Insertion Preference of Maize and Rice Miniature Inverted Repeat Transposable Elements as Revealed by the Analysis of Nested Elements

Ning Jiang and Susan R. Wessler
Departments of Botany and Genetics, University of Georgia, Athens, Georgia 30602

A 128-bp insertion into the maize \textit{waxy-B2} allele led to the discovery of \textit{Tourist}, a family of miniature inverted repeat transposable elements (MITEs). As a special category of nonautonomous elements, MITEs are distinguished by their high copy number, small size, and close association with plant genes. In maize, some \textit{Tourist} elements (named \textit{Tourist-Zm}) are present as adjacent or nested insertions. To determine whether the formation of multimers is a common feature of MITEs, we performed a more thorough survey, including an estimation of the proportion of multimers, with 30.2 Mb of publicly available rice genome sequence. Among the 6600 MITEs identified, \textgreater{}10\% were present as multimers. The proportion of multimers differs for different MITE families. For some MITE families, a high frequency of self-insertions was found. The fact that all 340 multimers are unique indicates that the multimers are not capable of further amplification.

\textbf{INTRODUCTION}

Transposable elements usually are divided into two classes. Class 1, the retroelements, including the long terminal repeat (LTR) retrotransposons, makes up the largest fraction of most plant genomes (reviewed by Kumar and Bennetzen, 1999). Retroelements are capable of attaining very high copy numbers in a relatively short period because the element-encoded mRNA, and not the element itself, forms the transposition intermediate. Class 2, the DNA elements, is characterized by short terminal inverted repeats (TIRs) and transposition via a DNA intermediate (reviewed by Kunze et al., 1997). Plant DNA elements (such as \textit{Ac/Ds}, \textit{Spm/dSpm}, and \textit{Mutator}) generally excise from one site and reinsert elsewhere in the genome. Class 2 elements can be further divided into two groups. Autonomous elements, such as \textit{Ac} and \textit{Spm} from maize, encode the products (transposase) necessary for their transposition (Baker et al., 1986; Yoder, 1990; Kunze et al., 1997). Nonautonomous elements, such as \textit{Ds} and \textit{dSpm}, usually are internally deleted versions of autonomous elements. As a result, they require the presence of the autonomous elements \textit{Ac} and \textit{Spm}, respectively, for their transposition.

As a result of their conservative mechanism of transposition, the copy number of class 2 element families is usually \textless{}100 per haploid genome. One exception to this generalization is miniature inverted repeat transposable elements (MITEs), a special category of nonautonomous elements that display very high copy number (in the thousands) and are uniformly short (usually \textless{}500 bp). In addition, most MITEs in plants have TIRs and insert into the TA dinucleotide or into a 3-bp trinucleotide (Bureau and Wessler, 1992, 1994a, 1994b; Mao et al., 2000; Zhang et al., 2000). Although first identified in several plant species, including maize (Bureau and Wessler, 1992, 1994a, 1994b), rice (Bureau and Wessler, 1994a, 1994b; Bureau et al., 1996), green pepper (Pozueta-Romero et al., 1996), and Arabidopsis (Casacuberta et al., 1998; Le et al., 2000), MITEs also are abundant in several animal genomes, including \textit{Caenorhabditis elegans} (Oosumi et al., 1995a, 1995b; Surzycki and Belknap, 2000), mosquito (Tu, 1997, 2001; Feschotte and Mouchès, 2000), fish (Izsvák et al., 1999), and human (Morgan, 1995; Smit and Riggs, 1996).

Another important feature of MITEs is their preference for insertion into low copy number sequences or genic regions (Tikhonov et al., 1999; Mao et al., 2000; Zhang et al., 2000). In addition, MITEs also were found, in several cases, to insert into each other. For instance, the first \textit{Stowaway} element was found as an insertion in a sorghum \textit{Tourist} element (Bureau and Wessler, 1994b), whereas in another case a \textit{Tourist} dimer was found in the same organism (Tikhonov et al., 1999). Such MITE multimers also were
reported in other organisms, including rice (Tarchini et al., 2000) and mosquito (Tu, 1997; Feschotte and Mouchès, 2000). Therefore, it was proposed that MITEs could be preferential targets for other MITEs (Feschotte and Mouchès, 2000).

Given the previously identified target site preference of MITEs and the frequent detection of MITE multimers, we wondered about the propensity of MITE insertion into other MITEs. Such a determination is possible only with a systematic comparison between the insertion frequency of MITEs into MITEs and of MITEs into other sequences. In this article, we report a detailed characterization of maize Tourist multimers and a comprehensive analysis of MITE multimers in rice, a species known to be particularly rich in MITEs (Bureau et al., 1996; Mao et al., 2000; Tarchini et al., 2000). The availability of 30.2 Mb of rice genomic sequence has enabled us to address questions about MITE multimers that could not be answered with the available maize sequence. The analysis of rice genomic sequence not only allows us to evaluate the prevalence of MITE multimers but also may provide new insight into the temporal order of amplification of different transposable elements in the rice genome.

RESULTS

Tourist Multimers in Maize

The first reported MITE was the B2 element, found as a 128-bp insertion into the maize waxy (wx) gene in the mutant wxB2 allele (Wessler and Varagona, 1985; Bureau and Wessler, 1992). Subsequent database searches revealed that this element belongs to a large family of related elements, called Tourist, whose members are associated with the noncoding regions of genes from maize, sorghum, barley, and rice (Bureau and Wessler, 1992, 1994a). Tourist multimers were discovered initially in a polymerase chain reaction (PCR) assay that was intended to identify additional B2-like (Tourist) elements in maize. Genomic DNA from maize inbred B79 was amplified with primers derived from the B2 TIR (14 bp) and 11 bp of internal element sequence. In addition to a product corresponding to the size of the B2 element (128 bp), we observed larger fragments that varied in size depending on the annealing temperature (Figure 1).

![Figure 1. The Identity of Multimers Resulting from PCR Amplification of B2-Tourist Elements in Maize Line B79.](image-url)

PCR products resolved by agarose gel electrophoresis and visualized by ethidium bromide staining were purified, cloned, and sequenced (see Methods). The annealing temperature for PCR was 60°C for the two samples at left and 55°C for the three samples at right. The identities of PCR bands are diagrammed at right. The positions of the insertion sites of the various Tourist subfamily members (Zm3, Zm11, Zm22, and Zm29) (Bureau and Wessler, 1992, 1994a; N. Jiang and S. Wessler, unpublished data) into B2 elements are shown along with the TSD. M, DNA molecular weight standard (bp).
PCR products from all size classes were cloned, and several were sequenced, revealing monomers, dimers, and a trimer (Figure 1). Four different multimers were found among the six dimer-sized clones that were sequenced. In contrast, the largest PCR fragment corresponded to a single trimer. All multimers contained a variety of elements that, like B2, are members of Tourist subfamily A (Bureau and Wessler, 1994a).

To exclude the possibility that the Tourist multimers were artifacts of PCR amplification, we used dimer and trimer products to probe a small insert library derived from B79 genomic DNA. Three of 11 sequenced clones contained Tourist multimers, thus confirming the presence of Tourist multimers in the genome.

**Insertion into Preexisting MITEs**

Among maize inbred lines, the insertion sites of MITEs frequently were polymorphic with respect to the presence or absence of an element at a particular locus (Casa et al., 2000; Zhang et al., 2000). Polymorphism of this type usually is associated with the recent spread of transposon families through the genome. In light of these findings, we designed a PCR assay to detect insertion site polymorphism within Tourist multimers. In this way, evidence might be obtained for the sequential insertion of one element into another. The locus harboring the Tourist trimer (Figure 1) was investigated for possible insertion polymorphism among different maize lines. Following the methodology described in Methods, we obtained B79 genomic sequence adjacent to one end of the trimer, revealing that another Tourist element (Tourist-Zm22) had inserted adjacent to the trimer with only an intervening target site duplication (TSD) (Figure 2). A locus-specific primer was designed from the sequence flanking Tourist-Zm22 and used to amplify B37 genomic DNA together with a B2 terminal primer (PB2r). The resulting PCR product, which harbored an additional Tourist element (Tourist-Zm3) (Figure 2), provided evidence for the progressive formation of multimers (tetramers from trimers).

**Nonrandom Insertion Sites**

Insertion sites within the sequenced multimers clearly were nonrandom. For ease of comparison, insertion sites have been calculated as the number of base pairs from the closest end of the target element to the first nucleotide of the TIR of the insertion element. For all insertions examined, this value corresponded to 27, 37, or 47 bp (Figure 1). To determine whether this periodicity was representative of the multimers in the maize genome, a two-step PCR assay was used to isolate additional multimers. In this assay (see Methods), the length of the PCR products reflects the position of the insertion sites within the multimers. That is, if the insertion sites are 10 bp apart, the PCR products will appear, more or less, linearly in the gel.
less, as a 10-bp “ladder” on the gel. Such a ladder was observed (Figure 3). Furthermore, sequencing of selected PCR products revealed that all contained a Tourist-Zm3 element inserted into another Tourist element at ∼10-bp intervals. The composition of some of the multimers is diagrammed in Figure 3. In addition, these data and the data from all previous multimer sequences are summarized in Table 1.

**MITE Multimers in Rice**

In the absence of a significant amount of maize genomic sequence, analysis of maize multimers is restricted to a description of the phenomenon and the characterization of a small fraction of the existing elements. A more thorough survey, including an estimate of the proportion of the multimers present, is possible for rice because a large amount of rice genomic sequence is available publicly (Yuan et al., 2001) and the rice genome contains thousands of MITEs (Mao et al., 2000; Tarchini et al., 2000).

**Prevalence of Multimers**

Computer searches were restricted to 30.2 Mb of complete bacterial artificial chromosome and P1-derived artificial chromosomes (PAC) sequences, of which 6.6 Mb was derived from pericentromeric regions (based on the most recent data on the location of rice centromeres; Harushima et al., 1998; Cheng et al., 2001). No significant differences were observed in the insertion patterns of MITEs between the sequences from chromosomal arms and those from

---

**Figure 3.** Autoradiograph of PCR Products Resolved on an Acrylamide Gel Showing an ∼10-bp Ladder.

The positions of primers used to obtain the PCR products from maize lines or plasmid DNA are indicated by horizontal arrows; the composition of multimers represented by the indicated PCR products also is diagrammed. Vertical arrows over some multimers represent Zm3 insertions. Lane 1, Zm3-B2 dimer containing plasmid; lane 2, B79 genomic DNA; lane 3, B73 genomic DNA; lane 4, recombinant inbred DNA from a B73 × Mo17 mapping population. The asterisk indicates that the primer was labeled with 32P.
pericentromeric regions (data not shown). Of the rice sequences queried, 6641 MITEs were detected (Table 2) with the RepeatMasker program (see Methods for details). This corresponds to 0.22 MITEs per kb of genomic DNA or 1 MITE per 4.5 kb. MITEs account for 1.54 Mb of DNA or 5.1% of the genomic sequence analyzed. These values are very close to those found in a previous study of MITEs from a 350-kb contig (Tarchini et al., 2000). MITEs grouped into 41 different families, of which 26 were reported previously and 15 were identified in this study (Bureau and Wessler, 1994a, 1994b; Bureau et al., 1996; Song et al., 1998; Zhang and Kochert, 1998; Tarchini et al., 2000; Turcotte et al., 2001) (see supplemental data for sequences of the new identified MITE families).

Of the 6641 MITEs, 732 (or 11%) are part of 340 multi-mers. These include 293 dimers, 35 trimers, nine tetramers, and 215 pentamers and higher-order multi-mers. The insertion sites of the MITEs are listed in Table 1.

Table 1. Insertion Sites within Tourist Dimers and Trimers

<table>
<thead>
<tr>
<th>Maize Line</th>
<th>Preexisting Element</th>
<th>Insertion Element</th>
<th>Insertion Site (bp)</th>
<th>TSD</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>B79</td>
<td>B2</td>
<td>Zm3</td>
<td>37</td>
<td>TGA</td>
<td>B2 PCR</td>
</tr>
<tr>
<td>B79</td>
<td>B2</td>
<td>Zm11</td>
<td>37</td>
<td>TAA</td>
<td>B2 PCR</td>
</tr>
<tr>
<td>B79</td>
<td>B2</td>
<td>Zm29</td>
<td>37</td>
<td>TGA</td>
<td>B2 PCR</td>
</tr>
<tr>
<td>B79</td>
<td>B2</td>
<td>Zm22</td>
<td>27</td>
<td>TTA</td>
<td>B2 PCR</td>
</tr>
<tr>
<td>B79</td>
<td>B2&lt;sup&gt;a&lt;/sup&gt; (1)</td>
<td>Zm3&lt;sup&gt;b&lt;/sup&gt; (2)</td>
<td>37</td>
<td>TGA</td>
<td>B2 PCR</td>
</tr>
<tr>
<td>B79</td>
<td>Zm&lt;sup&gt;b&lt;/sup&gt; (2)</td>
<td>Zm3&lt;sup&gt;b&lt;/sup&gt; (3)</td>
<td>47</td>
<td>TTA</td>
<td>B2 PCR</td>
</tr>
<tr>
<td>B73</td>
<td>B2</td>
<td>Zm3</td>
<td>27</td>
<td>TTA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>RIL22&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Zm3</td>
<td>Zm3</td>
<td>46</td>
<td>TAA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>B73</td>
<td>Zm3</td>
<td>Zm3</td>
<td>47</td>
<td>TAA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>B73</td>
<td>Zm3</td>
<td>Zm3</td>
<td>57</td>
<td>TTA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>B73</td>
<td>Zm13</td>
<td>Zm3</td>
<td>67</td>
<td>TTA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>B73</td>
<td>Zm24</td>
<td>Zm3</td>
<td>47</td>
<td>TCA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>B79</td>
<td>B2</td>
<td>Zm3</td>
<td>37</td>
<td>TGA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>RIL7</td>
<td>Zm24</td>
<td>Zm3</td>
<td>37</td>
<td>TGA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>Spanco</td>
<td>Zm13</td>
<td>Zm22</td>
<td>67</td>
<td>TTA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>Spanco</td>
<td>Zm24</td>
<td>Zm22</td>
<td>59</td>
<td>TAA</td>
<td>Two-step PCR</td>
</tr>
<tr>
<td>B79</td>
<td>B2</td>
<td>Zm28</td>
<td>58</td>
<td>TAA</td>
<td>Genomic library</td>
</tr>
<tr>
<td>B79</td>
<td>Zm13</td>
<td>Zm25</td>
<td>46</td>
<td>TGA</td>
<td>Genomic library</td>
</tr>
<tr>
<td>B79</td>
<td>Zm27&lt;sup&gt;b&lt;/sup&gt; (1)</td>
<td>Zm3&lt;sup&gt;b&lt;/sup&gt; (2)</td>
<td>47</td>
<td>TGA</td>
<td>Genomic library</td>
</tr>
<tr>
<td>B79</td>
<td>Zm3&lt;sup&gt;b&lt;/sup&gt; (2)</td>
<td>Unknown element&lt;sup&gt;b&lt;/sup&gt; (3)</td>
<td>47</td>
<td>TAA</td>
<td>Genomic library</td>
</tr>
</tbody>
</table>

<sup>a</sup> Insertion site is defined as the number of base pairs from the closest end of the target element to the first nucleotide of the TIR of the insertion element.
<sup>b</sup> Elements that form trimers. The numbers in parentheses indicate the order of insertion. Element 1 is the first preexisting element; element 2 is the element that inserted into element 1; element 3 is the element that inserted into element 2.
<sup>c</sup> RIL represents recombinant inbred lines from the B73 × Mo17 cross.

Table 2. Multimers Containing Rice MITEs

<table>
<thead>
<tr>
<th>Insertion of MITEs into</th>
<th>Insertion of Other Elements into</th>
</tr>
</thead>
<tbody>
<tr>
<td>MITEs</td>
<td>LTR Elements</td>
</tr>
<tr>
<td>Number of elements (kb) DNA</td>
<td>6,641</td>
</tr>
<tr>
<td>Insertion events</td>
<td>1,540</td>
</tr>
<tr>
<td>Insertion (kb)</td>
<td>0.251&lt;sup&gt;a&lt;/sup&gt; (1/4.0 kb)</td>
</tr>
</tbody>
</table>

<sup>a</sup> P < 0.05 compared with that for all genomic sequences by χ² test.
<sup>b</sup> P < 0.01 compared with that for all genomic sequences by χ² test.
<sup>c</sup> Indicates all sequences contain both elements and non-element sequences.
and three pentamers (the trimers and tetramers also contain non-MITE elements). These 387 MITEs inserted into other MITEs correspond to 387 MITEs per 1540 kb of MITEs, or an insertion frequency of MITEs into MITEs of 0.25 per kb or one MITE per 4 kb (Table 2). In contrast, there are very few insertions of MITEs into class 1 elements or into other class 2 elements, despite the fact that these elements constitute a much larger fraction of the genome. Although there is one MITE inserted per 4 kb of MITE DNA, there is only one MITE inserted per 330 kb of LTR retrotransposons and per 127 kb of other class 2 elements. These data indicate either a target site preference of MITEs for other MITEs or that MITE amplification preceded the amplification of the other elements in the genome. In the latter situation, it is envisioned that the bulk of the class 1 and non-MITE class 2 elements were not in the genome when most of the MITE families were undergoing amplification. In contrast, non-MITE elements show no discrimination for insertion into MITEs (Table 2); while the frequency of insertion into MITEs is one per 17 kb of MITEs, the insertion frequency into all genomic DNA is slightly higher at one per 14 kb.

### Table 3. Self-Insertion Preference of Some MITEs in Rice

<table>
<thead>
<tr>
<th>Elements</th>
<th>Castaway</th>
<th>Gaijin</th>
<th>Ditto</th>
<th>Stowaway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copies of the element in 30.2 Mb of genomic sequence</td>
<td>225</td>
<td>486</td>
<td>345</td>
<td>2690</td>
</tr>
<tr>
<td>Total size (kb)</td>
<td>76</td>
<td>71</td>
<td>89</td>
<td>561</td>
</tr>
<tr>
<td>Insertions by MITEs</td>
<td>29</td>
<td>23</td>
<td>53</td>
<td>104</td>
</tr>
<tr>
<td>Self-insertions</td>
<td>20 (69%&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>16 (70%&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>13 (25%&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>77 (71%&lt;sup&gt;a&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Insertion frequency by all MITEs&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.38&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.32</td>
<td>0.60&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.19</td>
</tr>
<tr>
<td>Self-insertion frequency&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.26</td>
<td>0.23</td>
<td>0.15</td>
<td>0.14</td>
</tr>
</tbody>
</table>

<sup>a</sup> Percentage of self-insertions.
<sup>b</sup> Insertion frequency equals insertions into the element divided by the total size (kb) of the element.
<sup>c</sup> P < 0.01 compared with the average insertion frequency (0.223) in the genome by χ² test.

The data presented in Table 2 reveal a slight preference for insertion of MITEs into MITEs. However, analysis of these data for individual MITE families indicates that this preference is not displayed by all families and is attributable largely to self-insertions. Four MITE families were analyzed in detail (Table 3). These families were chosen because they are abundant and they represent different groups of MITEs. Among the four families analyzed, Castaway, Gaijin, and Ditto are related to Tourist elements in maize, whereas Stowaway elements belong to another superfamily (see Discussion). As shown in Table 3, all of the Tourist-related elements have sustained more insertions per kb of DNA than has the genome as a whole (insertion frequencies of 0.38, 0.32, and 0.60, respectively, versus 0.22 for all DNA; Tables 2 and 3). In contrast, Stowaway has sustained insertions at approximately the same frequency as the rest of the genome. Although the increased insertions into Castaway and Gaijin elements can be accounted for completely by self-insertions, Ditto elements appear to attract a variety of

**Figure 4. A Cluster of Six Castaway Elements on a Rice P1-derived Artificial Chromosomes (PACs) Clone (GenBank Accession Number ap002844) from Chromosome 1.**

The distance from the end of the element to the insertion site is shown. Castaway-2 is a member of a subfamily of Castaway. The +3 indicates the inclusion of the 3-bp TSD that was generated upon insertion.
MITE Multimers Cannot Transpose

A MITE multimer can arise in at least two ways. The first is by the insertion of a MITE into another MITE, and the second is by amplification of a multimer. If a multimer is capable of transposition, several copies of the same multimer should be detected and the multimers should evolve similarly as single elements. Furthermore, these copies should be composed of the same elements in the same relative orientation and with the same insertion site and TSD. Among the 340 MITE multimers identified in this study, only three pairs of dimers share these structural features. However, the sequence similarity between the members of each dimer pair ranges from 65 to 72%, whereas at least one of the insertion elements in each dimer pair has homologs with >90% similarity in the same database. This striking discrepancy suggests that these dimer pairs resulted from independent insertions instead of amplification of dimers.

There was one exception involving a dimer composed of a MITE and a DNA element. This element (called Midway), initially found as an 850-bp insertion in a Stowaway-Os1 element, has 11-bp TIRs and an 8-bp TSD. A closer examination indicates that Midway harbors another Stowaway element (Stowaway-Os25). That there are three Midway/Stowaway composite elements in the database sharing 93 to 96% overall DNA sequence identity suggests that Midway can still transpose despite (or because of) the Stowaway-Os25 insertion.

DISCUSSION

Here, we report the characterization and quantification of MITE multimers in maize and rice. Although MITE multimers were first discovered in maize, limited genomic sequence precluded further analysis of these multimers. However, the high density of MITEs in the rice genome (Bureau et al., 1996; Mao et al., 2000; Tarchini et al., 2000) coupled with the availability of large amounts of genomic sequence facilitated a more comprehensive analysis of multimers in rice and has led to the following conclusions: (1) MITEs are numerically the most abundant transposable elements in the rice genome (one MITE per 4.5 kb); (2) >10% of rice MITEs are part of multimers, thus suggesting a preference for MITE insertion into MITEs; (3) an insertion preference is displayed by some, but not all, MITE families; (4) for the Castaway and Gaijin families, this preference is caused by a high frequency of self-insertions; in contrast, Ditto elements are targeted by many element families; (5) the frequency of MITE insertions into class 1 or other class 2 elements is surprisingly low; and (6) on the basis of our analysis of 30.2 Mb of rice sequence, nested MITE multimers arise from independent insertion events.

Self-Insertion Preference for Some MITE Families

As calculated in Table 2, the insertion frequency of all MITEs into other MITEs is slightly higher than the average value into the whole genome. However, there is a threefold variation in the frequency of MITE insertions into MITEs when individual families are examined (Table 3). More significantly, self-insertions constituted a major part of the multimers for several families. For Castaway, Gaijin, and Stowaway, self-insertions account for two-thirds of all insertions. These data indicate that the preferential insertion of MITEs into MITEs that is displayed by some families can be attributed, to a great extent, to self-insertions. One exception is the Ditto element. Among the rice MITEs, Ditto elements are targeted frequently by various types of elements, including other Ditto elements. In addition to being targeted 53 times by 12 families of MITEs, we detected five cases of insertions into flanking sequences.

Figure 5. The Frequencies of Self-Insertions and Insertions into Flanking Sequences for Four MITE Families in Rice.

Frequencies were calculated in the same way as in Table 2 (see Methods for details).
by four different LTR retrotransposons and 22 examples of MITEs inserted in adjacent (with an intervening TSD) sequences.

Composite elements, arising from self-insertion, have been reported previously in maize, in which double Ds and Ac elements were shown to be responsible for chromosome breakage and more complex rearrangements (McClintock, 1949; Courage-Tebbe et al., 1983; Döring and Starlinger, 1984; Weck et al., 1984; Döring et al., 1989; Michel et al., 1994). It was later hypothesized that chromosome breakage resulted from aberrant transposition of composite or adjacent Ds elements (English et al., 1993; Weil and Wessler, 1993). In contrast to the composite Ds elements that are still capable of transposition, the uniqueness of each MITE multimer suggests that self-insertion creates an inactive composite element. Inactivating self-insertions of the Tp1 element of Physarum polycephalum have been observed previously (Rothnie et al., 1991). It has been proposed that a preference for inactivating self-insertions minimizes deleterious effects on the host by providing a safe haven for insertion while simultaneously limiting the overall transposition frequency (Rothnie et al., 1991).

Regional versus Self-Insertion Preference

Previous studies indicate that some MITE families insert preferentially into genic regions (Mao et al., 2000; Zhang et al., 2000). A preference for genic regions also has been observed for the maize class 2 families Ac/Ds and Mutator (Chen et al., 1992; Cresse et al., 1995). Regional preferences have been demonstrated for many elements in a wide variety of species. For example, yeast Ty5 elements integrate preferentially into regions of silent chromatin at the telomeres and the mating loci (Zou et al., 1996), and for P elements, euchromatic sites, especially 5' regions of genes, are targeted more often than heterochromatin (Berg and Spradling, 1991; Liao et al., 2000).

Regardless of the mechanism responsible, an element with a regional preference is more likely to have a higher frequency of self-insertion than an element with no such preference. If the regional preference is the major factor leading to a high self-insertion frequency, comparable insertion frequencies are expected into elements and into their flanking genomic sequences. The availability of 30.2 Mb of rice sequence allowed us to test this assumption (Figure 5). For Castaway, Gaijin, and Ditto, the self-insertion preference is more likely to be caused by the targeting of preexisting elements than by a regional preference. In contrast, Stowaway elements show no significant difference between insertion into preexisting elements and insertion into flanking DNA, thus suggesting that the high ratio of self-insertions results from a regional preference. Alternatively, the presence of one Stowaway element may alter the flanking DNA in some manner, thereby creating a better target for future insertions. A similar effect was observed for the in vitro transposition of the C. elegans Tc1 element (Ketting et al., 1997). Interestingly, Stowaway elements, like Tc1, use TA dinucleotide targets.

The difference between Stowaway and the three other MITE families may indicate distinct integration mechanisms for different MITE families. Like the Tourist elements in maize, Castaway, Ditto, and Gaijin all create a 3-bp TSD upon insertion (Bureau et al., 1996). More importantly, the TIRs of Castaway, Ditto, and Gaijin are related to the TIR of Tourist elements in maize, suggesting that they may belong to the same superfamily. In contrast, Stowaway elements appear to belong to another superfamily based on their TIR and TSD (Bureau and Wessler, 1994b). Therefore, it is likely that these two superfamilies rely on distinct sources of transposases.

Target Site Preference in Maize Multimers

The potential to form secondary structures has been noted for several MITE families since the discovery of the Tourist family in maize (Bureau and Wessler, 1992; Izsvák et al., 1999). Given the occurrence of multimers among maize Tourist-B2 elements, we hypothesized that secondary structures might play a role in targeting. Consistent with this notion is the inability to detect MITE multimers involving two other maize MITE families (Hbr and mPf) lacking significant secondary structures (N. Jiang, Q. Zhang, X. Zhang, and S.R. Wessler, unpublished data). However, in the rice genome, multimer formation does not correlate with the potential to form significant secondary structures. In rice, the MITEs that sustained most insertions, Castaway and Ditto, are those without significant secondary structures (Bureau et al., 1996). In contrast, Stowaway elements usually have significant secondary structures but do not show a targeting bias. However, these data cannot rule out the possibility that small, local stem loops, such as the 14-bp palindrome targeted by P elements (Liao et al., 2000), might influence targeting of MITEs.

The analysis of MITE multimers in rice also was prompted by the discovery of nonrandom insertion sites among Tourist multimers in maize (Figures 1 and 2, Table 1). The 10-bp periodicity observed for Tourist multimers is reminiscent of the integration of human immunodeficiency virus. Integration of human immunodeficiency virus in vitro occurs preferentially into bent DNA in which the major groove is on the exposed face of the nucleosome (Pryciak and Varmus, 1992; Pruss et al., 1994). The 10-bp periodicity for Tourist multimers could be produced in a similar pattern (i.e., the transposition machinery attacks only major or minor grooves of the DNA double helix).

In rice, some “hot” spots for insertion were observed inside the sequence of some MITEs, and some of the insertion sites are ~10 bp apart. However, insertions that are not 10 bp apart also were observed. Because of the fact that the rice MITEs that sustained most insertions are much larger
than maize Tourist elements (maize Tourist, 130 bp; Ditto, 244 bp; Castaway, 364 bp), the distribution of insertion sites appears to be sporadic within rice MITEs. Thus, more rice multimers need to be examined to determine whether or not the 10-bp pattern is statistically significant. Alternatively, this feature may belong only to Tourist elements in maize.

To date, no autonomous element responsible for the transposition of MITEs has been available. The isolation of such elements and their associated protein(s) will ultimately facilitate the biochemical analysis of the various levels of targeting exhibited by MITE families.

Deficiency of MITE Insertions into Non-MITE Elements: Targeting Preference or Temporal Differences in Amplification?

A surprising and dramatic conclusion of the data presented in Table 2 is that MITEs have inserted into MITEs 80 times more often than have inserted into LTR retrotransposons and 32 times more often than they have inserted into other DNA elements (one MITE insertion versus 4, 330, and 127 kb, respectively). In contrast, the frequency of insertion of LTR retrotransposons and other DNA elements into MITEs is only slightly lower than the overall frequency of insertion of these elements into rice genomic DNA (one insertion per 17 kb of MITEs versus one insertion per 14 kb of genomic DNA).

Previous studies have noted a genic preference for maize class 2 elements, including members of the Ac/Ds and Mu-tator families (Chen et al., 1992; Cresse et al., 1995). Differences in chromatin density and/or the extent of DNA methylation between gene-rich and other regions of the genome have been proposed as possible target recognition mechanisms (Chen et al., 1992). A similar preference for genic regions has been demonstrated for members of the MITE families Hbr and mPiF (Casa et al., 2000; Zhang et al., 2000; X. Zhang, N. Jiang, and S.R. Wessler, unpublished data). In contrast, MITEs appear to be underrepresented in regions of the maize and barley genomes containing nested or clustered LTR retrotransposons (Tikhonov et al., 1999; Dubcovsky et al., 2001). Although MITEs may target gene-rich regions by the same or similar mechanisms as other class 2 elements, the analysis of MITE multimers in rice provides at least two alternate explanations for the observed (skewed) distribution. Enrichment for MITEs in genic regions and their apparent absence from retrotransposon clusters or domains could reflect a self-insertion preference coupled with avoidance of retrotransposon targets. Alternatively, a dearth of MITE insertions into non-MITE transposons also would result if the bulk of MITE amplification occurred before the amplification of LTR retrotransposons and other class 2 elements. To unambiguously distinguish between these seemingly mutually exclusive hypotheses, it will be necessary to identify an active MITE system that can be exploited to experimentally determine MITE target preference(s). In the mean time, we must rely on the comparative analysis of related genomes to provide clues to the mechanisms underlying the observed distributions of transposable elements.

METHODS

Plant Material, DNA Extraction, and Library Construction

Maize (Zea mays) lines B79 and B37 were obtained from the U.S. Department of Agriculture, Agricultural Research Service Plant Introduction Station at Ames, Iowa. Maize line B73 and recombinant inbred lines from a cross between B73 and Mo17 were provided by Michael Lee (Iowa State University, Ames). Maize line Spanco was provided by Andy Tull (University of Georgia, Athens). Plant DNA was extracted as described (McCouch et al., 1988). The small insert genomic library from B79 genomic DNA was constructed as described (Zhang et al., 2000).

Polymerase Chain Reaction and Gel Electrophoresis

Polymerase chain reaction (PCR) was performed as described (Bureau and Wessler, 1992) with annealing temperature ranging from 55 to 60°C, depending on the primers. Sequences of primers are available on request.

To clone the flanking sequence of the Tourist trimer in Figure 1, B79 genomic DNA was digested with Msel and ligated with adapters. The DNA then was amplified with a primer complementary to the adapter and primer P8, which contains the sequence at the junction of (Tourist) Zm3 and the B2-like element (Figure 2). To separate PCR products that resulted only from adapters and PCR products from the two primers, primer P8 was labeled with 32P, and the PCR products were loaded on 6% denaturing acrylamide-bisacrylamide gels and electrophoresed as described previously (Casa et al., 2000).

The two-step PCR assay described in Figure 3 involved amplification of genomic DNA with primers P1 and P2, followed by amplification of the PCR products with primers P2 and P3 (P3 was labeled with 32P). PCR products were resolved by PAGE, as described above.

Recovery of Gel Bands

DNA fragments were excised from radioactive gels by scratching the dried gel with yellow tips (Stumm et al., 1997; Elsevier Trends Journals Technical Tips online, http://tto.biomednet.com/cgi-bin/tto/pr), placing the tip in 20 μL of PCR reaction mix with relevant primers for 1 min before discarding, and reamplifying with the same cycling parameters as that of the original reaction. PCR products were resolved on 0.8% agarose gels, and fragments were excised, purified (QiAquick; Qiagen, Chatsworth, CA), and cloned (TA cloning kit; Invitrogen, Carlsbad, CA). DNA templates were sequenced at the Molecular Genetics Instrumentation Facility (University of Georgia).

DNA Sequence Analysis

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

Retrieval of Sequences

Completely sequenced rice (Oryza sativa) bacterial artificial chromosomes and P1-derived artificial chromosomes (PACs) were retrieved from the World Wide Web sites of different rice genomic projects, including groups in the United States (http://www.usricgenome.org/), Japan (http://rgp.dna.afric.go.jp/), Korea (http://bioserve.myongji.ac.kr/ricemac.html), the People’s Republic of China (http://www.ncgr.ac.cn/Ls/index.html), and Taiwan (http://genome.sinica.edu.tw/).

Screening for Transposable Elements

Transposable elements in rice sequences were searched with Repeat-Masker (http://ftp.genome.washington.edu/RM/webrepeatmaskerhelp.html). The grass repeats database in RepeatMasker was modified by adding sequences of other previously characterized transposable elements in maize and rice (references not listed in Results: Hirochika et al., 1992, 1996; SanMiguel et al., 1996; Kumekawa et al., 1999; Ohsubo et al., 1999) and new transposable elements identified in this study. New elements were found either by their similarity to known elements or by insertion into known elements. The rice genome sequences described above were used as query sequences in analysis with RepeatMasker using the modified grass repeats database at default settings. In the output of RepeatMasker, the annotation files display all of the matches and the positions of matches between the query sequences and any of the sequences in the repeats database.

Identification of Multimers

Potential multimers were first selected from the query sequences on the basis of the distance between two elements in the annotation files. For example, if one element is flanked by another element on both sides, the two elements probably form a dimer. The sequences of potential multimers were further analyzed manually with programs in GCC. If the ends of one element were located inside the sequence of another element, the elements were deduced to constitute a multimer; otherwise, elements were deduced to be monomeric.

Calculations

The insertion frequency was calculated by dividing the number of insertions by the size (in kilobases) of available sequences. For individual MITE families, the amount of DNA equals the size of the consensus element multiplied by the number of elements. If the length of the match was less than half of the consensus element, it was considered as half an element in calculating the amount of DNA. If a match was <30 bp, it was eliminated from consideration. The total amount of long terminal repeat (LTR) retrotransposons was approximated by multiplying the number of elements and solo-LTRs by their average lengths, which are 6.9 and 1.8 kb, respectively. The total amount of DNA representing DNA elements was estimated similarly, with an average size of 1.9 kb. The average size of LTR elements and other DNA elements was obtained by sampling an 880-kb region in chromosome 1 (71.8 to 73.5 cm).

In Figure 5, the length of flanking sequences was estimated by the number of elements multiplied by 2 and then by the range of flanking sequences, where 2 represents the fact that for each element there are flanking sequences on both sides. For example, 2690 Stowaway elements were detected in the 30.2-Mb rice genomic sequence, and 359 Stowaway insertions were observed in the range of 1.0 to 2.0 kb from another Stowaway element. In this case, the total length of available sequences = 2690 × 2 × (2.0 – 1.0) = 5380 kb, and the insertion frequency in this range of flanking sequences = 359 ÷ 5380 = 0.067 insertion per kb. Because the purpose of the analysis is to determine whether the high self-insertion frequency for some MITE families is caused by the targeting of preexisting elements or by a regional preference, adjacent insertions (only one target site duplication [TSD] between two elements) were not included. This type of insertion was not considered because it is not clear whether it is caused by the targeting for preexisting elements or for flanking sequences.

ACKNOWLEDGMENTS

We thank Arian Smit (Institute for Systematic Biology, Seattle, WA) and Phil Green (Washington University, St. Louis, MO) for providing the Repeat-Masker and cross_match programs, Zhirong Bao (Washington University) for valuable suggestions and discussions, Cedric Feschotte and Xiaoyu Zhang for critical reading of the manuscript, Alexander Nagel for communicating unpublished data, and Qiang Zhang and Liangjiang Wang for technical assistance. This study was supported by grants from the National Institutes of Health, the U.S. Department of Energy, and the National Science Foundation to S.R.W.

Received June 8, 2001; accepted August 22, 2001.

REFERENCES


