

TEOS1*, a novel transposable element family from *Oryza sativa

Ya-Ting Chao, Chun-Lin Su, Teh-Yuan Chow, Pei-Fang Lee, Chun-I Chung, Jiann-Jang Huang, Su-Mei Liu, and Yue-Ie C. Hsing*

Institute of Botany, Academia Sinica, Taipei, Taiwan 11529

(Received June 18, 2002; Accepted August 16, 2002)

Abstract. We have uncovered a novel group of transposable elements *TEOS1*, (Transposable Element of *Oryza sativa*) by repeat mining of rice genome sequence data. We first identified a *TEOS1* element, 6,896 bp in length, from a PAC clone of rice chromosome 5. The presence of long terminal repeats (LTRs) and other diagnostic features, such as a primer-binding site (PBS) and a polypurine tract (PPT), indicated that *TEOS1* is a group of LTR retrotransposon. The internal domain of *TEOS1* containing two arrays of tandem repeats and a predicated *Tf1* gene was characteristically distinct from those of other LTR retrotransposons previously described. *TEOS1* elements identified were flanked by 5-bp direct repeats; some identified as solo LTRs were flanked by 5-7 bp direct repeats. The PBS sequence was homologous to the 3' end region of tRNA^{Arg}. The deduced amino acid sequence of *Tf1* is dissimilar to that of any known mobility related proteins. *TEOS1* is probable to be trans-activated by other retrotransposons in the rice genome. We have estimated that there are about 300 *TEOS1* elements per haploid rice genome. The presence of repeats in *TEOS1* offers the potential to form secondary structure during replication and transcription, and to serve as a hot spot for transposon insertion, augmenting the dynamics of molecular interactions of the genomic sequences.

Keywords: *Oryza sativa*; Tandem repeat; *TEOS1*; Transposable element.

Abbreviations: GSS, Genomic survey sequence; HTG, Unfinished high throughput genomic sequence; LINE, Long interspersed repetitive element; LTR, Long terminal repeat; MITE, Miniature inverted repeat transposable element; PBS, Primer binding site; PPT, Polypurine tract; SINE, Short interspersed repetitive element; TE, Transposable element; TEOS, Transposable element of *Oryza sativa*; U5, Unique 5' RNA.

Introduction

Transposable elements (TE) are divided into two main classes according to their mechanism of transposition. Retrotransposons, termed as Class I elements, transpose through an RNA intermediate that is converted into extrachromosome DNA and integrated into the genome. The replication mode could increase their copy numbers rapidly and thereby greatly increase the size of the host genome (Kumar, 1996; SanMiguel and Bennetzen, 1998). In some plants, retrotransposons comprise over 50% of the nuclear DNA content (Pearce et al., 1996; SanMiguel et al., 1996; Suoniemi et al., 1996; Kossack and Kinlaw, 1999). According to the structural organization and the encoded ORFs, retrotransposons could be further typed as *cop*ia, *gypsy*, and the non-LTR retrotransposon. The *cop*ia and *gypsy* type retrotransposons are characterized by the flanking long terminal direct repeats (LTRs), and distinguished by the order of the domains in their polyprotein region.

Class I elements have been characterized in a wide range of plant taxa. The first Class I element identified in plants is *Tnt1* in tobacco (Grandbastien et al., 1989). Oth-

ers plant elements include *Tto1-Tto3* in tobacco (Hirochika, 1993), *BARE-1* in barley (Suoniemi et al., 1996), *Bs1* (Jin and Bennetzen, 1989) and *Opie-1* in maize (SanMiguel et al., 1996), soybean *SIRE-1* in soybean (Laten et al., 1998), and *Athila* (Pelissier et al., 1995) and *Tal 1* (Wright et al., 1996) in *Arabidopsis*.

Due to its small genome size of 430Mb, the smallest among the cereal grasses (Arumuganathan and Earle, 1991), and concerted efforts in its sequence analysis, several retrotransposon families have been well identified in the rice genome. These include *cop*ia type Ty1 (e.g. *Tos 1-20*) (Hirochika et al., 1992, 1996), *gypsy* type Ty 3 (e.g. RIRE3, RIRE7) (Kumekawa et al., 1999, 2001), and SINE (e.g. p-SINE1) (Mochizuki et al., 1992; Motohashi et al., 1996). Cell culture-induced activation of *Tos17* in rice has been applied to insertional mutagenesis analysis (reviewed in Hirochika, 2001). For example, the zeaxanthin epoxidase gene has been tagged using *Tos17* insertion (Agrawal et al., 2001). The feasibility of using *Tos17* in the PCR-screening strategy was demonstrated by screening for a mutant of the homebox gene OSH15 (Sato et al., 1999). Previous studies using computer-based sequence similarity searches have revealed the predominance of MITE elements in rice (Bureau and Wessler, 1994; Bureau et al., 1996; Mao et al., 2000; Turcotte et al., 2001). Many TEs are also identified in rice BAC end sequences (Mao et al., 2000) or genomic sequence (Turcotte et al., 2001).

*Corresponding author. Tel: 886-2-27899590 ext. 312; Fax: 886-2-27827954; E-mail: bohsing@gate.sinica.edu.tw

Rice genome sequencing project is under way by an international effort. Vast amount of sequence data would provide insights to observe previously unknown or rarely discovered repetitive elements. We report here a novel family of transposable elements, which was identified through *in silicon* repeat-mining from the currently available rice genomic sequences. These elements are characterized by their flanking LTRs and internal arrays of tandem repeats. These elements show no sequence similarity to any other transposable elements identified so far and might represent a novel class of transposable elements.

Materials and Methods

Sequencing of P0001A07

P0001A07 is a PAC clone of the HindIII PAC library constructed by members of the Japan Rice Genome Research Program (RGP) using the genomic DNA of the japonica rice Nipponbare with the vector pCYPAC2. Its PAC DNA was sheared (1.6-2 kb and 4.5-5 kb), ligated to a pUC18 vector, and transformed into *Escherichia coli*. DNA sequencing was performed using the automated ABI system with dye terminators and primers as described by the manufacturer. Shotgun clones were sequenced to generate at least 10-fold coverage. The sequences of shotgun clones were base-called and assembled using the Phred/Phrap/Consed system (Ewing et al., 1998; Ewing and Green, 1998; Gordon et al., 1998, 2001). In total, about 4000 reactions were carried out to generate the sequence of the whole PAC clone.

Analysis of Sequence Data

Annotation was based on a combination of the results of database searching and gene prediction programs. Standalone BLAST (Altschul et al., 1997) was used to compare the P0001A07 DNA sequences with the downloaded copies of the NCBI nr and EST databases and TIGR Rice Gene Index. Genscan (Burge and Karlin, 1997) was run with *Arabidopsis* and maize models to predict the exon structure of the genes. Artemis (Rutherford et al., 2000) was used as a sequence viewer for manual adjustment of the splicing sites of putative genes. The extracted amino acid sequences of the predicated genes were searched against nr database using BLASTP for confirmation. For each of the predicated proteins, the pIs and molecular weights were calculated using the PEPTIDESORT program of the GCG package (Devereux et al., 1984). PSORT (Nakai and Horton, 1999) tools were used to identify protein sorting signals. Protein domain characterization was performed using the InterPro database (Apweiler et al., 2001).

To reveal the presence of *Tos17* insertion sites along the rice genome sequence, Cluster FLAG (Fast Local Alignment for Gigabases) program was used to search against the NCBI GSS (Genome Survey Sequence) database (<http://www.ncbi.nlm.nih.gov/dbGSS/index.html>) which contained random "single pass read" genome survey sequences, cosmid/BAC/YAC end sequences, exon trapped genomic sequences, and Alu PCR sequences,

through the online service of the Bioinformatics Department of Biomedical Engineering Center (BMEC), Industrial Technology Research Institute (ITRI), TAIWAN (<http://flag.itri.org.tw/>).

Repeat-Mining for Discovery of TEOS1

Initial analysis of repetitive structure was performed using Miropeats (Parsons, 1995). The observed putative LTRs and the unit of short tandem repeats were identified using bl2seq program (Tatusova and Madden, 1999). The rice nr and HTG databases of GenBank were searched for homologues to the *TEOS1*. The genomic sequences that BLASTN hit originally were downloaded from the GenBank for further analysis and reexamined using Dotter (Sonnhammer and Durbin, 1995) to get a whole view of the sequence structure similarity. The putative target site duplication and PPT sites were examined manually through the sequence viewer of Artemis software. The PBS site was revealed by comparison with that of *RIRE2*. After counting the number of putative *TEOS1* elements, the total number of *TEOS1* was calculated by extrapolation. The searches were performed in GenBank release 124 and 126. Although the rice database grew very rapidly, the two extrapolations gave similar results.

Genomic Southern Blot

Maize (*Zea mays*), wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) seedlings were grown in 24-h darkness at room temperature for a week. Rice (*Oryza sativa* ssp. *japonica* cultivar TN67), *Arabidopsis* (*Arabidopsis thaliana* ecotype Columbia), tobacco (*Nicotiana glauca* cultivar W38), and soybean (*Glycine max* cultivar Shi-Shi) tissues were collected from greenhouse grown plants. Plant genomic DNA was prepared using a CTAB (cetyltrimethyl ammonium bromide) standard protocol (Doyle and Doyle, 1990). Genomic DNA (10 µg) from each plant species were digested overnight with EcoR I at 37°C. Gel electrophoresis and Southern blot analysis were carried out with standard protocols.

Probe was prepared by Dig-PCR probe synthesis kit (Roche) using plasmid DNA isolated from subclone 12H12 of P0001A07 (accession no. AC084218) as the template. The two primers used were Ur-5' (5'-ACGTCTTCACCG ACCGGCTTGC) and Ur-3' (5'-ATAAAGGCCAGCCGT GCAGGC) which located at the 5' and 3' end of each repeat unit. The filters were hybridized at 65°C overnight and washed in 5X SSC, 0.1% N-laurylsarcosine, 0.02% SDS, 1% blocking reagent. Hybridized signal was detected by DIG luminescent kit followed by exposure to X-ray film according to manufacturer's manual (Roche).

Results

Characterization of TEOS1

P0001A07 (accession no. AC084218), an 164-kb insert PAC clone, is located at 25 cM of rice chromosome 5 and is one of the materials of the genome sequencing project conducted in our lab. In the 79-kb to 87-kb region of this

PAC clone, we discovered a novel repetitive element, hereby named *TEOS1* (transposable element of *Oryza sativa*), by repeat-mining. This 6,896-bp element was flanked by 439-bp long terminal repeats (LTRs), and composed of two arrays of tandem repeat sequences separated by 3 spacer sequences (Figure 1). The LTRs of *TEOS1* began with TG and ended with CA, similar to those of the LTRs in most plant retrotransposons (Grandbastien, 1992). At both ends of the element, we found 5-bp direct repeats, presumably formed by duplication of target sites upon insertion. The occurrence of target site duplication is consistent with observations on other plant LTR retrotransposons (Grandbastien, 1998). Several features necessary for retroelement replication were also identified, for instances, the primer binding site (PBS) which was downstream to the 5' LTR and the polypurine tract (PPT) which was upstream to the 3' LTR. There are two base pairs separating the 5' LTR and PBS that complementary to the 3' end sequence of tRNA^{Arg}. *TEOS1* and *RIRE2* (Ohtsubo et al., 1999), a rice gypsy-type retrotransposon, have 80% sequence identities in the U5 region of the LTRs. The stem-loop structure in the U5 (Unique 5' RNA) region of the LTR and PBS exhibited homology (100% sequence identities) to those of *RIRE2*. Between PBS and PPT, there were a 5' array containing 10 tandemly arranged copies of 90-bp elements and a 3' array containing 12 copies of the same element. Sequence homology was high among these

90-bp repetitive elements. Most of them were perfect repeats, some contained one or few base substitutions. BLAST search (Altschul et al., 1997) revealed that these repetitive elements were homologs (86% ~ 100% sequence identities) to the tandem repeats near the telomeres of chromosome 2 (accession no. AF250385) and the rice panicle EST (accession no. AU182318). The predicted secondary structure using GCG package (Devereux et al., 1984) indicated that the tandem repeat region would form a stable giant stem-loop structure, which contained several hair-pin and internal loops (data not shown).

ORF in *TEOS1*

Three EST homologs were identified with the LTR sequence of *TEOS1* (> 95% sequence identities and an E-value of 1e-45) using BLASTN search. Two of them were from panicle EST at flowering stage and the other one from callus EST. Another homolog from panicle EST had significant match to the 90-bp tandem repeat region. Thus, at least one ORF in *TEOS1* was expected. The predicated gene, *Tf1*, spanned 1.2-kb and contained 2 exons. The first exon, with 55-bp in length, was small compared with the second one, 982 bp. The exon 2 was separated from exon 1 by an intron of 159-bp and was located in the 5' array of tandem repeats. There were four repetitive motifs (20-108, 109-197, 198-273, 274-345) in the deduced amino acid sequence of this region (Figure 2A). Each motif was encoded

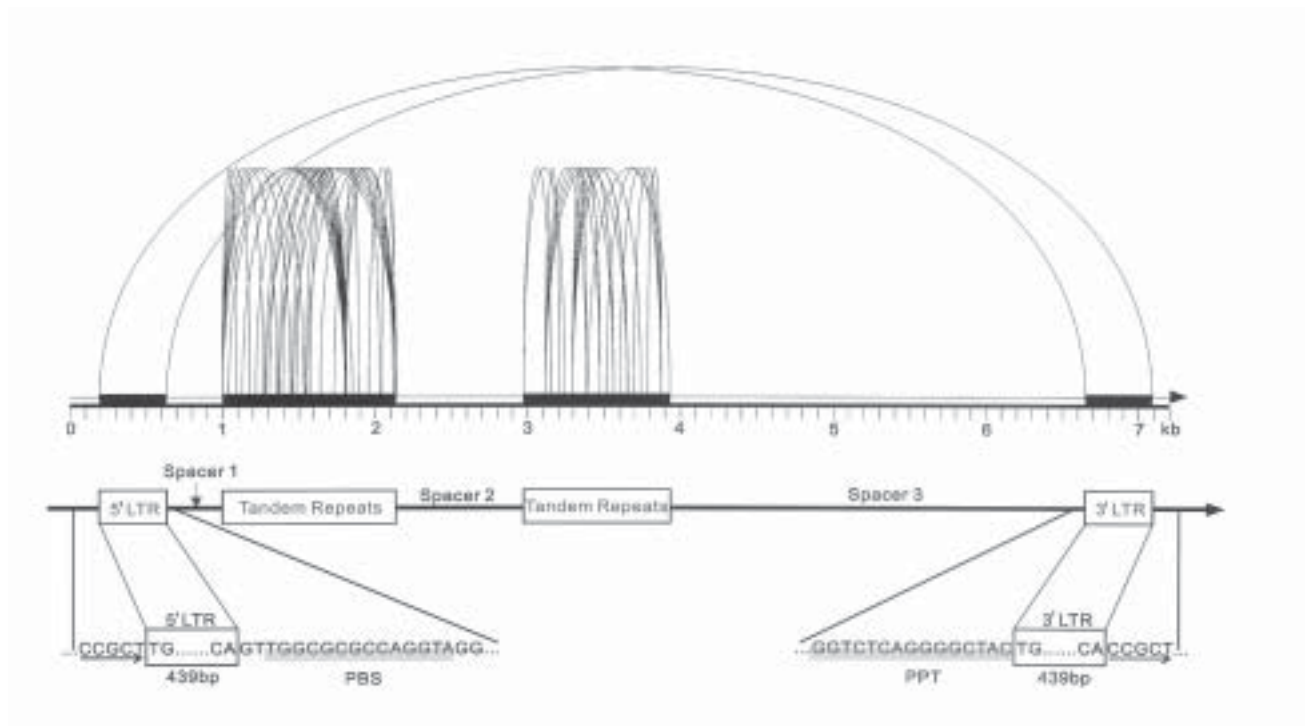


Figure 1. Repeat structures and organization of *TEOS1*. Panel A. Distribution of motifs within *TEOS1*. Sequence comparison is illustrated using Miropeats program with a threshold of 100. The arch lines are drawn between matching sites to indicate the tandem repeats. The two LTRs share an identity of 100% over 439 bp, and the tandemly arranged short repeats share at least 90% identity over at least 80 bp. Panel B. Nucleotide sequences of flanking the LTRs of *TEOS1*. LTRs are boxed, and the 5-bp target-site duplications are indicated by arrows.

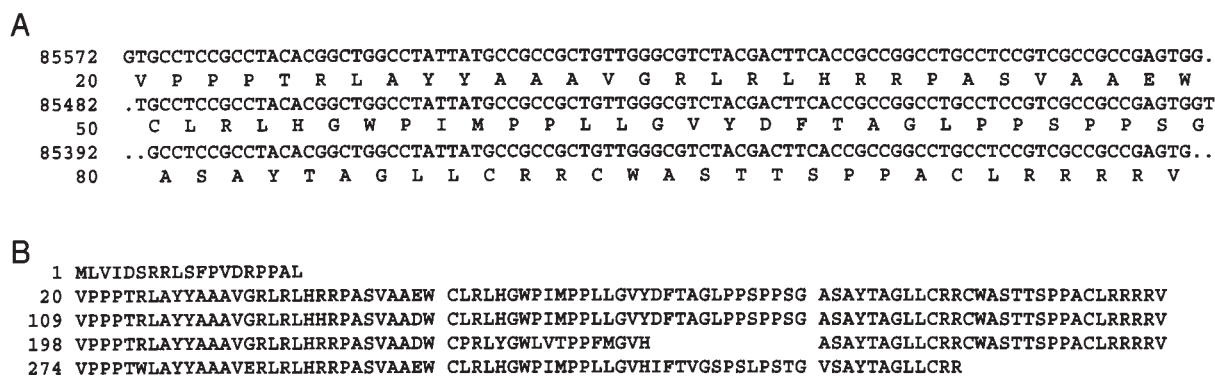


Figure 2. Partial sequence of *TEOS1* and *Tf1*. Panel A. The 3 copies of the 90-bp tandem repeats and their deduced amino acid residues in the second exon of the *Tf1* in *TEOS1*. The dots were introduced for better alignments. Nucleotide positions in P0001A07 and a.a. residue positions in *Tf1* protein are indicated at left. Panel B. The deduced amino acid sequence of the *Tf1* in *TEOS1*.

by three copies of the 90-bp tandem repeats with substitutions (R—W, G—E, E—D, H—Y, P—L, I—V, M—T, L—F, L—M, Y—H, and A—V). The first two motifs differed in only one residue (E vs. D, positions 48 and 137), while base pairs encoding 13 amino acids were deleted in the third and the fourth motif was truncated at the carboxyl end missing the sequence that would encode 17 amino acids. This is illustrated in Figure 2B. The putative *Tf1* protein contained 345 amino acids with a molecular mass of approximately 38 kDa and an isoelectric point of 11.8. The Cysteine-rich character indicated the potentiality of a complex secondary structure of the *Tf1* protein. PSORT search (Nakai and Horton, 1999) predicted the protein to be localized in the nucleus since there were three bipartite nuclear localization signals clustered between residues 90 and 272. We annotated data of other *TEOS1* elements in nr database and showed that homologues of *Tf1* protein were highly variable in size, ranging from 297 to 1187 residues in length, or an estimated molecular mass from 34 kDa to 134 kDa. Of the 23 putative *Tf1* proteins surveyed, all were shown to be basic proteins with high pI values on the basis of their translated amino acid compositions, and sixteen were predicated to be localized at nucleus, as each one of them contained at least 1 bipartite nuclear localization signal peptide in the sequence.

The *TEOS1* Family in Rice Genome

In order to estimate the frequency of *TEOS1*-like elements presented in the rice genome, we used the *TEOS1* sequence as a query to perform similarity search against nr and HTGs (unfinished high throughput genomic sequences) database. The results indicated that there were 117 *TEOS1* homologues, with 32 of them identified from nr, present in the current rice genome sequence database. Each member of the *TEOS1* family currently available from nr database was closely examined for their TE characteristics such as direct repeats, LTRs, PBS and PPT. We found one PPT present at the appropriate position in each of the 28 elements investigated. As to PBS, one site with homology to the 3' end of tRNA^{Arg} was intact and appropriately positioned in each element of the 22 out of the 32

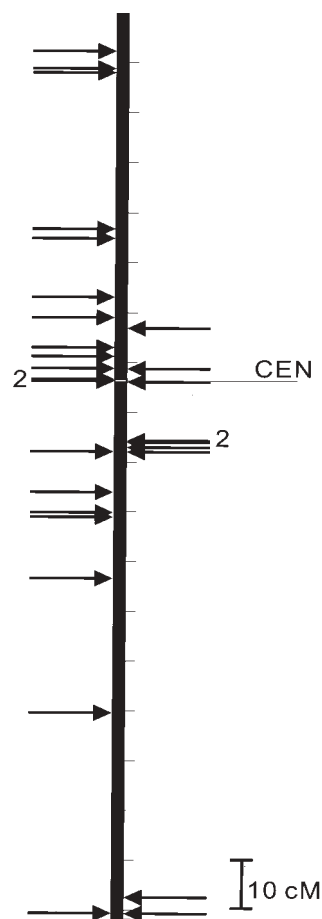


Figure 3. Distribution of *TEOS1* target loci on chromosome 1. The vertical line represent rice chromosome 1. Positions of PAC/BAC clones containing the *TEOS1* elements are indicated by the arrows. Arrows on the left side and right side indicate the *TEOS1* elements which are identified in nr and HTG database, respectively. Multiple clones in a small region are represented by the numbers attached to the arrows.

cases examined. The sequence of the flanking direct repeats of *TEOS1* was not conserved and the integration sites appeared to be random. High level of nucleotide sequence similarity among the *TEOS1* homologues suggested that *TEOS1* elements were still active in transposition until recent evolutionary time. As shown in Table 1, these 32 *TEOS1* are distributed in chromosomes 1, 2, 3, 5, 6 and 10, and over one half of them are in chromosome 1.

In GenBank release 126, there were 1291 released rice completed and unfinished BAC/PAC clones, with a total of 177 Mb sequenced, accounting for approximately 41% of the rice genome. Extrapolation to the 430-Mb of the whole rice genome, there are 300 *TEOS1* estimated in the total genome. It should be noted that in addition to *TEOS1*, we have identified 35 Solo LTRs. Four Solo-LTRs, each was 439 bp in length, from nr database were located in chromosomes 1 and 10. Their characteristic flanking, direct repeats are shown in Table 2. We examined the distribution of the 117 *TEOS1* so far uncovered, which is shown in Table 3. The copy number of *TEOS1* elements in each chromosome ranges from 2 in chromosome 12 to

28 in chromosome 1, without any in chromosomes 9 and 11. Since the sequence of chromosome 1 determination is almost completed, we were able to map the location of the 28 *TEOS1* elements in chromosome 1. This is shown in Figure 3. We note that the distribution of *TEOS1*s was uneven along the chromosome, with a tendency of gravitating towards, or expanding from, the centromeres and telomeres. While the probability of finding a member of the *TEOS1* family may be proportional to the extent of genome sequence information in a chromosome, the presence of sequence specific hot spots for *TEOS1* transposition cannot be ruled out.

The presence of LTR domains and arrays of tandem 90-bp repeats was the distinct structure feature of the family. As in the original *TEOS1* found in P0001A07, the short repeats in the arrays of other *TEOS1*-like elements were also homologues with AF250385 and AU182318. While tandem repeats had heretofore not been reported in other plant transposable elements, we found the arrays of repeats in *TEOS1* highly variable in their repeat numbers, ranging from 3 to 30, and in their length. For instance, the *TEOS1* in P0002B05 (accession no. AP003141) had only

Table 1. Characteristics of *TEOS1* family.

<i>TEOS1</i>	Accession no.	Chr.	Length (kb)	LTR (bp)	TSD ^a 5' / 3'	PBS complementary	Note
<i>TEOS1.1</i>	AC084218	5	6.9	439	CCGCT / CCGCT	3' end of tRNA ^{Arg}	
<i>TEOS1.2</i>	AP003504	1	2.9	439	N.A. / CCGAT	Not applicable	5' truncated
<i>TEOS1.3</i>	AP003258	1	3.5	400	GTTCT / N.A.	Not detected	3' truncated
<i>TEOS1.4</i>	AP003294	1	5.5	421	CACGC / N.A.	3' end of tRNA ^{Arg}	3' truncated
<i>TEOS1.5</i>	AP003141	1	5.8	417	AAGCT / AAGCT	3' end of tRNA ^{Arg}	
<i>TEOS1.6</i>	AP002869	1	6.1	417	CGTAT / CGTAT	3' end of tRNA ^{Arg}	
<i>TEOS1.7</i>	AP003250	1	6.7	438	ATACC / ATACC	3' end of tRNA ^{Arg}	
<i>TEOS1.8</i>	AP002970, AP003213	1	6.7	440	GGCGG / GGCGG	Not detected	
<i>TEOS1.9</i>	AP002860	1	6.8	442	ATAGT / ATAGT	3' end of tRNA ^{Arg}	
<i>TEOS1.10</i>	AP001278	1	6.8	417	TCGAG / TCGAG	3' end of tRNA ^{Arg}	
<i>TEOS1.11</i>	AP002871, AP002873	1	6.8	444	TCAAC / TCAAC	3' end of tRNA ^{Arg}	
<i>TEOS1.12</i>	AP003208	1	6.8	439	CCGAT / CCGAc	Not detected	
<i>TEOS1.13</i>	AP002843	1	7.2	452	gaaaT / atgtT	Not detected	
<i>TEOS1.14</i>	AP003328	1	7.2	441	GAAAT / GAAaA	Not detected	
<i>TEOS1.15</i>	AP003277	1	7.5	439	GGACG / GGACG	3' end of tRNA ^{Arg}	
<i>TEOS1.16</i>	AP002968, AP003204	1	8.1	442	CTATG / CTATG	3' end of tRNA ^{Arg}	
<i>TEOS1.17</i>	AP002482	1	8.2	441	AAAAT / AAAAT	3' end of tRNA ^{Arg}	
<i>TEOS1.18</i>	AP002525	1	8.5	439	TCAGG / TCAGG	3' end of tRNA ^{Arg}	
<i>TEOS1.19</i>	AP003255, AP003411	1	11.4	438	TCTAT / TCTAT	3' end of tRNA ^{Arg}	4.3 kb insertion at spacer 2
<i>TEOS1.20</i>	AP002913	1	20.6	443	CAGTC / CAGTC	3' end of tRNA ^{Arg}	13 kb insertion at spacer 3
<i>TEOS1.21</i>	AB013450	2	2.6	440	N.A. / GGCCA	Not applicable	5' truncated
<i>TEOS1.22</i>	AC084831	3	6.5	439	ATeGG / ATaGG	3' end of tRNA ^{Arg}	
<i>TEOS1.23</i>	AC090714	3	18.2	437	GGTTC / GGTC	3' end of tRNA ^{Arg}	
<i>TEOS1.24</i>	P0040B10	5	7.2	443	CACAT / CACAT	3' end of tRNA ^{Arg}	
<i>TEOS1.25</i>	AP001389	6	8.3	438	gTTCC / cTTCC	3' end of tRNA ^{Arg}	
<i>TEOS1.26</i>	AB023482	6	8.6	441	GATTC / GATTC	3' end of tRNA ^{Arg}	
<i>TEOS1.27</i>	AC091734	10	3.3	451	GTTCC / N.A.	Not detected	3' truncated
<i>TEOS1.28</i>	AC037197	10	4.5	437	CTTTC / N.A.	3' end of tRNA ^{Arg}	3' truncated
<i>TEOS1.29</i>	AC079936	10	6.5	440	cttca / agttt	Not detected	
<i>TEOS1.30</i>	AC078891	10	7.1	438	Aaaac / Attca	Not detected	
<i>TEOS1.31</i>	AC083945	10	7.4	439	GAAAC / GAAAC	3' end of tRNA ^{Arg}	
<i>TEOS1.32</i>	AC090441	10	12.1	438	GGTGG / GGTGG	3' end of tRNA ^{Arg}	4.2 kb insertion at spacer 2

^aTarget site duplications. Identities between repeats are indicated with uppercase lettering; mismatches are indicated with lowercase lettering. TSD was not applicable (N.A.) when *TEOS1* was truncated.

Table 2. Solo LTRs identified from nr database of GenBan.

Solo LTR	Accession no.	Chromosome	Length (kb)	Identity ^a	TSD ^b
<i>TL1.1</i>	AP003236, AP003281	1	439	93%	CCTTG / CCTTG
<i>TL1.2</i>	AP003504	1	439	95%	GTAGGCC / GTAGGCC
<i>TL1.3</i>	AC011806	10	439	93%	CCTTG / CCTTG
<i>TL1.4</i>	AC037197	10	439	94%	GGTAG / GGTAG

^a Sequence identity between Solo LTR and *TEOS1.1* LTR.

^b Target site duplications.

one array consisting 20 copies of short repeats. The variation of repeat number might be the result of unequal recombination.

Only one third of the *TEOS1* elements showed significant similarity in the spacer 1 region, while significant sequence similarities were noted in spacers 2 and 3 of most elements, respectively. Thus, spacer 1 sequence of *TEOS1* was less conserved than that of spacer 2 or 3. Among the 32 *TEOS1*s examined, three were shown to have large insertions in either spacer 2 or 3. Insertions of a 4.2 kb genomic sequence at spacer 2 occurred in both P0454H12 and OSJNBb0052C09 (accession no. AP003255 and AC090441). Besides, we found that the spacer 3 region of *TEOS1* in P0480E02 (accession no. AP002913) consisted of a 13 kb insertion which contained another retrotransposon.

In the BAC/PAC clones containing *TEOS1*-like elements, gene densities were estimated to range from one per 10.4 kb to one per 4.5 kb, mostly higher than one per 6 kb, an average gene density of gene-rich region for the rice genome (Mayer et al., 2001). Thus, the current available data indicated that *TEOS1* was prevalent in the gene-rich region, present close to coding regions that have been annotated. For example, we were able to locate an *TEOS1* element close to a potassium transporter gene (accession no. AF129485) (Rubio et al., 2000). We were able to identify the insertion of an *TEOS1* on the *Pib* gene (Wang et al., 1999, accession no. AB013450) in, Tohoku IL9, a blast resistant strain of rice. This fragment, 9.8 kb in length, contained a *Pib* gene 5' truncated by *TEOS1*.

TEOS1-Like Family in other Plant Genomes

Since the repetitive elements of different genomes might share only the structure similarity but not sequence identity, we scanned the genome sequence data of other plants with the Miropeats program. No similar structure was found in the current publicly available plant genomic sequences for *Arabidopsis*, maize and sorghum. In order to investigate the genomic organization of *TEOS1* within various species, the genomic Southern blot analysis was performed as a supplement to database similarity searches. The following species were included in this study: rice (*Oryza sativa* ssp. *japonica*), maize (*Zea mays*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), *Arabidopsis* (*Arabidopsis thaliana* ecotype Columbia), tobacco (*Nicotiana tabacum*), and soybean (*Glycine max*). Hybridization was performed using the probe specific to the 90-bp repeats within *TEOS1*. The *TEOS1* family,

Table 3. Chromosome distribution of *TEOS1* family.

Chromosome	Number of <i>TEOS1</i>
1	28
2	12
3	8
4	7
5	4
6	18
7	16
8	7
9	0
10	15
11	0
12	2
Total	117

as expected, produced ladder pattern of Southern signal in rice (data not shown). Only light and smearing background, discriminated by high stringency during probe hybridization, were observed in the lanes of wheat, maize and barley DNA, which were indications of similar but different transposable elements. No cross hybridization signal was detected in the dicot plant genomes that we had investigated. The genomic Southern blot analysis further confirmed that the repetitive sequences of *TEOS1* were unique in rice, as we did not find any homolog in other species from current data base search.

Discussion

Several approaches have been used to identify retrotransposons in rice. Experimentally, PCR primer sequences are deduced either from the highly conserved region of known retrotransposons, such as the internal reverse transcriptase domain (Hirochika et al., 1996; Kumekawa et al., 1999), or from the published PBS sites of plant retrotransposons such as complement of the 3' end of initiator tRNA^{Met} (for example, Hirochika et al., 1992). Computationally, FASTA (Pearson and Lipman, 1988) and/or BLAST based database searches (Mao et al., 2000; Turcotte et al., 2001) are used to mine elements with sequence similarity to the published TEs. The computer-based analysis has uncovered mainly short TEs in the rice genome, which are recognized as MITEs (Bureau et al., 1996).

In this study, we adopted a different approach, with which we identified *TEOS1*, a unique transposable element in rice. The approach took advantage of the presence of diagnostic sequences with specific features, i.e., the LTRs, target site direct repeats, PBS and tandem repeat structures, instead of relying on its sequence similarity with the published retrotransposons. In the Southern blot analysis, repeat sequence from *TEOS1* did not recognize well with other monocot genomic DNA and formed smear pattern rather than discrete bands in hybridization analysis. Further studies are required to determine whether other monocot genomes harbor the TEs similar to *TEOS1*'s repeat structures. It is interesting to note the absence of Southern signal in the dicot plants we investigated, as it reveals that the *TEOS1* family actively transposed after the divergence of the monocot and dicot plants.

The Arrays of Tandem Repeats were the Acquisition of TEOS1 Family

The PBS sequence in the *TEOS1* family, as well as that in RIRE2, is complementary to the 3' end of tRNA^{Arg}. The other known rice retrotransposons such as RIRE3, RIRE7 and *Tos17* have the tRNA^{Met} binding site as their PBS. The comparison between *TEOS1* and other rice retrotransposons is summarized in Figure 4. It has been observed that the LTRs in *gypsy* retrotransposon diverged greatly, even if the internal polyprotein regions were homologous (Kumekawa et al., 1999). Surprisingly, we noted that *TEOS1* and RIRE2 have high homology in the characteristic terminal regions of their LTRs. The presence of highly conserved LTR retrotransposon features (e.g. LTRs, PBS, PPT and target site duplication) in *TEOS1* suggests a similar transposition mechanism with that of the regular LTR retrotransposons, e.g. RIRE2. Although the interior region of *TEOS1* has coding capacity, the deduced amino acid sequence is dissimilar to that of any known mobility related proteins. The situation of the *TEOS1* is very similar to the maize *Bs1* retroelements (Jin and Bennetzen, 1989), the *Arabidopsis Athila* (Pelissier et al., 1995) and

Katydid-At1 retroelements (Witte et al., 2001) those lacked reverse transposase sequences and had the ability to transduce a cellular gene. The interior sequence of *TEOS1* was homologous with a rice panicle EST and encoded the non-retrotransposon ORFs containing a spliceable intron. This may indicate that the ORFs as well as the tandem repeat sequences present on *TEOS1* represent an unknown rice cellular gene or sequence that had been acquired by the ancestral *TEOS1*. The maize *Bs1* retroelements are mobile although *Bs1* codes for protein unrelated to its mobility. It was suggested that the mobility of *Bs1* elements have been provided *in trans* by another retroelements (Jin and Bennetzen, 1994). *TEOS1* appears to be similar in this regard. If the protein required for *TEOS1*'s mobility were not encoded by itself, the specialized sequence itself and/or its structure must have been recognized by the mobility related proteins. For *TEOS1*'s dispersion, we suggested that *TEOS1* was trans-activated by other retroelements, such as RIRE2, in the rice genome.

Full-Length TEOS1s and Solo LTRs

It was reported that the downstream region of PBS are poorly conserved among the members of a rice *gypsy* retrotransposon family (Kumekawa et al., 1999). The PBS was the initiation site for cDNA synthesis during duplication, its downstream region was the last region being reverse-transcribed and thus subjected to the accumulation of mutations. The spacer 1 of the rice *TEOS1*s also shows such tendency; only one third of the *TEOS1*s shows significant similarity in the spacer 1 region. It was indicated by the similarity among the sequences that there was relative difference in time of transposition. It was suggested that the sequence comparison of the 5' and 3' LTRs could be used to assess the relative age of TE insertions (Jordan and McDonald, 1999), since the 5' and 3' LTRs were generated from a single template during the reverse transcription process (Arkhipova et al., 1986). We showed that six out of 26 full-length *TEOS1* from nr database had 5' and 3' LTR with 100% identical nucleotide sequences.

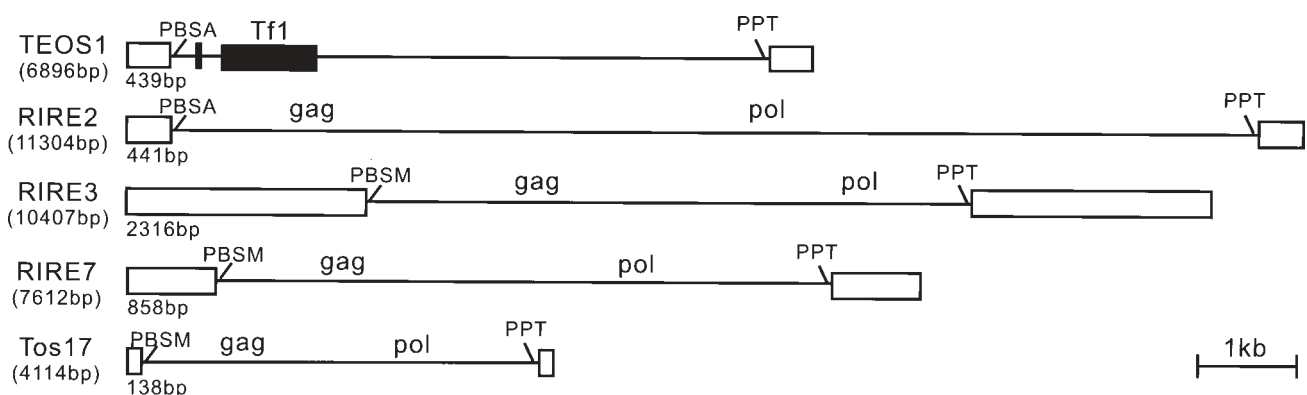


Figure 4. Organization of various rice retrotransposons. The full length of the elements and the length of LTR are indicated at left. The LTRs are indicated by open boxes, the exons of *Tf1* gene are indicated by solid boxes. PBSA and PBSM represent the primer-binding sites which are complementary to the 3' end of methionyl tRNA and arginyl tRNA, respectively. PPT indicates the polypurine tract. Genes encoding the gag protein and polypeptide are represented by gag and pol, respectively.

This is indicative that some *TEOS1*s had activated lately relative to the age of the *TEOS1* family.

The solo LTRs were suggested to be generated by the homologous recombination between two LTRs of a retrotransposon with a subsequent internal excision (Chaleff and Fink, 1980; Tschumper and Carbon, 1986). For the grass family, the conversion to solo LTRs occurred more rapidly than integration in barley (Shirasu et al., 2000). Very few solo LTRs have been found in the maize genome. It is possible that the scarcity is caused by lower recombination efficiency between the comparatively short LTRs of maize retroelements, e.g. maize *Ji* 1.2 kb; barley *BARE-1* 1.8kb; barley *sukkula* 4.9 kb (SanMiguel et al., 1996; Shirasu et al., 2000). However, the 439-bp LTRs of *TEOS1* in rice are shorter than those of the TEs from either maize or barley. Yet, several solo LTRs of *TEOS1* have been observed in the rice genome. Jordan and McDonald (1999) suggested that during the generation of solo LTRs, the excision of the internal sequence is an indication of a genomic turnover of the retrotransposon, and this process offered a means to repress the increase in genome size resulting from successive integrations of large retrotransposons (Shirasu et al., 2000). In both P0025A05 and OSJNBa0065H03 (accession no. AP003504 and AC037197), the insertion of a solo LTR on the opposite strand of the *TEOS1* was observed. Our results indicate that this mechanism of repressing the evolution of expanding genome size acted not only in grass plants with a large genome such as barley, but also in those with small genome such as the rice.

TEOS1s Might Act as a Hot Spot for *Tos17* Insertion

It has been shown that TEs prefer to insert into themselves or other existing TEs forming nested structures, such as in maize and rice genomes (SanMiguel et al., 1996; Suoniemi et al., 1997; Kumekawa et al., 1999). Such a nested TE insertion structure was not observed in the *TEOS1*s examined. However, our survey of Genomic Survey Sequence database (GSS) using Cluster FLAG revealed that many 3' flanking sequence of *Tos17* insertions had significant homology to genomic sequence locating at the 2-7 kb upstream from the 3' LTR of *TEOS1* at P0001A07, as showing in Figure 5A where red bars indicate the insertion sites of each independent *Tos17* mutant and blue bars indicate end sequences of rice BAC libraries (Mao et al., 2000). This region is corresponding to the RNA polymerase II gene. Southern analysis using RNA polymerase II genetic sequences as the probe indicated that this should be encoded by a single-copy gene (data not shown). As a result, the numerous GSS hits indicated that multiple *Tos17* insertions occurred in the vicinity of *TEOS1* in this specific case. Also shown in Figure 5B is the GSS hits in a nearby PAC sequence, P0431G05 (accession no. AC087551), of chromosome 5. There was no *TEOS1* in this PAC, and the GSS hits, mostly the BAC end sequences, were located randomly throughout the PAC sequence. It was reported that *Tos17* preferentially

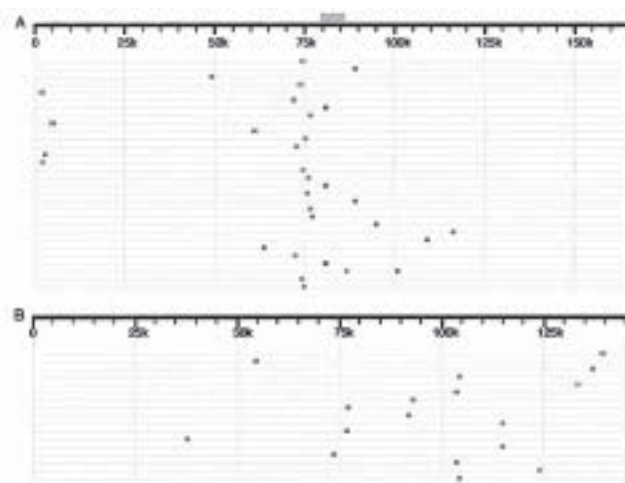


Figure 5. Scatter of matches for a FLAG search against the GSS database. Panel A. P0001A07 as a query, with *TEOS1* at 79 kb-87 kb. Panel B. P0431G05 (accession no. AC087551) as a query, no *TEOS1* was present in this clone. The sequence scale is shown at the top of the figure. Green solid box indicates the *TEOS1*. Red bars and blue bars represent the GSS matches corresponding to the 3' flanking sequences of *Tos17* insertion and the BAC-end sequences of CUGI Rice BAC library, respectively. The figures show the matches with E-value of less than e^{-232} .

integrated into gene-rich, low-copy-number regions of the rice genome (Yamazaki et al., 2001). The observed insertion hotspot might be resulted from the low-copy nature and the active expression of RNA polymerase at callus tissue when *Tos17* started jumping (Hirochika et al., 1996) of this region. Besides, since the *TEOS1* sequence has the potential to form stable secondary structure during DNA replication or transcription, as *TEOS1*s inserted themselves into the frequently transcribed region, they could serve as a hotspot for the insertion of other retrotransposons. Further studies on this mechanism would lead to a better understanding for the dynamic interactions of TE-TE and TE-gene in the rice genome.

Acknowledgments. The authors are grateful to RGP, Japan for providing the rice PAC clones and to Pien-Chien Huang, Department of Biochemistry and Molecular Biology, Johns Hopkins University for critical reading of the manuscript and editorial assistance. This research is supported by grants from the National Science Council, Council of Agriculture and Academia Sinica, Taiwan. This work is the effort of all members in the Academia Sinica Plant Genome Center.

Note added in proof. While our manuscript was under review, an LTR element comparable to *TEOS1* in this article was reported by Jiang, N., Z. Bao, S. Temnykh, Z. Cheng, J. Jiang, R. A. Wing, S.R. McCouch, and S.R. Wessler. (2002) Dasheng: A recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* 161: 1293-1305.

Literature Cited

- Agrawal, G.K., M. Yamazaki, M. Kobayashi, R. Hirochika, A. Miyao, and H. Hirochika. 2001. Screening of the rice viviparous mutants generated by endogenous retrotransposon Tos17 insertion. Tagging of a zeaxanthin epoxidase gene and a novel ostate gene. *Plant Physiol.* **125**: 1248-1257.
- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Apweiler, R., T.K. Attwood, A. Bairoch, A. Bateman, E. Birney, M. Biswas, P. Bucher, L. Cerutti, F. Corpet, M. D. Croning, R. Durbin, L. Falquet, W. Fleischmann, J. Gouzy, H. Hermjakob, N. Hulo, I. Jonassen, D. Kahn, A. Kanapin, Y. Karavidopoulou, R. Lopez, B. Marx, N.J. Mulder, T.M. Oinn, M. Pagni, and F. Servant. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**: 37-40.
- Arkhipova, I.R., A.M. Mazo, V.A. Cherkasova, T.V. Gorelova, N.G. Schuppe, and Y.V. Llyin. 1986. The steps of reverse transcription of *Drosophila* mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. *Cell* **44**: 555-563.
- Arumuganathan, K. and E.D. Earle. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**: 208-219.
- Bureau, T.E., P.C. Ronald, and S.R. Wessler. 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl. Acad. Sci. USA* **93**: 8524-8529.
- Bureau, T.E. and S.R. Wessler. 1994. Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses. *Proc. Natl. Acad. Sci. USA* **91**: 1411-1415.
- Burge, C. and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**: 78-94.
- Chaleff, D.T. and G.R. Fink. 1980. Genetic events associated with an insertion mutation in yeast. *Cell* **21**: 227-237.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387-395.
- Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.
- Ewing, B. and P. Green. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**: 186-194.
- Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175-185.
- Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**: 195-202.
- Gordon, D., C. Desmarais, and P. Green. 2001. Automated Finishing with Autofinish. *Genome Res.* **11**: 614-625.
- Grandbastien, M.A. 1992. Retroelements in higher plants. *Trends Genet.* **8**: 103-108.
- Grandbastien, M.A. 1998. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3**: 131-137.
- Grandbastien, M.A., A. Spielmann, and M. Caboche. 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* **337**: 376-380.
- Hirochika, H. 1993. Activation of tobacco retrotransposons during tissue culture. *EMBO. J.* **12**: 2521-2528.
- Hirochika, H. 2001. Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**: 118-122.
- Hirochika, H., A. Fukuchi, and F. Kikuchi. 1992. Retrotransposon families in rice. *Mol. Gen. Genet.* **233**: 209-216.
- Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa, and M. Kanda. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA* **93**: 7783-7788.
- Jin, Y.-K. and J.L. Bennetzen. 1989. Structure and coding properties of Bs1, a maize retrovirus-like transposon. *Proc. Natl. Acad. Sci. USA* **86**: 6235-6239.
- Jin, Y.-K. and J.L. Bennetzen. 1994. Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bs1 retroelement of maize. *Plant Cell.* **6**: 1177-1186.
- Jordan, I.K. and J.F. McDonald. 1999. Comparative genomics and evolutionary dynamics of *Saccharomyces cerevisiae* Ty elements. *Genetica* **107**: 3-13.
- Kossack, D.S. and C.S. Kinlaw. 1999. IFG, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. *Plant Mol. Biol.* **39**: 417-426.
- Kumar, A. 1996. The adventures of the Ty1-copia group of retrotransposons in plants. *Trends Genet.* **12**: 41-43.
- Kumekawa, N., N. Ohmido, K. Fukui, E. Ohtsubo, and H. Ohtsubo. 2001. A new gypsy-type retrotransposon, RIRE7: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. *Mol. Genet. Genomics* **265**: 480-488.
- Kumekawa, N., H. Ohtsubo, T. Horiuchi, and E. Ohtsubo. 1999. Identification and characterisation of novel retrotransposons of the gypsy type in rice. *Mol. Gen. Genet.* **260**: 593-602.
- Laten, H.M., A. Majumdar, and E.A. Gaucher. 1998. SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc. Natl. Acad. Sci. USA* **95**: 6897-6902.
- Mao, L., T.C. Wood, Y. Yu, M.A. Budiman, J. Tomkins, S. Woo, M. Sasinowski, G. Presting, D. Frisch, S. Goff, R.A. Dean, and R.A. Wing. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Res.* **10**: 982-990.
- Mayer, K., G. Murphy, R. Tarchini, R. Wambutt, G. Volckaert, T. Pohl, A. Dusterhoft, W. Stiekema, K.D. Entian, N. Terryn, K. Lemcke, D. Haase, C.R. Hall, A.M. van Dodeweerd, S.V. Tingey, H.W. Mewes, M.W. Bevan, and I. Bancroft. 2001. Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res.* **11**: 1167-1174.
- Mochizuki, K., M. Umeda, H. Ohtsubo, and E. Ohtsubo. 1992. Characterization of a plant SINE, p-SINE1, in rice genomes. *Jpn. J. Genet.* **67**: 155-166.
- Motohashi, R., E. Ohtsubo, and H. Ohtsubo. 1996. Identification of Tnr3, a suppressor-mutator/enhancer-like transposable element from rice. *Mol. Gen. Genet.* **250**: 148-152.
- Nakai, K. and P. Horton. 1999. PSORT: a program for detecting

- ing the sorting signals of proteins and predicting their sub-cellular localization. *Trends Biochem. Sci.* **24**: 34-35.
- Ohtsubo, H., N. Kumekawa, and E. Ohtsubo. 1999. RIRE2, a novel gypsy-type retrotransposon from rice. *Genes Genet. Syst.* **74**: 83-91.
- Parsons, J.D. 1995. Miropeats: graphical DNA sequence comparisons. *Comput. Applic. Biosci.* **11**: 615-619.
- Pearce, S.R., G. Harrison, D. Li, J.S. Heslop-Harrison, A. Kumar, and A.J. Flavell. 1996. The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. *Mol. Gen. Genet.* **250**: 305-315.
- Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**: 2444-2448.
- Pelissier, T., S. Tutois, J.M. Deragon, S. Tourmente, S. Genestier, and G. Picard. 1995. Athila, a new retroelement from *Arabidopsis thaliana*. *Plant Mol. Biol.* **29**: 441-452.
- Rubio, F., G.E. Santa-Maria, and A. Rodriguez-Navarro. 2000. Cloning of *Arabidopsis* and barley cDNAs encoding HAK potassium transporters in root and shoot cells. *Physiol. Plant.* **109**: 34-43.
- Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.-A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualisation and annotation. *Bioinformatics* **16**: 944-945.
- SanMiguel, P. and J.L. Bennetzen. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **81**: 37-44.
- SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765-768.
- Sato, Y., N. Sentoku, Y. Miura, H. Hirochika, H. Kitano, and M. Matsuoka. 1999. Loss-of-function mutations in the rice homeobox gene OSH15 affect the architecture of internodes resulting in dwarf plants. *EMBO J.* **18**: 992-1002.
- Shirasu, K., A.H. Schulman, T. Lahaye, and P. Schulze-lefert. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* **10**: 908-915.
- Sonnhammer, E.L.L. and R. Durbin. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1-10.
- Suoniemi, A., K. Ananthawat-Jonsson, T. Arna, and A.H. Schulman. 1996. Retrotransposon BARE-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. *Plant Mol. Biol.* **30**: 1321-1329.
- Suoniemi, A., D. Schmidt, and A.H. Schulman. 1997. BARE-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. *Genetica* **100**: 219-230.
- Tatusova, T.A. and T.L. Madden. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**: 247-250.
- Tschumper, G. and J. Carbon. 1986. High frequency excision of Ty elements during transformation of yeast. *Nucleic Acids Res.* **14**: 2989-3001.
- Turcotte, K., S. Srinivasan, and T. Bureau. 2001. Survey of transposable elements from rice genomic sequences. *Plant J.* **25**: 169-179.
- Wang, Z.X., M. Yano, U. Yamanouchi, M. Iwamoto, L. Monna, H. Hayasaka, Y. Katayose, and T. Sasaki. 1999. The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J.* **19**: 55-64.
- Witte, C.-P., Q.H. Le, T. Bureau, and A. Kumar. 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc. Natl. Acad. Sci. USA* **98**: 13778-13783.
- Wright, D.A., N. Ke, J. Smalle, B.M. Hauge, H.M. Goodman, and D.F. Voytas. 1996. Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics* **142**: 569-578.
- Yamazaki, M., H. Tsugawa, A. Miyao, M. Yano, J. Wu, S. Yamamoto, T. Matsumoto, T. Sasaki, and H. Hirochika. 2001. The rice retrotransposon Tos17 prefers low-copy-number sequences as integration targets. *Mol. Genet. Genomics* **265**: 336-344.

TEOS1，一個新的跳躍子家族

趙雅婷 蘇春霖 周德源 李佩芳 鐘君怡 黃建彰 劉淑美 邢禹依

中央研究院植物研究所

我們以電腦分析水稻第五條染色體的一個 PAC 殖系—P0001A07 的基因組序列，發現了一個新的跳躍子，並命名為 *TEOS1*。*TEOS1* 具有 LTR 反轉錄跳躍子的一般特徵，如 LTR、PBS 和 PPT，以及 5 個鹼基的 TSD 等等，而 PBS 和 tRNAArg 的 3' 端序列互補。它與其他已知的 LTR 反轉錄跳躍子最大的區別在於 *TEOS1* 的內部區域包含了兩組並列反覆序列。經由我們的解讀，*TEOS1* 的內部區域包含了一個 *Tt* 基因，其序列並沒有和已知的跳躍蛋白質有顯著的相似。經由資料庫搜尋我們推斷水稻基因組中至少有 300 個 *TEOS1*，另外還有以 solo LTR 型態出現的 *TEOS1*。根據我們的電腦分析，*TEOS1* 內部所包含的反覆序列很可能成為其他跳躍子插入水稻基因組的熱點。

關鍵詞：水稻；並列重覆；*TEOS1*；跳躍子。