Structural Diversity of a Novel LTR Retrotransposon, RTPOSON, in the Genus Oryza

Yu-Chia Hsu1, Chang-Sheng Wang2, Yann-Rong Lin3 and Yong-Pei Wu1

1Department of Agronomy, Chiayi Agricultural Experiment Station, Taiwan Agricultural Research Institute, Chiayi, Taiwan. 2Department of Agronomy, National Chung Hsing University, Taichung, Taiwan. 3Department of Agronomy, National Taiwan University, Taipei, Taiwan.

ABSTRACT: Retrotransposons with long terminal repeats (LTRs) are the most abundant transposable elements in plant genomes. A novel LTR retrotransposon named RTPOSON primarily occurs in the genus Oryza and in several species of the Poaceae family. RTPOSON has been identified in the Tyl-copia group of retrotransposons because two of its open reading frames encode an uncharacterized protein and UBN2_2 and zinc knuckle, respectively. More than 700 RTPOSONs were identified in Oryza genomes; 127 RTPOSONs with LTRs and gag-pol elements were classified into three subgroups. The subgroup RTPOSONsub3 contained the smallest DNA size and 97% (32/33) of RTPOSON elements were classified in this group. The insertion time of these RTPOSONs varied and their proliferation occurred within the last 8 Mya, with two bursting periods within the last 1.5–5.0 Mya. A total of 37 different orthologous insertions of RTPOSONs, with different nested transposable elements and gene fragments, were identified by comparing the genomes of Oryza sativa cv. Nipponbare and Oryza indica cv. 93–11. A part of intact RTPOSON elements was evolved independently after the divergence of indica and japonica. In addition, intact RTPOSON and homologous fragments were preferentially retained or integrated in genic regions. This novel LTR retrotransposon, RTPOSON, might have an impact on genome evolution, genomic innovation, and genetic variation.

KEYWORDS: LTR retrotransposon, RTPOSON, genome, Oryza, evolution, genetic variation

Introduction

Transposable elements (TEs) are mobile genetic sequences that can transpose themselves within genomes, and their activity produces structural changes in single genes or in overall genomes followed by altered spatial and temporal patterns of gene expression and function.1 Retrotransposons, also known as class I TEs, can translocate themselves through an RNA-mediated copy-and-paste mechanism for a rapid increase in copy numbers in plant genomes.2 A typical intact retrotransposon contains long terminal repeats (LTRs), a primer-binding site (PBS), a polypurine tract (PPT), and gag-pol encoding polyproteins for transposition in a genome. LTRs terminate, usually 5′-TG-3′ and 5′-CA-3′, and are flanked by 4–6-bp target site duplications (TSDs).3 Gag encodes a capsid-like protein to pack RNA, and pol encodes a polyprotein that contains the activities of protease (PR), reverse transcriptase (RT), RNaseH (RH), and integrase (INT) to facilitate the integration of double-stranded DNA into the genome.4 Accumulating evidence suggests that retrotransposon activities have a profound impact on various aspects of genome dynamics,5,6 such as genome size, chromosome rearrangement, chromatin formation, gene transcription, and evolution.7–10

LTR retrotransposons are the most abundant and most widely distributed mobile genetic elements in the plant kingdom. In fact, variable sizes of plant genomes are produced by the copy-and-paste proliferation of a few LTR elements and the aggressive purging of these proliferating LTRs via several mechanisms, including illegitimate and incomplete recombinations and double-strand break repairs by nonhomologous end joining.11 LTR retrotransposons represent a wide range of proportion of various genomes (eg, ~14% of the Arabidopsis genome,12 42% of the soybean genome,13 55% of the sorghum genome,14 and >85% of the maize genome).15

The genome size of Oryza sativa, a cultivated Asian species, is estimated to be ~389 Mb, of which 19% (72 Mb) contains retrotransposons.16 Nevertheless, the 10 other genomes of the Oryza genus exhibit substantial diversity in size, from 357 Mb for Oryza glaberrima (genotype type AA) to 960 Mb for the Oryza australiensis (genotype type EE), associated with the abundance of LTR retrotransposons.17–19 LTR retrotransposons may play important roles in genome expansion via amplified replication and also genome contraction by illegitimate and unequal homologous recombinations.20 More than 190 Mb of LTR-retrotransposon sequences have been deleted from the rice genome by recombination after their insertion during the
last eight million years ago (Mya), and similar findings have been reported in Arabidopsis, barley, and cotton.

The insertion positions of LTR retrotransposons could directly affect the regulation of neighboring genes at the transcriptional and posttranscriptional levels or result in a new combination of genes via alternative splicing. In Vitis vinifera, the insertion of the retrotransposon Gret1 upstream of the Myb-related gene VmybA1 and the consequent rearrangement of Gret1 resulted in the color change of pericarp from black to white and red. A similar phenomenon was found in the colors of the fruit flesh in Citrus sinensis, in which blood orange and Maro (I) were caused by an insertion of LTR retrotransposons in Myb transcription factors called Ruby and a recombination between the retrotransposons, respectively.

In foxtail millet, different insertions of TEs in granule-bound starch synthase 1 (GBSSI) led to the nonwaxy endosperm converted into a low-amylose and waxy endosperm. One waxy rice cultivar resulted from the insertion of an RIRE-like retrotransposon in the ninth exon of GBSSI. The insertion of an LTR retrotransposon, Renovator, in the promoter region of the rice blast resistance gene Pit caused a 34-fold increase in Pit expression after pathogen inoculation. Furthermore, one-sixth of the genes in the rice genome are associated with retrotransposons, and some of the retrotransposons are attributed to novel exons by altered splice sites, which may promote the evolution of new functions.

Rice is an appropriate model to study mobile elements because of its small genome size with various major types of TEs. This genomic resource allows for characterizing all rice TEs in silico, particularly LTR retrotransposons. A comparative genomics approach of the genomes of Nipponbare and 93-11 identified a family of novel LTR retrotransposons named RTPOSON. Here, we aimed to (1) identify all the elements related to RTPOSON in the Oryza genomes, (2) annotate the gene structure of RTPOSON, and (3) search for genes affected by RTPOSON or its homologs. In attempting to achieve these three aims, we elucidated the role of RTPOSON in gene and genome evolution and discussed various perspectives on the application of RTPOSON in functional genomics.

**Materials and Methods**

**Database of Oryza genomes.** A total of 13 rice species were investigated to uncover novel LTR retrotransposons, which are summarized in Table 1. The Asian cultivated rice O. sativa has two subspecies, ssp. japonica, Nipponbare, and ssp. indica, 93-11, whose genomes were analyzed. The African cultivated species O. glaberrima was also included. The other 11 species were wild rice species. They included the AA genome species Oryza nivara, Oryza rufipogon, Oryza barthii, Oryza meridionalis, and Oryza glumaepatula; the BB genome species Oryza punctata; the BBCC genome species Oryza minuta; the CC genome species Oryza officinalis; the FF genome species Oryza brachyantha; and the GG genome species Oryza granulate. All genome sequences used were retrieved from the Gramene database (http://www.gramene.org/), except for the Nipponbare genome sequence, which was downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov/).

**Identification and characterization of a novel LTR retrotransposon.** As shown in the workflow in Supplementary Figure 1, as a first step to identify a novel LTR in the genome of Oryza rice we used FastPCR Professional 6.5 or the LTR_Finder program (http://tlife.fudan.edu.cn/ltr_finder/) to search the Nipponbare genome with default

---

Table 1. Distribution of RTPOSON in 13 Oryza species.

| SPECIES | GENOME | GENOME SIZE (Mb) | COPY NUMBER |
|---------|--------|-----------------|-------------|
|         |        |                 | INTACT | FRAGMENT | SOLO-LTR | TOTAL |
| O. sativa ssp. japonica | AA | 389 | 15 | 37 | 41 | 93 |
| O. sativa ssp. indica | AA | 400 | 13 | 40 | 32 | 85 |
| O. nivara | AA | 448 | 14 | 26 | 30 | 70 |
| O. rufipogon | AA | 445 | 1 | 4 | 2 | 7 |
| O. glaberrima | AA | 354 | 16 | 18 | 33 | 67 |
| O. barthii | AA | 411 | 18 | 28 | 26 | 72 |
| O. meridionalis | AA | 435 | 5 | 27 | 41 | 73 |
| O. glumaepatula | AA | 464 | 10 | 32 | 41 | 83 |
| O. punctata | BB | 423 | 33 | 25 | 18 | 76 |
| O. minuta | BBCC | 1,124 | 1 | 2 | 7 | 10 |
| O. officinalis | CC | 653 | 1 | 11 | 13 | 25 |
| O. brachyantha | FF | 260 | 0 | 6 | 13 | 19 |
| O. granulata | GG | 862 | 0 | 6 | 19 | 25 |
| Total | | | 127 | 262 | 316 | 705 |

**Notes:** 1Reference from the study by Jacquemin et al. A limited number of sequences were searched and collected from rice chromosome 3 in Gramene website.
Structural diversity of a novel LTR retrotransposon.

Figure 1. The gene structure of the novel retrotransposon, RTPOSON. (A) Structural annotation for one of the intact RTPOSON elements from O. sativa ssp. indica, 93-11. The 5′- and 3′-LTRs are shown in the light blue arrows; TSD indicates the 5-bp target site duplication; PBS indicates the primer-binding site; PPT indicates the PPT; gag-pol is able to transcribe a polyprotein that is provided for the transposition activity of retrotransposon. The gene, gag-pol, contains two ORFs, of which ORF1 contains the uncharacterized protein and ORF2 contains two domains, UBN2_2 (pfam144227) for gag and zinc knuckle (pfam00098). (B) The sequence alignment of two LTRs of intact RTPOSON elements randomly selected from 93-11. The identical nucleotides are indicated by gray shading. (C) Homology matrix between the RTPOSON element and the tobacco Ty1-copia Tnt1 (X13777) internal coding region. The approximate regions corresponding to the gag, zinc knuckle, and PR domains of tobacco Ty1-copia encoded by pol are indicated below the dotted lines. The second long match region is similar to zinc knuckle of Tnt1.

parameters. The output LTRs were then manually inspected to filter out incorrectly predicted sequences and define the exact boundaries of retroelements. Significant hits were carefully inspected to examine the boundaries of each element and TSD. We uncovered a novel LTR retrotransposon that had not been previously reported and named it RTPOSON.

To search for homologous elements of RTPOSON in Oryza genomes, the complete RTPOSON2 (GenBank accession no. AP003073, Supplementary Table 1) was used individually as a query by BLASTN searches with a default threshold (E-value < 10^{-10}) against the NCBI and Gramene databases, including the Nipponbare, 93-11, O. nicara, O. rufipogon,
O. glaberrima, O. barthii, O. meridionalis, O. glumaepatula, O. punctata, O. minuta, O. officinalis, O. brachyantha, and O. granulate genomes.

The internal sequence of each RTPOSOn element was annotated for gag-pol and translated into amino acid sequences by using open reading frame (ORF) finder (http://www.ncbi.nlm.nih.gov/orf/gorf.html) and Vector NTI 7.0 (InforMax Inc.). The gene orders of gag-pol in RTPOSOn were determined by using BLASTX and tBLASTX (http://blast.ncbi.nlm.nih.gov) programs with the Tnt1 copia-like retroelement (GenBank accession no. X13777) as the reference sequence. Each of the intact elements defined in this study included intact LTRs and gag-pol (Table 1). The fragmental elements were defined as deficient in one LTR or eventually part of the internal regions. Solo LTRs were used to indicate the elements that contained an intact LTR sequence.

**Estimation of insertion time of RTPOSOn in Oryza genomes.** The insertion time of RTPOSOn in Oryza genomes was estimated by comparing the divergence of their 5′ and 3′ LTRs in the intact elements with TSD sequences, which is represented as being identical at the time of integration. For each intact copy, the two LTRs were aligned by using Jalview, and the nucleotide substitution (transitions + transversions) rate between the two LTRs was estimated by the Kimura two-parameter method by using MEGA version 6.0. The insertion time that was estimated by using the P-distance model.

**Phylogenetic analysis of RTPOSOns.** The phylogenetic tree of 127 homologous elements of RTPOSOns was constructed based on the conserved gag sequences, such as the gag sequence that was obtained from 2,230–2,489 bp in RTPOSOn2 (Supplementary Table 1), to reveal the evolutionary aspects. The 127 homologous elements of RTPOSOn included 33 elements from O. punctata, 18 elements from O. barthii, 16 elements from O. glaberrima, 15 elements from Nipponbare, 14 elements from O. nivara, 13 elements from 93-11, 10 elements from O. glumaepatula, 5 elements from O. meridionalis, and 1 element each from O. ruftpogen, O. minuta, and O. officinalis. The sequences used to build the phylogenetic trees are listed in Supplementary Table 1. The sequences were aligned by using Jalview with default options, and then underwent phylogenetic analysis. The phylogenetic tree was generated by the neighbor-joining method by using MEGA version 6.0 and 1,000 heuristic bootstrap replications by using the P-distance model.

**Results**

Identification and characterization of the RTPOSOn LTR retrotransposon in Oryza genomes. A novel LTR retrotransposon of 3 kb named RTPOSOn was found in the genomes of ssp. japonica cv. Nipponbare and ssp. indica cv. 93-11 according to the common characteristics of an internal putative gag-pol coding region flanking two same lengths of LTRs, and it is not annotated and identified in public repeat databases such as Repbase (http://www.girinst.org/) and Plant Repeat databases (http://plantrepeats.plantbiology.msu.edu/). The LTRs were 400 bp long and terminated by the consensus sequence 5′-TG...CA-3′. The internal putative gag-pol coding region was 2,857 bp long and consisted of two ORFs flanked by two conserved sites, the PBS and the PPT site (Fig. 1A). The PBS consisted of the sequence 5′-AGT-GGTATCAAGGTAAGG-3′, which complemented the 3′ sequence of tRNA (Met) and activated the transcription of mRNA. The PPT, with a conserved motif of 5′-CCA-AGGTGGAGTTT-3′, was speculated to prime the synthesis of the second-strand DNA. The two LTRs for each RTPOSOn had accumulated a few mutations that were found in most RTPOSOn (Fig. 1B). Upon searching the sequence of RTPOSOn in the rice expressed sequence tag (EST) database, the results indicated that LTR and gag-pol regions shared high identity (85%–99%; E-value < 10^{-10}) with EST. Comparison of all published LTR elements revealed that RTPOSOn and rwnha303 (CJ610081) of Triticum aestivum showed high identity (86%) in 60 bp of LTR regions.

Two ORFs, ORF1 and ORF2, which encode 235 and 279 amino acids, respectively, were predicted from RTPOSOn (Fig. 1A). ORF1 shared low identity (30%) with an uncharacterized protein. ORF2 contains UBN2_2 (pfam14227) and zinc knuckle domain (pfam00098), conserved domains for the LTR polyprotein or retrotransposon of the copia type. In addition, the homology matrix of deduced amino acid sequences between RTPOSOn and the copia-like retroelement Tnt1 (GenBank accession no. X13777) of tobacco revealed 29% protein similarity between the gag-pol region of RTPOSOn and Tnt1 (Fig. 1C). When combined and aligned with two flanking sequences, a typical TSD with the 5-bp short direct repeat was found in 1 kb for each site of RTPOSOn elements.

Several genomes of the genus Oryza showed both intact RTPOSOns and solo LTRs. The O. sativa l. ssp. japonica cv. Nipponbare genome exhibited 93 RTPOSOns and fragments, including 13 intact elements and 41 solo LTRs (Table 1). The genome of O. sativa l. ssp. indica cv. 93-11 showed 85 RTPOSOns and fragments, including 13 intact elements and 40 fragments of RTPOSOn and 32 solo LTRs. The ratio of solo LTRs to intact elements was similar in Nipponbare (2.7:1) and 93-11 (2.4:1). In addition, RTPOSOn was most commonly found in the other Oryza species, such as the AA genomes of O. nivara, O. glaberrima, O. barthii, O. meridionalis, and O. glumaepatula and the BB genome of O. punctata. Most of the intact RTPOSOn elements were identified in O. punctata. On the basis of screened whole-genome sequences from six Oryza species, namely, O. nivara, O. glaberrima, O. barthii, O. meridionalis, O. glumaepatula, and O. punctata, the highest and lowest
ratios of solo LTRs to intact elements were 8.2:1 and 1:1.8 in *O. meridionalis* and *O. punctata*, respectively (Table 1).

A search of ESTs in the GenBank database (http://www.ncbi.nlm.nih.gov/) revealed switchgrass (*P. virgatum*) and wheat (*T. aestivum*) with fragments of *RTPOSON* with significant similarity (E-value < $10^{-5}$). Meanwhile, nine and eight homologous fragments of *RTPOSON* were detected in the genomes of sorghum (*Sorghum bicolor*) and purple false brome (*Brachypodium distachyon*), respectively. No significant sequence matches were found in other genomes of *Arabidopsis* (*Arabidopsis thaliana*), soybean (*Glycine max*), barley (*Hordeum vulgare*), maize (*Zea mays*), papaya (*Carica papaya*), or grape (*V. vinifera*) after BLASTN searches. These results suggest that *RTPOSON* and its homologous elements are unique to some genomes of Poaceae and that most of the complete structures of *RTPOSON* reside in *Oryza* genomes.

To gain insight into the sequence diversity and evolutionary relationships of *RTPOSON* from different *Oryza* species, we used 127 intact elements identified from 10 *Oryza* species (Supplementary Table 1) to generate a phylogenetic tree. According to this phylogenetic tree and the gene structures of *RTPOSON*, the *RTPOSON* family could be classified into three different groups (Figs. 2A and B). The groups of *RTPOSON_sub1* and *RTPOSON_sub2* contained 24 (Supplementary Fig. 2) and 5 elements, respectively. The *RTPOSON_sub3* group contained most of the intact *RTPOSON* elements (93/127, 73.2%) (Supplementary Fig. 3) and 97% (32/33) of intact *RTPOSON* elements from *O. punctata* (Fig. 2A). *RTPOSON_sub1* group harbors structural features of LTRs and gag-pol polyprotein genes. For the other two groups, *RTPOSON_sub2* and *RTPOSON_sub3*, the LTRs and a part of gag-pol regions are homologous to those of *RTPOSON_sub1* by pair-compare using bl2 seq program (E-value < $10^{-10}$, Fig. 2B). *RTPOSON_sub1* (RTPOSON No. 4, Supplementary Table 1) and *RTPOSON_sub2* (RTPOSON No. 274, Supplementary Table 1) are closely related because they have similar DNA sizes and also share substantial sequence similarity, 81%–89%, in LTR regions and internal gag-pol
polyproteins. In contrast, \textit{RTPOSON\_sub3} (\textit{RTPOSON} No. 316, Supplementary Table 1) has a small size and shares relatively lower sequence similarity with \textit{RTPOSON\_sub1} and \textit{RTPOSON\_sub2} (Fig. 2B).

**Insertion of \textit{RTPOSONs} in \textit{Oryza} genomes.** The insertion times of LTR elements could be estimated by sequence divergence between the two LTRs because the left and right LTR diverged independently by accumulated mutations. The insertion times of 127 intact \textit{RTPOSONs} from \textit{Oryza} species were estimated by the average substitution rate ($r$) of $1.3 \times 10^{-8}$ substitutions/synonymous site/year.\textsuperscript{40} Many (53%) \textit{RTPOSON} elements were inserted into the genome from 1.5 to 3.5 Mya, and 32% (41/127) \textit{RTPOSON} elements were inserted from 3.5 to 6.0 Mya (Fig. 3A and Supplementary Table 1). Nevertheless, only a small proportion (6%) of the \textit{RTPOSON} elements could be dated as having occurred more recently than 1.5 Mya (6%) or longer than 6 Mya (6%). The greatest number of \textit{RTPOSON} elements, 27 elements (21%), was inserted during the time frame of 1.5–2 Mya (Fig. 3A). A low proportion (1%) of the \textit{RTPOSON} elements were inserted earlier than 8 Mya; thus, \textit{RTPOSON} might have had various periods of proliferation within the last 8 Mya, and it might still be active, or it might have had a relatively short period of proliferation from 1.5 to 3.5 Mya.

The insertion times of 15 intact \textit{RTPOSON} elements in Nipponbare ranged from 0 to 3.4 Mya; in contrast, the insertion times of 33 intact elements in \textit{O. punctata} ranged from 3.4 to 6.2 Mya. The \textit{RTPOSON} elements in the \textit{japonica} genome may be younger than those in the other genomes of \textit{Oryza} genus (Supplementary Table 1). The \textit{RTPOSON} element located on chromosome 1 of Nipponbare was identical to the element that was integrated into the genome of 93-11, which was recently found to have been inserted 0.82 Mya. The two elements share high identity (98%) and also have identical TSDs (5’-ATCTT-3’). In addition, the \textit{RTPOSON} elements from \textit{O. barthii} and \textit{O. nivara} were also identified as the younger
elements whose divergence time was 0.49 Mya, suggesting that the RTPOSON may have been activated recently in the genomes. According to the insertion times of the complete RTPOSON elements, significant two bursting periods seem to have occurred in the five Oryza species of the AA genome, namely, O. nivara, O. glaberrima, O. barthii, O. meridionalis, and O. glumaepatula, ∼3–5.0 Mya, whereas amplification in the O. punctata species of the BB genome occurred ∼3–5 Mya (Fig. 3B). These results provide an intriguing picture suggesting that the youngest elements are in japonica group and the oldest elements are in the O. punctata group, and thus, RTPOSONs in the japonica genome may be younger than those in the other genomes of Oryza genus (Fig. 3B).

To compare the orthologous insertions of RTPOSON in Nipponbare and 93-11, 31 unique insertions in Nipponbare and 6 insertions in 93-11 were identified. According to their characteristics, the insertions of RTPOSON could be divided into five categories: (1) different types of TE insertions (Figs. 4A–F); (2) insertions of additional copies of RTPOSON (Fig. 4G); (3) TEs or functional genes in RTPOSON (Figs. 4H–K); (4) truncated RTPOSON structure insertions (Fig. 4L); and (5) unique insertions of RTPOSON (Fig. 4M).

Table 2. Insertion sites of intact RTPOSON in japonica rice genome.

| INSERTION SITE       | NO. OF ELEMENT |
|----------------------|----------------|
| Intergenic region    | 64             |
| Gene                 | 11             |
| intron               | 9              |
| exon                 | 0              |
| 5′- or 3′-UTR        | 2              |
| Putative gene        | 9              |
| Within 1 kb flanking gene | 24         |
| Within other types of TEs | 19         |
| Total                | 127            |
been truncated or inserted independently into one of them (Figs. 4L, M). These findings provide evidence that a part of intact RTPOSON elements were activated and evolved independently after the divergence of indica and japonica.

**Effect of RTPOSON insertion on gene structure.** A total of 127 intact elements had potential retrotransposition activity because of the two same lengths of LTRs and gag-pol polyproteins. Many RTPOSONs (50.3%; 64/127) were found...
in intergenic regions (RTPOSON located between genes); the others (49.7%; 63/127) were integrated in the genomic region or flanked by TE s (Table 2). Different types of TE s identified as having been inserted by RTPOSON included Ty1-copia, Ty-gypsy, LINE, En-Spm, Mutator, and Pong. Eleven RTPOSON elements were transposed in various gene regions: nine RTPOSON elements in introns and two RTPOSON elements in untranslated regions (UTRs). However, none were transposed in exons. In addition, 24 (19%) RTPOSON elements were identified within 1 kb upstream or downstream of annotated genes. Overall, 27.6% of RTPOSONs in japonica genome were located within or near genes (RTPOSON inserted within 1 kb of genes).

The insertions of RTPOSON occurred in orthologous genes of nine Oryza species in various ways (Fig. 5). An RTPOSON element was found in the fourth intron of the STE kinase (LOC_Os02g44642, AK073040) of six Oryza species. Different structural components of RTPOSON were retained in the STE kinase of these Oryza species, such as the intact elements in japonica and indica, solo LTRs in O. glumaepatula and O. barthii, two LTRs in O. nivara, and gag-pol in O. glaberrima (Fig. 5A). The fragments of RTPOSON were distributed in the third intron of the disease-resistant gene RGH2B (LOC_Os10g22484, AK100740), which were identified in japonica, indica, and O. barthii (Fig. 5B). However, some RTPOSON elements were found in a specific species but not in others. For example, intact RTPOSON elements were found in the disease-resistant RPP13-like protein (LOC_Os10g04090, AK120449) of indica and in the glycosyl transferase (LOC_Os11g03160, AK105411) of O. meridionalis (Figs. 5C, D).

The retrotransposons embedded in genes would alter gene structures and gene functions, thus providing opportunities to promote gene evolution. A total of 47 LTRs were identified in the UTRs, exons, or introns of genes with significant similarity (E-value < 10^-5). Most LTRs were found in introns or UTRs and also in exons, such as the hypothetical conserved gene and EARLY flowering protein (LOC_Os08g22484, AK100740), which were identified in japonica, indica, and O. barthii (Fig. 5B). However, some RTPOSON elements were found in a specific species but not in others. For example, intact RTPOSON elements were found in the disease-resistant RPP13-like protein (LOC_Os10g04090, AK120449) of indica and in the glycosyl transferase (LOC_Os11g03160, AK105411) of O. meridionalis (Figs. 5C, D).

The retrotransposons embedded in genes would alter gene structures and gene functions, thus providing opportunities to promote gene evolution. A total of 47 LTRs were identified in the UTRs, exons, or introns of genes with significant similarity (E-value < 10^-5). Most LTRs were found in introns or UTRs and also in exons, such as the hypothetical conserved gene and EARLY flowering protein (LOC_Os08g22484, AK100740), which were identified in japonica, indica, and O. barthii (Fig. 5B). However, some RTPOSON elements were found in a specific species but not in others. For example, intact RTPOSON elements were found in the disease-resistant RPP13-like protein (LOC_Os10g04090, AK120449) of indica and in the glycosyl transferase (LOC_Os11g03160, AK105411) of O. meridionalis (Figs. 5C, D).

The LTR of RTPOSON is homologous to the 5′ UTR (89% identity) and also provides the partial sequence to exon1 in the EARLY flowering protein of seven Oryza species but not in O. punctata or O. brachyantha (Fig. 5E). Strikingly, the LTR of RTPOSON shared high identity (95%) with the 3′ UTR of dehydrogenase (LOC_Os10g41170, AK110882) in all nine analyzed species, even with variable sizes up to 302 bp. Furthermore, when the gag-pol of RTPOSON was used as a query sequence to identify the homologous fragments of genes, six homologous fragments of gag-pol were identified in the annotated genes. Except for both the MAP3Ka (LOC_Os02g44642, AK073040) and phosphoglucomutase precursor (LOC_Os06g28194, AK068502), found inserted with the intact RTPOSON, the others were found to contain gag-pol in intron regions (Supplementary Table 3).

**Discussion**

A novel retrotransposon conserved across the genus Oryza and other grass families. We uncovered a novel LTR retrotransposon, RTPOSON, of about 3 kb in length, that contains TG…CA at two LTRs (Fig. 1). A total of 705 RTPOSON elements were identified; 127 (~18%) RTPOSON elements were retained as intact structures of LTR retrotransposons in the species of Oryza genus, including O. nivara, O. rufipogon, O. glaberrima, O. barthii, O. meridionalis, O. glumaepatula, O. punctata, O. minuta, and O. officinalis (Table 1). An RTPOSON contains LTRs, conserved PBS and PPT, and two ORFs encoding the uncharacterized protein, UBN2_2, and zinc knuckle functional domains (Fig. 1). RTPOSON is a nonautonomous retroelement because it contains an incomplete pol, lacking components such as INT, RT, and RH. One explanation for why RTPOSON can transpose in rice genomes is that autonomous elements might provide the necessary proteins for nonautonomous elements to spread throughout the genomes, an explanation similar to that previously given for Daehong/RIRE2, CRR2/noaCRR2, Spil/RIRE3, and Squis/RER18.3,32,43

LTR retrotransposons in plants are characterized by LTRs that vary from a few hundred base pairs to several kilobases long. However, LTR retrotransposons are highly unstable in the Arabidopsis genome. Unequal homologous recombinations and illegitimate recombinations between the LTRs of the same elements or remote elements were considered as the prevalent mechanism for LTR-retrotransposon elimination in plants.20,21,44 In this study, the ratios of solo LTRs to intact elements in seven Oryza species ranged from 1.3:1 (O. barthii) to 8.2:1 (O. meridionalis; Table 1); hence, the RTPOSONs may lose their structures during different evolutionary times in each species. In the seven Oryza AA genome rice species, five of the Oryza rice species, namely, O. sativa spp. japonica (2.7:1), indica (2.4:1), O. nivara (1.9:1), O. glaberrima (2.0:1), and O. barthii (1.4:1), have similar ratios, but higher ratios were found in O. glumaepatula (4.1:1) and O. meridionalis (8.2:1). In addition, highly abundant levels of solo LTRs to intact elements were observed in O. minuta of the BBCC genome and O. officinalis of the CC genome, but not in O. punctata of the BB genome. Of note, much higher levels of intact elements than solo LTRs were found in the BB genome (Table 1). Relatively high truncated fragments from RTPOSON were identified in japonica, indica, and O. glumaepatula. Previous studies have indicated that most of the LTRs from 52 LTR-retrotransposon families had been rapidly deleted in cultivated rice genomes in the last 8–5 Mya.21,44 Of note, LTR sizes have been suggested to have an impact on solo LTR formation; meanwhile, fragments and solo LTRs were the most abundant elements uncovered (Table 1).
In addition to being identified in *Oryza* genomes, *RTPOSON* was identified in other species of grass family, such as switchgrass and wheat. Moreover, *RTPOSON* shared 86% sequence identity in 69 bp of LTRs between rice and wheat. In addition, the 221 bp of LTR region of *RTPOSON* was also identified in switchgrass and shared 69% identity. **RIRE1** was transferred in many species of the genus *Oryza* during a short evolutionary period.\(^1\) In plants, 65% of angiosperm species (26/40) with TEs have been found to harbor at least one case of horizontal transfer.\(^6\) *Route66*, an LTR retrotransposon, is highly conserved among rice, maize, and sorghum genomes; specifically, *Route66* has been transferred between Panicoideae (*Z. mays*, *S. bicolor*, *Saccharum robustum*) and several species of the genus *Oryza*.\(^7\) The multiple retrotransposons were homologous between Panicoideae and *Oryza* species, which might have been mediated by three main mechanisms, namely, plant-to-plant transfers, interspecific hybridization, and vector transfers.\(^45\),\(^47\)

**The impact of *RTPOSONs* on the *Oryza* genome.**LTR retrotransposons are abundant and highly variable in plant genomes and have played important roles in evolution, such as genome expansion, genome shaping, sequence duplication, recombination, and mutation.\(^6\) The accumulated mutations of LTRs occurred independently, which allowed for estimating the insertion times of retroelements and the divergence times of related species from common ancestors.\(^3\) In terms of the 127 intact *RTPOSONs* identified in this study, the estimated divergence time of ssp. *japonica* Nipponbare and ssp. *indica* 93-11 was 0.82 Mya according to the evolutionary rate of 1.3 × 10\(^{-9}\) substitutions/synonymous site/year. Similarly, the divergence time was 0.9–2.1 Mya for Nipponbare and 93-11, which was estimated by LTR sequence divergence of four retrotransposon families, namely, *bopi*, *bouba*, *Retrosat1*, and **RIRE8**.\(^8,18\)

Retrotransposon elements can proliferate within a relatively short evolutionary time if a particular family becomes active in the genomes. If so, such elements are amplified by thousands of copies, for a rapid genome expansion. Analysis of 41 LTR retrotransposons showed that elements were amplified within the past 5 Mya.\(^44\) Overall, 60% of the *O. australiensis* genome was amplified by >90,000 retrotransposon copies accumulated in the genome during the last three million years.\(^19\) In this study, 127 intact elements were classified into three subgroups. Most of the intact *RTPOSON* elements and intact *RTPOSON* elements from *O. punctata* were classified in *RTPOSON_sub3* group (Fig. 2), which suggests that numerous elements had smaller structures and retained conserved **gag** regions in the genome. Moreover, LTR retrotransposons might be silent in genomes during long-term evolution because *RTPOSON* elements of young evolutionary age were not found in the genome of *O. punctata*. The LTR retrotransposons that have proliferated through host defense are young and contain mostly full-length copies, whereas older elements are highly truncated and deleted with a half-life of <6 Mya, which suggests that LTR retrotransposons are born through transposition and die via random mutation and eventual deletion from genomes.\(^21,40\) In addition, >90% of *RTPOSON* elements accumulated in *Oryza* genomes during the last eight million years (Fig. 3). Amplifications were evident during the last 3.5 Mya in the AA genomes of *Oryza* species such as *O. nivara*, *O. glaberrima*, *O. barthii*, *O. meridionalis*, and *O. glumaepatula*, in contrast to the BB genome species *O. punctata*, in which amplification was found to have occurred earlier, during the last 3–5 Mya. In the grass genome, four groups of complete small LTR retrotransposons (SMARTs) were identified in *O. brachyantha* and had different insertion times, from 0 to 36.9 Mya, which suggests that SMARTs contain an ancient family and were activated recently. In the GG genome of *O. granulata*, nearly 25% of the genome size was expanded by amplification of *Gran3* subsequent to speciation.\(^8\) The divergence time between AA and BB genomes was calculated to have occurred ~3.8 Mya,\(^39\) and thus the genome of *O. punctata* might have undergone amplification of *RTPOSON* subsequent to speciation.

**The position effect of *RTPOSON in genes.*** A study of transposon insertion polymorphisms (TIPs) in *O. sativa* sub-species identified 14% of TIPs of the genomic DNA sequence differences between subpecies of *indica* and *japonica*, and >10% of TIPs were located in expressed gene regions and affected gene expression between the two subspecies.\(^51\) A total of 37 insertions of *RTPOSON* were identified in Nipponbare and 93-11 (Fig. 4); 31 and 6 intact *RTPOSONs* were found in Nipponbare and 93-11, respectively. Especially, the *RTPOSON* was only inserted in the 5′ UTR of disease-resistant RPP13-like protein 1 (LOC_Os10g04090) of 93-11, and hence this insertion may be responsible for the difference between *indica* and *japonica* (Fig. 5C). Two genes, *OsWRKY8* and *AtPUP11*, were found to have TEs inserted in the 3′ UTR and intron, respectively, resulting in TE insertion in the gene, which is responsible for the difference between Nipponbare and 93-11.\(^52\) The TEs in genes can change the splicing sites and also contributed part of the sequence in the transcripts. In addition, TEs located near critical genes can reduce gene expression by epigenetic silencing or antisense transcription.\(^52,53\) These genetic differences provide a number of cases exemplifying a wide spectrum of changes induced by retrotransposon insertions than might alter the gene regulation and expression of the host genes.

TEs can directly influence the regulation of expression of nearby genes at the transcriptional and posttranscriptional levels in plant genomes as well.\(^31,54\) The various inserted positions have resulted in various impacts on gene expression or gene silencing. In the rice genome, one-sixth of rice genes were reported to be associated with retrotransposons inserted in genes or promoter regions.\(^32\) In wheat, retrotransposons are reported to have altered the expression of adjacent genes, whereas read-through transcription from retrotransposons was found to be associated with the activation or silencing of flanking genes.\(^52\)
In fact, retrotransposons in genes also promote the binding of specific small RNAs (sRNAs) (transcriptional silencing) and result in transcripts for posttranscriptional silencing. In the rice genome, >400 sRNAs were found to be perfectly matched with the nonautonomous SMARTs; thus, sRNAs might be involved in the silencing of SMARTs and in the regulation of genes, in which SMARTs reside, or the nearby genes. More than 27% of RTPOSON elements were inserted within or near genes (Table 2). A total of 47 genes contained LTR homologous fragments in the UTR, exon, or intron of the gene. Some of the RTPOSON elements appear to be located in the 5′ UTR or intron of genes, such as the disease-resistant RPP13-like protein 1, STE kinase, and glycosyl transferase (Fig. 5). An analysis of one gene, EARLY flowering protein, showed that RTPOSON contributed the LTR sequence to the 5′ UTR and also provided an exon of gene splicing into the mRNA (Fig. 5E). In some cases, genes included the LTR retrotransposon as a part in the cDNA, such as genes for the hexose carrier protein, nonspecific lipid-transfer protein, transporter, GO9, Mov34 family protein, and aspartyl PR. By contributing alternative exons into the genes, retrotransposons may lead to the evolution of new gene functions.

Conclusion
Abundant genomic resources provide insights into the structural variations in the different genomes of the genus Oryza and help in better comprehending genome evolution. A novel retrotransposon RTPOSON with LTR, part of gene structures homologous with Tht1, was identified and classified as a type of Ty1-copia retrotransposons in the Oryza genome. RTPOSON retrotransposons have undergone amplification over the past 8 Mya in Oryza genomes. Three groups were classified from 127 RTPOSON elements that displayed different DNA sizes. In particular, the O. sativa spp. indica and ssp. japonica sub-species diverged from a common ancestor ~0.82 Mya, and a few RTPOSON elements were transposed independently after their divergence. Although RTPOSON is a nonautonomous element, it contributed to gene structures that differed among Oryza genomes; therefore, RTPOSON may have important functions in genome dynamics and species evolution.

Acknowledgments
The authors are thankful to Laura Smales for reviewing this manuscript.

Author Contributions
Conceived and designed the experiments: Y-CH, C-SW, Y-RL, Y-PW. Analyzed the data: Y-CH, C-SW, Y-RL, Y-PW. Wrote the paper: Y-CH, C-SW, Y-RL, Y-PW. All authors reviewed and approved of the final manuscript.

Supplementary Materials
Supplementary Figure 1. Workflow for identification of RTPOSON in the Oryza genome.

Diamonds represent steps involving the usage of bioinformatics tools. Solid arrows represent the general flow of information. Black arrows indicate the alternate flow of information.

Supplementary Figure 2. Phylogenetic tree from RTPOSON_sub1 of RTPOSON retrotransposons. The neighbor-joining tree was constructed based on gag sequences. The gag sequences were aligned by using MUSCLE, and then the tree was constructed by using MEGA version 6.0. Bootstrap values were calculated for 1,000 replicates. The scale bar indicates nucleotide sequence divergence.

Supplementary Figure 3. Phylogenetic tree from RTPOSON_sub3 of RTPOSON retrotransposons. The neighbor-joining tree was constructed based on gag sequences. The gag sequences were aligned by using MUSCLE, and then the tree was constructed by using MEGA version 6.0. Bootstrap values were calculated for 1,000 replicates. The scale bar indicates nucleotide sequence divergence.

Supplementary Table 1. Features of the RTPOSON elements in the Oryza genome.

Supplementary Table 2. LTR of RTPOSON homologous gene in Nipponbare genome.

Supplementary Table 3. gag-pol of RTPOSON homologous gene in Nipponbare genome.

REFERENCES
1. Rossi M, Araujo PG, Van Sluys MA. Survey of transposable elements in sugarcane expressed sequence tags (ESTs). Genet Mol Biol. 2001;24:1–4.
2. Jhang N, Feschotte C, Zhang X, et al. Using rice to understand the origin and evolution of an ancient retroposon as a part in the cDNA, such as genes for the hexose transporter, nonspecific lipid-transfer protein, transporter, GO9, Mov34 family protein, and aspartyl PR. 32 By contributing alternative exons into the genes, retrotransposons may lead to the evolution of new gene functions.

2. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flower plant Arabidopsis thaliana. Nature. 2000;408:796–815.
3. Ammiraju JS, Zuccolo A, Yu Y, et al. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. Plant J. 2007;52:342–51.
4. Jiang Y, Deng XW. A genome-wide transcriptional activity survey of rice transposable element-related genes. Genome Biol. 2007;8:R28.
5. Vitte C, Panaud O. Formation of solo-LTRs through unequal homologous recombination. Curr Opin Plant Biol. 2004;7:155–9.
6. Yim H, Liu J, Xu Y, et al. TaDEL1, a mutated Copia-like LTR retrotransposon with recent massive amplification in mustard. PLoS One. 2013;8:e68587.
7. Wong LH, Choo KH. Evolutionary dynamics of transposable elements at the centromere. Trends Genet. 2004;20:611–6.
8. Forde-Browne E, Drouaud J, Lucas H. Retrovirus-like end processing of the tobacco Tnt1 retrotransposon linear intermediates of replication. J Virol. 1997;71:4005–15.
9. Kumar A, Bennetzen JL. Plant retrotransposons. Annu Rev Genet. 1999;33:479–532.
10. Jiang N, Bao Z, Zhang X, et al. Pack-MULE transposable elements mediate gene evolution in plants. Nature. 2004;431:569–73.
11. Ammiraju JS, Zuccolo A, Yu Y, et al. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. Plant J. 2007;52:342–51.
12. Jiang Y, Deng XW. A genome-wide transcriptional activity survey of rice transposable element-related genes. Genome Biol. 2007;8:R28.
13. Vitte C, Panaud O. Formation of solo-LTRs through unequal homologous recombination. Curr Opin Plant Biol. 2004;7:155–9.
14. Yim H, Liu J, Xu Y, et al. TaDEL1, a mutated Copia-like LTR retrotransposon with recent massive amplification in mustard. PLoS One. 2013;8:e68587.
15. Wong LH, Choo KH. Evolutionary dynamics of transposable elements at the centromere. Trends Genet. 2004;20:611–6.
16. Forde-Browne E, Drouaud J, Lucas H. Retrovirus-like end processing of the tobacco Tnt1 retrotransposon linear intermediates of replication. J Virol. 1997;71:4005–15.
17. Kumar A, Bennetzen JL. Plant retrotransposons. Annu Rev Genet. 1999;33:479–532.
18. Jiang N, Bao Z, Zhang X, et al. Pack-MULE transposable elements mediate gene evolution in plants. Nature. 2004;431:569–73.
19. Ammiraju JS, Zuccolo A, Yu Y, et al. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. Plant J. 2007;52:342–51.
20. Jiang Y, Deng XW. A genome-wide transcriptional activity survey of rice transposable element-related genes. Genome Biol. 2007;8:R28.
21. Vitte C, Panaud O. Formation of solo-LTRs through unequal homologous recombination. Curr Opin Plant Biol. 2004;7:155–9.
22. Yim H, Liu J, Xu Y, et al. TaDEL1, a mutated Copia-like LTR retrotransposon with recent massive amplification in mustard. PLoS One. 2013;8:e68587.
23. Wong LH, Choo KH. Evolutionary dynamics of transposable elements at the centromere. Trends Genet. 2004;20:611–6.
24. Forde-Browne E, Drouaud J, Lucas H. Retrovirus-like end processing of the tobacco Tnt1 retrotransposon linear intermediates of replication. J Virol. 1997;71:4005–15.
25. Kumar A, Bennetzen JL. Plant retrotransposons. Annu Rev Genet. 1999;33:479–532.
26. Jiang N, Bao Z, Zhang X, et al. Pack-MULE transposable elements mediate gene evolution in plants. Nature. 2004;431:569–73.
27. Ammiraju JS, Zuccolo A, Yu Y, et al. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. Plant J. 2007;52:342–51.
