Mol Gen Genet (1999) 260: 593–602

ORIGINAL PAPER

N. Kumekawa · H. Ohtsubo · T. Horiuchi · E. Ohtsubo

Identification and characterization of novel retrotransposons of the *gypsy* type in rice

Received: 25 May 1998 / Accepted: 13 October 1998

Abstract We found that two DNA fragments, which were obtained from *Oryza sativa* L cv. IR36 by PCR using degenerate primers designed for amplification of a rice gene, showed homology with the *rt* gene encoding reverse transcriptase of the *Drosophila* retrotransposon *gypsy*. We named the element from which they originated *RIRE3* (for rice retrotransposon No. 3) and analyzed it further by isolating various clones containing segments of *RIRE3*. Nucleotide sequencing of the clones revealed that *RIRE3* has LTRs (2316 bp) and that the internal sequence (5775 bp) includes a large ORF with *gag* and *pol* regions; the *pol* region includes the *rt* gene as well as the *int* gene encoding integrase in this order, as in *gypsy*. Interestingly, the region upstream of *gag* in *RIRE3* contained another open reading frame, here called *orf0*, which does not exist in *gypsy* or in other retrotransposons related to it. In the course of characterizing *RIRE3*, we obtained a further clone, which showed less homology with the *pol* region of *RIRE3*. This clone was found to be derived from another *gypsy*-type retrotransposon (named *RIRE8*) containing the LTR sequence and *orf0* both of which were only weakly homologous to that in *RIRE3*. Further characterization of *RIRE8* revealed that there were actually two subtypes of *RIRE8* (named *RIRE8A* and *RIRE8B*), which show little homology to each other in the *orf0* region. Although the LTRs of *RIRE3* and *RIRE8* elements show very weak homology with each other, there exists a conserved sequence at their termini. We therefore carried out PCR using primers which hybridize to the *rt* gene of *RIRE3*, and total genomic DNA from various monocot and dicot plants as templates, and found that a family of *RIRE3* elements was present in all plants tested.

Key words Retrotransposon · *gypsy* · *Oryza sativa* · LTR · *orf0*

Communicated by H. Saedler

N. Kumekawa · H. Ohtsubo · E. Ohtsubo (✉)
Institute of Molecular and Cellular Biosciences
The University of Tokyo
Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-0032, Japan
e-mail: eohtsubo@ims.u-tokyo.ac.jp
Fax: +81-3-56843269

T. Horiuchi
Department of Engineering, Soka University
Tangi 1-236, Hachiohji, Tokyo 192-8577, Japan

Introduction

Genome sizes differ greatly in plants, ranging from that of *Arabidopsis thaliana* – about 145 Mb – to that of lily – about 10^5 Mb (Arumuganathan and Earle 1991). It is believed that the number of structural genes does not differ much, and thus that the variation in genome size depends on differences in the amounts of repeated sequences (Graham 1995; Bennetzen 1996). Earlier studies on animal systems revealed that the repeated sequences include retrotransposons with long terminal repeats (LTRs) like those in proviruses, the integrated forms of retroviruses, which make copies of themselves via mRNA intermediates (for reviews, see Bingham and Zachar 1989; Varmus and Brown 1989). LTRs contain a promoter and a terminator for transcription of a retrotransposon. The two LTRs flank the internal region which contains *gag* and *pol* regions: *gag* codes for Gag protein which forms a virus-like particle by encapsulating the mRNA of the retrotransposon (Hansen et al. 1992); the *pol* region includes the genes *rt*, *rh* and *int*, encoding reverse transcriptase, RNase H, and integrase, respectively, which are required for cDNA synthesis and integration of the cDNA into host chromosomes (for reviews, see Bingham and Zachar 1989; Varmus and Brown 1989). The proteins encoded by the *pol* region are synthesized as a polyprotein, which is cleaved by the action of a protease encoded by the *pro* gene, which is also found in the *pol* region. In the regions immediately adjacent to the LTRs, there exist PBS (primer-binding site) and PPT (polypurine tract) sequences, which are cis elements that are essential for cDNA synthesis.

Retrotransposons have been classified into two types, based on their homology with either *copia* (Mount and Rubin 1985) or *gypsy* (Marlor et al. 1986) of *Drosophila melanogaster*. The two types of retrotransposons do not show extensive homology and differ in the *pol* region, in which the order of *rt* and *int* genes is reversed. *gypsy*-type retrotransposons are similar to retroviruses in structure and sequence, and are thus assumed to be ancestors of retroviruses (Xiong and Eickbush 1990).

Two types of retrotransposons also exist in plants (Bennetzen 1996; SanMiguel et al. 1996; Kunze et al. 1997). During a search for cytochrome P450 genes in rice (*Oryza sativa* L cv. IR36) by PCR using degenerate primers, we identified two PCR-amplified fragments that showed homology with the *rt* gene of *gypsy*. In this paper, we report that these fragments are portions of a *gypsy*-type retrotransposon in rice, named *RIRE3*, with LTRs of an abnormally large size, and that *RIRE3* is ubiquitously present as a gene family in angiosperms. We also demonstrate that there exist other retrotransposons, named *RIRE8A* and *RIRE8B*, which are related to *RIRE3*, with homology in the *pol* region but with poor homology in the LTRs. These retrotransposons are rather closely related to *del1* of lily (Smyth et al. 1989) but not to other plant retrotransposons including *Grande1* of teosinte (Martínez-Izquierdo et al. 1997). *RIRE* elements have an ORF in the region preceding the *gag*-*pol* segment, which is not present in any other *gypsy*-type retrotransposons previously identified. Sequences which are specific for *RIRE* elements are found in the terminal regions of their LTRs. It should thus be possible to identify *gypsy*-type retrotransposons related to the *RIRE* elements by examining their terminal sequences. We also report that *RIRE* elements often insert into one another to form a nested structure; indeed, one copy of *RIRE3* was found to harbor a new retrotransposon, *RIRE7*.

Materials and methods

Plants and preparation of genomic DNA

Plants used are listed in Table 1. Genomic DNA was extracted from a plant (2–5 g) in a buffer containing 100 mM TRIS-HCl, 50 mM EDTA, 500 mM NaCl pH 8.0 as described by Ohtsubo et al. (1991). Some DNA preparations were further purified by CsCl centrifugation according to Lichtenstein and Draper (1985), and the DNA solution was dialyzed against TE buffer (10 mM TRIS-HCl, 1 mM EDTA pH 8.0). After ethanol precipitation, DNA was dissolved in 200–400 μl of TE buffer to a final concentration of 50 μg/ml. The optical density of each DNA solution was measured using a GeneQuant II spectrophotometer (Pharmacia Biotech) at 260 nm, and the DNA concentration was calculated by assuming that 1 OD is equivalent to 50 μg of DNA/ml.

Reagents and enzymes

Reagents used were obtained from Wako Junyaku or Bio Rad. Restriction endonucleases used were *Eco*RI (Takara), *Hin*dIII and *Bam*HI (New England Biolabs). The various kinds of DNA polymerase used are described below.

Synthetic oligonucleotides

Oligonucleotides used as probes or primers for PCR are listed in Table 2. These oligonucleotides were synthesized by the *β*-cyanoethylphosphoamidide method using an OLIGO1000 M DNA synthesizer (Beckman).

PCR

PCR was carried out using 0.2 μg of total rice DNA as the template, 1 μM of each primer, and 2.5 units of DNA polymerase in a Perkin Elmer-Cetus Thermal Cycler. When we used rTaq DNA polymerase (Takara), 30 cycles of amplification were carried out under the following conditions; denaturation for 1 min at 94° C, annealing for 1 min at 50 or 45° C, and DNA synthesis for 2 min at 72° C. The frequency of mutations induced by PCR under the above condition is 1 per 642 bp (Tenzen et al. 1994). When we used LA-Taq DNA polymerase (Takara), 30 cycles of amplification were carried out under the following conditions; denaturation for 30 s at 96° C, annealing for 1 min at 50° C, and DNA synthesis for 5 min at 72° C.

To obtain fragments of *RIRE3*-related retrotransposons from various plant species, we used 2.5 units of AmpliTaq Gold DNA polymerase (Perkin Elmer Cetus) and carried out PCR under the conditions recommended by the supplier.

Cloning of PCR-amplified fragments

The PCR-amplified fragments were treated with Taq polymerase (Takara) according to Clark (1988), and were ligated to the linearized pCR2.1 vector using a TA cloning kit (Invitrogen) as recommended by the supplier. The sample DNA (5 ng) was transformed into *E. coli* XL1-blue MRF′ (Stratagene).

Table 1 Plants used

Plant	Strain	Source/reference
Rice	*Oryza sativa* L. cv. IR36	Motohashi et al. (1997)
Rice	*Oryza sativa* L. cv. Nipponbare	Motohashi et al. (1997)
Barley	*Hordeum vulgare* Translocation TS$1(T1-2)	NIAR[a]
Wheat	*Triticum aestivum* 11C LA 6345	NIAR[a]
Tobacco	*Nicotiana tabacum* SR1	NIAR[a]
Maize	*Zea mays* Dt.	NIAR[a]
Arabidopsis	*Arabidopsis thaliana* Landsberg	K. Okada (NIBB)[a]
Morning glory	*Pharbitis nil* (ray white flowers)	A commercial line
Mulberry	*Morus bombycis* Koidz	K. Oshigane (Soka University)

[a] NIAR, National Institute of Agrobiological Resources; NIBB, National Institute for Basic Biology

Table 2 Synthetic oligonucleotides used

Primer[a]	Sequence (5′→3′)[b]	Position[c]	Coordinates[d]
Degenerate primers			
CR1-1	GTKTTYGGTAARGGWGTT		
CR4-2	TGAMAGTARGCAAARTKTTC		
Primers or probes used for cloning *RIRE3*			
R3down2 (1)	TTCTATTTACGCTTCCGCTTG	LTR	2193′-2213′
R3gag3 (2)	GTTAGCTCTTTCTGTTTGCTGG	*gag*	1202-1223
R3gag8 (3)	TTGTCACCATCTGAAAACCCCA	*gag*	1230-1209
R3n1 (4)	ATCCGGATTTCCAACGTCTG	*gag*	1937-1956
R3gag7 (5)	CAGATTCAGCAAGGGACTCGT	*gag*	2176-2196
R3gag 9 (6)	GGAGATATTCCGAAGACCGCA	*pro*	3234-3254
R3RT3 (7)	GATTCATTGGTGAGGCCGAA	*rt*	3338–3319
R3-5′Pri (8)	TGCTGATGGTCTTCCTCT	*rt*	3430-3413
R35′Pro	CTTGGATAAGTTTGTTGTGG	*rt*	3359-3378
R3-3′Pri (9)	GAATCCACTATCATCCTG	*rh*	4105-4122
R3-3′Pro	GACTTAGAGCGTCGGCG	*rh*	4155-4139
R3rh1 (10)	GGCTTTTTCGACTTAGAGCGT	*rh*	4164-4144
3′npr (11)	AGACAGATGGACAGACCG	*int*	4967-4989
3′nIIpr (12)	TGTGTTCCATGTGTCGCAA	*int*	5478-5496
R3LTR1 (13)	CGCGTCCCAAATCGACTCTA	LTR	4′-23′
R3LTR2 (14)	ATCCTCGAAATTGTGCAAAG	LTR	30′-49′
RTLTR3	TCAATGTAAAGCCTCCAAC	LTR	62′-81′
Primers/probes used for cloning *RIRE8*			
R8down1 (1′)	ATTGTGTGCCTGGGCTGATC	LTR	2865′-2884′
pro5′-1 (2′)	GAAAGGATACATTAGACCCAGT	*pro*	3307-3328
pro5′-2 (3′)	CTACCGAGAATTGATGACCTG	*pro*	3440-3460
3′npr (4′)	AGACAGATGGACAGACCG	*int*	5251-5268
Int3′-1 (5′)	CCACCGAAGTCTAAAGCACAA	*int*	5326-5306
Int3′-2 (6′)	TCGTAAGGAGCCATCTGAAG	*int*	5407-5388
3′nIIpr (7′)	TGTGTTCCATGTGTCGCAA	*int*	5762-5780
R8downII	GACGAGGGTTTACACACATG	LTR	2889′-2908′

[a] The oligonucleotides CR1-1 and CR4-2 are homologous to the conserved regions of P450 genes including the heme-binding domain. Primers used for nucleotide sequencing are not listed in this table. Numbers in *parentheses* are alternative designations of primers used for PCR. The approximate positions of these primers are shown in Fig. 2

[b] In the degenerate primer sequences, Y stands for C or T, R for A or G, K for G or T, W for A or T, and M for C or A

[c] LTR, long terminal repeat; *pro*, protease; *rt*, reverse transcriptase; *rh*, RNase H; *int*, integrase

[d] Coordinates of the sequences of *RIRE3* and *RIRE8A*. Numbers marked with ′ refer to LTR sequences, others to the internal region

Nucleotide sequencing

DNA sequencing was also carried out by the dideoxynucleotide chain-termination method (Sanger et al. 1977; Messing 1983) using a Model 373S DNA sequencer (Applied Biosystems) as follows. Sequencing reactions were carried out using dye-labeled primers (M13-21, RV) with an ABI PRISM Dye Primer Cycle sequencing kit with AmpliTaq DNA polymerase FS (Perkin Elmer Cetus); alternatively, sequencing reactions were carried out using synthetic oligonucleotide primers and an ABI PRISM Dye Terminator Cycle sequencing kit with AmpliTaq DNA polymerase FS. The samples were electrophoresed in a 4.25% Longranger gel (Takara).

Computer analysis of nucleotide sequences

Alignments of sequence data were carried out using the programs HarrPlot v. 1.2.2 and GENETYX-MAC v. 7.3 (Software Development). Searches for homology of nucleotide and amino acid sequences were performed by running the programs FASTA, BLAST and MP search against the sequences in the GenBank and Swiss-Prot databases. Multiple sequences were aligned using the CLUSTAL W program (version 1.7), and a dendrogram of the sequences was constructed by using the categories model in Protdist, followed by the program Neighbor in the PHYLIP version 3.572 software (Thompson et al. 1994; Felsenstein 1995).

Southern analysis

DNA fragments were electrophoresed in a 0.7 or 1.8% agarose gel and stained with ethidium bromide. DNA was then transferred to a nylon membrane (Hybond-N$^+$; Amersham) by alkaline blotting under the conditions recommended by the membrane supplier. The filter was preincubated at 65° C for 1 h in 5 ml/100 cm^2 of a hybridization solution containing 6× SSC (0.9 M NaCl, 0.09 M trisodium citrate), 5× Denhardt's solution [0.1% Ficoll (Pharmacia), 0.1% polyvinylpyrrolidon (Nakarai), 0.1% (w/v) bovine serum albumin (Seikagaku Kogyo)], 0.2% (w/v) sodium dodecyl sulfate (SDS), and 100 μg of sonicated salmon sperm DNA (Sigma)/ml. Then the solution was replaced by the same hybridization solution containing the ^{32}P-labeled oligonucleotide (or cloned DNA) at 10^5 cpm/ml, and hybridization was carried out for 12–15 h at 65° C. The filter was washed sequentially in 2× SSC, 1× SSC and 0.1× SSC, each containing 0.1% SDS, at a temperature calculated for each of the probe sequences in Table 2 as described by Wahl et al. (1987). DNA fragments that hybridized with the ^{32}P-labeled probe were visualized using a Bioimaging Analyzer BAS1000 or BAS1500 (Fuji).

Oligonucleotide probes were labeled at the 5′ ends with [γ-^{32}P]ATP (Amersham, 185 TBq/nmol) using T4 polynucleotide kinase (Takara). The cloned DNA used as the probe was labeled with [α-^{32}P]dCTP (Amersham, 185 TBq/nmol) using a Random Primer DNA labeling kit Ver. 2 (Takara).

Accession numbers

The nucleotide sequence data will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases under the accession numbers AB014738-AB014756.

Results

Identification and characterization of a *gypsy*-type retrotransposon in rice, *RIRE3*

We carried out PCR using degenerate primers in order to amplify fragments of the p450 genes in *O. sativa* L cv. IR36 and found that two clones, 3-31 and 3-43, contained a sequence that was homologous to the *rt* gene of *gypsy* and the related element *Ty3* of yeast (Hansen et al. 1988) (Fig. 1A). We then carried out a Southern hybridization analysis using clone 3-31 as a probe against restriction digests of total DNA from *O. sativa* L cv. Nipponbare and *O. sativa* L cv. IR36. The probe hybridized to many *Hin*dIII or *Bam*HI fragments from these rice strains (data not shown). These results indicated that there exist in the rice genome repeated sequences which appear to represent *gypsy*-type retrotransposons, which have not previously been found in this plant. We named this element *RIRE3* (rice retroelement No. 3).

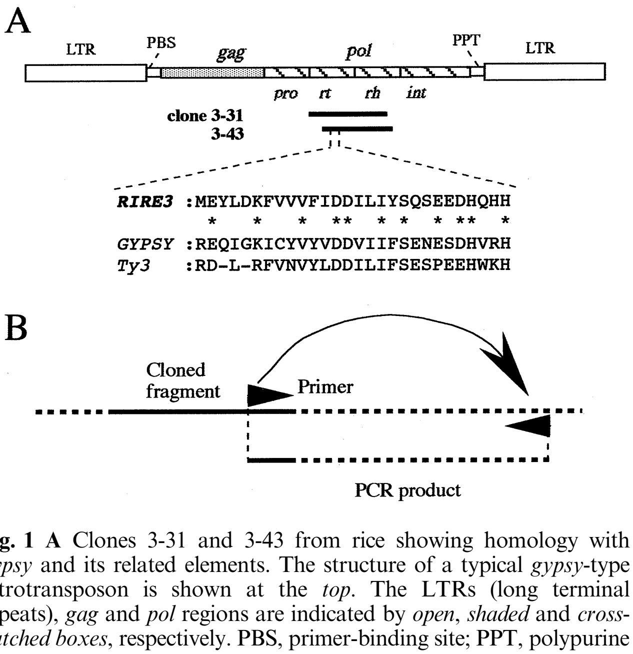

A
LTR
PBS
gag
pol
PPT
LTR
pro
rt
rh
int
clone 3-31
3-43
RIRE3 :MEYLDKFVVVFIDDILIYSQSEEDHQHH
* * * ** * * * ** *
GYPSY :REQIGKICYVYVDDVIIFSENESDHVRH
Ty3 :RD-L-RFVNVYLDDILIFSESPEEHWKH
B
Cloned fragment
Primer
PCR product

Fig. 1 A Clones 3-31 and 3-43 from rice showing homology with *gypsy* and its related elements. The structure of a typical *gypsy*-type retrotransposon is shown at the *top*. The LTRs (long terminal repeats), *gag* and *pol* regions are indicated by *open*, *shaded* and *cross-hatched boxes*, respectively. PBS, primer-binding site; PPT, polypurine tract; *pro*, protease ORF; *rt*, reverse transcriptase ORF; *rh*, RNase H ORF; *int*, integrase ORF. The positions of clones 3-31 and 3-43 are shown by *horizontal bars*. Relevant amino acid sequences encoded by *rt*, genes of retrotransposons – *gypsy* of *D. melanogaster* (Marlor et al. 1986) and *Ty3* of *Saccharomyces cerevisiae* (Hansen et al. 1988) – are shown together with that of *RIRE3* at the *bottom*. *Asterisks* indicate homologous amino acids encoded by *RIRE3* and *gypsy*. **B** One-primer PCR method. The primer used can hybridize not only with the legitimate target sequence but also with an illegitimate target by mis-annealing, to give rise to a PCR-amplified fragment containing the region near the legitimate sequence

To determine the entire nucleotide sequence of *RIRE3*, we obtained clones containing sequences adjacent to each end of the fragment in clone 3-31 by the one-primer PCR method. This method is based on the assumption that the sequence adjacent to one end of the fragment of *RIRE3* can be amplified by PCR if a primer which hybridizes with an *RIRE3* sequence also happens to hybridize nonspecifically with a sequence flanking the *RIRE3* fragment (Fig. 1B). We found that each of the primers used (Table 2) gave rise to the amplified fragments that hybridized with an *RIRE3* probe. Cloning and nucleotide sequencing of these fragments revealed that they contained an additional segment of the *pol* region not present in the original *RIRE3* fragment, 3-31 (fragment I, Fig. 2A). We carried out either one-primer PCR using each of the primers that hybridize with the newly obtained fragments or conventional PCR using two relevant primers to amplify appropriate regions and were finally able to obtain fragments that

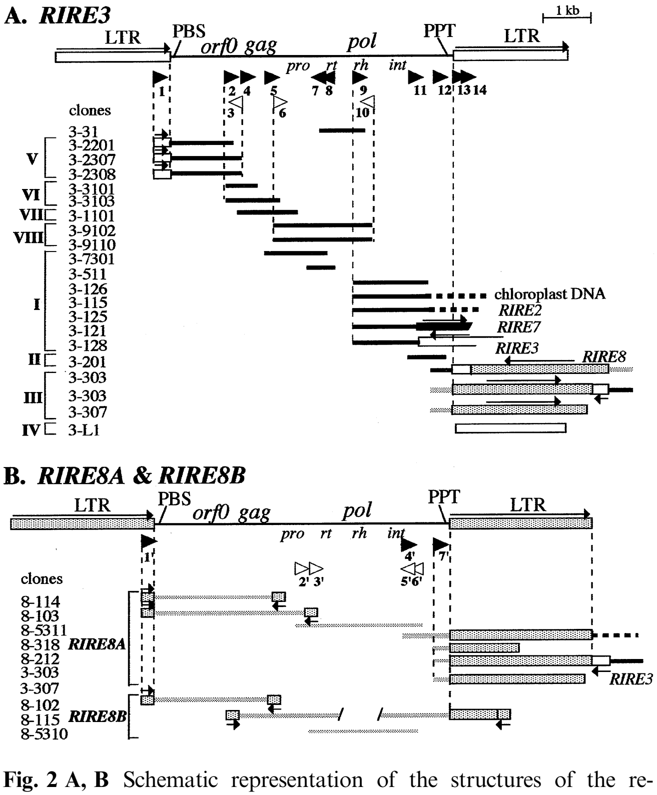


Fig. 2 A, B Schematic representation of the structures of the retrotransposons *RIRE3* and *RIRE8* and their deletion derivatives. The structures of *RIRE3* and *RIRE8* are shown at top of panels **A** and **B**, respectively. *Filled and open arrowheads* beneath the diagrams indicate the positions of primers used for one-primer PCR and for two-primer PCR, respectively (see Table 2 for primers). *Horizontal bars* indicate positions of clones used to determine the sequences of retrotransposons. *Thick dotted lines* adjacent to the horizontal bars indicate non-retrotransposon sequences. LTR sequences of *RIRE3* (*open boxes*), *RIRE8* (*shaded boxes*), and *RIRE7* (*solid boxes*) are shown. *Thin horizontal arrows* indicate the orientations of the LTR sequences. In **A**, the *Roman numerals* indicate groups of the segments recovered in cloning experiments designed to determine the complete structure of *RIRE3*. The DNA segment in clone 3-303 is represented in two different orientations

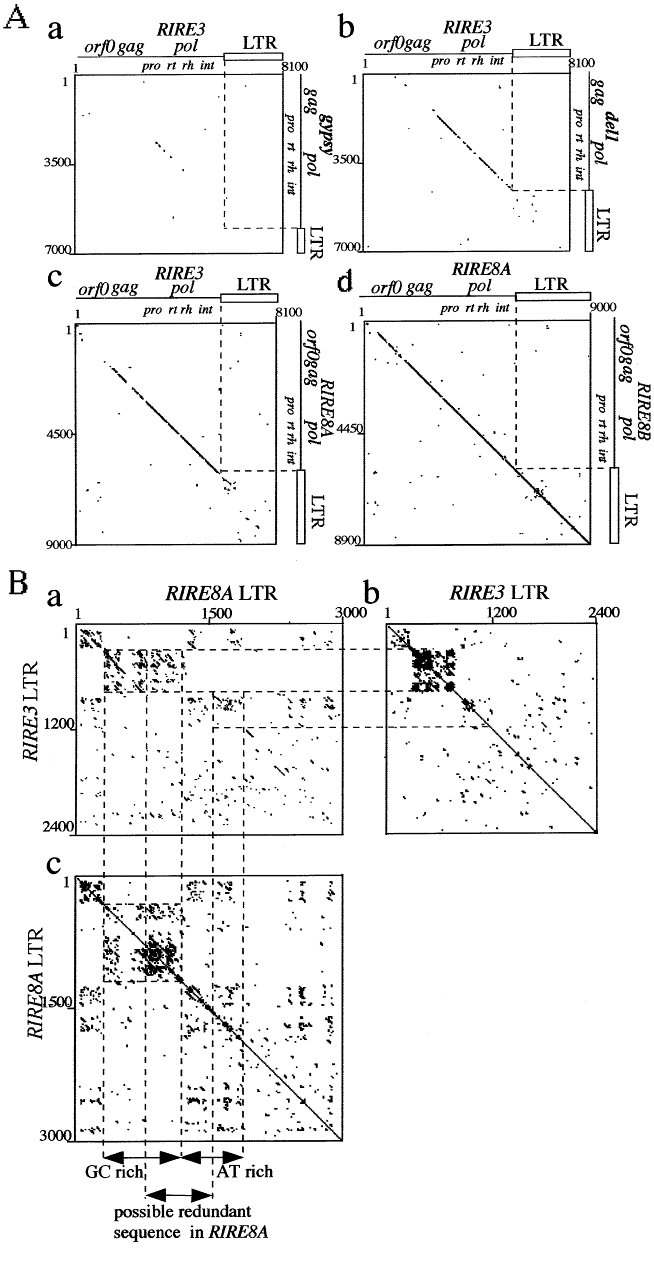


Fig. 3 Dot matrix comparisons of the nucleotide sequences of *gypsy*-type retrotransposons (**A**) and the LTRs of *RIRE8A* and *RIRE3* (**B**). **A** *RIRE3* vs. *gypsy* (**a**), *RIRE3* vs. *del1* (**b**), *RIRE3* vs. *RIRE8A* (**c**) and *RIRE8A* vs. *RIRE8B* (**d**). Structural features of retrotransposons are explained in Fig. 1. *Dots* are placed at locations of nucleotide sequence identity when more than 29 nucleotides out of 50 are identical. The structures of retrotransposons with the LTR at the 5′ end and the internal region are shown at *top* and on the *right* of each plot. **B** *RIRE3* LTR vs. *RIRE8A* LTR (**a**), *RIRE3* LTR vs. *RIRE3* LTR (**b**), and *RIRE8A* LTR vs. *RIRE8A* LTR (**c**). The *broken lines* indicate GC- or AT-rich sequences and a possible redundant sequence in *RIRE8A*. *Dots* are placed at locations of the sequence identity when more than 25 nucleotides out of 50 are identical

covered the entire *RIRE3* sequence (Fig. 2 A, fragments II-VIII). A consensus sequence for *RIRE3* derived by comparing the nucleotide sequences of the fragments showed homology with the *pol* region of *gypsy* (Fig. 3 Aa). The internal region (5775 bp) of *RIRE3* contained an ORF consisting of the genes *gag*, *pro*, *rt*, *rh* and *int*, in that order, as in *gypsy* and *Ty3* (Fig. 4). The amino acid sequence of the reverse transcriptase encoded by *RIRE3* shows significant homology to those of *gypsy* and *Ty3* (Fig. 5B). This confirms that *RIRE3* is a *gypsy*-related retrotransposon. *RIRE3* has significant homology with the *gypsy*-type retrotransposon *del1* of lily in the *pol* region, as shown in Fig. 3Ab (see also Fig. 5B), but did not have such homology with other plant retrotransposons, including *Grande1* of teosinte, either in the *pol* region or in the *gag* region.

Interestingly, the region preceding the *gag* gene in the internal region is large and contains an additional ORF, named *orf0*, and encoding a 253 amino-acid protein (Figs. 3A and 4), which is not present in other *gypsy*-type retrotransposons. PBS and PPT sequences of *RIRE3* are shown together with those of other *gypsy*-type retrotransposons in Fig. 5A. The PBS sequence is homologous to $\mathrm{tRNA}^{\mathrm{Met}}$, like that in *del1*.

It should be noted here that the nucleotide sequence of each of the cloned fragments differed from that of the consensus sequence at, on average, 1.6% of the sites, which is higher than the frequency of base substitutions (0.16%) induced by PCR. This indicates that these fragments are not all derived from the same *RIRE3* copy, but originate from many divergent copies of *RIRE3* in the rice genome. These mutations alter codons, but 98.5% of them, however, result in synonymous codons. It is important to mention that non-synonymous codons, including termination codons generated by 0.5% of the mutations, as well as the synonymous codons, do not appear in the consensus sequence of *RIRE3*. The consensus sequence is probably that which is most closely related to an active copy of *RIRE3*. It is worthwhile mentioning here that we tried but failed to obtain clones containing the entire *RIRE3* sequence with two LTRs – which should be amplifiable by PCR – probably because such clones are unstable, as they may be subject to homologous recombination between the two LTR sequences (Roeder and Fink 1980).

Among the clones obtained, however, one had a nucleotide sequence that differed from the consensus sequence by about 28% in the region coding for *int* (Fig. 2, clones 3-303 and 3-307). The nucleotide sequence was found to be associated with an LTR sequence with very poor homology to that of *RIRE3* in the region downstream of the PPT sequence. We named this sequence *RIRE8*, and it will be described below. Several other clones were found to have an *RIRE3* segment connected to a retrotransposon or non-retrotransposon sequence, in the region downstream of *int* or within the LTR of *RIRE3* (Fig. 2A). We describe these in a later section.

RIRE8 and members of its subfamily

To determine the entire sequence of *RIRE8*, we carried out one-primer PCR and cloned and sequenced fragments which covered the entire sequence of *RIRE8*

(Fig. 2B). Among the clones obtained, we noticed that there are two subtypes of *RIRE8*, and named them *RIRE8A* and *RIRE8B* (Fig. 2B). These have the same LTR sequences, but their *pol* regions differ by about 19.5% (Fig. 3Ad). Like *RIRE3*, *RIRE8A* and *RIRE8B* contain *orf0* in the region preceding *gag*; *orf0* shows less conservation than any other regions in the nucleotide sequence but all copies encode Orf0 proteins that are similar to each other in amino acid sequence (Figs. 3Ac, d and 5C; see also Fig. 4). Some clones contained an *RIRE8* sequence, which was joined to an LTR of another *RIRE8* in one or the other orientation relative to the *RIRE8* sequence (see Fig. 2B). We describe these clones in a later section.

The LTRs of *RIRE3* and *RIRE8A* are 2316 and 2948 bp long, respectively – the largest LTRs so far identified in retrotransposons. These LTRs are very poorly conserved in nucleotide sequence (see Fig. 3B). *RIRE8A* is about 600 bp larger than *RIRE3*, possibly due to an insertion in the 770–1530 bp region of *RIRE8A*. This sequence can be divided into GC-rich and AT-rich regions (Fig. 3B): the GC-rich region contains repeats of the sequence CGCCGT which appear with an irregular spacing of 19–38 bp. Although the LTR sequences of the two retrotransposons differ from each other, possible signal sequences (AATAAA and TATAA) for transcriptional termination and initiation, respectively, are found within the first 240 bp of the LTRs of both elements, separated by a space of about 70 bp (data not shown).

Nested structures seen in *RIRE3* and *RIRE8*

As described above, among the clones obtained during characterization of *RIRE3* and *RIRE8* were several that contained an *RIRE* sequence interrupted by the LTR sequence of another copy of the *RIRE* element (see Fig. 2). These clones appear to be derived from nested structures of the type shown in Fig. 6, and thus PCR using a single primer that hybridizes with an end region of LTR generates fragments containing various segments of the *RIRE* elements. In clone 3-121, however, a *RIRE3* sequence was interrupted by a sequence beginning with 5′-TG, which is characteristic for retrotransposons. We assume that the interrupting sequence is also a retrotransposon, which we have named *RIRE7* (see Fig. 6), that permitted amplification of the fragment in clone 3-121 by one-primer PCR. In fact, we have identified the LTR sequence of *RIRE7*, 858 bp in length, which is followed by the PBS sequence with homology with the 13 nt 3′ end nuclnteotides of $\mathrm{tRNA}^{\mathrm{Met}}$ (data not shown).

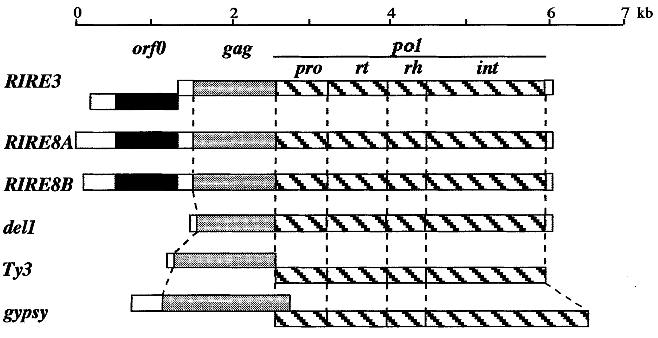


Fig. 4 Schematic representations of the internal regions of the *gypsy*-type retrotransposons. The *boxes* labeled *orf0*, *gag*, *pro*, *etc*., represent the coding regions for the corresponding proteins. *orf0*, *gag* and *pol* regions are indicated by *solid*, *shaded* and *cross-hatched boxes*, respectively. Note that *orf0* is specifically present in *RIRE* elements. Boxes at different levels are in different reading frames. With *gag* assigned the 0 reading frame, the boxes immediately below it are in the −1 frame. Sequence relationships between some of these elements are shown in Figs. 3 and 5

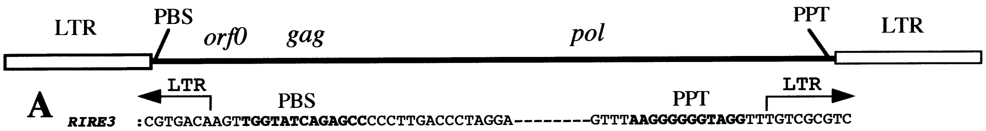
Among the clones identified, clone 3-125 contained a recombinant DNA segment between *RIRE3* and another retrotransposon, *RIRE2* (H. Ohtsubo, unpublished result), such that the *RIRE3* sequence is attached to a central region of *RIRE2* (see Fig. 2A). Clone 3-115 contained another recombinant DNA segment between an *RIRE3* sequence and a chloroplast DNA (ribosomal protein S11) (see Fig. 2A). There are no homologous sequences at the recombinational junctions between the two different sequences (data not shown). Thus, the mechanism involved in generation of these recombinants is unknown at present.

RIRE3 elements as ubiquitous components of plant genomes

To determine whether the *RIRE3* elements identified in this paper are distributed widely in plants, we carried out PCR to isolate fragments (797 bp in length) of a region in the *rt* sequence by using a pair of primers (R3-3′Pro and R3-5′Pro; Table 2), which hybridize with the *rt* sequence of *RIRE3* but not with the *rt* sequence of *RIRE8* elements or *del1,* with total DNA from each of the plants listed in Table 1 as the template. Monocotyledonous plants, such as foxtail millet, maize, wheat and barley, as well as dicotyledonous plants, such as mulberry, morning glory, tobacco and *Arabidopsis thaliana*, generated fragments, whose sizes were the same as that generated from the rice strain *O. sativa* L cv. IR36 or the rice strain *O. sativa* L cv. Nipponbare. After cloning and sequencing the fragments, we constructed a phylogenetic tree based on the amino acid sequences deduced from the nucleotide sequences, which clearly showed that *RIRE3* is ubiquitously distributed as a gene family in monocots and dicots (Fig. 7).

Discussion

Many retrotransposons have been identified in rice, but all of these have been found to be related to *copia* (Hirochika et al. 1992; Hirochika and Hirochika 1993;



A

RIRE3	:CGTGACAAGT**TGGTATCAGAGCC**CCCTTGACCCTAGGA--------GTTT**AAGGGGGGTAGG**TTTGTCGCGTC
RIRE8A	:CGTGACAAGT**TGGTATCAGAGCCA**CCTTGACCTTAGGA--------GTTT**AAGGGGGGTAGG**TTTGTCACACC
RIRE8B	:CGTGACAAGT**TGGTATCAGAGCC**CC-TTGACCCTAGGT--------GTTT**AAGGGGGGTAGG**TTTGTCACACC
del1	:GTGTATAAAG**TGGTATCAGAGC**TTTAGGTTCCAGGTAA--------CTTTT**GGGGGGAAGAAT**TATGAAACC
GYPSY	:TTATATAATTTGGCGCCCAACCAACAATCTGAACCCAC--------TAAACTT**GAGGGGGGAG**TAGTTAACAA
Ty3	:TACAACA-CCTGGTAGCGTTAAAGGTTACTAATTGTTC--------CCCT**GAGAGAGAGGAAGA**TGTTGTATC
tRNA^{Met}:	**TGGTATCAGAGCCAGGTT**-------

B

```
           *                  *    *    **         **          *   *   **  *     * * *
RIRE3 :YKRPYRMAANELAE-VKKQVDEQLQKGYIRPGTSPWGAPVIFVEKK------DKTKRMCVDYRALNEVTIKNKYPLPRID  73
RIRE8A:HKRPYRMAANELAE-VKKQVDDLLQKGYIRPSTSPWGAPVIFVEKK------DHTQRMCVDY-ALNEVTIKNKYPLPRID  72
RIRE8B:H-RPYRMAANELAE-VKRQVDDLLQKGYIRLSSSPWGAPVIFVEKK------DHT-RMCVDYRALNDVTIKNKYPLPRID  71
del1  :SITPYHMAPSELQE-LKEQLEDLLNKGFIRGSTSPWGAHVLFDPKK------DDSKRMCIDYK-LNSVTVKNKYPLPRID  72
Ty3   :L-QPYHVTEKNEQE-INKIVQKLLDNKFIVPSKSPCSSPVVLVPKK------DGTFRLCVDYRTLNKATISDPFPLPRID  72
gypsy :YSRAYP-TLMGVSDFVNNEVKQLLKDGIIRPSRSPYNSPTWVVDKKGTDAFGNPNKRLVIDFRKLNEKTIPDRYPMPSIP  79

               *  *   **   ***      *   * *    *  *     *** *    *                 *  *
RIRE3 :DLFDQLKGAKVFSKIDLRSGYHQLRIREXDIPKTAFTTRYGLYEFTVMSFGLTNAPAFFMNLMNKVFMEYLDKFVVVFID 153
RIRE8A:DLFDQLEGATVFSKIDLRSGYHQLRIREEDIPKTAFTTRYGLFECTVMSFGLTNAPAFFMNLMNKVFMEYLDKFVVVFID 152
RIRE8B:DLFDQLKGATVFSKIDLRSGYHQLRIKEEDIP-TAFTTRYGLFECTVMSFGLTNAPTFFMNLMNKVFMEYLDKFVVVFID 150
del1  :DLFDQLNGAY-FSKIDLRFRYHQLRIR-ADIPKTAFRTRYGHYEFLVMPFGLTNVPTAFMNLMNRVFREYLDKFIVVFVD 150
Ty3   :NLLSRIGNAQIFTTLDLHSGYHQIPMEPKDRYKTAFVTPSGKYEYTVMPFGLVNAPSTFARYMADTFRD-L-RFVNVYLD 150
gypsy :MILANLGKAKFFTTLDLKSGYHQIYLAEHDREKTSFSVNGGKYEFCRLPFGLRNASSIFQRALDDVLREQIGKICYVYVD 159

          * *     *  *    *  *          *  *       **                       *        **
RIRE3 :DILIYSQSEEDHQHHLRLVLGKLREHQLYAKLSKCEFWLSEVKFLGHVISAKGVAVDPETVTAVTDWKQPKTVTQVRSFL 233
RIRE8A:DILIYSKTKEEHEEHLRLALEKLREHQLYAKFSKCEFWLSEVKFLGHVISSGGVAVDPSNVESVLSWKQPKTVSEIRSFL 232
RIRE8B:NILIYSRTKEEHEEHLRLALEKLREHQLYAKFSKCEFWLSEVKFLGHVISAGGVAVDPSNVESVTNWKQPKTVSEIRSFL 230
del1  :YVLIYSRTQKDHEHHLRISLQLLRNNQLYAKLSKCEFWMEKVKFLGHVVSREGIVVDPVKVKAVMNWELPKNIFEIRSFL 230
Ty3   :DILIFSESPEEHWKHLDTVLERLKNENLIVKKKKCKFASEETEFLGYSIGIQKIAPLQHKCAAIRDFPTPKTVKQAQRFL 230
gypsy :DVIIFSENESDHVRHIDTVLKCLIDANMRVSQEKTRFFKESVEYLGFIVSKDGTKSDPEKVKAIQEYPEPDCVYKVRSFL 239

       *   ***  *     **                                    *   *      *
RIRE3 :GLAGYYRSCIENFSKIARPMTQLLKKEEKFV-----------WSPQCEKAFKTLKEKLVSSPV-LILPD            290
RIRE8A:GLAGYYRRFIENFSKIARPMTRLLQKEVKYK-----------WTEGCEQSFQELKKRLVTAPV-LILPD            289
RIRE8B:GLAGYYRRFIENFSKIAKPMTRLL-KDVKYK-----------WSEECEQSFQELKNRLISAPI-LISAD            286
del1  :GLAGYYRRFIKGFSKLAALMTQLTKKGENFN-----------WTKKYQNSFDELK-RLTTVPV-LTIPI            286
Ty3   :GMINYYRRFIPNCSKIAQPI-QLFICD-KSQ-----------WTEKQDKAIEKLKAALCNSPV-LVPFN            285
gypsy :GLASYYRVFIKDFAAIARPITDILKGENGSVSKHMSKKIPVEFNETQRNAFQRLRNILASEDVILKYPD            308
```

C

```
       *        *    *** *** *    *      *     *              **    * *    *    ***
RIRE3 :MAGSVTRRGL-DMTGFVLELRGMCVVLGYPHGVDYQARPLPEQEGEDAEPHAGWEVTAVILSGSPERTSLAVTAGGDSFP  79
RIRE8A:M-PRHSPRNLIEVLGFVRELRLMAFEVGYEVAPEY-RQ-IP-HDPDKEKCRV-R-VT--LASDSEDLPSFKFEAGGRSYC  72
RIRE8B:M-PRYKPEYLFGVKGFVQELRTMSFVIGFGTAPFY-AQ-IP-HDRHEKKCRV-K-VT--LNSNSEDIPSVMFEAGGRNYI  72

         **  *  *** *       *      ***      *          *  **  *   *  * *            *
RIRE3 :AACQSAALLAIGTLHQRYPDELQHSPYRYHPRRGGARDYATFRDASS-EDDATIVHLARMVEAYDAARIDFHQMVR-RGM 156
RIRE8A:HACQEAALVAIGELRQHFEEELDSSAFQYHPHKPHGQDYGTYTCPDG-EESATLMHAVHMLSAMDAVVVERDKAAHDREN 151
RIRE8B:HSCQEVARIAIGELRDRYSDQLADTEYRYHPRQPQGSDRGSYLETEGIENDATTRHLVEMLWAMDESRAETVLAAQDRED 152

         *  **  *      * * *                  *       * *           *                 *
RIRE3 :VENNLKILELRQENLQLKKDLDAVEAQLHQLKIAQGEDCRPQRRRVCRSQKITARKSTSRPELVRQSLAWTCFVETPRAE 236
RIRE8A:W-NRGKICKLKSKVYRLQKELAELKGRTPLPTSKLRLTPRKRTCPPPRLQLASKIRVMG--EAVPDHAEPVIDIISDEEE 229
RIRE8B:R-NRGKICKLEDKVDRLEKELAALKGEAPPQKARIRLTARKRALFVLRYQLAPKVRVVEK-EVAPTLADPPVIIISDDEE 231

                 *       **         *
RIRE3 :PAPV-VPQEGEASGVGSTEDALL---LTFRPGPSQR                                            268
RIRE8A:EE---DPEERE-PATPEEEDTSLRP--DARRK----                                            254
RIRE8B:EESKRNHSKIEWVATQDDEDEPNEPSIDLRCSEELE                                            266
```

Fig. 5A–C Nucleotide sequences of PBS and PPT (**A**), and amino acid sequences encoded by the *rt* gene (**B**) and *orf0* (**C**) in retrotransposons. The structure of an *RIRE* element is schematically shown at the *top*. In **A**, the sequence complementary to the 3′ region of methionyl tRNA (tRNA^{Met}), which initiates (-) strand DNA synthesis, is shown in *bold type*. Possible PPT sequences are also shown in *bold type*. In **B**, amino acid sequences encoded by the *rt* genes of *gypsy* of *D. melanogaster* (Marlor et al. 1986), *Ty3* of *S. cerevisiae* (Hansen et al. 1988) and *del1* of *Lilium henryi* (Smyth et al. 1989) are shown together with those encoded by the *RIRE* elements. In **C**, amino acid sequences encoded by *orf0* of the *RIRE* elements (see Fig. 6) are shown. In **B** and **C**, *asterisks* indicate conserved amino acids

Noma et al. 1997). In this paper, we have identified and characterized three elements named *RIRE3*, *RIRE8A* and *RIRE8B* as retrotransposons that are related to *gypsy*. This demonstrates that the rice genome contains not only *copia*-type retrotransposons but also several kinds of *gypsy*-type retrotransposons. We demonstrated here that *RIRE3* is distributed in both monocot and dicot plants, indicating that the *gypsy*-type of retrotransposon is also ubiquitously distributed in plants. The phylogenetic tree constructed for the *RIRE3*-related elements shows that the elements do not fall into two clusters corresponding to dicots and monocots (see Fig. 7). This suggests that *RIRE3*-related retrotransposons are transferred horizontally. It is however as yet unknown whether horizontal transfer is the major mechanism of distribution of the retrotransposons or not, although this possibility has been suggested for some *copia*-type retrotransposons (Voytas et al. 1992; Hirochika and Hirochika 1993). Further phylogenetic analysis needs to be done in this respect by identifying and characterizing fragments from more species of plants, using primers that hybridize not only with *RIRE3* but also with other *gypsy*-type retrotransposons.

In the internal regions of the *RIRE* elements, the PBS sequences are complementary to the 3′-end sequence of tRNA^{Met}, as in the *gypsy*-type retrotransposon *del1* of lily (see Fig. 5A). It is interesting to note here that the plant *gypsy*-type retrotransposon *Grande1* – which is not

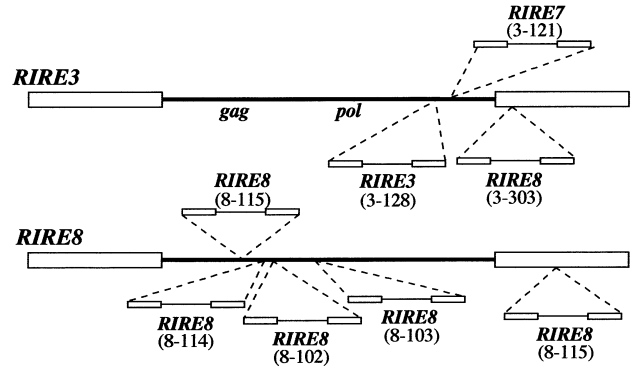


Fig. 6 Nested structures of *RIRE3* and *RIRE8*. *RIRE8*, *RIRE7* or *RIRE3* itself can be found inserted in *RIRE3* at the positions indicated. *RIRE8* can also be inserted in itself. The *RIRE* elements can be inserted in either orientation: the elements shown above *RIRE3* or *RIRE8* are inserted in the same orientation; those underneath *RIRE3* or *RIRE8* are inserted in the inverse orientation. Clones used to determine the positions of the nested elements are shown in *parentheses* (see Fig. 2)

closely related to *RIRE* elements – and those in animals have PBS sequences with homology to the 3′-end sequences of tRNA other than tRNA^{Met} (Bingham and Zachar 1989). We observed that the internal region of *RIRE3* contained one large ORF consisting of *gag* and *pol* regions, indicating that *RIRE3* synthesizes a polyprotein comprising Gag and Pol. We have shown that the *pol* regions of the three *RIRE* elements are more highly homologous than the *gag* regions (see Fig 3A; the degree of nucleotide sequence homology in the *pol* regions between *RIRE3* and *RIRE8A,* for example, is about 68%, whereas that in the *gag* regions is about 57%). The regions upstream of the *gag*-*pol* regions in the *RIRE* elements were even less homologous than the *gag* regions but contained a small ORF, here called *orf0*, which is not present in *del1* or other *gypsy*-type retrotransposons. Although the nucleotide sequences of *orf0*s in three *RIRE* elements are not well conserved (see Fig. 3A; the nucleotide sequence homology in the *orf0* regions between *RIRE3* and *RIRE8A*, for example, is about 49%), the amino acid sequences encoded by *orf0*s are well-conserved (see Fig. 5C). This indicates that *orf0* encodes a protein that has a function which, however, is not known at present. The region between *orf0* and *gag* contains a termination codon(s) in all three frames, suggesting that the *orf0* product is not synthesized as a polyprotein together with Gag and Pol. The expression pattern of *orf0* is currently under investigation.

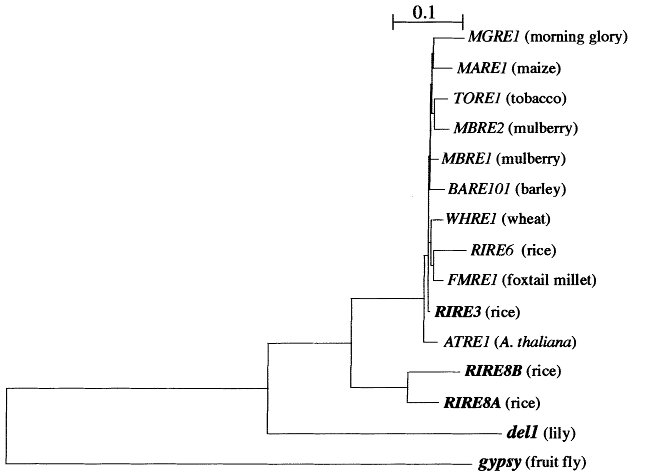


Fig. 7 Dendrogram of the *gypsy*-type retrotransposons in various plants, based upon portions of the amino acid sequence of reverse transcriptase. The tree is unrooted, and branch lengths are proportional to the amino acid distance (in percent divergence) between the node connecting all sequences. The scale bar represents a distance of 0.1. Amino acid sequences derived from the primer sequences were omitted from the calculations for construction of the tree, because they may not represent the authentic sequences

Although the two subtypes of *RIRE8*, namely *RIRE8A* and *RIRE8B*, had the same LTR sequences and homologous *gag*-*pol* regions, their *orf0* regions are very poorly conserved (see Fig. 3A; the nucleotide sequence homology in the *pol*, *gag* and *orf0* regions between *RIRE8A* and *RIRE8B* is about 80%, 77% and 49%, respectively). This indicates that the region downstream of PBS tends to vary in retrotransposons. PBS is the initiation site for cDNA synthesis in retrotransposons, and thus the regions downstream of these sites are reverse-transcribed last, and this may cause accumulation of mutations in these regions by a mechanism which is presently unknown.

LTRs in *RIRE3* and *RIRE8* have diverged greatly, although their internal regions are homologous, particularly the *pol* regions. It has been observed that the *copia*-type retrotransposons, such as *RIRE1*, *BARE-1* and *Wis-2-1A*, which are closely related to one another, have LTRs that are greatly divergent (Noma et al. 1997). This suggests that LTRs in retrotransposons generally can diverge very frequently, presumably because LTRs are non-coding regions. The sizes of LTRs in *RIRE3* and *RIRE8* are slightly different but are much larger than those of other retrotransposons so far identified. As mentioned in the Results section, microsatellite-like sequences exist in different regions, which might have caused the LTR sequences to lengthen.

As described in this paper, the *RIRE* elements show homology with *del1* of lily but did not with other plant retrotransposons, such as *Grande1* of teosinte, whose entire sequence has been reported. The *RIRE* elements and *del1*, which have very long LTR sequences, appear to show partially homologous sequences at the terminal regions of their LTRs, which begin with TG (Fig. 8). Note here that these homologous sequences are not seen in *Grande1*, which has short LTR sequences. We have recently identified and characterized another *RIRE* element, named *RIRE2*, with short LTR sequences, which is closely related to *Grande1* (H. Ohtsubo and E. Ohtsubo, unpublished results). *RIRE2* and *Grande1* appear to have partially homologous sequences in the terminal regions of their LTRs, indicating that closely related to one another have their own characteristic terminal sequences.

It should thus be possible to identify an insertion sequence which begins with 5′-TG as a retrotransposon related to *RIRE3* and *RIRE8* by comparing its terminal sequence with those shown in Fig. 8, without knowing the coding regions for reverse transcriptase and integrase. For example, one LTR of the barley retrotransposon *BARE-1* contains an insertion sequence, 3130 bp in length (Manninen and Schulman 1993). This sequence begins with 5′-TG and ends with CA-3′ and is inserted such that it is flanked by the 5-bp sequence CCTAG. The terminal regions of the insertion sequence, which we wish to call *BARE-100* here, show significant homology with those of *RIRE3* and the elements related to it (see Fig. 8), suggesting that the insertion sequence is a *gypsy*-type retrotransposon of barley. If this is the case, the 3130-bp segment, which is too short to be an intact retrotransposon and has no long terminal repeat sequences, must be a solo-LTR, which is assumed to have been generated by recombination between two LTRs of a retrotransposon originally inserted into *BARE-1*. The 5-bp sequences at the insertion sites are the target sequence duplicated upon insertion of *BARE-100*. As another example, a retrotransposon-like sequence, called *RIRE4*, which interrupts a gene in rice, begins with 5′-TG (Y. Iida, E. Ohtsubo and H. Ohtsubo, unpublished results). The terminal sequences show significant homology with those of *RIRE3* and the elements related to it (see Fig. 8), suggesting that *RIRE4* is another *gypsy*-type retrotransposon in rice.

As described in this paper, *RIRE3* and *RIRE8* often serve as sites for the insertion of these or different retrotransposons (see Fig. 6). A *copia*-type retrotransposon *RIRE1* of rice is present in the *O. australiensis* genome in an extraordinary number of copies and is often inserted into itself to form a nested structure (Nakajima et al. 1996; Noma et al. 1997). It has been reported that the *Adh-1-F* locus of *Zea mays* contains retrotransposons of more than 10 types, which are nested in one another (SanMiguel et al. 1996). It seems likely that retrotransposons, once inserted at some loci, may be used as targets for new insertions, thus reducing the chance of insertions into genes that are essential for the growth of plant cells.

Thus we identified another retrotransposon element, named *RIRE7*, which appeared to be inserted into an *RIRE3* sequence (see Fig. 6). We found by computer analysis of the databases a cDNA clone (accession numbers C28873 and D39833), from a cDNA library made from mRNA of etiolated shoots of cv. Nipponbare, that showed homology with the LTR (858 bp in length) of *RIRE7* in the region starting at nucleotide position 518 bp. It is assumed that this cDNA is derived from an mRNA which is expressed – probably upon etiolation of shoots – from a possible promoter in the region between nucleotide positions 481 to 415 bp that contains a TATA box . We do not know at present whether *RIRE7* belongs to the *gypsy* or *copia* family.

```
RIRE3    :TGTCGCGTCCCAAATCGACTCTAAAATATATCCTCGAAATT----
         :TGTCACGACCGGAAATAACCCAACGGGCGTTCCTTACGTGC----
RIRE8    :TGTCACACCCTAAAAATCCAAAATATATAAATTGTTGTTTA----
         :TGTCACGCCCGGAATTTCTATCCAAAATTCCAAACGCTTAC----
del1     :TATGAAACCCTGAATTTTCGCATAAAACTATGAGTTACCGT----
         :TTTCACGACCCGACATTTTATATAAAAAACACCGGGTGTGA----
BARE-100:TGTGACAGCCCGATGCCGACGTTCCAGAAGATTCCCCCCTT----
         :TGTCACGCCCAAGATGCGACCCTATCCTCGATTTGGCACGA----
RIRE4    :TGTCACGTCTGAAAATTCACTAGTAATTTCCGAACTTATTT----
```

Fig. 8 Comparison of nucleotide sequences at the terminal regions of LTRs of *gypsy*-type retrotransposons. Nucleotides shown in *bold type* are those frequently seen at each position. The sequences of *RIRE3* and *RIRE8* LTRs were determined in this work. Nucleotide sequences of *del1* of lily (Smyth et al. 1989) are shown. *BARE-100* and *RIRE4* are described in the text

After submission of this paper, we noted that Wright and Voytus (1998) have described *Tat1*, an element related to a group of *Arabidopsis thaliana Ty3/gypsy* retrotransposons. *Tat1* and the elements related to it are closely related to *Grande1*, suggesting that they are distantly related to the *RIRE* elements described in this paper. Suoniemi et al. (1998) have reported that *gypsy*-like retrotransposons are present in some monocot and dicot plants, including rice. They analyzed the sequences of an *rt-int* segment in these elements and reported them to be closely related to *del1*. Since the central regions of all of the sequences, except one from barley, are missing, it is difficult to guess how closely they are related to the *RIRE* and other elements described in this paper.

Acknowledgments We thank Dr. K. Okuno (National Institute of Agricultural Resources) for providing us seeds of the rice strain used in this study. We also thank our colleague Dr. S. Tsuchimoto for valuable discussion. This work was supported by a Grant-in Aid for Scientific Research from Ministry of Education, Science and Culture of Japan, and by a grant for Development of Methods of Rice Genome Analysis and Use of Accurate Genomic Map of Rice from the Ministry of Agriculture, Forestry and Fisheries of Japan.

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–219

Bennetzen JL (1996) The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol 4:347–353

Bingham PM, Zachar Z (1989) Retrotransposons and the FB Transposon from *Drosophila melanogaster*. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 485–502

Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res 16:9677–9686

Felsenstein J (1995) PHYLIP (Phylogeny Inference Package) and manual, version 3.57c. Department of Genetics, University of Washington, Seattle

Graham M (1995) Cereal genome evolution:pastoral pursuits with 'Lego' genomes. Curr Opin Genet Dev 5:717–724

Hansen LJ, Chalker DL, Sandmeyer SB (1988) Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol Cell Biol 8:5245–5256

Hansen LJ, Chalker DL, Orlinsky KJ, Sandmeyer SB (1992) *Ty3 GAG3* and *POL3* genes encode the components of intracellular particles. J Virol 66:1414–1424

Hirochika H, Hirochika R (1993) *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68:35–46

Hirochika H, Fukuchi A, Kikuchi F (1992) Retrotransposon families in rice. Mol Gen Genet 233:209–216

Kunze R, Saedler H, Lönnig W-E (1997) Plant transposable elements. Adv Bot Res 27:331–470

Lichtenstein C, Draper J (1985) Genetic engineering of plants. In:Glover DM (ed) DNA cloning: a practical approach, vol 2. IRL Press, Washington, DC, pp101–109

Manninen I, Schulman AH (1993) *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 22:829–846

Marlor RL, Parkhurst SM, Corces VG (1986) The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. Mol Cell Biol 6:1129–1134

Martínez-Izquierdo JA, García-Martínez J, Vicient CM (1997) What makes *Grande1* retrotransposon different? Genetica 100:15–28

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Motohashi R, Mochizuki K, Ohtsubo H, Ohtsubo E (1997) Structures and distribution of *p-SINE1* members in rice genomes. Theor Appl Genet 95:359–368

Mount SM, Rubin GM (1985) Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol 5:1630–1638

Nakajima R, Noma K, Ohtsubo H, Ohtsubo E (1996) Identification and characterization of two tandem repeat sequences (TrsB and TrsC) and a retrotransposon (*RIRE1*) as genome-general sequences in rice. Genes Genet Syst 71:373–382

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E (1997) *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet Syst 72:131–140

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

Roeder GS, Fink GR (1980) DNA rearrangements associated with a transposable element in yeast. Cell 21:239–249

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Smyth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the gypsy group of *Drosophila*. Proc Natl Acad Sci USA 86:5015–5019

Suoniemi A, Tanskanen J, Schulman AH (1998) *Gypsy*-like retrotransposons are widespread in the plant kingdom. Plant J 13:699–705

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

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

Varmus H, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington DC, pp 53–108

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

Wahl GM, Meinkoth JL, Kimmel AR (1987) Northern and Southern blots. Methods Enzymol 152:572–581

Wright DA, Voytas DF (1998) Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana* Ty3/*gypsy* retrotransposons that encode envelope-like proteins. Genetics 149:703–715

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362