Theor Appl Genet (1997) 95:359–368

R. Motohashi · K. Mochizuki
H. Ohtsubo · E. Ohtsubo

Structures and distribution of *p-SINE1* members in rice genomes

Received: 15 June 1996 / Accepted: 22 November 1996

Abstract We determined copy numbers of *p-SINE1*, a short interspersed element (SINE) in rice, and found that *p-SINE1* was distributed in the *Oryza sativa* genome with an average spacing of about 1 member per 70 kb. We identified 31 *p-SINE1* members located at different loci by cloning and characterized them by sequencing. Comparison with a consensus sequence derived from their sequences revealed that all of the *p-SINE1* members contained base substitution mutations at various positions. In addition to the substitutions, some members contained deletions, insertions and tandem duplications of a few bases or of a large DNA segment. These mutations seem to have occurred to inhibit transcription from each *p-SINE1* member by RNA polymerase III. PCR using a pair of primers that hybridize with the sequences flanking each *p-SINE1* member revealed that many of the *p-SINE1* members were present at corresponding loci in strains belonging to all rice species carrying the AA genome. Several of them were, however, present at corresponding loci in strains belonging to a limited number of species or in a limited number of the strains belonging to a rice species. These *p-SINE1* members are supposed to be useful in the identification and classification of various rice strains with the AA genome. Simple tandem repeats of a trinucleotide (CAT) or dinucleotide (AG) sequence existed in the flanking regions of 2 *p-SINE1* members. Such repeats, called microsatellite DNA, varied in number even in the cultivated rice strains examined, suggesting that microsatellite DNA is useful for the identification and classification of cultivars.

Key words Retroposon · SINE · Target site duplication · *Oryza sativa* · AA genome

Communicated by J. Beckmann

R. Motohashi · K. Mochizuki · H. Ohtsubo · E. Ohtsubo (✉)
Institute of Molecular and Cellular Biosciences,
The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Introduction

SINEs (short interspersed elements) are retroposons which make their copies by the retroposition of RNA made by RNA polymerase III. The human *Alu* sequence was the first SINE to be identified and characterized (Schmid and Jelinek 1982; Lee et al. 1984; Nelson et al. 1989; Batzer et al. 1990, 1994). Many other kinds of SINEs have been found in mammals, fish and insects: rodent *B1* and *B2* (Krayev et al. 1982), canoidae *Can* (Coltman and Wright 1994), artiodactyl *PRE1* (Kaukinen and Varvio 1992), equine *ERE-1* (Sakagami et al. 1994), salmon *HapI* (Murata et al. 1993), squid *SK* (Ohshima et al. 1993) and chironomid *CP1* (He et al. 1995). Several kinds of SINEs have even been found in plants: *p-SINE1* in rice (Umeda et al. 1991; Mochizuki et al. 1992), *TS* in tobacco (Yoshioka et al. 1993) and *SIbn* in *Brassica napus* (Deragon et al. 1994). It is thus evident that SINEs are ubiquitous elements in higher eukaryotes.

We have previously identified seven numbers of *p-SINE1* located at different loci in the *Oryza sativa* genome. These contain a promoter for RNA polymerase III, like the SINEs of the animal system, and an AT-rich region with a variable length of T-rich pyrimidine tract at the 3′ end (Umeda et al. 1991; Mochizuki et al. 1992). Direct repeats of target sequences are present at the regions flanking each *p-SINE1* member. Some *p-SINE1* members are present at corresponding loci in strains belonging to two closely related species, *O. sativa* and *O. rufipogon*, but not in strains belonging to the other rice species with the AA genome. These *p-SINE1* members are assumed to have been inserted into the respective loci by retroposition during divergence of the rice species (Mochizuki et al. 1992; Hirano et al. 1994). These *p-SINE1* members have been shown to be useful for identifying and classifying various rice strains carrying the AA genome (Mochizuki et al. 1993; Ohtsubo et al. 1993).

In this paper, we report first that there is an extraordinary number of copies of *p-SINE1* in the *Oryza sativa* genome. We then report the identification and characterization of many members of *p-SINE1*, which contained various kinds of mutations including those caused by a drastic change in a large DNA segment. These mutations seem to have occurred to inactivate each *p-SINE1* member which is not transcribed by RNA polymerase III. Some of the *p-SINE1* members would be useful for classifying various strains of rice, which self pollinate to give offspring, and for inferring their relationships.

Materials and methods

Rice strains

The rice strains used are listed in Table 1. Total DNA from these strains was prepared as described previously (Ohtsubo et al. 1991).

Polymerase chain reaction (PCR), inverse PCR (IPCR) and cloning of the amplified fragments

PCR and IPCR were carried out as described previously (Tenzen et al. 1994) using a relevant pair of primers (Table 2). The PCR products were cloned essentially as described by Tenzen et al. (1994), except that they were ligated to the linear pCR™II vector DNA supplied with the TA cloning kit (Invitrogen) and introduced into INVα F′ cells (Invitrogen).

Screening of clones from a genomic library

The *Eco*RI-digested total DNA from *O. sativa* L. cv. 'IR36' was ligated with the *Eco*RI-digested DNA of plasmid pUC118 (Vieira and Messing 1987) and then transformed into *Escherichia coli* XLI-BLUE MRF′ cells; the ampicillin-resistant transformants were selected on plates containing 100 μg/ml of ampicillin. Colony blotting and hybridization was essentially carried out as described by Sambrook et al. (1989) using the *p-SINE1* probe, Mc3con (Table 2). The probe DNA was labeled at its 5′ ends with γ-[^{32}P]ATP (Amersham, 185 TBq/nmol) using phage T4 polynucleotide kinase (Takara).

Table 1 A list of rice strains used for genomic testing

Species	Strain	Reference or source
O. sativa L.	Nipponbare	Mochizuki et al. 1993
ecosp. japonica cv.	Koshihikari	K. Okuno
	Sasanishiki	Mochizuki et al. 1993
	T65	Mochizuki et al. 1993
O. sativa L.	IR36	K. Okuno
ecosp. indica cv.	108	Mochizuki et al. 1993
	C340	Mochizuki et al. 1993
	C5924	Ohtsubo et al. 1991
O. glaberrima	W025	Y. Sano
	GMS1	Mochizuki et al. 1993
	W440	Mochizuki et al. 1993
O. longistaminata	W1451P1	Mochizuki et al. 1993
O. meridionalis	W1625	Mochizuki et al. 1993
O. glumaepatula	W1192	Y. Sano

DNA sequencing

Sequencing reactions were carried out using the *Taq* cycle sequencing kit for Shimadzu DNA sequencer ver.2 (Takara), and DNA sequences were analyzed in a 4.5% Long Ranger gel (AT Biochem) using an automatic sequencer DSQ-1000 (Shimadzu). Alternatively, sequencing reactions were carried out using the Dye-termination sequencing kit (Applied Biosystems), and DNA sequences were analyzed in a 6% polyacrylamide gel using an automatic sequencer 373A (Applied Biosystems).

Slot-blot hybridization

Solutions of total rice DNA at concentrations of 200, 100, 50, 25, 12.5, 6.25, 1.56, 0.78, 0.39, 0.19 or 0.09 μg/ml in a total volume of 100 μl were prepared. Slot-blot hybridization was carried out using the [^{32}P]-labeled MC3con DNA as a probe, as described previously (Tenzen et al. 1994), except that hybridization was carried out 55°C, and the filter was washed in 2 × SSC, 0.1% SDS for 10 min at 55°C followed by washing twice in 0.1 × SSC containing 0.1% SDS for 10 min at 55°C. The copy number was estimated by using the DNA of plasmid pKAY3 or pKAY5, a pUC118 derivative carrying an insert with *p-SINE1*-r3 or *p-SINE1*-r5, respectively (Mochizuki et al. 1992), as control. The size of the haploid rice genome was considered to be 4.3×10^8 bp (Arumuganathan and Earle 1991).

Accession numbers

The nucleotide sequence data reported appears in the DDBJ, EMBL and GenBank nucleotide sequence databases under the accession numbers, D85045 ~ D85067.

Results

Copy numbers of *p-SINE1* in rice genomes

Previous Southern hybridization analysis indicated that *p-SINE1* is dispersed in the chromosomes of *O. sativa* and *O. glaberrima* (Umeda et al. 1991). We determined the copy number of *p-SINE1* in *O. sativa* L. cv 'Nipponbare' and *O. glaberrima* GMS1 by slot-blot hybridization to be about 6,500 and 3,000 per haploid genome, respectively (not shown), which indicated that the *p-SINE1* members were distributed at an average spacing of about 1 member per 70 kb in the *O. sativa* genome. These members must, however, be assumed to be distributed at random in the rice genome, and some of them therefore, closely linked to each other in one of two orientations in the rice chromosomes. We thus tried to amplify the fragments containing the regions flanked by 2 *p-SINE1* members by PCR using primers cp13 and cp23, either singly or as a pair; these primers hybridize with sequences located inside the *p-SINE1* entities (see Table 2). Gel electrophoresis revealed that any one of the primers produced PCR-amplified fragments, showing length polymorphism, from the total DNA of three strains belonging to *O. sativa* and *O. glaberrima* (Fig. 1). Some fragments seemed to be

Table 2 Synthetic oligodeoxyribonucleotides used in this study. Oligodeoxyribonucleotides Mc3con, S61 and S62 were used as probes to identify *p-SINE1* sequence in the clones obtained. Oligodeoxyribonucleotides cp13 and cp23 were used as primers for IPCR. Pairs of oligodeoxyribonucleotides, F25 and R25, F29 and R29, etc., were used as primers for the PCR to amplify the fragments containing *p-SINE1*-r25, *p-SINE1*-r29 etc., respectively. Lowercase letters in the sequence indicate linker sequences containing the *Eco*RI site. Numbers are coordinates to the *p-SINE1* consensus sequence defined as 1-122 (see Fig. 2A), taking the 5′ end of *p-SINE1* as position 1

Oligonucleotide	Sequence (5′ → 3′)	Position
Mc3con	CCCAGGGGTCTTCCGGCTAGCTCCACAA GGTGGTGGGCTAG	9–49
S61	AAGAAACGCCCAGGGGTCTTCCAG	1–24
S62	GACGCGAATATTAGGGAAGG	98–117
cp13	ccggaattcTCGGCTAGCCCACCACCTTG	34–53
cp23	ccggaattcGGTTCGAAGCCTCACCCCT	58–76
F21	TCATTGTTTCTCGCCTT	
F25	GGATGGCTTCAGCAGGATCA	
R25	TTCTGACAGGGAATCAAATG	
F29	CTACACTGCTAGTGGTGCTG	
R29	TTCACCAACTCTGTCAAATG	
F30	CCACATAAGTGCTATGTAGT	
R30	GGGCTCCGTCTAGTATACCG	
F32	AGTACAGAAGGTAATCACGT	
R32	AACTGACTCTTATTAGACTGG	
R32-CTA	TATGGTATTGAGGAGCAGAT	
F34	TTGGACCATTCTTGCACA	
R34	AGTAATCACTGCACTTGA	
F38	GTAGCAGAGTTCTGCACTAG	
R38	CACAATATTAAGGTCAGCTT	
F102	GATCAAACAGGGTCATT	
R102	GATCTGCTGATGTGCCTC	
F103	GATCTCTATGTACTCATGT	
R103	GATCCAACTGGCTGTTGTCT	

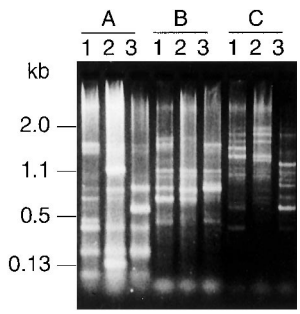


Fig. 1A–C An agarose gel (2%) showing PCR products. PCR products were obtained using primers cp13 and cp23 (**A**), or primers cp13 (**B**) and cp23 (**C**), and electrophoresed. Templates used for PCR were total DNA prepared for *O. sativa* 'IR36' (*lane 1*), *O. sativa* Nipponbare (*lane 2*), *O. glaberrima* GMS1 (*lane 3*)

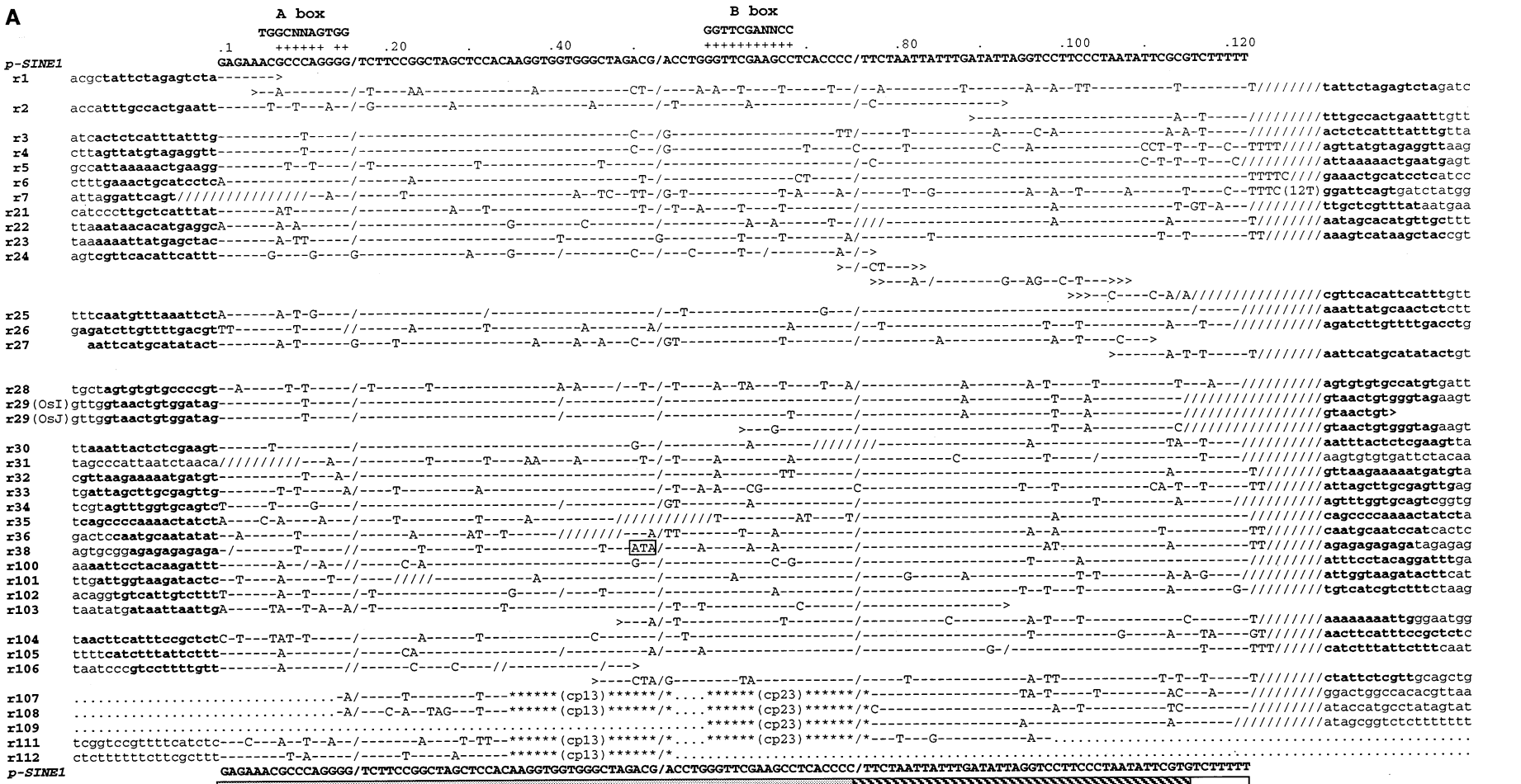
generated in common from the rice strains examined, suggesting that some *p-SINE1* members were positioned closely at the same loci in their chromosomes in one of two orientations.

Identification and characterization of *p-SINE1* members

p-SINE1 members isolated by IPCR

To identify and characterize *p-SINE1* members located at various loci in *O. sativa* or *O. glaberrima*, we carried out IPCR using primers cp13 and cp23 to amplify the fragments containing the end regions of *p-SINE1* (Table 2). These primers hybridize with the *p-SINE1* sequence and prime DNA synthesis towards the outside of the sequence. We also used total DNA which had been digested with *Mbo*I and circularized as a template. We obtained 13 clones of the PCR-amplified fragments which hybridized with a *p-SINE1* probe, S61 or S62 (Table 2). Nucleotide sequencing revealed that 7 clones had the two end sequences of *p-SINE1* that flanked chromosomal sequences. We then carried out PCR to isolate the fragments containing the entire *p-SINE1* sequence by using a relevant pair of primers that hybridized with the chromosomal sequences (see Table 2) and by using total DNA as a template. Nucleotide sequencing of the 7 clones of the PCR-amplified fragments revealed that they actually contained an entire *p-SINE1* sequence, these were named *p-SINE1*-r100–r106 (Fig. 2A).

The other 6 clones contained only the sequence of one end of *p-SINE1* which was connected with a chromosomal sequence. We tried to identify these *p-SINE1* members with an entire sequence by PCR using a primer which hybridized with the chromosomal sequence flanking a *p-SINE1* member in each of the 6 clones. This method (which we call one-primer PCR) is based on the assumption that the fragments containing the entire *p-SINE1* sequence could be amplified by PCR if the primer used happened to hybridize nonspecifically with the other chromosomal sequence flanking the *p-SINE1* member with an entire sequence. We found that only 1 primer, F21, (Table 2) gave rise to the PCR-amplified fragment that hybridized with a *p-SINE1* probe Mc3con (Table 2). Nucleotide sequencing revealed that the fragment in fact contained the entire *p-SINE1* sequence, which we named *p-SINE1*-r21 (Fig 2A). The other 5 members with an end portion of the *p-SINE1* sequence were then named r107–r109, r111 and r112 as shown in Fig. 2A.


A
A box
TGGCNNAGTGG
B box
GGTTCGANNCC
p-SINE1
GAGAAACGCCCAGGGG/TCTTCCGGCTAGCTCCACAAGGTGGTGGGCTAGACG/ACCTGGGTTCGAAGCCTCACCCC/TTCTAATTATTTGATATTAGGTCCTTCCCTAATATTCGCGTCTTTTT
r1
r2
r3
r4
r5
r6
r7
r21
r22
r23
r24
r25
r26
r27
r28
r29(OsI)
r29(OsJ)
r30
r31
r32
r33
r34
r35
r36
r38
r100
r101
r102
r103
r104
r105
r106
r107
r108
r109
r111
r112
p-SINE1
B
Strain
p-SINE1-r102
Microsatellite DNA region
ATTCGCGTCTGTTTTGTTGTCGTCGTCTTTCTAAGCGTGCGTGCTTGCAAAGGAAAGGAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGCACATCAGCAGATCA
O. sativa ecosp. indica IR36
O. sativa ecosp. indica 108
O. sativa ecosp. indica C340
O. sativa ecosp. indica C5924
O. sativa ecosp. japonica T65
O. glaberrima W025
O. glaberrima GMS1
O. glaberrima W440

p-SINE1 members isolated by screening from a genomic library

To identify and characterize more *p-SINE1* members, we made a library of plasmid pUC118 containing the *Eco*RI fragments of *O. sativa* L. cv. 'IR36' and screened the clones which hybridized with the *p-SINE1* probe Mc3con. Nucleotide sequencing of the 17 positive clones obtained revealed new *p-SINE1* members, which we named *p-SINE1*-r22–r38 (Fig. 2A).

The nucleotide sequences of the *p-SINE1* members identified above and previously were similar to one another and could not be classified into subfamilies (Fig. 2A). A consensus sequence of 122 bp in length was derived from the nucleotide sequences of all the *p-SINE1* members (Fig. 2A), and each *p-SINE1* member showed about 90% homology with this consensus sequence. Every *p-SINE1* member, except for one, was flanked by a direct repeat of a sequence between 9 and 17 bp in length which is believed to be the target sequence duplicated upon retroposition. The consensus sequence consisted of two regions: The GC-rich region (65% GC), which contained the sequences homologous to those of the internal promoter, "A box" and "B box", for RNA polymerase III; and the AT-rich region (70% AT) with a T-rich pyrimidine tract at the 3′ end (see Fig. 2A). A computer search revealed that the homologous sequences of *p-SINE1* members (or the consensus sequence) did not show extensive homology with the other SINEs previously identified in various organisms. However, they did have features characteristic of SINEs, such as the GC-rich region containing the RNA polymerase III promoter, and the AT-rich region and a T-rich tract preceding the GC-rich region.

PCR using primers that hybridized with the sequences flanking each *p-SINE1* member generated fragments forming one band in a gel upon electrophoresis (see representative results shown in Figs. 3 and 4). This indicated that each *p-SINE1* member was inserted as a single-copy sequence in the rice genome. However, PCR using a pair of primers flanking *p-SINE1*-r25 was an exception in that it generated fragments forming two bands from the rice strains belonging to *O. sativa* (Fig. 5A). Southern hybridization analysis revealed that the large fragment, but not the small one, contained the *p-SINE1* sequence (Fig. 5A). The nucleotide sequencing of the two fragments confirmed this finding (data not shown), suggesting that *p-SINE1*-r25 was inserted in one of two homologous sequences (or genes) present in the *O. sativa* genome. We used the sequences flanking each *p-SINE1* member as query sequences of the databases in a computer search and found that no homologous genes and repetitive sequences to the flanking sequences existed.

Mutations observed in *p-SINE1* members

Comparison of the sequences of the *p-SINE1* members with the consensus sequence revealed that many base substitution mutations existed within each member. Most of the substitutions were transitions from G/C to A/T (Fig. 2A). In addition to the substitutions, some members contained other mutations such as deletions, insertions and tandem duplications of a few bases or of a large DNA segment (Fig. 2A). Of those with mutations of a large DNA segment, r7 and r31 seemed to have a deletion in the 5′-end region of *p-SINE1* (Fig. 2A). In the case of *p-SINE1*-r7 a direct repeat sequence 9 bp length was seen in the flanking regions; this was the shortest one found and suggested that in *p-SINE1*-r7 a DNA segment had been deleted that included not only the 5′-end region of the *p-SINE1* sequence but also a portion of the direct repeat sequence of about 14 bp in length which was assumed to have been duplicated upon retroposition. In the case of *p-SINE1*-r31, no direct repeat sequences were seen in the flanking regions, suggesting that this member had a deletion of not only the 5′end region of the *p-SINE1* sequence, but also of a region which included the entire direct repeat sequence that was duplicated upon retroposition. Alternatively, a large sequence may have been inserted into the 5′-end region of *p-SINE1*-r31 resulting in separation of the *p-SINE1* sequence being separated into two parts.

p-SINE1-r103 was found to contain a tandem duplication of a large DNA segment, whereas *p-SINE1*-r38 contained an insertion of a DNA segment 1,536 bp in length in the sequence 5′-ATA-3′ within the *p-SINE1* sequence (Fig. 2A). The insertion sequence, named *Tnr3*, was shown to carry imperfect inverted repeats of about 13 bp in length at its termini that begin with 5′-CACTA-3′ (Motohashi et al. 1996). Furthermore, the subterminal regions of *Tnr3* contained 35 copies of

Fig. 2A,B **A** Nucleotide sequences of *p-SINE1* members and their flanking regions. A consensus sequence derived from the aligned sequences of the *p-SINE1* members is shown in *boldface uppercase letters*. "*A box*" and "*B box*" are the promoter sequences for RNA polymerase III. *Lowercase letters* indicate sequences flanking the *p-SINE1* members. Direct repeats in the flanking sequences are indicated by *boldface letters*. In each *p-SINE1* member, nucleotides identical to those in the consensus sequence are shown by *dashes*, and deletions are indicated by *slashes*. The sequences duplicated in tandem are depicted by an *arrowhead*(*s*) as shown. Two sequences of a *p-SINE1* member, r29, present in *O. sativa* IR36 (*OsI*) and *O. sativa* Nipponbare (*OsJ*) are shown. *p-SINE1*-r38 actually has an insertion of transposon *Tnr3* at the sequence ATA (*boxed*), which appeared as a target site duplication (Motohashi et al. 1996). *Asterisks* are the positions of primer cp13 and cp23 used to obtain the *p-SINE1* members by IPCR. The consensus sequence consists of two regions; the GC-rich region with homology with the tRNA coding sequence (a *solid box*) and an AT-rich region (a *cross-hatched box*) with a T-rich pyrimidine tract at the 3′end (an *open box*). Note that *p-SINE1* members have a variable length of the pyrimidine tract. **B** Nucloetide sequences of the region containing simple tandem repeats of the AG sequence located downstream of *p-SINE1*-r102 in various rice cultivars (see Table 1)

short DNA sequences 15 bp in length as direct or inverted repeats, indicating that *Tnr3* is a transposable element belonging to the *En*/*Spm* family (Motohashi et al. 1996).

Presence or absence of *p-SINE1* members in various rice strains

We carried out PCR using a relevant pair of primers that hybridized with the sequences flanking each *p-SINE1* member to see if that particular *p-SINE1* was present (or absent) at the corresponding locus in a set of strains belonging to various rice species with the AA genome (*O. sativa* ecosp. indica 'IR36', *O. sativa* ecosp. japonica Nipponbare, *O. glumaepatula* W1192, *O. meridionalis* W1625, *O. glaberrima* W025 and *O. longistaminata* W1451P1; see Table 1). When we used primers hybridizable with the sequences flanking *p-SINE1*-r30 or -r34, some rice strains generated PCR fragments which were hybridized with *p-SINE1* probe Mc3con (Table 2). However, the other strains generated fragments which were smaller in size than those with *p-SINE1* and which were not hybridized with the *p-SINE1* probe (Fig. 3; Table 3). These results indicated that *p-SINE1*-r30 or -r34 was present at corresponding loci in some rice strains but not in others.

Similarly, another *p-SINE1* member, r29, was found to be present at the corresponding locus in strains belonging to all rice species except *O. longistaminata* (Fig. 4A; Table 3). The PCR-amplified fragments from Nipponbare and other strains belonging to *O. sativa* ecosp. japonica were, however, slightly larger in size that those from strains belonging to *O. sativa* L. ecosp. indica, including 'IR36', and those strains of other rice species (Fig. 4A; Table 3). Cloning and sequencing of the PCR-amplified fragments from Nipponbare revealed that there was a tandem duplication of a large

Table 3 The presence or absence of *p-SINE1* in various rice strains. The presence (+) or absence (−) of a *p-SINE1* member at the corresponding locus in the strains of rice carrying the AA genome was determined by PCR for the generation of fragments with or without the *p-SINE1* member, respectively, using a relevant pair of primers that hybridize with the regions flanking the members

Strain	*p-SINE1*[a]					
	r30	r34	r29	r103	r38	r25
O. sativa L. ecosp. japonica cv.						
Nipponbare	+	+	+*	+*	+†	+/−
Koshihikari	+	+	+*	+*	+†	+/−
Sasanishiki	+	+	+*	+*	+†	+/−
O. sativa L. ecosp. indica cv.						
IR36	+	−	+	+*	+†	+/−
108	+	−	+	+*	+†	+/−
C340	+	+	+	+*	+†	+/−
C5492	+	+	+	+*	+†	+/−
O. glaberrima						
W025	−	+	+	+*	+†	+
GMS1	−	+	+	+*	+†	+
O. longistaminata						
W1451 P1	−	−	−	+*		+
O. meridionalis						
W1625	−	−	+	+*	+	−
O. glumaepatula						
W1192	−	+	+	+*	+†	+/−

[a] +* indicates generation of the PCR fragment with *p-SINE1* having a tandem duplication, +† indicates generation of the PCR fragment with *p-SINE1* containing transposon *Tnr3*, blank indicates no generation of the PCR fragments

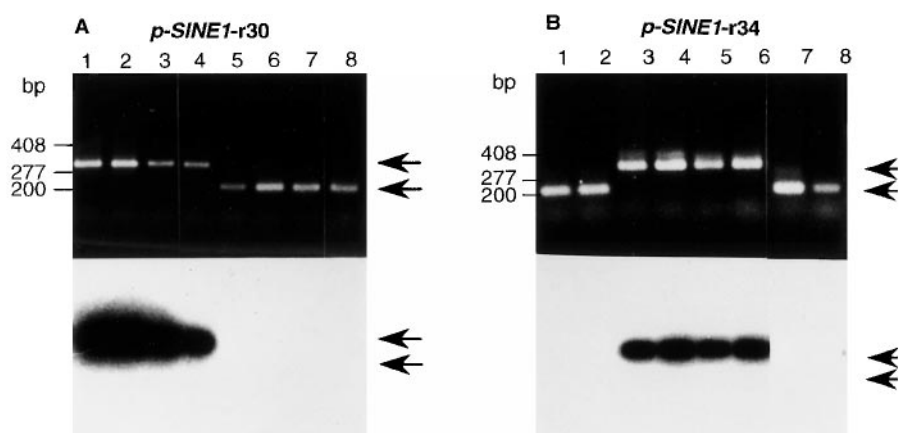


Fig. 3A,B Agarose gels (2%) showing the PCR-amplified fragments with or without *p-SINE1*-r30 (**A**) or *p-SINE1*-r34 (**B**). The *upper portion* of each panel shows an ethidium bromide-stained gel, and the *lower portion* shows an autoradiogram after Southern hybridization with the [^{32}P]-labeled probe Mc3con (Table 2). *Arrows* indicate positions of two fragments (317 and 187 bp in length) with and without *p-SINE1*-r30, respectively, and of two fragments (333 and 198 bp in length) with and without *p-SINE1*-r34, respectively. Total DNA used as templates for PCR were prepared from *O. sativa* 'IR36' (*lane 1*), *O. sativa* 108 (*lane 2*), *O. sativa* C340 (*lane 3*), *O. sativa* Nipponbare (*lane 4*), *O. glaberrima* WO25 (*lane 5*) *O. glumaepatula* W1192 (*lane 6*), *O. meridionalis* W1625 (*lane 7*) *O. longistaminata* W1451 P1 (*lane 8*)

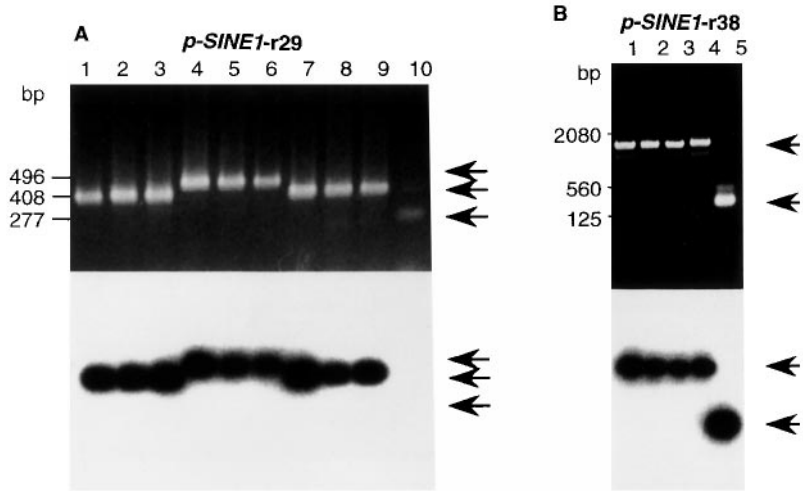


Fig. 4A, B Agarose gels (2%) showing the PCR amplified fragments with or without *p-SINE1*-r29 (**A**) or *p-SINE1*-r38 (**B**). The *upper portion* in each panel shows an ethidium bromide-stained gel, and the *lower portion* shows an autoradiogram after Southern hybridization with the [^{32}P]-labeled probe Mc3con (Table 2). In **A**, total DNA used as templates were prepared from *O. sativa* 'IR36' (*lane 1*), *O. sativa* 108 (*lane 2*), *O. sativa* C340 (*lane 3*), *O. sativa* Nipponbare (*lane 4*), *O. sativa* Koshihikari (*lane 5*), *O. sativa* Sasanishiki (*lane 6*), *O. glaberrima* W025 (*lane 7*), *O. glumaepatula* W1192 (*lane 8*), *O. meridionalis* W1625 (*lane 9*), *O. longistaminata W* 1451P1 (*lane 10*). *Arrows* indicate positions of three fragments, 464, 404 and 261 bp in length. In **B** total DNA used as templates were prepared from *O. sativa* 'IR36' (*lane 1*) *O. sativa* Nipponbare (*lane 2*), *O. glaberrima* W025 (*lane 3*), *O. glumaepatula* W1192 (*lane 4*), *O. meridionalis* W1625 (*lane 5*). Arrows indicate position of two fragments, 1815 and 276 bp in length

DNA segment that started from the middle of *p-SINE1* to a site in the 3′ region (Fig. 2A). Interestingly, a *p-SINE1* member, r103, with a tandem duplication of a different DNA segment within *p-SINE1* (see Fig. 2A) was present in all of the rice strains examined, unlike *p-SINE1*-r29 (Table 3).

p-SINE1-r38 was present at the corresponding locus in the rice strains examined (Fig. 4B; Table 3). However, *p-SINE1*-r38 in strains belonging to *O. sativa*, *O. glaberrima* and *O. glumaepatula* contained transposon *Tnr3* within the *p-SINE1* sequence, while *p-SINE1*-r38 in *O. meridionalis* did not (Fig. 4B; Table 3).

As already described *p-SINE1*-r25 is supposed to be present in one of the two homologous sequences (or genes) in *O. sativa* 'IR36'. All of the strains belonging to *O. sativa* and *O. glumaepatula* generated two PCR-amplified fragments; however, only the large fragment was hybridized with the *p-SINE1* probe, indicating that *p-SINE1*-r25 was present in one of the two homologous sequences in these rice strains, as in *O. sativa* 'IR36' (Fig. 5A; Table 3). The strains belonging to two African rice species, *O. glaberrima* and *O. longistaminata*, however, generated only the large fragment with *p-SINE1*-r25 (Fig. 5A; Table 3). This suggested that the two homologous sequences contained *p-SINE1*-r25 or that the homologous sequence which would give rise to the small fragment with no *p-SINE1*-r25 was deleted in the strains belonging to the African rice species. The strain belonging to *O. meridionalis* generated only the small fragment without *p-SINE1*-r25 (Fig. 5A; Table 3), suggesting that both of the homologous sequences did not contain *p-SINE1*-r25 or that the homologous sequence which would give rise to the large fragment with *p-SINE1*-r25 was deleted in the strain examined.

The other *p-SINE1* members were present in all of the rice strains examined, since the generated PCR-amplified fragments were of the same size and hybridized with the *p-SINE1* probe (data not shown). This indicated that these members were present at the corresponding loci in the rice strains.

Length polymorphism caused by a variation in the number of simple tandem repeats

p-SINE1-r32 was present at corresponding loci in all of the rice strains examined. The PCR-amplified fragments containing *p-SINE1*-r32 from some strains, however, showed length polymorphism (data not shown). Nucleotide sequencing of the PCR-amplified fragments revealed that they contained simple tandem repeats, called microsatellite DNA, of a trinucleotide sequence (CAT) at the region, 200 bp from the 3′ end of *p-SINE1*-r32, and that this length polymorphism was caused by a variation in the number of the repeats. To demonstrate this variation more clearly and to determine the number of the repeats, we carried out PCR using primers (F32 and R32-CAT; Table 2) which hybridized with the sequences flanking the microsatellite DNA and analysed the PCR-amplified fragments by electrophoresis in a polyacrylamide gel. As shown in Fig. 5B, polymorphic fragments were generated, which enabled us to determine the number of the CAT repeats

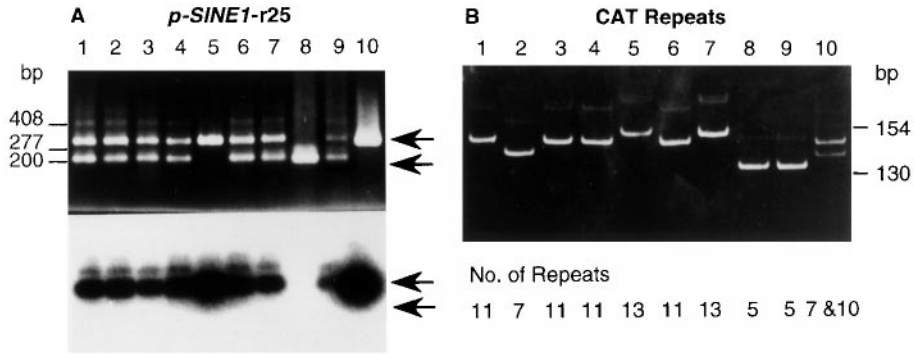


Fig. 5A An agarose gel (2%) showing the PCR-amplified fragments with or without *p-SINE1*-r25. The *upper portion* of each panel shows an ethidium bromide-stained gel, and the *lower portion* shows an autoradiogram after Southern hybridization with the [^{32}P]-labeled probe Mc3con (Table 2). Total DNA used as templates were prepared from *O. sativa* 'IR36' (*lane 1*), *O. sativa* 108 (*lane 2*), *O. sativa* C340 (*lane 3*), *O. sativa* Nipponbare (*lane 4*), *O. glaberrima* W025 (*lane 5*), *O. sativa* Sasanishiki (*lane 6*), *O. sativa* Koshihikari (*lane 7*), *O. meridionalis* W1625 (*lane 8*), *O. glumaepatula* W1192 (*lane 9*), *O. longistaminata* W1451 P1 (*lane 10*). *Arrows* indicate positions of two fragments, 332 and 206 bp in length, with and without *p-SINE1*-r25, respectively. **B** An ethidium bromide-stained polyacrylamide gel (10%) showing the PCR-amplified fragments with simple tandem repeats of the CAT sequence. Total DNA used as templates for PCR were prepared from *O. sativa* 'IR36' (*lane 1*), *O. sativa* 108 (*lane 2*), *O. sativa* C340 (*lane 3*), *O. sativa* C5492 (*lane 4*), *O. sativa* Nipponbare (*lane 5*), *O. sativa* Koshihikari (*lane 6*), *O. sativa* Sasanishiki (*lane 7*), *O. glaberrima* GMS1 (*lane 8*), *O. glaberrima* W025 (*lane 9*), *O. longistaminata* W1451 P1 (*lane 10*). The number of copies of the CAT sequence in each rice strain examined is shown at the *bottom* of the figure

in the rice strains. A strain belonging to *O. longistaminata* generated two fragments with 7 and 10 copies of the CAT sequence (Fig. 5B). The strain used here has been growing perennially and does not produce any fertile seeds, unlike the other rice strains (Y. Sano personal communication; see also Mochizuki et al. 1993). This indicated that the two fragments were generated from a pair of homologous chromosomes probably originating from different strains.

Length polymorphism was also seen in the PCR-amplified fragments containing *p-SINE1*-r102. Cloning and sequencing of the fragments revealed that length polymorphism was caused by variation in other simple tandem repeats or microsatellite DNA of the dinucleotide sequence AG in the region downstream of the 3′ end of *p-SINE1*-r102 (Fig. 2B).

Discussion

p-SINE1 as a retroposon in plant

We determined the copy number of *p-SINE1* in *O. sativa* to be about 6,500 per haploid genome, leading to the assumption that *p-SINE1* members are distributed in the rice genome at an average spacing of about 1 member per 70 kb. *Alu* (about 300 bp in length) is a human SINE (Schmid and Jelinek 1982) which is supposed to be derived from the *7SL RNA* gene. On the basis of its copy number, which is 500,000 per haploid genome, *Alu* is assumed to be distributed at an average spacing of about 1 *Alu* per 5 kb (Batzer et al. 1990). Although *p-SINE1* appeared less frequently in the host genome than *Alu*, various lengths of fragments containing the regions flanked by 2 *p-SINE1* members, which were closely linked in one of two orientations in the rice chromosomes, could be amplified by PCR using a primer(s) that hybridizes with a sequence(s) inside *p-SINE1*. Length polymorphism may be a useful means by which to classify various rice strains and to infer their relationships.

We have identified and characterized 31 *p-SINE1* members at different loci in the rice genome. Direct repeat sequences about 14 bp in length appeared at regions flanking almost all of the *p-SINE1* members. These are supposed to be the target sequences duplicated upon retroposition. The GC contents of these sequences were about 41%, suggesting that *p-SINE1* tended to be inserted into the AT-rich regions like introns, since the GC contents of exons and introns in the rice *wx* gene, for example, are about 61% and 36%, respectively (Umeda et al. 1991). *Alu* family elements are known to be often inserted into microsatellite DNA as targets (Rogaev 1989). In our study, however, only one member, *p-SINE1*-r38, was found to be inserted into the AG repeat sequences (see Fig. 2A).

The *p-SINE1* members, with a few exceptions identified here and previously, were present at corresponding loci in all of the rice strains with the AA genome that were examined. This suggests that most *p-SINE1* members have been inserted into the respective loci in an ancestral rice strain before divergence of the rice species with the AA genome, whereas the exceptional members have been inserted into the respective loci in several rice strains during divergence of the rice species with the AA genome. We assume that all of these members have originated from a *p-SINE1* sequence and accumulated mutations dependent on the length of time after insertion; the original *p-SINE1* might have a sequence(s),

which is the same as or very close to the 122-bp consensus sequence with features characteristic for *SINEs*, such as the GC-rich region containing the RNA polymerase III promoter, and the AT-rich region and T-rich tract preceeding the GC-rich region (see Fig. 2A). Comparison of the *p-SINE1* sequences with the consensus sequence revealed that those *p-SINE1* members which were specifically present in a limited number of the rice strains with the AA genome had substitution mutations at a frequency of about 6% on average (see *p-SINE1*-r2, -r6, -r25, -r29, -r30 and -r34 in Fig. 2A), whereas the other *p-SINE1* members had sustitution mutations at a frequency of about 11% on average. This may support the assumption described above.

Various mutations occurring in the *p-SINE1* members

All of the *p-SINE1* members contained base substitution mutations, most of which were transitions from G/C to A/T (see Fig. 2A). We have previously discussed that a mechanism may exist that preferentially introduces a transition from C to T in interspersed sequences like *p-SINE1* in rice (Mochizuki et al. 1992). We have observed that these members also acquired deletions, tandem duplications and insertions of a few bases. All of these mutations seem to have occurred to inhibit the transcription of each *p-SINE1* member by RNA polymerase III. In fact, almost all of the *p-SINE1* members acquired mutations in the "A-box" sequence of a promoter for RNA polymerase III (see Fig. 2A).

Of the *p-SINE1* members, r35 and r36 had a deletion of a DNA segment, and r29 and r103 contained a tandem duplication of a DNA segment. These mutations occurred at the 5′region of the "B-box" sequence of a promoter for RNA polymerase III (see Fig. 2A). Another member, *p-SINE1*-r38, contained transposon *Tnr3* inserted into the sequence located at the 5′ region of the "B box" (see Fig. 2A). These findings suggest that the 5′ region of the "B box" in *p-SINE1* contains mutational hot spots. This result is interesting when we consider the results of *Alu-Alu* recombination which often occurs at the 3′ region of the "A box" and at the 5′ region of the "B box" of the promoter in the left arm of the *Alu* sequence (Lehrman et al. 1987).

As mentioned above, *p-SINE1*-r38 contained an insertion of *Tnr3*. We have previously reported that *p-SINE1*-r4 in many rice strains contain another transposon, *Tnr2* (Mochizuki et al. 1993). This means that *p-SINE1* becomes a good target for transposition because there is no deleterious effect of any insertions into *p-SINE1* on the growth of plant cells. Expanding on this connection, we mentioned in the Results section that a *p-SINE1* member, r31, may have had a large sequence inserted into the 5′-end region of the *p-SINE1* sequence, explaining why this member has an abnormal structure at its 5′ end (see Fig. 2A). The sequence joined with *p-SINE1*-r31 is 5′---CA-3′ (see Fig. 2A), which is a sequence at the ends of long terminal repeat sequences (LTR) of retrotransposons identified in plants (Grandbastien 1992). This suggests that *p-SINE1*-r31 has acquried an insertion of a retrotransposon. In fact, the sequence of the region which ends with CA is partially homologous to that of the retrotransposon, named *RIRE2*, recently identified in our laboratory (unpublished results).

p-SINE1 members useful for classification of rice strains with the AA genome

Among the *p-SINE1* members identified, *p-SINE*1-r30 was present at the corresponding locus in strains belonging to *O. sativa* but was not present at the corresponding locus in strains belonging to other rice species with the AA genome (see Table 3). *p-SINE1*-r30 may, therefore, be useful in distinguishing strains belonging to *O. sativa* from other closely related species. In this respect, it is interesting that we have found *p-SINE1*-r30 to be present at the corresponding locus in all of the 35 *O. sativa* strains examined and in about 30% of the *O. rufipogon* strains which are thought to be most closely related to those of *O. sativa* (our unpublished results).

p-SINE1-r34 was not present at the corresponding locus in a limited number of strains belonging to *O. sativa* ecosp. indica. Also, *p-SINE1* members, r25, r29 and r30, were not present in strains belonging to particular rice species. Mochizuki et al. (1993) stated that the 2 *p-SINE1* members (r2 and r6), which are not present at corresponding loci in some rice strains, could be useful for classifying various rice strains with the AA genome. We believe, therefore, that if the *p-SINE1* members mentioned above are sought out at corresponding loci in the rice strains, these strains can be classified and their relationships inferred in more detail.

We have described here that *p-SINE1*-r29, with a tandem duplication of a large DNA segment, was present in strains of *O. sativa* ecosp. japonica, whereas *p-SINE1*-r29, without the tandem duplication, was present in other rice species. *p-SINE1*-r29 may thus be useful for distinguishing strains of *O. sativa* ecosp. japonica from the *O. sativa* ecosp. indica. However, it is likely that the tandem duplication may be resolved by homologous recombination to the original non-tandem sequence. As described in the Results section, however, *p-SINE1*-r103 with a tandem duplication of a different *p-SINE1* segment was present in all rice strains with the AA genome, indicating that the tandem duplication is very stable and not subject to homologous recombination.

We have also described here that microsatellite DNA consisting of simple tandem repeats of a trinucloetide sequence were present in the 3′ regions of 2 *p-SINE1* members, r32 and r102, respectively, and that the number of the repeats varied in the cultivars of rice

examined. The microsatellites may thus be used for identification and classification of the rice cultivars.

Acknowledgments We thank Drs. K. Okuno (National Institute of Agrobiological Resources) and Y. Sano (National Institute of Genetics) for providing us with seeds of the rice strains. We also thank Dr. S. Tsuchimoto for critical reading of the manuscript and for a useful suggestion. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, and by a grant Pioneering Project in Biotechnology from the Ministry of Agriculture, Forestry and Fisheries of Japan.

References

Arumuganathan K, Earle DE (1991) Nuclear DNA content of some important plant species, Plant Mol Biol Rep 9 : 208–218

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL (1990) Structure and variability of recently inserted *Alu* family members. Nucleic Acids Res 18 : 6793–6798

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human-specific polymorphic *Alu* insertions. Proc Natl Acad Sci USA 91 : 12288–12292

Coltman DW, Wright JM (1994) *Can* SINEs: a family of tRNA-derived retroposons specific to the superfamily Canodiea. Nucleic Acids Res 22 : 2726–2730

Deragon J-M, Landry BS, Pelissier T, Tutois S, Tourmente S, Picard G (1994) An analysis of retroposition in plants based on a family of SINEs from *Brassica napus*. J Mol Evol 39 : 378–386

Grandbastien M-A (1992) Retroelements in higher plants. Trends Genet 8 : 103-108

He H, Rovia C, Recco-Pimentel S, Liao C, Edström J-E (1995) Polymorphic SINEs in *Chironomids* with DNA derived from the R2 insertion site. J Mol Biol 245 : 34–42

Hirano H, Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E, Sano Y (1994) Retroposition of a plant SINE into the *Wx* locus during evolution of rice. J Mol Evol 38 : 132–137

Kaukinen J, Varvio S-L (1992) Artiodactyl retroposons: association with microsatellites and use in SINEmorph detection by PCR. Nucleic Acids Res 20 : 2955–2958

Krayev SA, Markusheva TV, Kramerov DA, Ryskov AP, Skryabin KG, Bayev AA, Georgiev GP (1982) Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res 10 : 7461–7475

Lee MG-S, Loomis C, Cowan NJ (1984) Sequence of an expressed human β-tubulin gene containing ten *Alu* family members. Nucleic Acids Res 12 : 5823–5836

Lehrman MA, Goldstein JL, Russell DW, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by *Alu-Alu* recombination in a subject with a familial hypercholesterolemia. Cell 48 : 827–835

Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E (1992) Characterization of a plant SINE, *p-SINE1*, in rice genomes. Jpn J Genet 57 : 155–166

Mochizuki K, Ohtsubo H Hirano H-Y, Sano Y, Ohtsubo E (1993) Classification and relationships of rice strains with AA genome by identification of transposable elements at nine loci. Jpn J Genet 68 : 205–217

Motohashi R, Ohtsubo E, Ohtsubo H (1996) Identification of *Tnr3*, a Suppressor-Mutator/Enhancer-like transposable element from rice. Mol Gen Genet 250 : 148–152

Murata S, Takasaki N, Saitoh M, Okada N (1993) Determination of the phylogenetic relationships among pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc Natl Acad Sci USA 90 : 6995–6999

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT (1989) *Alu* polymerase chain reaction. A method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86 : 6686–6690

Ohshima K, Okada N (1994) Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposon in the octopus. J Mol Biol 243 : 25–27

Ohshima K, Koishi R, Matsuo, M, Okada N (1993) Several short interspersed repetitive elements (SINEs) in distant species may have originated from a common ancestral retrovirus: Characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons. Proc Natl Acad Sci USA 90 : 6260–6264

Ohtsubo E, Mochizuki K, Tenzen T, Ohtsubo H (1993) A simple method to classify rice strains with AA genome and infer their relationships by identification of transposable elements at various loci. Gamma Field Symp 32 : 71–83

Ohtsubo H, Umeda M, Ohtsubo E (1991) Organization of DNA sequences highly repeated in tandem in rice genomes. Jpn J Genet 66 : 241–254

Rogaev EI (1989) Simple human DNA-repeats associated with genomic hypervariability, flanking the genomic retroposons and similar to retroviral sites. Nucleic Acids Res 18 : 1879–1885

Sakagami M, Ohshima K, Mukoyama H, Yasue H, Okada N (1994) A novel tRNA species as an origin of short interspersed repetitive elements (SINEs). J Mol Biol 239 : 731–735

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: Laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Schmid CW, Jelinek WR (1982) The *Alu* family of dispersed repetitive sequences. Science 216 : 1065–1070

Tenzen T, Matsuda Y, Ohtsubo H, Ohtsubo E (1994) Transpostion of *Tnr1* in rice genomes to 5′-PuTAPy-3′sites, duplicating the TA sequence. Mol Gen Genet 245 : 441–448

Umeda M, Ohtsubo H, Ohtsubo E (1991) Diversification of the rice *Waxy* gene by insertion of mobile DNA elements into introns. Jpn J Genet 66 : 569–586

Vieira J, Messing J (1987) Production of single stranded plasmid DNA. Methods Enzymol 153 : 3–11

Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to tRNAs. Proc Natl Acad Sci USA 90 : 6562–6566