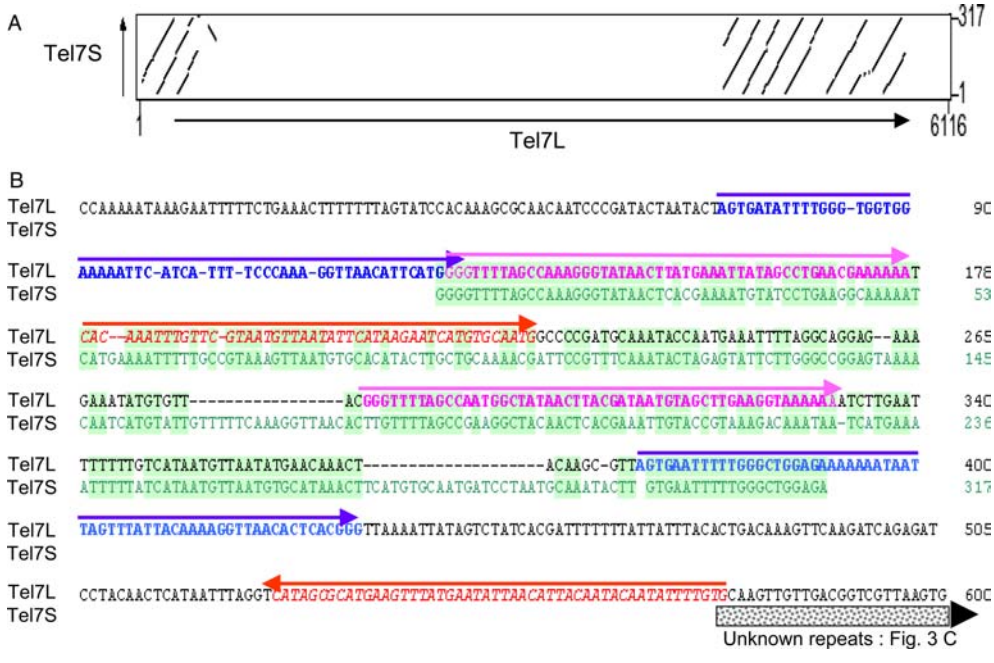


Fig. 2 Sequence analysis of TATR7 in the Tel7L clone. **a** Dotplot of Tel7L (6,116 bp) and Tel7S (317 bp) represents that Tel7S is a part of TATR7 in Tel7L clone. **b** The sequence alignment shows that the TATR7 is similar to Tel7S and consists of a mosaic array of several small repeat units. Three repeat units—*blue*, *pink*, and *red*, respectively—are designated in different *colors and arrows* with direction. A long terminal repeat (LTR) retrotransposon-like sequence which was redundant in rice genome and followed by the TATR7 repeat (*speckled box*). Each repeat unit was aligned under Tel7L sequence

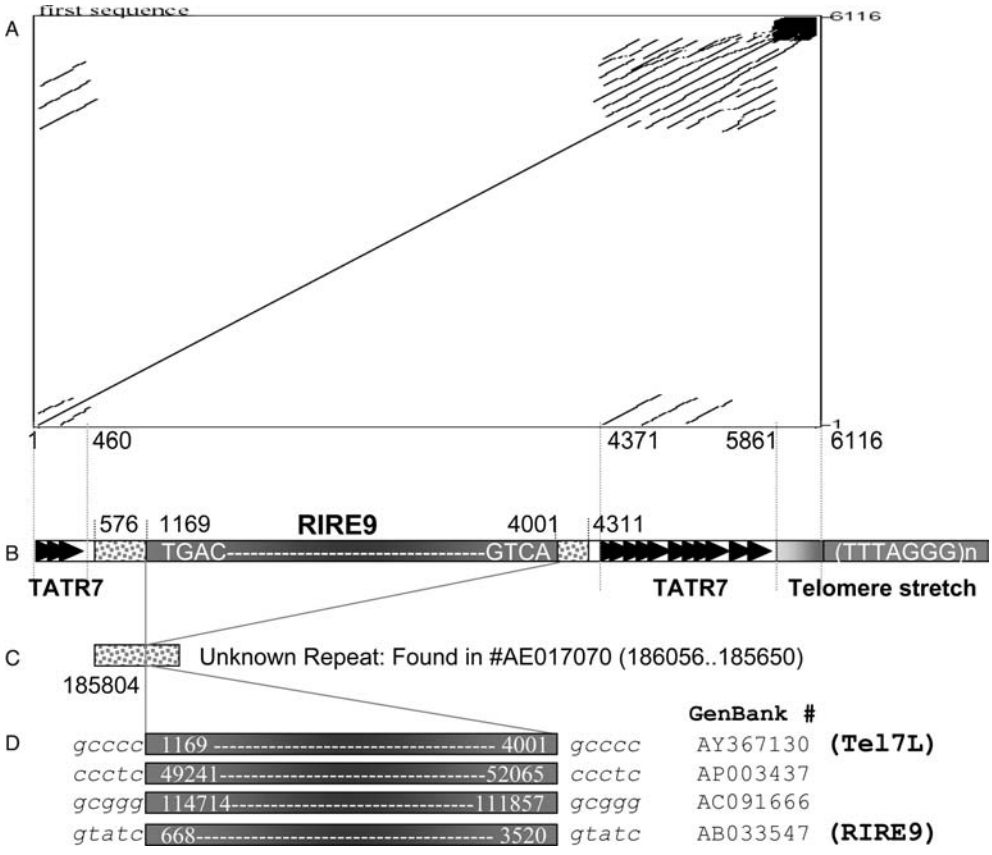


Fig. 3 Schematic representation of the Tel7L sequence. **a** Dotplot shows the appearance of tandem repeats, TATR7, next to the telomere stretch. **b** BLAST revealed that two elements (*speckled and shaded boxes*) are nested insertions into the TATR7. **c** The unclassified element (*speckled box*) shows significant similarity (85%) with several rice genome sequences, such as GenBank accession no. AE017070 inserted into TATR7, and a RIRE9 element (*shaded box*) is a nested insertion into the unclassified element. **d** A few hundred members of the RIRE9 element occur in rice genome. Two other members and a part of the original RIRE9 (668...3520) show 98% sequence identity with unique flanking 5-bp TSD sequences. The positions of corresponding nucleotide sequences are represented as *white numbers* in the *gray boxes* and their GenBank accession numbers are shown at the *right*. The RIRE9 has a 4-bp inverted terminal repeat beginning with TGAC and ending with GTCA

of a total 3,852-bp RIRE9) is the complete element, resembling a solo-LTR of a retrotransposon, where the terminus begins with TG and ends with CA (Fig. 3c) . A number of identical RIRE9 members (2,283 bp long) were also identified with the flanking 5-bp TSD sequence (Fig. 3d), but, no complete structure of this LTR retrotransposon was identified in the IRGSP genome sequence (http://rgp.dna.affrc.go.jp/IRGSP/, December 2004), suggesting that it is a novel structure.

Tel10S (pb106I21)

Clone pb106I21, which measures 6,207 bp and ends with a 180-bp telomere repeat array, genetically maps to the terminal end of the short arm of chromosome 10 (named: Tel**10S**) using two STS markers, 10S-A and 10S-B (Table 2). Tel10S features segments of divergent GC composition, ranging from 20% to 80% in each 50-bp window. A distinct telomere-associated tandem repeat was identified in the Tel-10S (named: TATR10s) that is followed by the 180-bp telomere repeat stretch (Fig. 4b). No significant sequence similarity with the TATR10s was detected in GenBank, suggesting that the repeat is unique to the short arm telomere region of chromosome 10. The remaining part of the sequence, i.e., without the TATR10s, is redundant in the IRGSP genome sequence. BLAST and subsequent analysis of similar sequences revealed that the first 1,417 bp is from part of an LTR retrotransposon, the complete structure of which was identified from the sequence of GenBank accession no. AL662960 (RIRE3-like retrotransposon). This RIRE3-like retrotransposon contains 3,165 bp of 99% identical LTR sequences and an internal sequence encoding 371 amino acids of the gag protein (Fig. 4c). A putative TNP2-like transposase gene (GenBank accession no. AAN37398.1, 3e-99) follows these unique tandem repeats, TATR10s (Tn in Fig. 4b). A signature sequence such as a terminal inverted repeat (TIR) or target site duplication (TSD) was not identified due to significant sequence degeneracy and disruption by the retrotransposons. The TATR10s consists of three repeat units, one copy of RepA (42 bp), four copies of RepB (100 bp), and three copies of RepC (89 bp) (Fig. 5). The arrays of RepA and RepB are organized in an inverted form with an approximate 2-kb interval, forming a putative stem loop structure as shown in Fig. 5. A hypothetical protein coding gene of 70 amino acids is predicted between 4,589 and 5,200 bp in the loop structure based on gene prediction using RICE GAAS (http://ricegaas.dna.affrc.go.jp/).

Tel3L (pc311K23)

The clone pc311K23 has a 435-bp insert with 218 bp of telomere repeat sequence. A set of primers, one from a degenerate telomere repeat and the other from a unique sequence, amplified a distinctly dominant band that was mapped at the terminal end of the long arm of

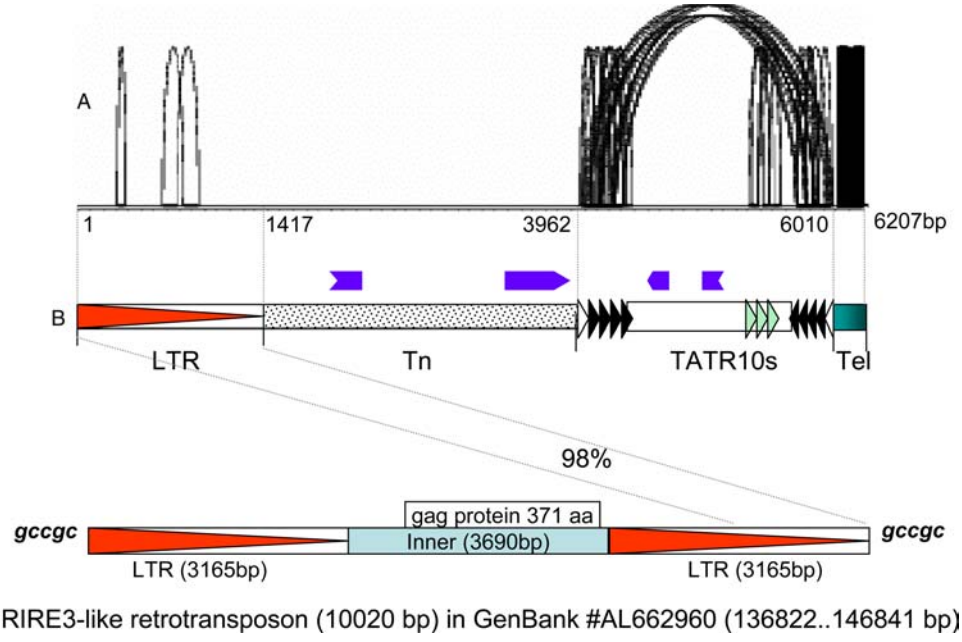
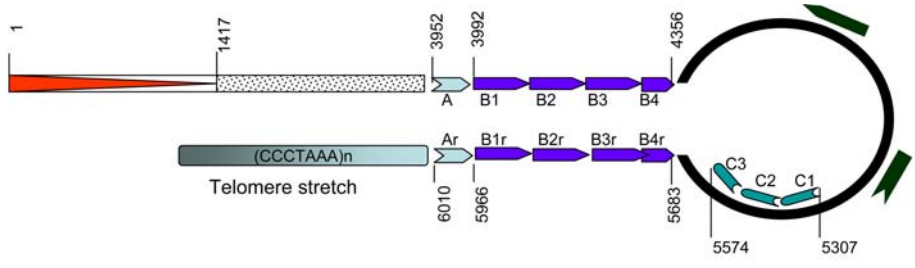


Fig. 4 Sequence annotation and elucidation of TATR10s in the Tel10S clone. **a** MIROPEAT output shows tandem and inverted repeats (TATR10s). **b** Sequence annotation reveals that two elements, a retrotransposon (LTR) and a transposable element (Tn), are followed by the TATR10s. The putative TNP2-like transposase gene is deduced, and their coding regions are denoted on the Tn. **c** Complete structure of a RIRE3-like LTR retrotransposon was found in the rice genome sequence, GenBank accession no. AL662960 (136,822...146,841). The retrotransposon contains a total of 10,020 bp with a 99% identical 3,165-bp LTR and encodes a 371-amino acid gag protein in the 3,690-bp internal sequence

chromosome 3 (Tel**3L**). The sequence (219–435 bp) preceding the telomere array showed significant sequence similarity (93%) with two uncharacterized rice sequences (GenBank accession nos. AP004673 and AP005187) that are not telomeric (Fig. 6).

Fig. 5 Stem loop structure of Tel10S and the sequence of TATR10s. The TATR10s arrays, an inverted array of one copy of repeat A (42 bp) and four copies of repeat B (100 bp), occurred next to the telomere stretch. The inverted array of tandem repeats resembles a stem loop structure. Another tandem repeat array, three copies of repeat C, occurred inside the loop. The original sequence position of tandem repeats in TATR10s is at on the *right*



Repeat A (3952..3993) / Repeat A reverse (telomere stretch ..5967)

```
A1                                                gattgttTATTTATCATAG--AGTTTTACAGTTGATTATATATGATTTTTAAAA
A1R       ccctaaatcctaaaccctaaaccctaatccctaatccctaaaccgtaatccataatgATATTATCATAGCTAGTTTTACGGTTGATTATATATGACTTTTAAAA
```

Repeat B (3994bp..4356bp) / Repeat B reverse (5966bp..5626bp)

```
B1    TTTTAATGTAATTTTAACATCTTTTTAA-TTTATTTTACTTATTTTTATGTTTTTTATT-TGTTTAGCGGACTTAATCGGCTCGTGTTTAATTATATATGATTTTAATA
B2    TTTCTTATAATTTTGAACCTCCTTTTAA-TGGATTTCATTTATTTTTATGAGTTTTTTTATTTTTAAGAGACTTAATTGGCCTGTGGTTAATCATATACAATTTTAATA
B3              TTTTTAACCTCTTTTTAAGTTTTTTTTACTTATTTTTATGA--TTTTTTATTTTTAATGGACTTAATCGGCCCATATTTAATTATATACGATTTTAGT-
B4              TTTTTAACCGC-TTTTAA--TGTTTTTAATTATTTTTATGATTTTTATAATTTTTAATGCAGTTA----------------------------------

B1R             TTTTTTGCAATTTTTGTA-TCTTTTTAATATACTTT----ACTTTTTTTATTTTTATTGAACTTAATCGTCCAATGGTTAATTATGTACGATTTTAATA
B2R -GTTTTTTATAAATTTGAACC TTTTTTAT-TGTATCTTACTTCTTTT-ATGATTCTTTTT----TTAATAAACTTAATCGGTCCTTGGTTAATAGTATACGATTTTAATA
B3R             ----------------------CTTTTTATAATTTTTGAGCACCTTTTTTATT-TTAATGCACTTAATCAACGCTTATTTAATTATATACGATTTTAATA
B4R           TTTTTTAACCTTTTTTTTAAAAGTTTCTACTTATTTTTAACA-TGTTTTTAATGTTTTTACAAGTGCT----------------------------------
```

Repeat C (5297bp..5584bp)

```
C1    TTTTATGATTACATTATACATATA-GGTTGCATATATTTTTCTTA-GTAAGTGAAATTTAGAAGTGCTTCGTAGTTGTCGACAATGTGGATA-TA-TATA
C2              ACAATATACATCTA-GGTTGCATAT-TT---CTTAAGTTGGTGAAATTTGGAAGTTCTTCGTAGTTGTCGATAATGTGGACAATA-TGTA
C3              ACAATATACACGTA-GATTGCATAT-TTTTTCTTAAATAGTTGAAATTTAGATGAGCTTCGAAGTTGTTGACAATGTGGATAATAATCTATGCGGATATA
```

Tel2S (pc098E15)

The pc098E15 clone (insert size: 895 bp) contains a 223-bp array of telomere repeats (Table 1). The remainder of the sequence showed significant sequence similarity (90%) with a full-length expressed sequence tag (EST; GenBank accession no. AK108592, 1–595 bp) that is a functionally unknown gene on chromosome 7 (GenBank accession no. AP005199). A set of primers designed from the degenerate telomere stretch and from the redundant sequence of pc098E15 amplifies a distinctly dominant band which was mapped to the terminal end of chromosome 2 (Tel**2S**). This region has been

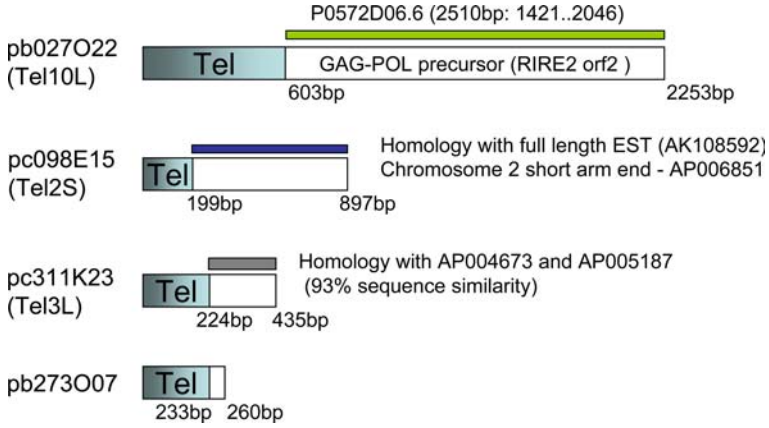


Fig. 6 Illustration of four telomeric clones without a unique TATR sequence. The telomere stretches are presented on the *left* as *blue boxes* for each clone. Sequences following the telomere stretch (*white boxes*) are represented based on the rice homologous sequence *in or following* each *box*

recently uncovered and integrated into the end of the short arm of chromosome 2 by the Japanese Rice Genome Program (RGP; GenBank accession no. AP006851) (Fig. 6).

Tel10L (pb027O22)

Clone pb027O22 (insert size: 2,257 bp) contains a 603-bp stretch of telomere repeat sequence (Table 1). The remaining sequence showed a significant sequence similarity (approx. 85%) with part of the expressed polyprotein gene, the gag-pol precursor of a RIRE2 (GenBank accession no. P0572D06.6: 1,421–2,046 bp). The LTR sequence of the retrotransposon was not identified in the sequence of pb027O22 (Fig. 6). This clone mapped at the terminal end of the long arm of chromosome 10 (Tel**10L**). Difficulties in mapping this clone because of the repetitive sequence following the telomere repeats were overcome by applying a primer set, one from the degenerate telomere repeat and one from the repetitive subtelomeric sequence. Based on several rounds of experiments, the primer set **10L** provides dominant polymorphism (Table 2).

Discussion

Genetic mapping of telomere clones

The TrsA (350 bp, GenBank no. D1453; Ohtsubo et al. 1994) was positioned by FISH on the ends of the long arms of chromosomes 11L and 12L in *japonica* rice and on eight chromosomal ends in *indica* rice (Ohmido and Fukui 1997; Ohmido et al. 2000, 2001). Three telomere-associated sequences, GenBank accession nos. D16335 (101 bp), D16336 (278 bp), and D16337 (313 bp), were mapped on chromosomes 5S, 12S, and 11L, respectively, in *japonica* rice (Ashikawa et al. 1994). The D16336 sequence was mapped on chromosome 11S in the Kasalath (*indica*) genome with a non-allelic form located on chromosome 12S in the Nipponbare (*japonica*) genome. However, none of these maps provide exact sequence information in terms of telomere repeat sequences as do the seven telomere-associated clones identified in this study. Intensive efforts were expended to develop polymorphic STS markers using the sequences of telomere clones. Three of these markers, 10SA, 7LA, and 7LB, were designed from chromosome-specific telomere-associated tandem repeats (TATR10s and TATR7). Other markers, for 2S, 3L and 10L were obtained by using one primer designed from the degenerate telomere arrays and the other from the subtelomere sequence. All of the STS markers analyzed were inherited in a dominant manner in Nipponbare (*O. sativa* ssp. *japonica*) and null in Kasalath (*O. sativa* ssp. *indica*). An STS survey with additional *japonica* and *indica* germplasm will help to clarify the divergent presence of the telomere-associated sequences between these subspecies.

Sequence characteristics of telomere-associated sequences

The unique chromosome-specific telomere-associated tandem repeats TATR7 and TATR10s—in the Tel7L and Tel10S clones, respectively—are interrupted by other transposons such as RIRE3 and RIRE9 that are dispersed into several hundreds of copies throughout the rice genome. Polymorphic subtelomere regions have been shown to serve as hot spots for the nested insertion of non-LTR retroelements, such as long interspersed nucleotide elements (LINEs) and short interspersed nucleotide elements (SINEs), as in the *Trypanosoma brucei* (Bringaud et al. 2002) and *Chlorella* genomes (Higashiyama et al. 1997; Noutoshi et al. 1998). The accumulation of transposable elements in subtelomere regions has been postulated to account for both chromosome stability and genome rearrangements (Zhang and Peterson 1999; Bringaud et al. 2002; Lonnig and Saedler 2002; Barry et al. 2003). FISH analyses using telomere clones Tel10S and Tel7L revealed strong

hybridization on many chromosomal regions. Our data based on sequence and FISH analyses showed that the rice subtelomere regions also contain various transposons.

An interesting point is the occurrence of an intact EST sequence adjacent to the telomere array in pc098E15 (Fig. 6). This sequence showed 90% similarity with the 5′ region of a 2,284 bp full-length EST (one of the 28 K full-length cDNA clones: GenBank accession no. AK108592, 1–595 bp). This non-telomeric EST is located as a single exon gene on chromosome 7 (GenBank accession no. AP005199.3: gene no. P0627E10.30) and has not been functionally characterized (Kikuchi et al. 2003). Based on the sequence information of the Japanese RGP (GenBank accession no. AP006851), the end of the short arm of chromosome 2 contains the complete sequence of the 2,284-bp full-length EST. An expressed polyprotein gene, the gag-pol precursor of a retrotransposon RIRE2, is followed by a telomere repeat array in pb027O22 (Fig. 6). A similar organization was also found in the telomere region of *Arabidopsis* of the short arm of chromosome 1 where an expressed gene sequence without an intron (GenBank accession no. At1g81020) is followed by a telomere repeat array, suggesting that this may be one mechanism for maintaining telomere structure in plants.

Sequence characteristics of TATR7 and TATR10s

Both novel telomere-associated repeats TATR10s and TATR7 contained many poly-adenylation signal sequences, strong stop codon sequences, and polyA tail-like sequences, all of which are characteristics of SINE and LINE elements. However, no similarity was found with classified elements. Subtelomeric repeats have been shown to confer a capacity for gene diversification, especially for "contingency" (virulence factor) genes, which have very important roles in parasite and in mammalian host genomes (Chiurillo et al. 1999, 2000, 2002a, 2002b; del Portillo et al. 2001; Scherf et al. 2001; Barry et al. 2003). The TATRs are interrupted by subsequent insertions of retrotransposons or transposable elements. Transposons are able to mediate large-scale genome reorganization by virtue of their ability to induce chromosomal rearrangements such as deletions, duplications, inversions, reciprocal translocations (reviewed in Zhang and Peterson 1999; Lonnig and Saedler 2002), and small-scale gene evolution (Song et al. 1998; Witte et al. 2001; Bringaud et al. 2002). The TATR7 sequences appear to be located at both ends of chromosome 7 and seem to be dispersed in the large region because it is interrupted by other transposons. The presence of TATR7 at both ends of chromosome 7 might have occurred by transposon-mediated recombination between the ends. Such a feature has also been observed on the ends of *Arabidopsis* chromosome 5 in which 700 bp of unique telomere-associated repeats have been identified (Kotani et al. 1999).

The most probable function of repetitive subtelomere sequences is to prevent telomere shortening, such as in the case of telomerase activity loss (Lundblad and Blackburn 1993). The formation of telomere loops (T-loops) is one of the broadly known chromosomal end features that protects against degradation of telomere ends. Loops are created by tucking G-overhangs (3′ telomeric single-strand overhangs) back into the duplex region of telomeres through interactions with TRF2 (Stansel et al. 2001). The various sizes of the T-loops, which range from 1 kb to 25 kb, are found in many organisms and in vitro (Murti and Prescott 1999; Munoz-Jordan et al. 2001; Stansel et al. 2001). T-looping controls gene activation in yeast (de Bruin et al. 2001). However, no G-overhangs have been identified in some of the chromosome ends of *Arabidopsis* by primer extension/nick translation (PENT) assays, implying that two distinct telomere architectures exist in plants (Riha et al. 2000). The folding loop structure has been detected in rice chromosome ends based on a high-resolution fiber FISH study using a TrsA subtelomeric sequence (Ohmido et al. 2001). The subtelomeric repeat sequence of Tel10S has the potential to form a stem loop structure such as the one shown in Fig. 5, which resembles T-loops. If a chromosome end does not contain the G-overhangs, as reported by Riha et al. (2000), the stem loop structure may function as a backup mechanism for protecting the end from degradation. The telomere repeat can be elongated by homologous recombination in the yeast cell without telomerase, and tens of kilobases of subtelomeric repeats can be rapidly amplified by unequal crossovers, which has also been observed in non-homologous recombination. These large blocks of tandem arrays of subtelomeric sequences may help stabilize the telomeres by promoting a heterochromatin-like structure (reviewed in McEachern et al. 2000; Kojima et al. 2002).

Subtelomeric sequence as an identity for homologous chromosome pairing

None of the TATRs identified in this study showed significant similarity with the known rice subtelomere repeat, TrsA1. The complex structure of subtelomere regions with various units of tandem or interspersed repeats has been shown to represent chromosome identity in several mammalian and lower eukaryotic organisms (Higashiyama et al. 1997; Myler et al. 1999; Kojima et al. 2002; Sunkin et al. 2002). In plant genomes, very little information on subtelomere sequence is available (Richards et al. 1992) compared with the abundance of data obtained from numerous functional studies with telomeres and related proteins (Riha et al. 2000, 2001; Yu et al. 2000; Chen et al. 2001; Gallego and White 2001; Bundock and Hooykaas 2002; Riha and Shippen 2003). The two TATRs found on chromosomes 7 and 10 showed totally different and unique repeat sequences. Chromosome 7 represented symmetric ends, while the end of the short arm of chromosome 10 showed a stem loop like-structure, as discussed above. GC composition

in the subtelomere sequence showed quite distinctive and biased distribution, and the sequence units of the degenerate telomere repeat array were distinctive to each telomere clone. These unique features may contribute to chromosomal identity in the pairing of homologous chromosomes during meiosis and mitosis. The sequence information obtained in this study will be used in future investigations to identify BAC and PAC clones proximal to the subtelomere physical gap to help complete the total rice genome sequence. This may provide us with a further understanding of chromosomal structure in terms of telomere function.

Acknowledgements This work was supported by grants from the United States Department of Agriculture Cooperative Research, Education and Extension Service, the National Science Foundation and the Department of Energy (Principal Investigator—RAW).

References

Ashikawa I, Kurata N, Nagamura Y, Minobe Y (1994) Cloning and mapping of telomere-associated sequences from rice. DNA Res 1:67–76

Barry JD, Ginger ML, Burton P, McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? Int J Parasitol 33:29–45

Bishop R, Gobright E, Nene V, Morzaria S, Musoke A, Sohanpal B (2000) Polymorphic open reading frames encoding secretory proteins are located less than 3 kilobases from *Theileria parva* telomeres. Mol Biochem Parasitol 110:359–371

Bringaud FN, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, Berriman M, Hall N, Donelson JE, Baltz T (2002) A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. Eukaryot Cell 1:137–151

Broun P, Ganal MW, Tanksley SD (1992) Telomeric arrays display high levels of heritable polymorphism among closely related plant varieties. Proc Natl Acad Sci USA 89:1354–1357

Bruin D de, Zaman Z, Liberatore RA, Ptashne M (2001) Telomere looping permits gene activation by a downstream UAS in yeast. Nature 409:109–113

Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. Genome Res 10:129–136

Bundock P, Hooykaas P (2002) Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in Arabidopsis MRE11 mutants. Plant Cell 14:2451–2462

Chen CM, Wang CT, Ho CH (2001) A Plant gene encoding a myb-like protein that binds telomeric GGTTTAG repeats in vitro. J Biol Chem 276:16511–16519

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 14:1–10

Cheng Z, Presting GG, Buell CR, Wing RA, Jiang J (2001) High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics 157:1749–1757

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704

Chiurillo MA, Cano I, Da Silveira JF, Ramirez JL (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. Mol Biochem Parasitol 100:173–183

Chiurillo MA, Beck AE, Devos T, Myler PJ, Stuart K, Ramirez JL (2000) Cloning and characterization of *Leishmania donovani* telomeres. Exp Parasitol 94:248–258

Chiurillo MA, Peralta A, Ramirez JL (2002a) Comparative study of *Trypanosoma rangeli* and *Trypanosoma cruzi* telomeres. Mol Biochem Parasitol 120:305–308

Chiurillo MA, Santos MRM, Da Silveira JF, Ramirez JL (2002b) An improved general approach for cloning and characterizing telomeres: the protozoan parasite *Trypanosoma cruzi* as model organism. Gene 294:197–204

Cho YG, Blair MW, Panaud O, McCouch SR (1996) Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. Genome 39:373–378

del Portillo HA, Fernandez-Becerra C, Bowman S, Oliver K, Preuss M, Sanchez CP, Schneider NK, Villalobos JM, Rajandream MA, Harris D, Pereira da Silva LH, Barrell B, Lanzer M (2001) A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. Nature 410:839–842

Der-Sarkissian H, Vergnaud G, Borde YM, Thomas G, Londono-Vallejo JA (2002) Segmental polymorphisms in the proterminal regions of a subset of human chromosomes. Genome Res 12:1673–1678

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

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X et al (2002) Sequence and analysis of rice chromosome 4. Nature 420:316–320

Gallego ME, White CI (2001) RAD50 function is essential for telomere maintenance in *Arabidopsis*. Proc Natl Acad Sci USA 98:1711–1716

Ganal MW, Lapitan N, Tanksley SD (1991) Macrostructure of the tomato telomeres. Plant Cell 3:87–94

Gasser SM (2000) A sense of the end. Science 288:1377–1379

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

Han CG, Frank MJ, Ohtsubo H, Ohtsubo E (2000) New transposable elements identified as insertions in rice transposon Tnr1. Genes Genet Syst 75:69–77

Higashiyama T, Noutoshi Y, Fujie M, Yamada T (1997) Zepp, a LINE-like retrotransposon accumulated in the Chlorella telomeric region. EMBO J 16:3715–3723

International Rice Genome Sequencing Project (IRGSP) (2005) The map based sequence of the rice genome. Nature (in press)

Jong JH de, Zhong X-B, Fransz PF, Wennekes-van Eden J, Jacobsen E, Zabel P (2000) High resolution FISH reveals the molecular and chromosomal organisation of repetitive sequences of individual tomato chromosomes. Chromosomes Today 13:267–275

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 301:1849

Kilian A, Kleinhofs A (1992) Cloning and mapping of telomere-associated sequences from *Hordeum vulgare* L. Mol Gen Genet 235:153–156

Kim HR, Yang TJ, Kudna DA, Wing RA (2004) Construction and application of genomic DNA libraries. In: Christou P, Klee H (eds) Handbook of plant biotechnology. Wiley, New York, pp 71–80

Knight SJL, Flint J (2000) Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. J Med Genet 37:401–409

Knight SJL, Lese CM, Precht K, Kuc J, Ning Y, Lucas S, Regan R, Brenan M, Nicod A, Martin Lawrie N, Cardy DLN, Nguyen H, Hudson TJ, Riethman H, Ledbetter DH, Flint J (2000) An optimized set of human telomere clones for studying telomere integrity and architecture. Am J Hum Genet 67:320–332

Kojima KK, Kubo Y, Fujiwara H (2002) Complex and tandem repeat structure of subtelomeric regions in the Taiwan cricket, *Telogryllus taiwanemma*. J Mol Evol 54:474–485

Kotani H, Hosouchi T, Tsuruoka H (1999) Structural analysis and complete physical map of *Arabidopsis thaliana* chromosome 5 including centromeric and telomeric regions. DNA Res 31:381–386

Lese CM, Fantes JA, Riethman HC, Ledbetter DH (1999) Characterization of physical gap sizes at human telomeres. Genome Res 9:888–894

Lonnig W, Saedler H (2002) Chromosome rearrangements and transposable elements. Annu Rev Genet 36:389–410

Lundblad V, Blackburn EH (1993) An alternative pathway for yeast telomere maintenance rescues est1-senescence. Cell 73:347–360

Mao L, Devos KM, Zhu L, Gale MD (1997) Cloning and genetic mapping of wheat telomere-associated sequences. Mol Gen Genet 254:584–591

McEachern MJ, Krauskopf A, Blackburn EH (2000) Telomeres and their control. Annu Rev Genet 34:331–358

Munoz-Jordan J, Cross GAM, de Lange T, Griffith JD (2001) T-loops at trypanosome telomeres. EMBO J 20:579–588

Murti KG, Prescott DM (1999) Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. Proc Natl Acad Sci USA 96:14436–14439

Myler PJ, Audleman L, de Vos T, Hixson G, Kiser P, Lemley C, Magness C, Rickel E, Sisk E, Sunkin S, Swartzell S, Westlake T, Bastien P, Fu G, Ivens A, Stuart K (1999) Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes. Proc Natl Acad Sci USA 96:2902–2906

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

Noutoshi Y, Arai R, Jujie M, Yamda T (1998) Structure of the Chlorella Zepp retrotransposon: nested Zepp clusters in the genome. Mol Gen Genet 259:256–263

Oh CS, Lee SJ, Yoon DB, Suh JP, Ahn SN (2004) QTLs for domestication-related and agronomic traits in temperate japonica weedy rice. Korean J Breed 36:20–30

Ohmido N, Fukui K (1997) Visual verification of close disposition between a rice A-genome specific DNA sequence (TrsA) and the telomere sequences. Plant Mol Biol 35:963–968

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

Ohmido N, Kijima K, Ashikawa I, de Jong JH, Fukui K (2001) Visualization of the terminal structure of rice chromosomes 6 and 12 with multicolor FISH to chromosomes and extended DNA fibers. Plant Mol Biol 47:413–421

Ohtsubo H, Ohtsubo E (1994) Involvement of transposition in dispersion of tandem repeat sequences (TrsA) in rice genomes. Mol Gen Genet 245:449–455

Parsons JD (1995) Miropeats: graphical DNA sequence comparisons. Comput Appl Biosci 11:615–619

Richards EJ, Chao S, Vongs A, Yang J (1992) Characterization of *Arabidopsis thaliana* telomeres isolated in yeast. Nucleic Acids Res 20:4039–4046

Riethman H (1997) Closing in on telomeric closure. Genome Res 7:853–855

Riethman HC, Xiang Z, Paul S, Morse E, Hu XL, Flint J, Chi HC, Grady DL, Moyzis RK (2001) Integration of telomere sequences with the draft human genome sequence. Nature 409:948–951

Riha K, Shippen DE (2003) Ku is required for telomeric C-rich strand maintenance but not for end-to-end chromosome fusions in *Arabidopsis*. Proc Natl Acad Sci USA 100:611–615

Riha K, McKnight TD, Fajkus J, Vyskot B, Shippen DE (2000) Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. Plant J 23:633–641

Riha K, McKnight TD, Griffing LR, Shippen DE (2001) Living with genome instability: plant responses to telomere dysfunction. Science 291:1797–1800

Röder MS, Lapitan NL, Sorrells ME, Tanksley SD (1993) Genetic and physical mapping of barley telomeres. Mol Gen Genet 238:294–303

Rozen S, Skaletsky HJ (2000) PRIMER3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

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

Scherf A, Figueiredo LM, Freitas-Junior LH (2001) Plasmodium telomeres: a pathogen's perspective. Curr Opin Microbiol 4:409–414

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PIPMAKER—a web server for aligning two genomic DNA sequences. Genome Res 10:577–586

Singh K, Multani DS, Khush GS (1996a) Secondary trisomics and telotrisomics of rice: origin, characterization, and use in determining the orientation of chromosome map. Genetics 143:517–529

Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS (1996b) Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). Proc Natl Acad Sci USA 93:6163–6168

Song WY, Pi LY, Bureau TE, Ronald PC (1998) Identification and characterization of 14 transposon-like elements in the noncoding regions of members of the Sa21 family. Mol Gen Genet 258:449–456

Stansel RM, de Lange T, Griffith JD (2001) T-llp assembly in vitro involves binding of TRF2 near the 3′ telomeric overhang. EMBO J 20:5532–5540

Sunkin SM, Kiser P, Myler PJ, Suart K (2002) The size difference between Leishmania major riedlin chromosome one homologues is localized to sub-telomeric repeats at one chromosomal end. Mol Biochem Parasitol 109:1–15

The Rice Chromosome 10 Sequencing Consortium (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. Science 300:1566–1569

The Rice Chromosome 3 Sequencing Consortium (2005) Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. Genome Research (in press)

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98:13778–13783

Wu KS, Tanksley SD (1993) Genetic and physical mapping of telomeres and macrosatellites of rice. Plant Mol Biol 22:861–872

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

Xiang Z, Morse E, Hu XL, Flint J, Chi HC, Grady DL, Moyzis RK, Riethman HC (2001) A sequence-ready map of the human chromosome 1q telomere. Genomics 72:105–107

Yang TJ, Yu Y, Nah G, Atkins M, Lee S, Frisch D, Wing RA (2003) Construction and utility of 10 kb libraries for efficient clone gap closure for rice genome sequencing. Theor Appl Genet 107:652–660

Yang TJ, Yu Y, Frisch D, Lee S, Kim HR, Kwon SJ, Park BS, Wing RA (2004) Construction of various copy number plasmid vectors and their utility for genome sequencing. Genom Inform 2:153–158

Yu EY, Kim SE, Km JH, Ko JH, Cho MH, Chung IK (2000) Sequence-specific DNA Recognition by the Myb-like domain of plant telomeric protein RTBP1. J Biol Chem 275:24208–24214

Zhang J, Peterson T (1999) Genome rearrangements by nonlinear transposons in maize. Genetics 153:1403–1410

Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, Wing RA, Han B (2004) Structural features of the rice chromosome 4 centromere. Nucleic Acids Res 32:2023–2030

Zhong X-B, Fransz PF, Wennekes-Eden J, Ramanna MS, van Kammen A, Zabel P, de Jong JH (1998) FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato. Plant J 13:507–517