Dasheng and *RIRE2*. A Nonautonomous Long Terminal Repeat Element and Its Putative Autonomous Partner in the Rice Genome[1]

Ning Jiang, I. King Jordan, and Susan R. Wessler*

Department of Plant Biology, University of Georgia, Athens, Georgia 30602 (N.J., S.R.W.); and National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894 (I.K.J.)

Dasheng is one of the highest copy number long terminal repeat elements and one of the most recent elements to amplify in the rice (*Oryza sativa*) genome. However, the absence of any significant coding capacity for retroviral proteins, including *gag* and *pol*, suggests that *Dasheng* is a nonautonomous element. Here, we have exploited the availability of 360 Mb of rice genomic sequence to identify a candidate autonomous element. *RIRE2* is a previously described *gypsy*-like long terminal repeat retrotransposon with significant sequence similarity to *Dasheng* in the regions where putative cis factors for retrotransposition are thought to be located. *Dasheng* and *RIRE2* elements have similar chromosomal distribution patterns and similar target site sequences, suggesting that they use the same transposition machinery. In addition, the presence of several *RIRE2*-*Dasheng* element chimeras in the genome is consistent with the copackaging of element mRNAs in the same virus-like particle. Finally, both families have recently amplified members, suggesting that they could have been co-expressed, a necessary prerequisite for *RIRE2* to serve as the source of transposition machinery for *Dasheng*. Consistent with this hypothesis, transcripts from both elements were found in the same expressed sequence tag library.

Transposable elements fall into two classes based on their transposition mechanisms. DNA or class II elements are characterized by short terminal inverted repeats and transposition via a DNA intermediate (Kunze et al., 1997). Because of their conservative mechanism of transposition, the copy number of DNA element families is usually less than 100 per haploid genome. In contrast, RNA or class I elements are capable of attaining very high copy numbers in a relatively short period of time because the element-encoded mRNA, and not the element itself, forms the transposition intermediate. Based on their structural features, class I elements are further divided into two subclasses. Non-long terminal repeat (LTR) elements, including long interspersed nuclear elements and short interspersed nuclear elements, are the most abundant class I elements in mammalian genomes (Smit, 1996; Lander et al., 2001). LTR retrotransposons, on the other hand, are the most abundant elements in plants and compose a significant fraction of the genome (Kumar and Bennetzen, 1999). In the grass clade, the differential amplification of LTR retrotransposon has been shown to be largely responsible for the significant difference in genome size among members (Chen et al., 1997; Dubcovsky et al., 2001).

LTR retrotransposons are either *copia* like or *gypsy* like, depending on the order of retroviral domains encoded by the *gag* and *pol* genes (Xiong and Eickbush, 1990). The products of *gag* comprise the major structural proteins of the virus-like particle (VLP), which is involved in maturation and packaging of element RNA and other proteins. As is typical for many retroelements, translation of element-encoded RNA generates a long poly-protein precursor, which is specifically cleaved by a *pol*-encoded protease, releasing mature proteins. In addition to the protease, the *pol* gene also encodes reverse transcriptase (RT), RNase H, and integrase, proteins involved in the synthesis and integration of the element DNA into the host genome. In addition to these coding regions, cis sequences within the element are also necessary for transposition. The LTR usually contains initiation and termination sites for transcription. Immediately internal to the LTR is the primer binding site (PBS) and the polypurine tract (PPT; Fig. 1), which are necessary for the initiation of the synthesis of element DNA from the RNA intermediate (Boeke and Corces, 1989).

Although the vast majority of previously reported LTR retrotransposons are *copia* or *gypsy* like, an increasing number of elements lacking any significant retroviral domains cannot be classified in this way. These elements are considered to be nonautonomous because they require the products from another element in trans to amplify in the genome. One such element is *MaLR*, an ancient element family with the highest copy (approximately 240,000 copies) among LTR elements in the human genome (Smit, 1993;

[1] This study was supported by the U.S. Department of Energy (grant to S.R.W.) and by the National Science Foundation (grant to S.R.W.).

* Corresponding author; e-mail sue@dogwood.botany.uga.edu; fax 706–542–1805.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.015412.

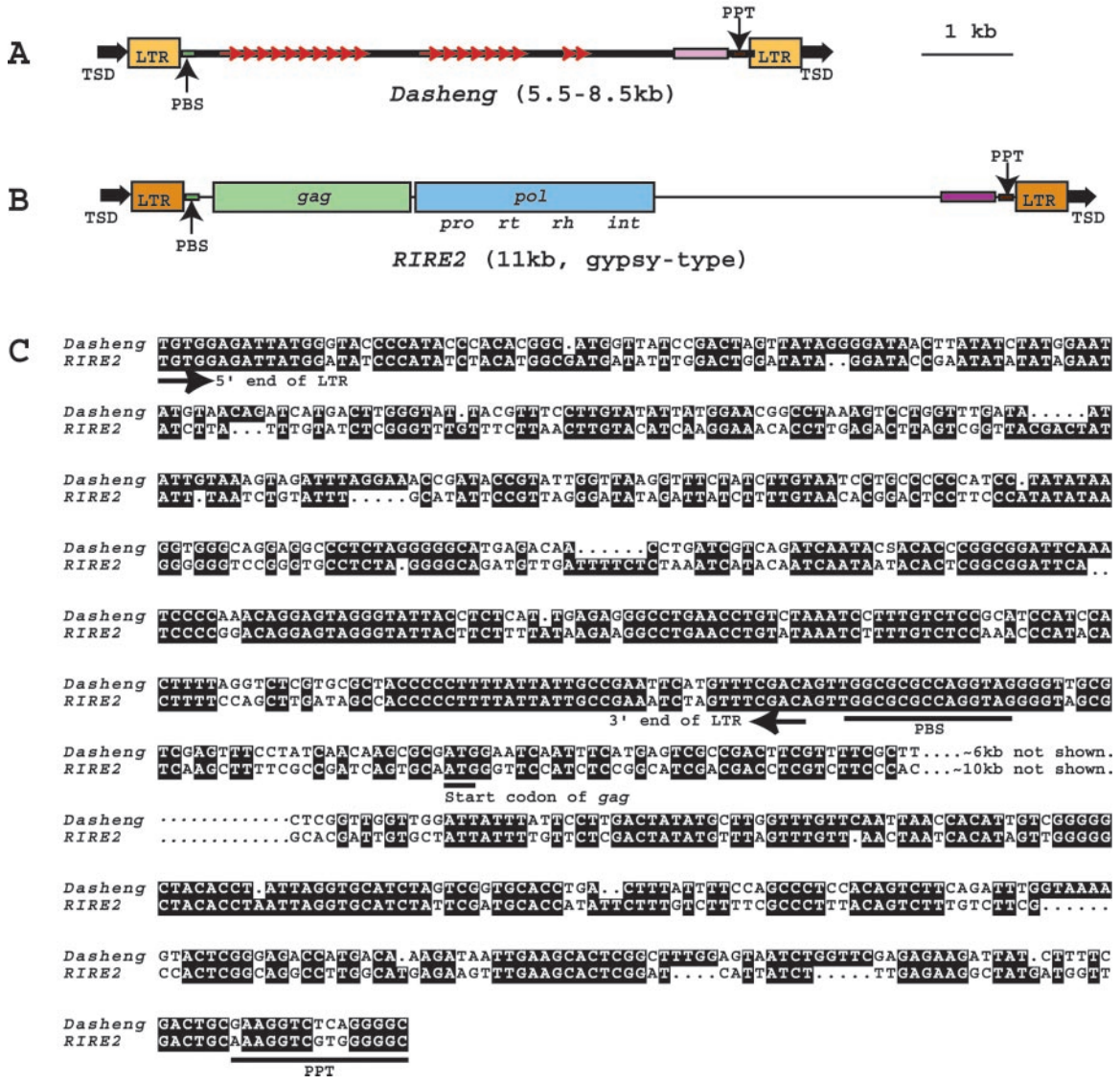

Figure 1. Comparison of *Dasheng* and *RIRE2*. LTRs are shown as yellow or orange boxes. Open reading frames (ORFs) in *RIRE2* are shown as green and blue boxes. Regions of sequence similarity (between *Dasheng* and *RIRE2*) upstream of the PPT are shown as purple boxes. pro, Protease; rt, RT; rh, RNase H; int, integrase. A, Typical *Dasheng* element. Red arrows represent the 89- to 90-bp tandem repeat. B, Typical *RIRE2* element. C, Sequence comparison of the related regions of *Dasheng* and *RIRE2*. The 3′ LTR and the internal sequence that is not related are not shown. The positions of the PBS, PPT, and the start codon of *gag* in *RIRE2* are indicated. Comparisons are based on consensus sequences (see "Materials and Methods").

Lander et al., 2001). *Bs1*, a maize (*Zea mays*) LTR element with one to five copies, was first detected as a new insertion into *adh1* (Johns et al., 1985; Jin and Bennetzen, 1989). *Zeon-1* is another maize nonautonomous LTR element, with a copy number of up to 32,000 (Hu et al., 1995; Meyers et al., 2001). Despite these and several other examples, virtually nothing is known about how nonautonomous LTR retroelements originate or the trans-encoded products necessary for their transposition.

As mentioned above, LTR retrotransposons make up a significant fraction of most plant genomes, and account for at least 15% of the relatively small rice (*Oryza sativa*) genome (Tarchini et al., 2000; Turcotte et al., 2001). The availability of a substantial amount of rice genomic sequence provides a unique opportunity to identify both nonautonomous LTR retrotransposons and their autonomous partners. A previous study reported the discovery of *Dasheng*, a high-copy number nonautonomous LTR element that amplified very recently in the rice genome (Jiang et al., 2002). In this study, 360 Mb of publicly available rice cv Nipponbare genomic sequence has been searched for candidate autonomous LTR retrotransposons that may have provided the proteins necessary for *Dasheng* to spread throughout the rice ge-

nome. To our knowledge, this study provides, for the first time, several lines of evidence implicating a distinct but related LTR retrotransposon family (*RIRE2*) as the source of the transposition machinery for a nonautonomous LTR element (*Dasheng*).

RESULTS AND DISCUSSION

Sequence Similarity between *RIRE2* and *Dasheng*

Candidate autonomous elements were selected by searching public databases of rice genomic sequence for LTR retrotransposon families that displayed significant sequence similarity to *Dasheng*. This search led to the identification of a single candidate, *RIRE2*, a previously described rice LTR retrotransposon (Ohtsubo et al., 1999; Fig. 1B). *RIRE2* is a *gypsy*-type element that, unlike *Dasheng*, encodes all of the protein products necessary for retrotransposition. The two elements have similar LTRs, PBS and PPT (Fig. 1C). In addition, sequences downstream of the PBS and upstream of the PPT displayed substantial sequence similarity.

The extent of similarity is striking given that LTRs, as noncoding DNA, usually evolve much more rapidly than other regions of a retrotransposon. Even within a single element family, LTRs can diverge by almost 50%, whereas families of closely related elements (as determined by RT sequence comparisons) often have LTRs with no detectable sequence similarity (Arkhipova et al., 1995; Jordan and McDonald, 1998). One explanation for the conservation of LTR sequences is that *RIRE2* and *Dasheng* must be coexpressed if *Dasheng* is to gain access to the *RIRE2*-encoded retrotransposition machinery. Because the transcriptional regulatory sequences reside in the LTR, it follows that selection would favor the conservation of LTR sequences in autonomous and nonautonomous partners. Sequences adjacent to both the 5′ and 3′ LTRs also show significant similarity, and this too may reflect the conservation of cis sequences required for retrotransposition. This is certainly true for the PPT and the PBS. In addition, cis-packaging signals, which are essential for VLP mRNA recognition, should also be conserved. Although not well studied in LTR retrotransposons, cis-packaging signals have been investigated in several retroviruses, including human immunodeficiency virus type 1 and murine leukemia virus. In both cases, these signals are located upstream or near the start codon of *gag* and interact directly with the Gag proteins (Adam and Miller, 1988; Clever et al., 1995). In addition, a 17-bp sequence upstream of the PPT was shown to be involved in murine leukemia virus packaging (Yu et al., 2000). Analogous regions appear to be conserved in *Dasheng* and *RIRE2* (Fig. 1), perhaps because they enable the mRNA of the two elements to be packaged by *RIRE2*-encoded proteins (see below).

Distribution of *RIRE2* and *Dasheng* on Rice Chromosomes

During retrotransposition, element cDNA, generated by reverse transcription, is integrated into host chromosomes through interactions with the element-encoded integrase. Given this mechanism, it should follow that if *RIRE2* integrase is recognizing *Dasheng* cDNA and directing its integration, the two elements should have similar targeting preference as reflected in chromosomal distribution patterns.

As noted previously, *Dasheng* has a striking distribution, with more than one-half the elements in the family clustered in pericentromeric heterochromatin (Jiang et al., 2002). Since the *Dasheng* study, there has been a tremendous increase in the amount of rice genomic DNA in public databases. For this reason, the distribution of both *Dasheng* and *RIRE2* in rice cv Nipponbare was carried out. The distribution of these elements in complete or nearly complete chromosomes (1, 2, 4, 6, 8, and 10) at the time of this analysis is shown in Figure 2. It is clear from these data that the distribution of both families is remarkably consistent in terms of clustering in pericentromeric regions.

Similarity of *Dasheng* and *RIRE2* Insertion Sites

Although *Dasheng* and *RIRE2* have similar distribution patterns, clustering in pericentromeric regions is a feature of other LTR retrotransposons. For example, another rice LTR retrotransposon, *RIRE8*, is clustered in pericentromeric regions (Nonomura and Kurata, 2001), as are most of the retrotransposons in Arabidopsis (Arabidopsis Genome Initiative, 2000). To test whether the distribution of *Dasheng* and *RIRE2* results from similarities in their targeting mechanisms or only reflects a general feature of rice LTR elements, the target sequences of the two elements were determined and compared.

Analysis of 240 *Dasheng* and 194 *RIRE2* insertion sites clearly demonstrates that neither element inserts randomly into the chromosome and, more importantly, their consensus insertion sites are almost identical (Tables I and II). Both *Dasheng* and *RIRE2* show an overall bias for AT-rich insertion sites (56% and 58% AT content, respectively), which is attributable to the strong bias for A or T in a few positions (see below). More significantly, of the 15 nucleotides including and surrounding the 5-bp target site that is duplicated upon insertion (designated T1–T5 in Tables I–III), 10 of 15 for *Dasheng* and 11 of 15 for RIRE2 were not random (tested by χ^2 test). For example, both *Dasheng* and *RIRE2* show a strong bias for A or T at positions −3 and +3, whereas a strong bias against T and a strong bias against A were observed at T1 and T5, respectively. For all positions examined, the two elements show a similar bias except for positions −5, T1, and T2. At these positions, *RIRE2* shows a slightly stronger bias for A or T, which could

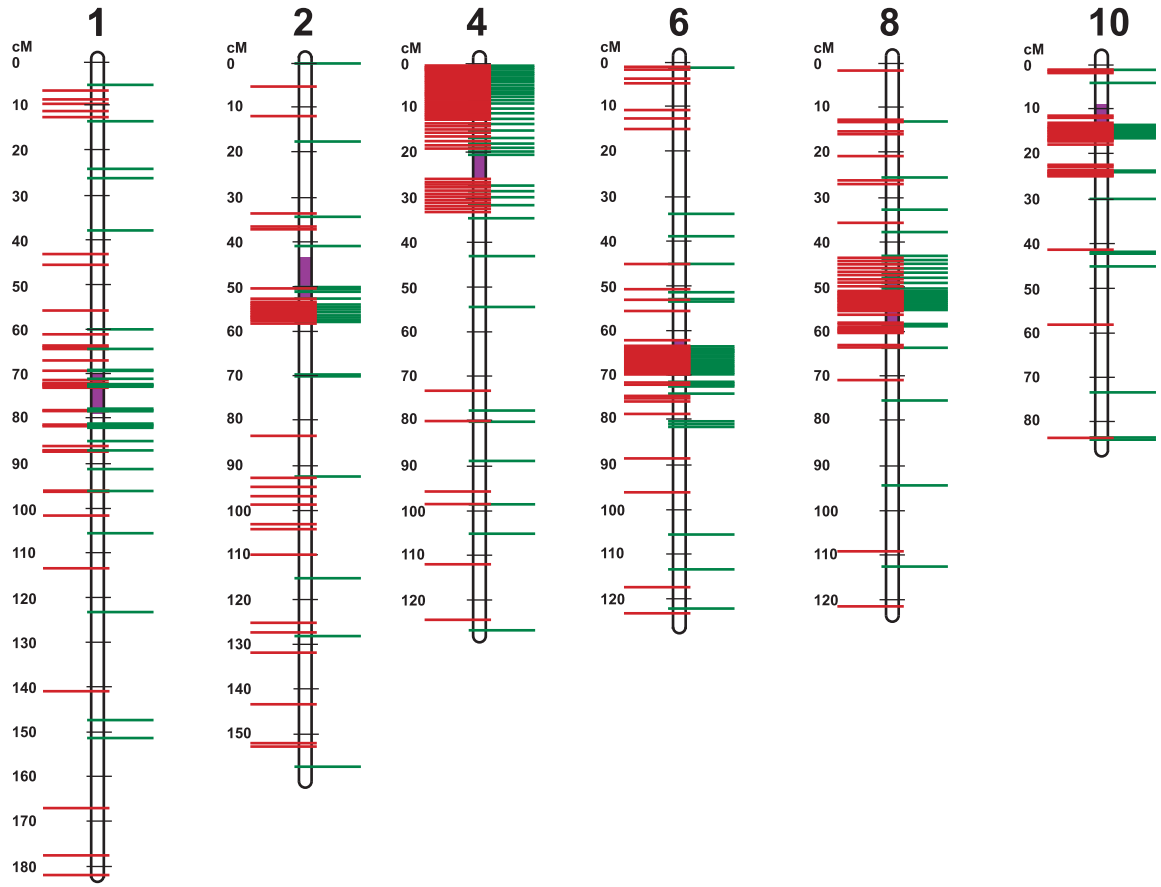

Figure 2. Chromosomal distribution of *Dasheng* and *RIRE2* elements. Only the finished or nearly finished chromosomes are shown. *Dasheng* and *RIRE2* insertions are in red and green, respectively. Each insertion is represented by a horizontal line. The position of each insertion was based on the position of the relevant bacterial artificial chromosome (BAC) or Pi-derived artifical chromosome (PAC) on the genetic map (see "Materials and Methods"), as shown by the black bars and numbers (cM) to the left of each chromosome. The position of all centromeres (in purple) is from Harushima et al. (1998), except the centromere of chromosome 10 (Cheng et al., 2001).

be explained by the fact that *RIRE2* insertions are generally older than *Dasheng* (see below). In this case, *RIRE2* elements with GC-rich TSDs are more likely to be excluded from the survey because of the frequent C to T transition, which leads to unmatched TSD (see "Materials and Methods").

As mentioned above, *RIRE8* is a previously described rice LTR retrotransposon (*gypsy* type) that localizes to pericentromeric and centromeric regions. However, unlike *RIRE2*, it is not a candidate autonomous element because it shares no significant sequence similarity with *Dasheng*. As such, although its chromosomal distribution is similar to *RIRE2* and *Dasheng*, it should display a different target site consensus (Table III). In contrast to *Dasheng* and *RIRE2*, the target sequences for *RIRE8* appear more random.

Table I. *The target sequences of 240 Dasheng elements*[a]

Position	−5	−4	−3	−2	−1	T1	T2	T3	T4	T5	+1	+2	+3	+4	+5	Total	%
A	72	58	95	70	59	80	63	77	78	17	59	63	82	57	46	976	27
T	64	49	98	79	53	22	72	81	54	71	64	78	119	68	70	1,042	29
C	41	67	14	44	55	83	57	45	60	66	63	55	25	53	66	794	22
G	63	66	33	47	73	55	48	37	48	86	54	44	14	62	58	788	22
For	–	–	AT	AT	–	AC	–	AT	A	G	–	AT	AT	–	–	AT	–
Against	C	–	GC	GC	–	T	–	GC	–	A	–	GC	GC	–	–	GC	–

[a] T1 to T5 represents the 5-bp target site duplication (TSD); −5 to −1 and + 1 to + 5 represent the 5-bp flanking TSD on each side. A bias is considered to be significant if a nucleotide in a certain position appears significantly more or less frequent than what expected in a random pattern (i.e. 25%) by χ^2 test ($P < 0.01$).

Table II. *The target sequences of 194 RIRE2 elements*[a]

Position	−5	−4	−3	−2	−1	T1	T2	T3	T4	T5	+1	+2	+3	+4	+5	Total	%
A	71	44	77	68	50	85	51	75	84	21	44	61	80	46	41	898	31
T	46	54	72	52	43	12	78	45	40	55	50	56	76	47	59	785	27
C	31	44	15	38	52	54	26	37	34	46	58	33	25	45	47	586	20
G	46	52	30	36	49	43	39	37	36	72	41	44	13	56	47	641	22
For	A	–	AT	AT	–	A	AT	AT	A	G	–	AT	AT	–	–	AT	–
Against	C	–	GC	GC	–	T	GC	GC	–	A	–	GC	GC	–	–	GC	–

[a] The indication of positions and the analysis are the same as in Table I (see footnote a).

No significant preference for AT- or GC-rich regions was found (49% versus 51%). Furthermore, significant bias was observed in only five of the 15 positions, and these sites differ from the *Dasheng*/*RIRE2* consensus.

Chimeric *Dasheng*/*RIRE2* Elements

The data thus far are consistent with a scenario whereby *RIRE2* and *Dasheng* mRNAs are reverse transcribed in the same VLP. The simultaneous presence of distinct but related mRNAs in VLPs has been shown to result in the formation of chimeric elements during reverse transcription, through a process called template switching (Boeke et al., 1986; Jordan and McDonald, 1999). Evidence for the simultaneous presence of *Dasheng* and *RIRE2* mRNAs in the same VLP might be preserved in the genome as chimeric elements (Fig. 3).

Analysis of 318 *RIRE2* elements revealed four apparent chimeras where a fragment of *Dasheng* is contiguous with *RIRE2* sequence (Fig. 3). It is unlikely that these chimeras are the result of genome rearrangements because, in each case, the *RIRE2* element appears to be normal except for the presence of *Dasheng* sequence. The TSDs that flank these elements indicated that they resulted from insertion instead of recombination (Fig. 3). Similarly, it is unlikely that these structures resulted from the insertion of *Dasheng* into *RIRE2* for several reasons. First, in all cases, only a fragment of *Dasheng* is present in the *RIRE2* element. There are other cases in the rice genome where a full-length *Dasheng* element has apparently inserted into *RIRE2*. For example, a *Dasheng* element, flanked by a perfect TSD, was found in *RIRE2* at 121,909 bp in BAC clone OSJNBa0016D02 (accession no. AP004731). Furthermore, these are the only examples where a fragment of *Dasheng* is found within another transposable element. That is, the association between *RIRE2* and *Dasheng* sequences appears to be unique, again suggesting a specific interaction between these two LTR element families.

Phylogenetic Relationship and Dating of *Dasheng* and *RIRE2* Insertions

Four Monophyletic Groups of Related Elements

Based on intraelement LTR identity of family members, it was previously reported that *Dasheng* is one of the most recent elements to amplify in the rice genome (Jiang et al., 2002). If *RIRE2* has provided the machinery necessary for *Dasheng* retrotransposition, it follows that they had to be co-expressed. For this to happen, at least some *RIRE2* elements must be as young as *Dasheng* elements. To estimate the age of the two element families, sequences homologous to *Dasheng* and *RIRE2* LTRs were retrieved and aligned, and the alignments used to reconstruct their evolutionary histories (see "Materials and Methods").

The resulting phylogeny has four distinct monophyletic groups (Fig. 4). One consists entirely of closely related *Dasheng* sequences, whereas another consists solely of *RIRE2* sequences. Both of these groups include solo LTR sequences as well as the LTRs of full-length elements. Between these two groups, and basal to the *Dasheng* group, is an intermediate group that has both an element with *Dasheng* internal sequence and an element with *RIRE2* internal sequence. Finally, there is a fourth monophyletic group of relatively distantly related solo LTRs with similarity to both *Dasheng* and *RIRE2* elements. Because this group consists only of solo LTR sequences (i.e. no family-specific internal sequences), it was not

Table III. *The target sequences of 120 RIRE8 elements*[a]

Position	−5	−4	−3	−2	−1	T1	T2	T3	T4	T5	+1	+2	+3	+4	+5	Total	%
A	31	20	34	19	24	34	24	40	32	28	18	28	37	22	18	409	22
T	23	23	53	28	16	28	52	26	39	28	34	19	37	24	41	471	26
C	34	33	9	39	20	19	25	29	18	35	40	49	27	36	29	442	25
G	32	44	24	34	60	39	19	25	31	29	28	24	19	38	32	478	27
For	–	G	T	–	G	–	T	–	–	–	–	C	–	–	–	–	–
Against	–	–	C	–	T	–	–	–	–	–	–	–	–	–	–	–	–

[a] The indication of positions and the analysis are the same as in Table I (see footnote a).

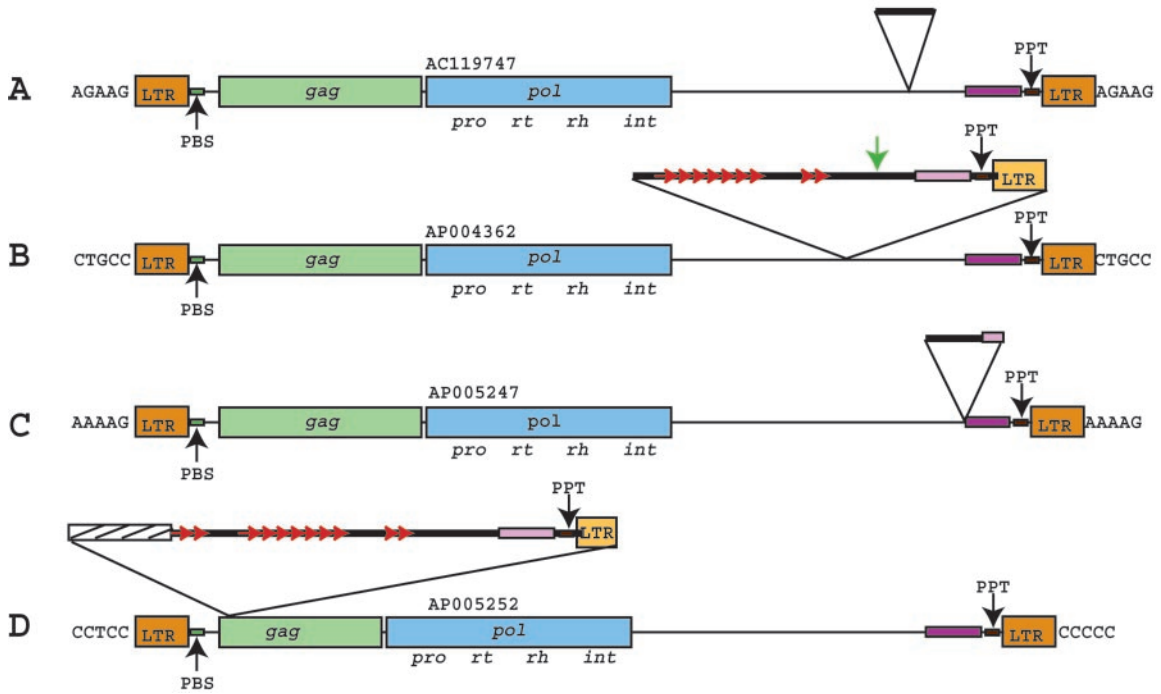

Figure 3. *RIRE2-Dasheng* chimeric elements. The TSD of each element is shown along with the accession number of the BAC or PAC that contains each element. To simplify the comparison, an insertion of an 11-kb LTR element in the *Dasheng* fragment in B is shown as a vertical arrow (in green). The hatched box in D indicates a fragment with unknown sequence identity; such a fragment is frequently located in the 5′ region of *Dasheng* (Jiang et al., 2002).

possible to definitively classify them as either *Dasheng* or *RIRE2*. However, this group is clearly the most divergent of the four, and its distance from the other groups indicates that it represents a related but unique family of elements that has become extinct since diverging from the *Dasheng-RIRE2* common ancestor.

Young RIRE2 Elements

Because the 5′ and 3′ LTRs of individual full-length retrotransposons are identical at the time of insertion, their divergence has been used to estimate when insertion occurred (SanMiguel et al., 1998; Bowen and McDonald, 2001). Based on an analysis of 4 times as many elements as the previous study, the *Dasheng* family still appears to be very young. Fifty-nine of 238 full-length *Dasheng* elements (25%) have identical LTRs and are estimated to have inserted within the last 150,000 years (see "Materials and Methods"). Although most *RIRE2* insertions are more ancient than full-length *Dasheng* elements (data not shown), 39 of 182 full-length *RIRE2* elements (21%) have identical LTRs. In addition, all 39 appear to have perfect ORFs encoding proteins for retrotransposition. Furthermore, several putative transcripts from *Dasheng* and *RIRE2* (LTRs and internal regions) were found in the expressed sequence tag database (http://www.ncbi.nih.gov/dbEST/index.html; the search was performed on September 27, 2002) in a variety of rice cDNA clones, suggesting that both elements are still transcribed. More importantly, the fact that both *Dasheng* and *RIRE2* hits were found in the same cDNA library (accession nos. AU078092 and AU078094) suggests that they were co-expressed.

Dating the Four Monophyletic Groups

The relative time of insertion of monophyletic groups of LTR sequences can also be determined from the level of sequence divergence. This approach relies on the estimation of a consensus sequence, which approximates the common ancestor of the group, and comparison of extant sequences in the group with this putative ancestral sequence. The average divergence between the group ancestor and all of its sequences is calibrated with the substitution rate to estimate the time elapsed since the elements of the group last shared a common ancestor (i.e. the age of the group). The *Dasheng* group, at 7.1 million years old, is by far the youngest group (Fig. 5). This value is consistent with the dating based on LTR comparisons of full-length elements and on the high level of *Dasheng* insertion site polymorphism between rice ssp. *indica* and ssp. *japonica* (up to 80%; Jiang et al., 2002). In addition, the fact that *Dasheng* is widespread in the *Oryza* genus suggests that this family originated earlier than or around the time of speciation, approximately 5 to 10 million years ago (Kellogg, 2001; Jiang et al., 2002). The next oldest groups are

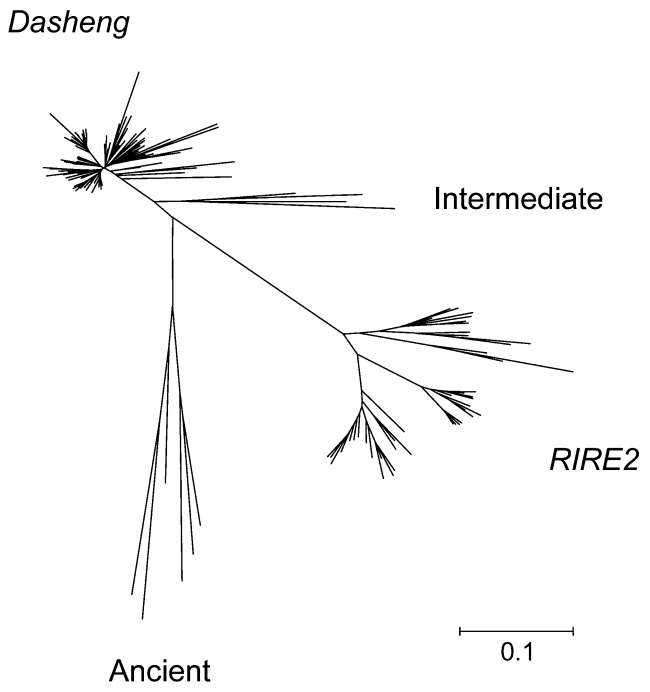

Figure 4. Phylogenetic tree based on the LTR sequence of *Dasheng* and *RIRE2*. LTR nucleotide sequences homologous to *Dasheng* and *RIRE2* were aligned using the ClustalX program. Phylogenetic reconstruction used the neighbor-joining method with Kimura-2 parameter distances implemented in the MEGA program (see "Materials and Methods").

RIRE2 and the intermediate group, respectively. The intermediate group appears to be slightly older, but the difference from *RIRE2* is not significant (Student's *t* test). Both groups are more than twice as old as the *Dasheng* group, indicating that they evolved before *Dasheng*. The fourth group is by far the most ancient, consistent with the hypothesis that it represents a unique but extinct family of elements.

A Scenario for the Origin and Spread of the Dasheng Family

The topological relationships and ages of the four phylogenetically distinct groups of elements (Figs. 4 and 5) suggest that there was a bifurcation early in *RIRE2* evolution. One of the resulting groups became the modern *RIRE2* family, whereas the other gave rise to the intermediate group where we hypothesize the nonautonomous *Dasheng* elements originated. This first *Dasheng* element may have been formed by template switching during an aberrant retrotransposition event when heterologous RNA templates (*RIRE2* and something else) were accidentally packaged in a single VLP. Alternatively, *Dasheng* may have originated at the chromosomal level because of ectopic recombination. Based on their low copy numbers, these ancestral *Dasheng* elements appear to have been largely quiescent or otherwise unsuccessful in the genome. The *RIRE2* elements of this intermediate group also may have been doomed to extinction. At about the same time, the ancestors of the modern *RIRE2* group were successfully amplifying in the genome. However, sometime after the initial formation of the *Dasheng* retrotransposons in the intermediate group (approximately 10–13 million years), a subset of these nonautonomous elements became active and spread. This young and prolific *Dasheng* group appears to still be active, or has been recently active. As discussed above, it is most likely that these active but nonautonomous *Dasheng* elements utilized the enzymatic machinery of *RIRE2*. In support of this is the fact that although the *RIRE2* group is older than the *Dasheng* group, there are still many recently transposed *RIRE2* elements, suggesting it is temporally compatible for *RIRE2* products to act in trans on *Dasheng* transcripts.

Concluding Remarks

Several lines of evidence presented in this study indicate that the previously described *RIRE2* family of LTR retrotransposons is the only candidate autonomous partner of the nonautonomous *Dasheng* family currently in the rice genome. By this, we mean that *RIRE2* is the probable source of enzymes for *Dasheng* retrotransposition, and may have been the source of *Dasheng* itself. In the absence of demonstrated activity for either family, our evidence remains circumstantial. The identification of strains harboring active elements or establishment of systems where the elements can be activated will ultimately facilitate our

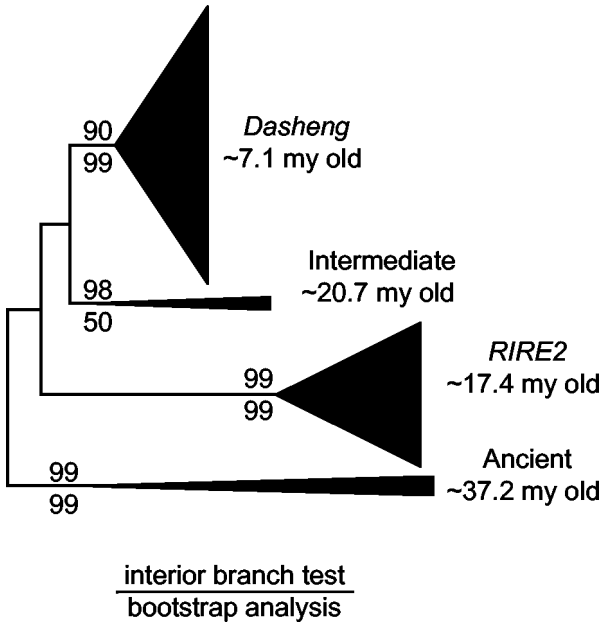

Figure 5. Dating the four monophyletic groups. Each groups is shown as a triangle, where the horizontal distance is proportional to the age and the vertical distance is proportional to the number of sequences in the group. Support for the internal branches of the phylogeny was assessed using 100 bootstrap replicates and the interior branch test (see "Materials and Methods").

understanding about whether and how *RIRE2* serves as the autonomous element for *Dasheng*.

MATERIALS AND METHODS

DNA Sequence Analysis

Rice (*Oryza sativa*) sequences used in this study are publicly available (Nipponbare sequence) at http://rgp.dna.affrc.go.jp. Sequence analysis (pair-wise comparisons, multiple sequence alignments, sequence assembling, and formatting) was performed with programs in the University of Wisconsin Genetics Computer Group program suite (version 10.1) accessed through Research Computing Resources (University of Georgia, Athens). The ORFs were defined using the ORF finder at http://www.ncbi.nlm.nih.gov/gorf/gorf.html.

Chromosomal Distribution of *Dasheng* and *RIRE2*

Sequences used to determine chromosomal locations of *Dasheng* and *RIRE2* were downloaded as described above on August 8, 2002 (360 Mb including overlaps). Based on the frequency of detecting duplicates (same element appearing multiple times in overlapping regions) for *Dasheng* and *RIRE2* elements, the redundancy of the sequence at the time of analysis was about 25%, corresponding to about 270 Mb of nonredundant sequence. The consensus sequence of *Dasheng* and *RIRE2* used for retrieving elements (with WU-BLASTN2.0, http://blast.wustl.edu) in this study was built based on the elements from 100 Mb of rice sequence (Jiang et al., 2002; also see below), and the cutoff E value is e-25. For full-length and truncated elements, *Dasheng* and *RIRE2* were distinguished by their internal sequence. For solo LTRs (approximately 10%), only those falling into *Dasheng* or *RIRE2* groups (Fig. 4) were considered. The positions of insertions were determined by the chromosomal position of BACs and PACs containing the elements (http://rgp.dna.affrc.go.jp, http://www.usricegenome.org, http://www.gramene.org, and http://www.genome.arizona.edu/fpc/rice/WebChrom).

Target Site Sequence of *Dasheng*, *RIRE2*, and *RIRE8*

LTR hits and 20 bp of flanking sequence were retrieved from the database. Resulting sequence files were then masked with *Dasheng* and *RIRE2* LTR sequence, and the TSD and flanking sequences were recorded. The same approach is not suitable for *RIRE8* because of its long LTR (approximately 3 kb) and high level of sequence divergence among family members. For *RIRE8*, elements were randomly chosen from RepeatMasker (Smit and Green, XXXX; http://ftp.genome.washington.edu/RM/webrepeatmaskerhelp.html) output of rice genomic sequences, and target sequences were recorded. Only *Dasheng*, *RIRE2*, and *RIRE8* elements (solo LTRs) with identical TSDs were considered for the determination of target sites.

Screening for Chimeric Elements

The sequence context for each element was obtained by retrieving the element plus 10 kb of flanking sequence per end and masked with a comprehensive database of rice repeats (Jiang and Wessler, 2001, Jiang et al., 2002). An element fragment present in another element and not flanked by a TSD was considered a chimeric element (see text).

Phylogenetic Analysis

LTR nucleotide sequences homologous to *Dasheng* and *RIRE2* were aligned using the ClustalX program (Thompson et al., 1997) with default options. The phylogeny of these sequences was reconstructed using the neighbor-joining method (Saitou and Nei, 1987) with Kimura-2 parameter distances (Kimura, 1980) implemented in the MEGA program (Kumar et al., 2001). Support for the internal branches on the phylogeny was assessed using 100 bootstrap replicates (Felsenstein, 1985) and the interior branch test (Nei and Kumar, 2000).

Dating Element Insertions

Full-length elements were aged (as in SanMiguel et al., 1998) by comparing their 5′ and 3′ LTRs. Kimura-2 parameter distances (K) between 5′ and 3′ LTRs of individual elements were calculated using MEGA. An average substitution rate (r) of 6.5×10^{-9} substitutions per synonymous site per year for grasses (Gaut et al., 1996) was used for calculations. The time (T) since element insertion was estimated using the formula: $T = K/2r$. Phylogenetic groups (families) of elements were dated (as in Kapitonov and Jurka, 1996; Costas and Naveira, 2000) by calculating the average Kimura 2-parameter distance (K) from sequences in a group to the consensus sequence of that group and calibrating with the average synonymous substitution rate (r) for grasses. Fifty percent consensus sequences were determined from group-specific alignments using the EMBL consensus sequence server (http://www.bork.embl-heidelberg.de/Alignment/consensus.html). The age of each group (T), or time since divergence from the group's common ancestor, was estimated using the formula: $T = K/r$.

ACKNOWLEDGMENT

We thank Zhirong Bao (Washington University, St. Louis) for help with DNA sequence analysis.

Received September 30, 2002; returned for revision October 4, 2002; accepted October 8, 2002.

LITERATURE CITED

Adam MA, Miller AD (1988) Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. J Virol **62:** 3802–3806

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815

Arkhipova IR, Lyubomirskaya NV, Ilyin YV (1995) *Drosophila* Retrotransposons. R.G. Landes Company, Austin, TX

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol **43:** 403–434

Boeke JD, Styles CA, Fink GR (1986) *Saccharomyces cerevisiae* SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol Cell Biol **6:** 3575–3581

Bowen NJ, McDonald JF (2001) *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res **11:** 1527–1540

Chen MP, SanMiguel P, de Oliveira AC, Woo SS, Zhang H, Wing RA, Bennetzen JL (1997) Microcolinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. Proc Natl Acad Sci USA **94:** 3431–3455

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

Clever J, Sassetti C, Parslow TG (1995) RNA secondary structure and binding sites for *gag* gene products in the 5′ packaging signal of human immunodeficiency virus type 1. J Virol **69:** 2101–2109

Costas J, Naveira H (2000) Evolutionary history of the human endogenous retrovirus family ERV9. Mol Biol Evol **17:** 320–330

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol **125:** 1342–1353

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA **93:** 10274–10279

Harushima Y, Jano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A et al. (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics **148:** 479–494

Hu WM, Das OP, Messing J (1995) *Zeon-1*, a member of a new maize retrotransposon family. Mol Gen Genet **248:** 471–480

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002) *Dasheng*: a recently amplified non-autonomous LTR element that is a major component of pericentromeric regions in rice. Genetics **161:** 1293–1305

Jiang N, Wessler SR (2001) Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. Plant Cell **13:** 2553–2564

Jin YK, Bennetzen JL (1989) Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. Proc Natl Acad Sci USA **86:** 6235–6239

Johns MA, Mottinger J, Freeling M (1985) A low copy number, *copia*-like transposon in maize. EMBO J **4:** 1093–1102

Jordan IK, McDonald JF (1998) Evolution of the *copia* retrotransposon in the *Drosophila melanogaster* species subgroup. Mol Biol Evol **15:** 1160–1171

Jordan IK, McDonald JF (1999) Phylogenetic perspective reveals *Ty1/Ty2* hybrid elements in the *Saccharomyces cerevisiae* genome. Mol Biol Evol **16:** 419–422

Kapitonov V, Jurka J (1996) The age of *Alu* subfamilies. J Mol Evol **42:** 59–65

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol **125:** 1198–1205

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **16:** 111–120

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet **33:** 479–532

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe

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

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. (2001) Initial sequencing and analysis of the human genome. Nature **409:** 860–921

Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res **11:** 1660–1676

Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York

Nonomura K, Kurata N (2001) The centromere composition of multiple repetitive sequences on rice chromosome 5. Chromosoma **110:** 284–291

Ohtsubo H, Kumekawa N, Ohtsubo E (1999) *RIRE2*, a novel *gypsy*-type retrotransposon from rice. Genes Genet Syst **74:** 83–91

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:** 406–425

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet **20:** 43–45

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev **6:** 743–748

Smit AF (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res **21:** 1863–1872

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell **12:** 381–391

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:** 4876–4882

Turcotte K, Srinivasan S, Bureau T (2001) Survey of transposable elements from rice genomic sequences. Plant J **25:** 169–179

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

Yu SS, Kim JM, Kim S (2000) The 17 nucleotides downstream from the *env* gene stop codon are important for murine leukemia virus packaging. J Virol **74:** 8775–8780