Dasheng: A Recently Amplified Nonautonomous Long Terminal Repeat Element That Is a Major Component of Pericentromeric Regions in Rice

Ning Jiang,* Zhirong Bao,† Svetlana Temnykh,‡ Zhukuan Cheng,§ Jiming Jiang,§ Rod A. Wing, Susan R. McCouch‡ and Susan R. Wessler*,[1]**

**Departments of Plant Biology and Genetics, University of Georgia, Athens, Georgia 30602, †Department of Genetics, Washington University, Saint Louis, Missouri 63110, ‡Department of Plant Breeding, Cornell University, Ithaca, New York 14853, §Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706 and **Clemson University Genomics Institute, Clemson, South Carolina 29634*

Manuscript received December 5, 2001
Accepted for publication April 23, 2002

ABSTRACT

A new and unusual family of LTR elements, *Dasheng*, has been discovered in the genome of *Oryza sativa* following database searches of ~100 Mb of rice genomic sequence and 78 Mb of BAC-end sequence information. With all of the *cis*-elements but none of the coding domains normally associated with retrotransposons (*e.g.*, *gag*, *pol*), *Dasheng* is a novel nonautonomous LTR element with high copy number. Over half of the ~1000 *Dasheng* elements in the rice genome are full length (5.6–8.6 kb), and 60% are estimated to have amplified in the past 500,000 years. Using a modified AFLP technique called transposon display, 215 elements were mapped to all 12 rice chromosomes. Interestingly, more than half of the mapped elements are clustered in the heterochromatic regions around centromeres. The distribution pattern was further confirmed by FISH analysis. Despite clustering in heterochromatin, *Dasheng* elements are not nested, suggesting their potential value as molecular markers for these marker-poor regions. Taken together, *Dasheng* is one of the highest-copy-number LTR elements and one of the most recent elements to amplify in the rice genome.

TRANSPOSABLE elements (TEs) have been divided into two classes, class 1 or RNA elements and class 2 or DNA elements. An RNA intermediate and a replicative mechanism of transposition are involved in the transposition of class 1 elements (Lewin 1997). RNA elements can be further divided into several groups, including long terminal repeat (LTR) retrotransposons, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). RNA elements are capable of attaining very high copy numbers because the element-encoded mRNA, not the element itself, forms the transposition intermediate.

LTR retrotransposons make up the largest fraction of most plant genomes (Kumar and Bennetzen 1999). The LTRs usually contain the initiation and termination sites of a transcript that encodes at least two genes, *gag* and *pol*. The products of these genes are involved in the different steps of retrotransposition, including reverse transcription and integration (Lewin 1997; Figure 1). Immediately internal to the LTR is the primer binding site (PBS) and the polypurine tract (PPT). Both are important *cis*-elements that are necessary for the initiation of the synthesis of element DNA from the RNA intermediate. LTR elements are classified into two types on the basis of the order of their encoded genes: *Ty1/copia* and *Ty3/gypsy* elements (Xiong and Eickbush 1990). Both are prevalent in plant genomes (Voytas *et al.* 1992; Suoniemi *et al.* 1998).

Differential amplification of LTR retrotransposons has been shown to be largely responsible for the C-value paradox in members of the grass clade (Chen *et al.* 1997; SanMiguel and Bennetzen 1998; Dubcovsky *et al.* 2001). The C-value paradox refers to the lack of correlation between the genome size and the biological complexity of an organism (Thomas 1971). For example, rice (*Oryza sativa*) and barley (*Hordeum vulgare*) have roughly the same number of genes and a largely conserved gene order (Moore *et al.* 1995; Dubcovsky *et al.* 2001). The 11-fold difference in the size of their genomes (430 *vs.* 4800 Mb) is due, in part, to the fact that retrotransposons comprise more than half of the barley genome and only 14% of the rice genome (Vicient *et al.* 1999; Tarchini *et al.* 2000). With the International Rice Genome Sequencing Program scheduled for completion in less than a year (Myers 2001), new insights about the identity and frequency of different TE families will emerge.

Despite its small genome, rice is still a model organism for the study of transposable elements. The genome of *O. sativa* contains all of the major types of elements found in the larger grass genomes, including retrotransposons, miniature inverted repeat transposable elements (MITEs), and other DNA elements (Bureau *et al.* 1996; Mao *et al.* 2000; Tarchini *et al.* 2000; Turcotte *et al.* 2001). Furthermore, the availability of several well-

[1]*Corresponding author:* Departments of Plant Biology and Genetics, University of Georgia, Athens, GA 30602.
E-mail: sue@dogwood.botany.uga.edu

characterized wild relatives provides the material necessary to analyze the impact of TEs on genome evolution and speciation. *O. sativa* is composed of two cultivated subspecies (*indica* and *japonica*) with thousands of diverse cultivars distributed worldwide. The genus Oryza has >20 species whose evolutionary relationships have been the subject of several phylogenetic analyses (Uozo *et al.* 1997; Ge *et al.* 1999; Sharma *et al.* 2000).

In this study, database searches of ~100 Mb of rice genomic sequence and 78 Mb of bacterial artificial chromosome (BAC) end sequence led to the identification of a new and unusual family of LTR elements called *Dasheng*. *Dasheng* is a very recently amplified family of 800–1300 nonautonomous elements, making it one of the most recently amplified and highest-copy-number families in rice. The family also includes ~16% solo LTRs. Like many other high-copy-number LTR elements, *Dasheng* elements are concentrated in the gene-poor pericentromeric regions of the chromosomes, which might be the reason for *Dasheng*'s success in the small genome of rice. The availability of large amounts of genomic sequence and an almost completely assembled chromosome 1 has allowed us to address questions regarding the distribution and timing of insertion events and to test models that explain the formation of solo LTRs.

MATERIALS AND METHODS

Plant material and DNA extraction: A doubled haploid (DH) mapping population (Guiderdoni *et al.* 1992; Huang and Kochert 1994) was used in conjunction with an existing SSR mapping data set (Temnykh *et al.* 2001) to map *Dasheng* elements. This population consisted of a subset of 96 doubled haploid lines derived via anther culture from the inter-subspecific cross between IR64 (*O. sativa* ssp. *indica*) and Azucena (tropical *japonica*). Other rice cultivars and wild species were obtained from the McCouch lab (Cornell University) and Gary Kochert (University of Georgia). Plant DNA was extracted as described (McCouch *et al.* 1988).

Genetic mapping: Transposon display was performed as described in Casa *et al.* (2000) to generate segregation patterns in the DH population with the following modifications. The element-specific primers were derived from the LTR sequence of *Dasheng* and the reaction was performed with rice DNAs. The final annealing temperature for selective amplification was 58° with ^{33}P-labeled *Dasheng* primer. Sequences of primers are available upon request. DNA fragments from transposon display were excised and cloned as described (Casa *et al.* 2000). DNA templates were sequenced by the Molecular Genetics Instrumentation Facility (University of Georgia).

The gel images of transposon display with DNAs from the DH mapping population were scored manually for presence/absence of polymorphic bands corresponding to *Dasheng* elements. The *Dasheng* markers were integrated into the SSR framework map using the Kosambi mapping function and MapMaker 3.0 software (Lander *et al.* 1987). Markers with a ripple of LOD > 2.0 were integrated into the framework maps and those mapping with LOD < 2.0 were assigned to the most likely intervals.

Fluorescence *in situ* hybridization analysis: Fluorescence *in situ* hybridization (FISH) analysis was performed as previously described (Jiang *et al.* 1995) using Nipponbare and *indica* cultivar Zhongxian 3037. The *Dasheng* probe (Figure 1) was labeled with biotin-16-UTP and detected using a fluorescein-5-isothiocyanate (FITC)-conjugated antibiotin antibody (Vector Laboratories, Burlingame, CA). Propidium iodide in an antifade solution was used to counterstain the chromosomes. Chromosome and FISH signal images were captured using a SenSys charge-coupled device camera (Photometrics, Tucson, AZ) and analyzed using IPLab Spectrum software (Signal Analytics, Vienna, VA).

DNA sequence analysis: DNA sequence analyses (pairwise comparisons, multiple sequence alignments, sequence assembling, and formatting) were performed with programs in the University of Wisconsin Genetics Computer Group program suite (version 10.1) accessed through Research Computing Resources, University of Georgia.

Identification of repetitive sequences from BAC ends: All sequences in the rice BAC end database (*O. sativa* cv. Nipponbare) were downloaded from the website of Clemson University Genome Institute (http://www.genome.clemson.edu) for the initial analysis (August 1999). An all *vs.* all comparison was performed with the sequences using WUBLASTN (http://blast.wustl.edu) with parameters $M = 5$, $N = -11$, $Q = 22$, $R = 11$, −kap E = 0.001, −hspmax 5000). Groups with highest intragroup similarities (>95%) were further characterized with BLAST search in the National Center for Biotechnology Information (NCBI) server (http://www.ncbi.nlm.nih.gov).

Identifying transposable elements in genomic sequence: The sequences of rice BACs and PACs were downloaded from the web sites of different rice genome projects including groups in the United States (http://www.usricegenome.org/), Japan (http://rgp.dna.affrc.go.jp/), Korea (http://bioserve.myongji.ac.kr/ricemac.html), People's Republic of China (http://www.ncgr.ac.cn/Ls/index.html), and Taiwan (http://genome.sinica.edu.tw/). Completely sequenced PACs or BACs and those in annotation and finishing (phase 2) were used as query sequences to search for transposable elements with RepeatMasker (A. Smit and P. Green, http://ftp.genome.washington.edu/RM/webrepeatmaskerhelp.html) as described (Jiang and Wessler 2001).

Copy-number determination: The copy number of *Dasheng* was estimated in three ways:

1. By blasting BAC ends using LTR sequence as a query. Using this method, copy number = matches in BAC ends × 430 Mb (rice genome size) ÷ the size of the BAC ends database (in megabases).
2. By probing a rice BAC library [derived from Nipponbare (Mao *et al.* 2000)] with a 500-bp fragment located between the third tract of direct repeats and the PPT (see Figure 1). Using this method, copy number of *Dasheng* elements = (number positive clones ÷ number of BACs screened) × 430 Mb ÷ average size of BACs (in megabases). The raw value, estimated to be 700 elements per haploid genome, was corrected for the number of solo LTRs (16%), BACs and PACs containing two or more elements (12% of the positive clones), and truncated elements (30%). The corrected copy number was 900–1300 (depending on the percentage of truncated elements detected).
3. By screening the genomic sequence with RepeatMasker followed by manual examination. The copy number = number of elements in genomic sequences × 430 Mb ÷ total size of the genomic sequence screened.

The copy number of other rice LTR elements (elements reported previously and those identified in this study) was estimated by blasting the BAC end database and GenBank (NCBI BLAST server) with LTR sequences. Low score matches ($e > 10^{-30}$) from GenBank were checked manually to deter-

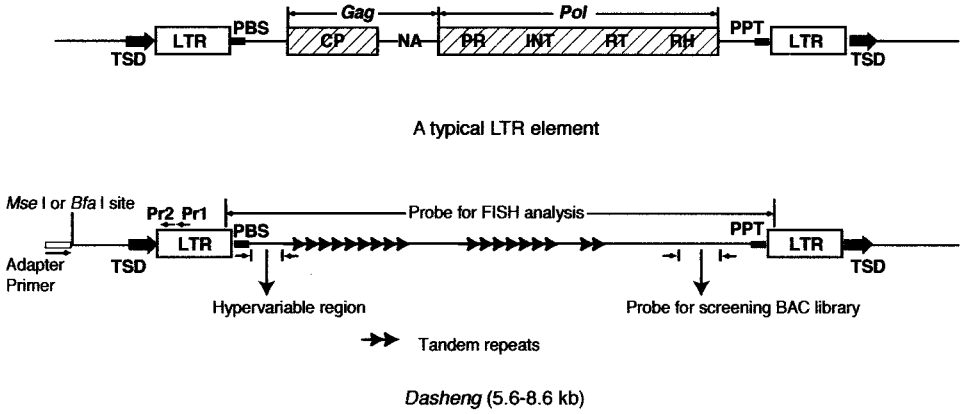


FIGURE 1.—Comparison of *Dasheng* with a typical autonomous LTR retrotransposon. Coding regions are shown as hatched boxes. CP, capsid-like proteins; PR, protease; INT, integrase; RT, reverse transcriptase; RH, RNase H. The relative order of RT, RH, and INT varies with different types of elements (see text). Other sequences indicated are: LTR, long terminal repeat; PBS, primer binding site; PPT, polypurine tract; NA, nucleic acid binding moiety. Arrows above the LTR of *Dasheng* indicate the positions of transposon-specific primers (Pr1 and Pr2) for transposon display. These would be used for PCR with the adapter primer shown (see Figure 2 and text).

mine whether the matches represented the element. The copy number for each element = matches for this element × 430 Mb ÷ total size of the rice genomic sequence in GenBank.

Phylogenetic analysis and aging of elements: LTR nucleotide sequences homologous to *Dasheng* and *RIRE2* were aligned using GCG (see above). Tree production and bootstrap analyses were performed using PAUP version 4.0. Sequence similarities and standard error were calculated with MEGA program (KUMAR *et al.* 2001). Full-length elements were aged (as in SANMIGUEL *et al.* 1998) by comparing their 5′ and 3′ LTR sequences. 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 to calibrate the ages of the elements. The time (T) since element insertion was estimated using the following formula: $T = K \div 2r$. 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 distribution of *Dasheng* elements in genomic sequences: The distribution of *Dasheng* on chromosome 1 of Nipponbare was constructed according to the positions of PACs and BACs that contained *Dasheng* elements (http://rgp.dna.affrc.go.jp/). Estimates of physical:genetic distance and insertion frequency were based on the data provided by the Rice Genome Research Program (http://rgp.dna.affrc.go.jp/) at the time of analysis. DNA density for chromosomal arms and pericentromeric regions was calculated from the total DNA of three contigs (20.2–34.5 cM and 40–50 cM in arms, 60–70 cM in pericentromeric regions) on chromosome 1. The borders of pericentromeric regions were defined as 15 cM from the center of the centromere on each arm. The position of the centromere was according to HARUSHIMA *et al.* (1998) and CHENG *et al.* (2001a). The remainder of the chromosome was defined as arms. Physical:genetic distance equals the physical length of DNA in base pairs divided by the map units covered. Insertion frequency equals the number of elements found in a certain region divided by the physical length of DNA in that region. The total amount of DNA was the size of all the clones minus overlap.

RESULTS AND DISCUSSION

A nonautonomous LTR element with very high copy number: To identify repeat sequences that might be novel transposable elements, we performed an all *vs.* all comparison with BAC end sequences of rice (*O. sativa* ssp. *japonica* cv. Nipponbare; see MATERIALS AND METHODS for details). Several groups of BAC ends were distinguished by their high within-group sequence similarity (~95%). The sequence of each group was then used as a query to perform further searches in GenBank. Significant matches for all groups were found in an 8.6-kb segment of a PAC clone from rice chromosome 6 (GenBank accession no. AB023482). This region has the structural features of an LTR retroelement including a long terminal repeat (441 bp with 99.5% sequence

TABLE 1

Polymorphism detected in the IR64 × Azucena mapping population

Adapter primer/enzyme combination	No. of amplified fragments			% polymorphic
	Monomorphic	Polymorphic	Total	
*Dasheng-Bfa*I + A	8	43	51	84.3
*Dasheng-Bfa*I + C	14	45	59	76.3
*Dasheng-Bfa*I + G	9	47	56	83.9
*Dasheng-Mse*I + A	18	58	76	76.3
*Dasheng-Mse*I + G	9	36	45	80.0
*Dasheng-Mse*I + T	15	53	68	77.9
Total	73	282	355	79.4

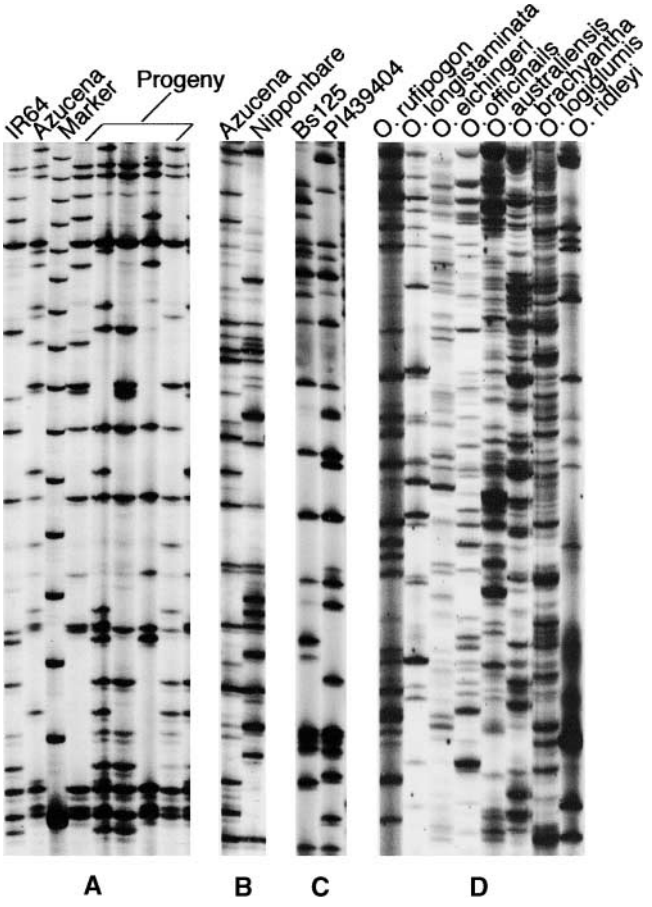


FIGURE 2.—Autoradiograph of *Dasheng* display with DNAs from *O. sativa* and other Oryza species. For all reactions, the transposon-specific primer was Pr2. (A) *Dasheng* display with DNAs from the IR64 × Azucena doubled haploid mapping population using adapter primer *Mse*I + T; (B) *Dasheng* display with DNAs from Azucena and Nipponbare, two *japonica* cultivars, using adapter primer *Bfa*I + C; (C) *Dasheng* display with DNAs from Bs125 and PI1439404, two *indica* cultivars, using adapter primer *Mse*I + A; (D) *Dasheng* display with DNAs from eight other Oryza species using adapter primer *Mse*I + TG.

similarity), an adjacent putative PBS and PPT, and a 5-bp target site duplication flanking the LTR (Figure 1). The 441-bp LTR is related (65–70% sequence similarity) to the LTR of *RIRE2*, a previously described *Ty3/gypsy* type LTR element in rice (OHTSUBO *et al.* 1999). In addition, the two elements also have similar PBSs and PPTs that differ only at 1 or 2 out of 15 nucleotides.

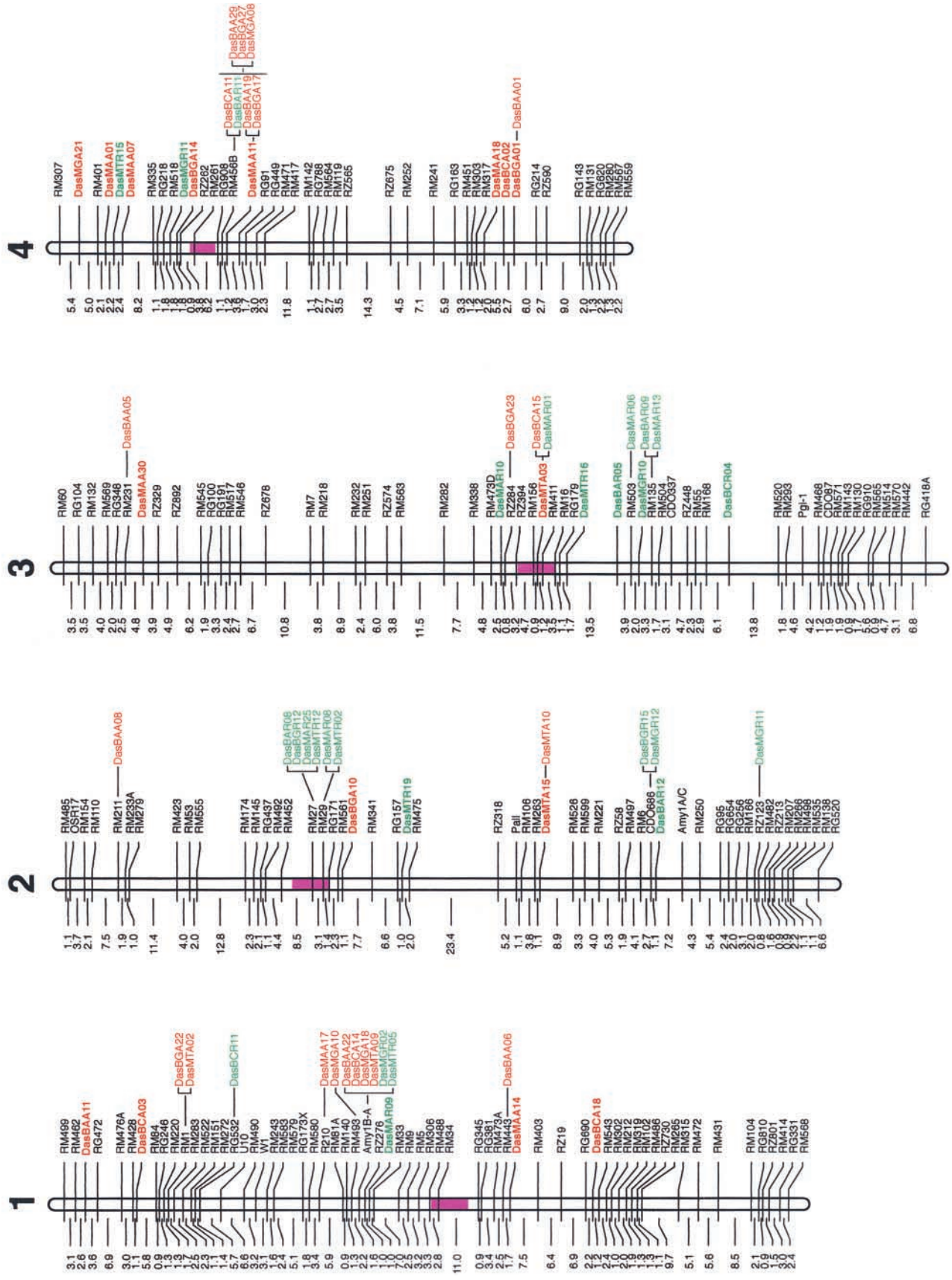
Despite having structural features of LTR retrotransposons, the 7.8-kb region between the LTR contains only very short open reading frames with no similarity to known proteins. Instead, ~4 kb of this region is composed of tandem repeats of an 89- to 90-bp unit (Figure 1). The other part of the internal region includes a hypervariable domain (of 0–1.2 kb) located between the PBS and the first tract of tandem repeats (Figure 1). The lack of coding capacity suggests that this element is most likely nonautonomous.

The copy number of this element family (named *Dasheng*) was estimated in three ways (see below and MATERIALS AND METHODS for details). On the basis of the prevalence of the LTR sequence in BAC ends [150 hits in 78 Mb of *Hin*dIII- and *Eco*RI-digested sequences ($e \leq 10^{-15}$)], we estimate that ~800 copies of *Dasheng* are in the genome of cv. Nipponbare. To test whether the prevalence of *Dasheng* in BAC ends is representative of the rest of the genome, a BAC library of the cv. Nipponbare genome was screened with a *Dasheng* probe. This experiment led to a copy number estimate of 900–1300. In contrast, a search of ~100 Mb of publicly available assembled genomic sequence led to a copy number determination of 470 per haploid genome or approximately one element per megabase. The two- to threefold difference in the values obtained from BAC screening and BAC end sequences *vs.* genomic sequence may be due to the fact that the latter is biased toward gene-rich regions, whereas several LTR retrotransposon families are enriched in pericentromeric regions of the genome (MILLER *et al.* 1998; LANGDON *et al.* 2000; NONOMURA and KURATA 2001; also see below).

The copy number of *Dasheng* was also compared with that of other LTR elements in rice. This was done by querying the BAC end and genomic sequence databases with LTRs from several high-copy-number rice elements previously described and elements identified in this study (see MATERIALS AND METHODS). As with the searches using *Dasheng* sequences as queries, the results were inconsistent from one database to the other. The average values obtained (from BAC ends and from genomic sequence) were, in descending order of copy number, *Retrosat2* (1080; GenBank accession no. AF111709), *Bajie* (730; identified in this study), *RIRE4* (730; KUMEKAWA *et al.* 1999), *SZ-19* (725; identified in this study), *Dasheng* (635), *RIRE8* (620; KUMEKAWA *et al.* 1999), *RIRE3* (510; KUMEKAWA *et al.* 1999), *RIRE2* (420; OHTSUBO *et al.* 1999), *RIRE9* (115; HAN *et al.* 2000; LI *et al.* 2000), and *RCS1* (90; DONG *et al.* 1998).

In a prior study, dot blot hybridization led to a copy

FIGURE 3.—Genetic map of rice (*O. sativa*) with *Dasheng* and framework markers. *Dasheng* markers from Azucena and IR64 are in red and green, respectively. *Dasheng* markers with a ripple of LOD > 2.0 were integrated into the framework map (in boldface type). *Dasheng* markers that cosegregate with a framework marker with absolute linkage are connected to this framework marker by a horizontal or slanted line. Vertical lines indicate possible intervals for *Dasheng* markers that are mapped with low LOD scores. Centromeres are indicated by purple boxes. The position of centromeres in this map is based on TEMNYKH *et al.* (2001), except for that in chromosome 10, which is based on CHENG *et al.* (2001a). Also shown for chromosome 11 is a diagram of the distribution of heterochromatic regions (indicated by solid ovals; the open circle in the middle represents the centromere; CHENG *et al.* 2001b).


1
RM499 RM462 DasBAA11 RG472 RM476A RM428 DasBCA03 RM84 RG246 RM220 DasBGA22 DasMTA02 RM1 RM283 RM522 RM151 RM272 RG532 DasBCR11 RM490 W1 RM243 RM583 RM579 RG173X RM580 R210 RM81A RM140 RM493 Amy1B-A RZ276 DasMAR09 RM33 RM9 RM5 RM306 RM488 RM34 DasMAA17 DasMGA10 DasBAA22 DasBCA14 DasMGA18 DasMTA09 DasMGR02 DasMTR05 RG345 RG381 RM473A RM443 DasMAA14 DasBAA06 RM403 RZ19 RG690 DasBCA18 RM543 RM302 RM212 RM319 RM102 RM486 RZ730 RM265 RM315 RM472 RM431 RM104 RG810 RZ801 RM414 RG331 RM568
2
RM485 OSR17 RM154 RM110 RM211 DasBAA08 RM238A RM279 RM423 RM53 RM555 RM174 RM145 RG437 RM492 RM452 DasBAR06 DasBGR12 DasBAR12 DasMTR12 DasMAR06 DasMTR02 RM27 RM29 RG171 RM561 DasBGA10 RM341 RG157 DasMTR19 RM475 RZ318 Pall RM106 RM263 DasMTA15 DasMTA10 RM526 RM599 RM221 RZ58 RM497 RM6 CDO686 DasBAR12 DasBGR15 DasMGR12 Amy1A/C RM250 RG95 RG654 RG256 RM186 RZ123 RM482 RZ213 RM207 RM498 RM535 RM138 RG520 DasMGR11
3
RM60 RG104 RM132 RM569 RG348 RM231 DasBAA05 DasMAA30 RZ329 RZ892 RM545 RG100 RG191 RM517 RM546 RZ678 RM7 RM218 RM232 RM251 RZ574 RM563 RM282 RM338 RM473D DasMAR10 DasBGA23 RZ284 RZ394 RM156 DasMTA03 DasBCA15 DasMaMR01 RM411 RM16 RG179 DasMTR16 DasBAR05 RM503 DasMAR06 DasMGR10 DasBAR09 RM135 DasMAR13 CDO337 RZ448 RM55 RM168 DasBCR04 RM520 RM293 Pgi-1 RM468 CDO87 RM571 RM143 RM130 RG910 RM565 RM514 RM570 RM442 RG418A
4
RM307 DasMGA21 RM401 DasMAA01 DasMTR15 DasMAA07 RM335 RG218 RM518 DasMGR11 DasBGA14 RZ262 RM261 RM458 RM458B DasBCA11 DasBAR11 DasBAA19 DasBGA17 DasMAA11 DasBAA29 DasBGA27 DasMGA08 RG91 RG449 RM471 RM417 RM142 RG788 RM264 RM119 RZ565 RZ675 RM252 RM241 RG163 RM451 RM303 RM317 DasBAA18 DasBCA02 DasBGA01 DasBAA01 RG214 RZ590 RG143 RM131 RG680 RM587 RM559

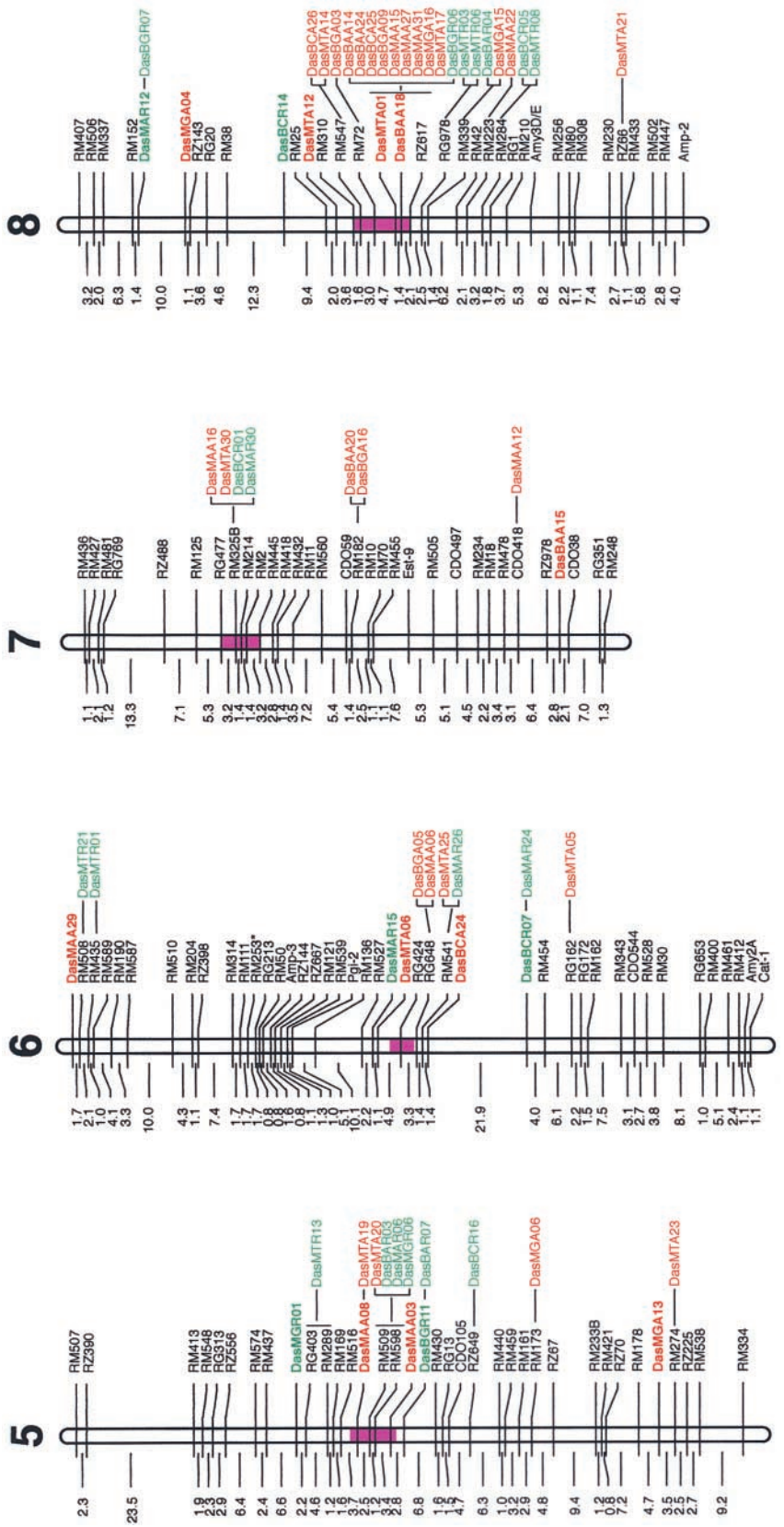


FIGURE 3.—*Continued.*

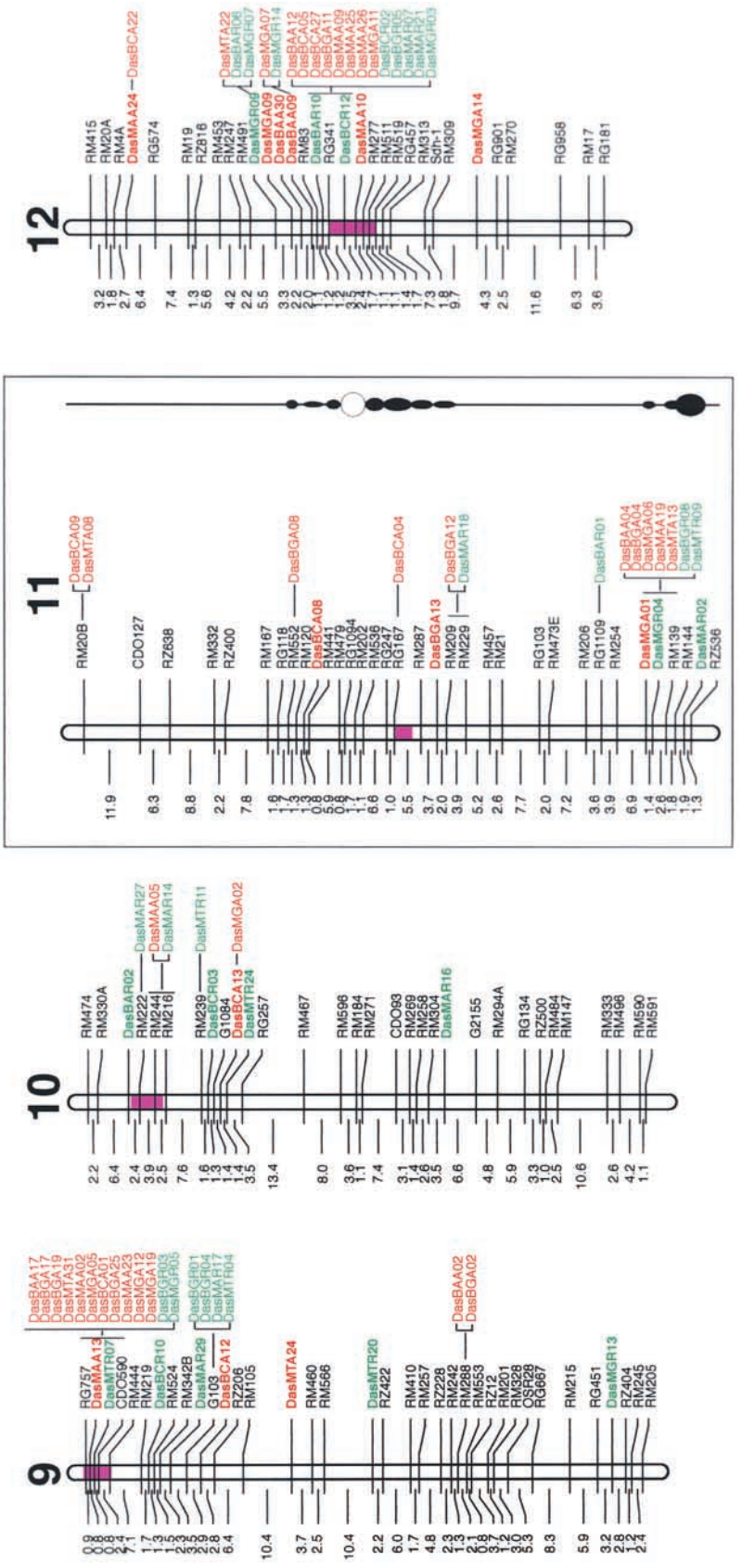


FIGURE 3.—*Continued.*

number determination for the *RIRE2* family of 10,000 in IR36 (Ohtsubo *et al.* 1999). In contrast, we found that the number of hits using *RIRE2* sequences was no higher than that found for *Dasheng*. The striking discrepancy may be due to the presence of distantly related families, a frequent cause of copy number overestimation when employing hybridization methods (Meyers *et al.* 2001).

The chromosomal location of *Dasheng* elements: *Genetic mapping of Dasheng elements:* To determine the chromosomal distribution of *Dasheng*, family members were mapped using a technique called transposon display, which is a modification of the amplified fragment length polymorphism (AFLP) procedure that generates PCR products anchored in a transposable element and a flanking restriction site (Waugh *et al.* 1997; Van den Broeck *et al.* 1998; Casa *et al.* 2000). The number of fragments amplified in one reaction can be adjusted by adding extra bases to the adapter primer (so-called selective bases); fewer fragments will be detected with more selective bases (Vos *et al.* 1995). Transposon display has the added advantage of detecting solo LTRs since the transposon-specific primers are located within the LTR (Figure 1, Pr1 and Pr2).

Dasheng primers were designed so as not to recognize the related *RIRE2* elements. Insertion site polymorphism, as defined by the presence of a PCR product in one parent but not in the other, was high for the parents (IR64 and Azucena), varying from 76.3 to 84.3% for different adapter primer/enzyme combinations in this inter-subspecific cross (Table 1; Figure 2A). High levels of polymorphism were also detected within *indica* and *japonica* subspecies (Figure 2, B and C), indicating that *Dasheng* elements can serve as a valuable marker system. Several wild species of rice were also tested with the same enzyme but they needed more selective bases (*Mse*I + T for *O. sativa*; *Mse*I + TG for wild species; Figure 2, A and D). The multiple fragments detected in the wild species indicate that *Dasheng* is also abundant in these genomes (Figure 2D).

The chromosomal location of polymorphic bands was determined by integrating segregation patterns into a previously constructed framework map consisting of 432 SSRs (Temnykh *et al.* 2001). The map was based on a population of doubled haploid lines, derived from a cross between IR64 (*indica*) and Azucena (*japonica*). In this study, a total of 215 bands (128 from Azucena and 87 from IR64) from six primer-enzyme combinations were assigned to all 12 chromosomes (Figure 3). Cloning and sequencing of 20 bands provided confirmation that all fragments were amplified from element-containing loci. For this reason, the mapped bands will be referred to as *Dasheng* markers.

Dasheng markers cluster around all centromeres and on the long arm of chromosome 11. We define a cluster as three or more elements mapping to the same site or to adjacent loci with an average distance of <1 cM. On the basis of this definition, >50% (120) of the elements were clustered in regions that account for only 3% of the total map distance. The largest clusters of markers were on chromosomes 4, 8, 9, and 12, which all correspond to small chromosomes containing prominent blocks of highly condensed chromatin (Fukui and Iijima 1991). The correlation of *Dasheng* clusters and the distribution of heterochromatin is best seen in chromosome 11, where significantly more elements were observed in the distal region on the long arm than in the pericentromeric region (Figure 3). The distal region of chromosome 11 is one of the most heterochromatic regions in the rice genome (Cheng *et al.* 2001b).

FISH analysis: The mapped elements represent only ~20% of the *Dasheng* family. FISH analysis was performed to ascertain whether the entire family shows similar clustering. To this end, an internal fragment of *Dasheng* (Figure 1) was used as a FISH probe with chromosomes prepared from cv. Nipponbare and Zhongxian 3037, an *indica* cultivar (Figure 4). In agreement with the mapping results, the majority of the FISH signal concentrated in pericentromeric regions with the most intense signals located on several small chromosomes. One of the small chromosomes with an intense signal was unambiguously identified as chromosome 4 on the basis of its distinctive arm ratio. This chromosome was previously found to contain one of the most heterochromatic regions in the rice genome (Cheng *et al.* 2001b). The absence of an exceptionally large cluster of elements on our genetic map could be due to the use of different strains for the genetic mapping *vs.* the cytogenetic analysis. Alternatively, since the genetic map reflects only polymorphic insertion sites, most of the elements on chromosome 4 may not be polymorphic in this mapping population.

Distribution of Dasheng on chromosome 1: At the time of this study, ~30% of rice genomic sequence was publicly available, including almost the entire chromosome 1. To provide a direct physical measure of how densely clustered the elements are on chromosome 1, the positions of all *Dasheng* were determined from the genomic sequence. The actual distribution of *Dasheng* elements on a single chromosome permits a determination of whether the apparent clustering of *Dasheng* in pericentromeric regions on the genetic map might instead be an artifact of the lower recombination rate in these regions (Moore and Sherman 1975). In other words, 1 cM may contain far more DNA around the centromeres, and this would give the appearance of clustering on a genetic map even if the insertion frequency is the same, in physical terms, as that in the gene-rich chromosome arms. To test this notion, we dissected chromosome 1 into pericentromeric regions and chromosome arms and calculated both physical:genetic distance and insertion frequency of *Dasheng* elements in different regions (see materials and methods for details). Consistent with the low recombination ratio in pericentromeric

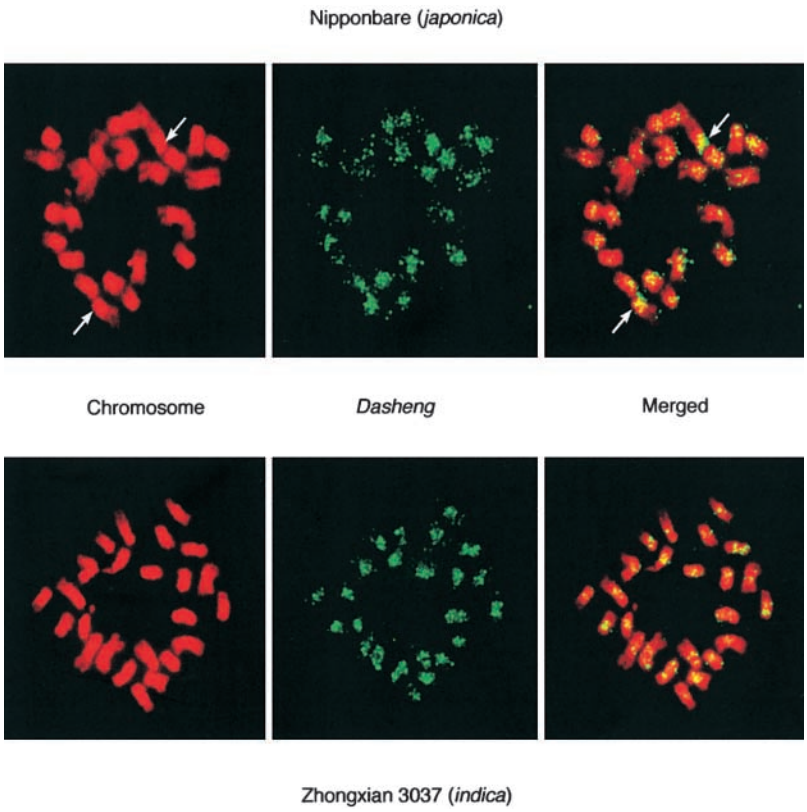


FIGURE 4.—FISH analysis of *Dasheng* distribution in rice mitotic chromosomes. *Dasheng* probes were detected by fluorescein isothiocyanate-conjugated antibiotin antibody (green); chromosomes were stained with propidium iodide (red). Arrows point to the strong signal of *Dasheng* on the short arm of chromosome 4 in Nipponbare (see text for details).

regions, the ratio of physical:genetic distance was roughly three times higher (660 kb DNA/cM) in pericentromeric regions compared to chromosome arms (206 kb DNA/cM; $P < 0.05$). However, more significant is the variation in the insertion frequency, which was about five times higher in the pericentromeric regions than in the arms (1.9 *vs.* 0.4 elements/Mb DNA, $P < 0.01$). These data confirm the higher density of *Dasheng* elements in pericentromeric regions.

***Dasheng* elements are not nested:** LTR retrotransposons are commonly found in large clusters in the genomes of grasses. In many instances, these clusters are composed of LTR retrotransposons inserted into other members of the same family (like *BARE-1*; SHIRASU *et al.* 2000) or into elements of other families (SANMIGUEL *et al.* 1998). Two rice LTR elements, *RIRE3* and *RIRE8*, were previously found to be nested (KUMEKAWA *et al.* 1999). Although the density of *Dasheng* in the rice genome (~1000 copies/430 Mb; this study) is comparable to that of *BARE-1* in barley (14,000 copies/4800 Mb; VICIENT *et al.* 1999), nested insertions of *Dasheng* elements were not observed. Only 6 out of 109 *Dasheng* elements are located within 10 kb of another *Dasheng* element, and the shortest distance between two *Dasheng* elements was 1.6 kb. As such, the clustering of *Dasheng* is unlikely to be due to a self-insertion preference, as has been observed for some retrotransposons and MITEs (HIGASHIYAMA *et al.* 1997; JIANG and WESSLER 2001).

Since the pericentromeric regions are enriched in repetitive sequences, including transposable elements (DONG *et al.* 1998; LANGDON *et al.* 2000; NONOMURA and KURATA 2001), the clustering of *Dasheng* could also be attributed to an insertion preference for other repetitive DNA, such as microsatellites or other transposable elements (CHRISTENSEN *et al.* 2000). To address this question, sequences flanking all *Dasheng* elements in the database were used as queries in computer-assisted searches. Of the 109 elements, 19 were found within an identifiable transposable element, 24 were located within 100 bp of an element, and about one-half were associated with low-copy-number sequences. Among this latter group, none showed significant similarity with a comprehensive database of rice TEs (N. JIANG and S. WESSLER, unpublished data).

In addition, unlike *RCS1*, *RIRE3*, and *RIRE8*, three other high-copy-number LTR retrotransposons in rice (DONG *et al.* 1998; KUMEKAWA *et al.* 1999; LANGDON *et al.* 2000; NONOMURA and KURATA 2001), *Dasheng* elements were not flanked by the *RCS2* centromere repeat (DONG *et al.* 1998), indicating that *Dasheng* is not a centromeric component. However, over half of the 215 *Dasheng* markers described in this study are located in pericentromeric regions. Since *Dasheng* elements do not specifically insert into other repetitive sequences, these markers may prove useful in the construction of fine structure maps of rice pericentromeric regions and isolation of genes buried in heterochromatic regions. Other cloning strategies frequently miss such genes.

Recent amplification of *Dasheng*: *Evidence from LTR similarity:* Since the LTR of a single retrotransposon is

identical upon insertion (Lewin 1997), sequence divergence between LTRs provides a measure of the time of insertion when an estimate of the nucleotide substitution rate is available (SanMiguel *et al.* 1998; Bowen and McDonald 2001). The average substitution rate in the *adh1* and *adh2* loci of grasses (6.5×10^{-9} substitutions per synonymous site per year) has been used to estimate the time of insertion of maize retrotransposons (Gaut *et al.* 1996; SanMiguel *et al.* 1998). In this study, a search of the 100 Mb of publicly available rice sequence led to the identification of 109 *Dasheng* elements of which 60 were full length (56%), 32 were truncated (28%), and 17 were solo LTRs (16%). Among the 60 full-length elements, 35 (58%) have >99.5% LTR similarity, with 15 being identical. In the discussions that follow, these elements are referred to as "recent." LTR sequence similarity of the other 25 elements varies from 92.7 to 99.1%. By using the same base substitution rate as SanMiguel *et al.* (1998), we estimate that the *Dasheng* elements with >99.5% LTR identity (58% of the available full-length elements) inserted within the last 500,000 years. This is a conservative estimate because LTRs evolve more rapidly than coding regions like *adh1* and *adh2* (SanMiguel *et al.* 1998) and because reverse transcription is known to be an error-prone process. Based on a comparison with other high-copy-number LTR elements in rice, the *Dasheng* family has the highest ratio of elements with identical LTRs (15 out of 109; N. Jiang and S. Wessler, unpublished data). Thus, *Dasheng* may have amplified more recently than all other high-copy-number elements in the rice genome.

A hypervariable region and tandem repeats: A phylogenetic tree was constructed on the basis of the LTR sequences of *Dasheng* elements and some *RIRE2* elements (Figure 5) and used to evaluate whether other structural features of *Dasheng* correlated with recently amplified elements. Of particular interest were a hypervariable region, the tandem repeats, and the solo LTRs.

As mentioned above, the hypervariable region is located between the PBS and the first tract of tandem repeats (Figure 1). This region consists of a common sequence shared by many or a few elements (no sequence is shared by all elements) and a unique sequence. A similarly organized region of sequence heterogeneity was reported for the *Stonor* elements of maize (Marillonnet and Wessler 1998). Interestingly, more than half of the recent elements (19 out of 35) have the same sequence in this region (see Figure 5 for branch lengths of elements labeled with an asterisk), suggesting that the recent amplification of *Dasheng* could be due largely to the transposition of elements in this subgroup. In addition, the average length of the tandem repeat region in this group is significantly longer than that of other elements (3.7 *vs.* 2.2 kb, $t < 0.001$). These data do not permit a determination of whether younger elements have more repeats or if repeats are gradually deleting from the older elements.

Evolution of *Dasheng* elements: *Targeted insertion vs. negative selection:* Having both complete sequences of the element and precise chromosomal locations permits a preliminary determination of whether the clustering of *Dasheng* elements in the pericentromeric region is due to targeted insertion into the gene-rich arms or postinsertion selection. If *Dasheng* shows no target site preference, but elements are lost over time from the arms, the arms should contain more recent insertions than the pericentromeric regions. However, no significant difference is seen in the number of recent insertions in arms *vs.* pericentromeric regions. In chromosomal arms, 6 out of 27 (22%) full-length elements have identical LTRs, whereas in pericentromeric regions, 9 out of 33 (27%) have identical LTRs (χ^2 test; $P > 0.50$).

Origin of solo LTRs: Solo LTRs are believed to arise from intraelement recombination between transiently paired LTRs (Parket *et al.* 1995). Recently, the formation of solo LTRs in barley has been proposed as a mechanism that can reverse genome expansion (Shirasu *et al.* 2000). With 16% of the sequenced elements present as solo LTRs, the *Dasheng* family provides a unique opportunity to address questions about the formation of solo LTRs by analyzing both their age and their distribution relative to full-length family members.

At least two models can account for the formation of solo LTRs. If it is a stochastic process, older insertions are more likely to have undergone recombination and the distribution of solo LTRs should be no different than the distribution of full-length elements. On the other hand, since solo LTRs arise from intraelement recombination, the frequency of solo LTR formation might correlate with regional recombination frequencies. In this case, there would be relatively fewer solo LTRs in the centromeric and pericentromeric regions where recombination rates are much lower than those in the gene-rich chromosomal arms (Moore and Sherman 1975). As can be seen in Figure 5, solo LTRs are associated more often with the longer branches, indicating that they are older, on average, than the full-length elements (0.0437 *vs.* 0.0187, $t < 0.001$). In addition, the ratio of solo LTRs to full-length elements in the arms (8:27) is only slightly higher than that in pericentromeric regions (6: 33), and the difference is not significant ($P > 0.10$ by χ^2 test). Taken together, these data suggest that solo LTR formation in the *Dasheng* family is probably a stochastic process.

Concluding remarks: In this study, we characterized an unusual LTR element in rice. As a special category of LTR elements, *Dasheng* is distinguished by its (1) lack of coding capacity, (2) presence of long tracts of tandem repeats, (3) clustering in heterochromatic regions, (4) high copy number, and (5) recent amplification.

Since *Dasheng* is among the rice elements of highest copy number and most recent amplification, it is of great interest to know if members of the *Dasheng* family are still capable of retrotransposition. To date, activity

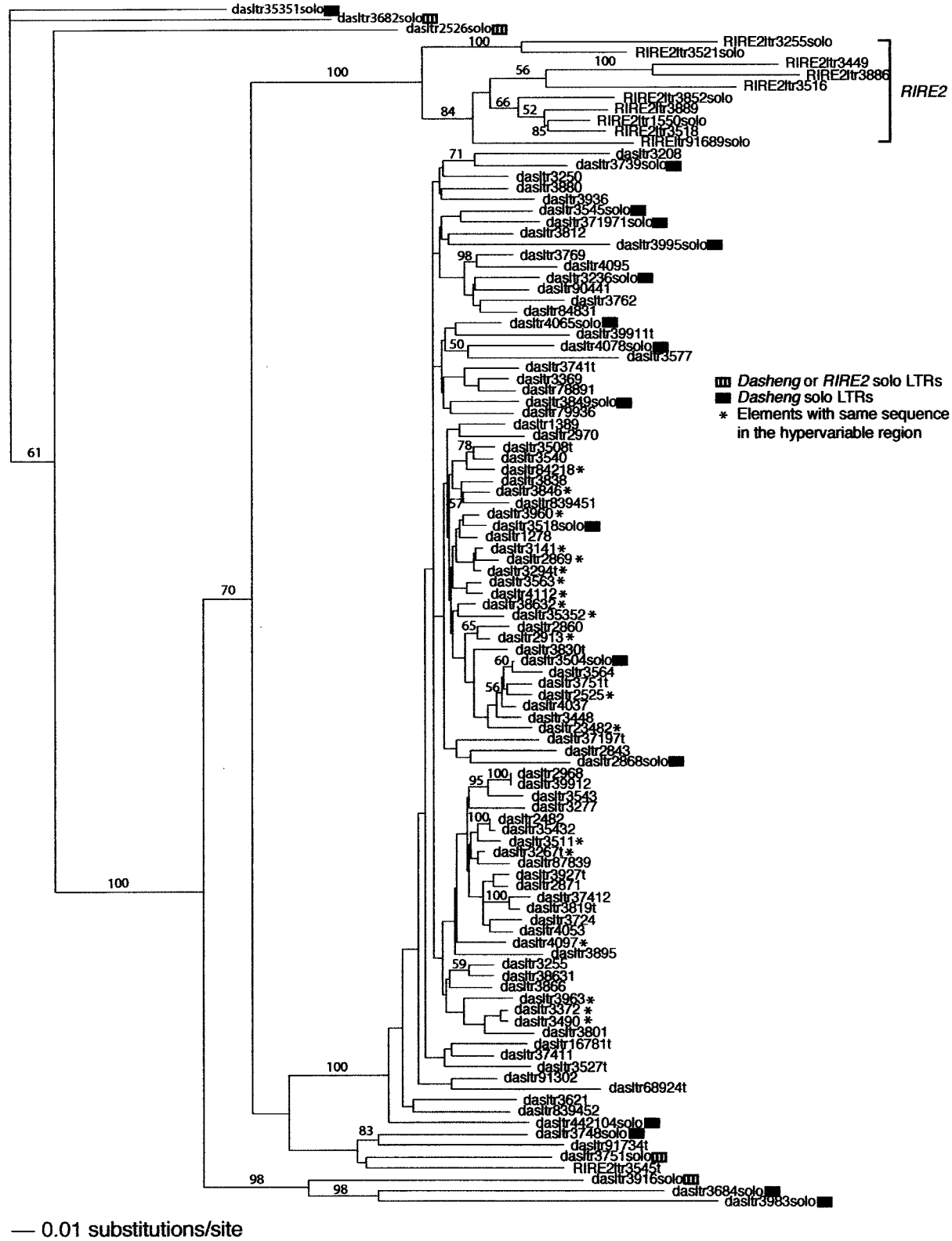


FIGURE 5.—Phylogenetic analysis of LTR sequences of *Dasheng* and some *RIRE2* elements using the neighbor-joining algorithm from distance matrices. Branch length is proportional to genetic distance. Bootstrap values >50 are indicated as a percentage of 1000 replicates. "*Dasheng* or *RIRE2* solo LTRs" indicates that the sequence similarity of these solo LTRs to the *Dasheng* consensus is not significantly different from that of *RIRE2* consensus by *t*-test.

has not been demonstrated for any of the high-copy-number LTR retrotransposons in rice. The only active rice elements (such as *Tos17*) are present in less than five copies and are activated to retrotranspose by tissue culture (Hirochika *et al.* 1996; Agrawal *et al.* 2001; Yamazaki *et al.* 2001). In fact, although LTR retrotransposons are the major component of most plant genomes, the only high-copy-number LTR retrotransposon with demonstrated activity is *BARE-1* from barley (Suoniemi *et al.* 1996; Jaaskelainen *et al.* 1999). As such, it is the only genomic component known to be capable of contributing significantly to genome size variation between populations and related species in plants (Kalendar *et al.* 2000). Like *BARE-1* in barley, *Dasheng* is a major component of the genome of cultivated rice, *O. sativa*. In addition, a preliminary survey indicates that *Dasheng* is probably abundant in all species of the genus Oryza (Figure 2). For these reasons it will be important for future studies to determine whether *Dasheng* elements are still capable of transposition.

The origin of the *Dasheng* is also of interest since it is a nonautonomous class 1 element. Nonautonomous transposable elements are widespread in eukaryotic organisms. For DNA elements and non-LTR retrotransposons, the copy number of nonautonomous elements is usually much higher than that of the corresponding autonomous element (Kapitonov and Jurka 1999; Feschotte and Mouchès 2000; International Human Genome Sequencing Consortium 2001). Unlike other classes of nonautonomous elements, only a few high-copy-number LTR elements have been characterized. The only other plant element is the maize *Zeon-1*, which has a copy number of 6000–32,000 (Meyers *et al.* 2001). However, unlike *Dasheng*, *Zeon-1* is one of the oldest elements in the maize genome (Hu *et al.* 1995; SanMiguel *et al.* 1998). As such, it will be difficult, if not impossible, to deduce what autonomous element could be responsible for the amplification of *Zeon-1*. In contrast, the availability of most of the rice genome sequence coupled with the recent amplification of *Dasheng* facilitates a comprehensive analysis of autonomous elements that could have given rise to *Dasheng*. At this time, the *RIRE2* family with its related LTRs, as well as the presence of some recently amplified members, is the best candidate. Further studies are underway to establish additional connections between these two LTR element families.

We thank I. King Jordan (National Institutes of Health and National Library of Medicine, NCBI, Bethesda, MD) and Nathan J. Bowen (National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD) for assistance in sequence and phylogenetic analysis and Xiaoyu Zhang for critical reading of the manuscript. This study was supported by grants from the U.S. Department of Energy (DEFG02-94ER20135) and the National Science Foundation (DBI-0077709) to S.R.W.

LITERATURE CITED

Agrawal, G. K., M. Yamazaki, M. Kobayashi, R. Hirochika, A. Miyao *et al.*, 2001 Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion. Tagging of a zeaxanthin epoxidase gene and a novel OsTATC gene. Plant Physiol. **125:** 1248–1257.

Bowen, N. J., and J. F. McDonald, 2001 *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res. **11:** 1527–1540.

Bureau, T. E., P. C. Ronald and S. R. Wessler, 1996 A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wildtype rice genes. Proc. Natl. Acad. Sci. USA **93:** 8524–8529.

Casa, A. M., C. Brouwer, A. Nagel, L. Wang, Q. Zhang *et al.*, 2000 The MITE family heartbreaker (*Hbr*): molecular markers in maize. Proc. Natl. Acad. Sci. USA **97:** 10083–10089.

Chen, M. P., P. SanMiguel, A. C. de Oliveira, S. S. Woo, H. Zhang *et al.*, 1997 Microcolinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. Proc. Natl. Acad. Sci. USA **94:** 3431–3455.

Cheng, Z., G. G. Presting, C. R. Buell, R. A. Wing and J. Jiang, 2001a 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.

Cheng, Z., R. Buell, R. A.Wing, M. Gu and J. Jiang, 2001b Towards a cytological characterization of the rice genome. Genome Res. **11:** 2133–2141.

Christensen, S., G. Pont-Kingdon and D. Carroll, 2000 Target specificity of the endonuclease from the *Xenopus laevis* non-long terminal repeat retrotransposon, *Tx1L*. Mol. Cell. Biol. **20:** 1219–1226.

Dong, F., J. T. Miller, S. A. Jackson, G. L. Wang, P. C. Ronald *et al.*, 1998 Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc. Natl. Acad. Sci. USA **95:** 8135–8140.

Dubcovsky, J., W. Ramakrishna, P. J. SanMiguel, C. S. Busso, L. Yan *et al.*, 2001 Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol. **125:** 1342–1353.

Feschotte, C., and C. Mouchès, 2000 Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. Mol. Biol. Evol. **17:** 730–737.

Fukui, K., and K. Iijima, 1991 Somatic chromosome map of rice by imaging methods. Theor. Appl. Genet. **81:** 589–596.

Gaut, B. S., B. R. Morton, B. C. McCaig and M. T. Clegg, 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.

Ge, S., T. Sang, B.-R. Lu and D.-Y. Hong, 1999 Phylogeny of rice genomes with emphasis on origins of allotetroaploid species. Proc. Natl. Acad. Sci. USA **96:** 14400–14405.

Guiderdoni, E., E. Galinato, J. Luistra and G. Vergarro, 1992 Anther culture of tropical *japonica* × *indica* hybrids of rice (*Oryza sativa* L.). Euphytica **62:** 219–224.

Han, C. G., M. J. Frank, H. Ohtsubo and E. Ohtsubo, 2000 New transposable elements identified as insertions in rice transposon *Tnr1*. Genes Genet. Syst. **75:** 69–77.

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

Higashiyama, T., Y. Noutoshi, M. Fujie and T. Yamada, 1997 *Zepp*, a LINE-like retrotransposon accumulated in the *Chlorella* telemetric region. EMBO J. **16:** 3715–3723.

Hirochika, H., K. Sugimoto, Y. Otsuki and M. Kanda, 1996 Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA **93:** 7783–7788.

Hu, W. M., O. P. Das and J. Messing, 1995 *Zeon-1*, a member of a new maize retrotransposon family. Mol. Gen. Genet. **248:** 471–480.

Huang, H., and G. Kochert, 1994 Comparative RFLP mapping of an allotetraploid wild rice species (*Oryza latifolia*) and cultivated rice (*O. sativa*). Plant Mol. Biol. **25:** 633–648.

International Human Genome Sequencing Consortium, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860–921.

Jaaskelainen, M., A. H. Mykkanen, T. Arna, C. M. Vicient, A. Suoniemi *et al.*, 1999 Retrotransposon *BARE-1:* expression of

encoded proteins and formation of virus-like particles in barley cells. Plant J. **20:** 413–422.

Jiang, J., B. S. Gill, G. L. Wang, P. C. Ronald and D. C. Ward, 1995 Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc. Natl. Acad. Sci. USA **92:** 4487–4491.

Jiang, N., and S. R. Wessler, 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.

Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo and A. H. Schulman, 2000 Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA **97:** 6603–6607.

Kapitonov, V. V., and J. Jurka, 1999 Molecular paleontology of transposable elements from Arabidopsis thaliana. Genetica **107:** 27–37.

Kumar, A., and J. L. Bennetzen, 1999 Plant retrotransposons. Annu. Rev. Genet. **33:** 479–532.

Kumar, S., K. Tamura, I. B. Jakobsen and M. Nei, 2001 *MEGA2: Molecular Evolutionary Genetics Analysis Software*, Arizona State University, Tempe, AZ.

Kumekawa, N., H. Ohtsubo, T. Horiuchi and E. Ohtsubo, 1999 Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol. Gen. Genet. **260:** 593–602.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

Langdon, T., C. Seago, M. Mende, M. Leggett, H. Thomas *et al.*, 2000 Retrotransposon evolution in diverse plant genomes. Genetics **156:** 313–325.

Lewin, B., 1997 *Genes VI*. Oxford University Press, New York.

Li, Z. Y., S. Y. Chen, X. W. Zheng and L. H. Zhu, 2000 Identification and chromosomal localization of a transcriptionally active retrotransposon of *Ty3-gypsy* type in rice. Genome **43:** 404–408.

Mao, L., T. C. Wood, Y. Yu, M. A. Budiman, J. Tomkins *et al.*, 2000 Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. **10:** 982–990.

Marillonnet, S., and S. R. Wessler, 1998 Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics **150:** 1245–1256.

McCouch, S. R., G. Kochert, Z. H. Yu, G. S. Khush, W. R. Coffman *et al.*, 1988 Molecular mapping of rice chromosomes. Theor. Appl. Genet. **76:** 815–829.

Meyers, B. C., S. V. Tingey and M. Morgante, 2001 Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. **11:** 1660–1676.

Miller, J. T., F. Dong, S. A. Jackson, J. Song and J. Jiang, 1998 Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150:** 1615–1623.

Moore, C. W., and F. Sherman, 1975 Role of DNA sequences in genetic recombination in the iso-1-cytochrome c gene of yeast. I. Discrepancies between physical distances and genetic distances determined by five mapping procedures. Genetics **79:** 397–418.

Moore, G., T. Foote, T. Helentjaris, K. Devos, N. Kurata *et al.*, 1995 Was there a single ancestral cereal chromosome? Trends Genet. **11:** 81–82.

Myers, T., 2001 Rice genome consortium will finish ahead of schedule. Nature **409:** 752.

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

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

Parket, A., O. Inbar and M. Kupiec, 1995 Recombination of *Ty* elements in yeast can be induced by a double-strand break. Genetics **140:** 67–77.

SanMiguel, P., and J. L. Bennetzen, 1998 Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. **81:** 37–44.

SanMiguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L. Bennetzen, 1998 The paleontology of intergene retrotransposons of maize. Nat. Genet. **20:** 43–45.

Sharma, S. D., S. R. Dhu and P. K. Agarwal, 2000 Species of genus *Oryza* and their interrelationships, pp. 311–346 in *Rice Breeding and Genetics*, edited by J. S. Nanda. Science Publisher, Enfield, NH.

Shirasu, K., A. H. Schulman, T. Lahaye and P. Schulze-Lefert, 2000 A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res. **10:** 908–915.

Suoniemi, A., A. Narvanto and A. H. Schulman, 1996 The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol. Biol. **31:** 295–306.

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

Tarchini, R., P. Biddle, R. Wineland, S. Tingey and A. Rafalski, 2000 The complete sequence of 340 kb of DNA around the rice Adh1-adh2 region reveals interrupted linearity with maize chromosome 4. Plant Cell **12:** 381–391.

Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour *et al.*, 2001 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11:** 1441–1452.

Thomas, C. A., 1971 The genetic organization of chromosomes. Annu. Rev. Genet. **5:** 237–256.

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

Uozo, S., H. Ikehashi, N. Ohmido, H. Ohtsubo, E. Ohtsubo *et al.*, 1997 Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus Oryza. Plant Mol. Biol. **35:** 791–799.

Van den Broeck, D., T. Maes, M. Sauer, J. Zethof, P. De Keukeleire *et al.*, 1998 Transposon display identifies individual transposable elements in high copy number lines. Plant J. **13:** 121–129.

Vicient, C. M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav *et al.*, 1999 Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell **11:** 1769–1784.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee *et al.*, 1995 AFLP, a new technique for DNA fingerprinting. Nucleic Acids Res. **23:** 4407–4414.

Voytas, D. F., M. P. Cummings, A. Konieczny, F. M. Ausubel and S. R. Rodermel, 1992 *copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA **89:** 7124–7128.

Waugh, R., K. McLean, A. J. Flavell, S. R. Pearce, A. Kumar *et al.*, 1997 Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. **253:** 687–694.

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

Yamazaki, M., H. Tsugawa, A. Miyao, M. Yano, J. Wu *et al.*, 2001 The rice retrotransposon *Tos17* prefers low-copy-number sequences as integration targets. Mol. Gen. Genet. **265:** 336–344.

Communicating editor: C. S. Gasser