Theor Appl Genet (1994) 87:657–667

The generation of *Mutator* transposable element subfamilies in maize

J. L. Bennetzen, P. S. Springer

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received: 8 March 1993 / Accepted: 17 May 1993

Abstract. The mobile DNAs of the *Mutator* system of maize (*Zea mays*) are exceptional both in structure and diversity. So far, six subfamilies of *Mu* elements have been discovered; all *Mu* elements share highly conserved terminal inverted repeats (TIRs), but each sub-family is defined by internal sequences that are apparently unrelated to the internal sequences of any other *Mu* subfamily. The *Mu1/Mu2* subfamily of elements was created by the acquisition of a portion of a standard maize gene (termed MRS-A) within two *Mu* TIRs. Beside the unusually long (185–359 bp) and diverse TIRs found on all of these elements, other direct and inverted repeats are often found either within the central portion of a *Mu* element or within a TIR.

Our computer analyses have shown that sequence duplications (mostly short direct repeats interrupted by a few base pairs) are common in non-autonomous members of the *Mutator*, *Ac/Ds*, and *Spm(En)* systems. These duplications are often tightly associated with the element-internal end of the TIRs. Comparisons of *Mu* element sequences have indicated that they share more terminal components than previously reported; all subfamilies have at least the most terminal 215 bp, at one end or the other, of the 359-bp *Mu5* TIR. These data suggest that many *Mu* element subfamilies were generated from a parental element that had termini like those of *Mu5*. With the *Mu5* TIRs as a standard, it was possible to determine that elements like *Mu4* could have had their unusual TIRs created through a three-step process involving (1) addition of sequences to interrupt one TIR, (2) formation of a stem-loop structure by one strand of the element, and (3) a subsequent DNA repair/gene conversion event that duplicated the insertion(s) within the other TIR. A similar repair/conversion extending from a TIR stem into loop DNA could explain the additional inverted repeat sequences added to the internal ends of the *Mu4* and *Mu7* TIRs. This same basic mechanism was found to be capable of generating new *Mu* element subfamilies. After endonucleolytic attack of the loop within the stem-loop structure, repair/conversion of the gap could occur as an intermolecular event to generate novel internal sequences and, therefore, a new *Mu* element subfamily. Evidence supporting and expanding this model of new *Mu* element subfamily creation was identified in the sequence of MRS-A.

Key words: Transposable elements – Mutation – Evolution – DNA repair – Gene conversion

Communicated by H. K. Dooner
Correspondence to: J. L. Bennetzen

Introduction

The *Mutator* (*Mu*) transposable element system of maize is exceptional in many of its most basic attributes. The system, only recently discovered by Robertson (1978), conditions a very high forward mutation rate in any maize line in which it is active. This high mutagenicity, otherwise known as *Mutator* activity, has only been observed in plants derived from one maize line (Bennetzen et al. 1993). It is not known why the high activity of this system arose or was maintained only in this one line. Mutant alleles isolated from *Mutator* stocks are generally associated with the insertion of various 1-kb to 5-kb elements (reviewed in Bennetzen et al. 1993 and in Chandler and Hardeman 1992). The *Mu* transposable elements all induce

9-bp-long flanking direct duplications of host DNA and also have the same approximately 200-bp terminal inverted repeats (TIRs). However, the internal sequences of these elements fall into at least six classes that are unrelated at a hybridizational or DNA sequence level. The origin of these "subfamilies" of *Mu* elements has not yet been explained.

Some *Mu* elements can transpose at germinal frequencies exceeding an average of once per element per generation (Alleman and Freeling 1986; Bennetzen et al. 1987; Hardeman and Chandler 1989). Although a high percentage of the mutations isolated from *Mutator* stocks exhibit somatic instability (Robertson 1978), germinal reversion of these mutations is generally rare (Britt and Walbot 1991; Brown et al. 1989). Hence, a strong causal correlation between excision and transposition has not been observed.

The multiple *Mu* element subfamilies are variable in dispersal, as well as in internal structure. The first *Mutator* components cloned, the related elements *Mu1* and *Mu2*, are both the most active and the most limited in their range within the *Zea* (reviewed in Bennetzen et al. 1993). The copy number of *Mu1*/*Mu2* elements is much lower in maize lines lacking *Mutator* activity than it is in the rare maize lines that exhibit *Mutator* activity (0–4 versus 10–60) (Bennetzen 1984; Bennetzen et al. 1987; Chandler et al. 1986). Other *Mu* element subfamiles, like *Mu4* or *Mu5*, exhibit a somewhat higher copy number in *Mutator* lines than in standard maize lines, but they appear to have a few (4–12) copies in most maize lines (Talbert et al. 1989). The higher activity of *Mu1*/*Mu2* elements, their limited dispersal, and their relatively low degree of internal and terminal sequence variability have suggested that the *Mu1*/*Mu2* subfamily has been more recently generated than any other *Mu* subfamily (Bennetzen et al. 1993). In this regard, Chandler and coworkers have identified and cloned a sequence, *Mu*-Related Sequence A (MRS-A), with extensive internal homology to *Mu1* and *Mu2* (Talbert and Chandler 1988). MRS-A has no *Mutator* inverted termini and is a portion of a *Mu*-unrelated gene. *Mu2* was apparently generated by the 'capture' of a segment of chromosomal DNA by *Mu* ends. Similarly, Oishi and coworkers (K. Oishi, personal communication) have identified and cloned sequences related to the *Mu3* internal region but without *Mu* TIRs.

Recently, genetic and molecular evidence has accumulated to indicate the existence of an 'autonomous' *Mutator* element (Chomet et al. 1991; Qin and Ellingboe 1990; Robertson and Stinard 1989). This element can both transpose and transactivate the transposition of other *Mu* elements. This autonomous element, or elements structurally quite similar to it, has been cloned (Chomet et al. 1991; Hershberger et al. 1991; Qin et al. 1991). By sequence criteria, the putative autonomous element *Mu9* (Hershberger et al. 1991) contains some internal sequences unrelated to any previously identified *Mu* element.

Through detailed structural analyses of the sequenced *Mu* elements, we have been able to develop a consistent model for the generation and variation of new *Mu* element subfamilies. Our analyses also uncovered evidence supporting this model of element generation by DNA gap repair and gene conversion.

Materials and methods

Sequence information

The DNA sequences of *Mu1*, *Mu2*(*Mu1.7*), *Mu4*, *Mu5*, *Mu7*(*rcy*), *Mu8*, *Adh1*, *Ac*, *Spm*(*En*), *Ds*(*wx-m5*), *Ds*(*bz-m4*), *Ds2*, and *dSpm*(*I*) have all been published (Barker et al. 1984; Dennis et al. 1984; Fleenor et al. 1990; Gierl et al. 1985; Klein et al. 1988; Merckelbach et al. 1986; Muller-Neumann et al. 1984; Pereira et al. 1986; Pohlman et al. 1984; Schnable et al. 1989; Talbert et al. 1989; Taylor and Walbot 1987; Weil et al. 1992) and were acquired through the GenBank and EMBL databases. The DNA sequence of *Mu3* was kindly provided by K. Oishi, University of Arizona, Tucson, Arizona. The DNA sequence of *Mu9* was generously supplied prior to publication (Hershberger et al. 1991) by V. Walbot, Stanford University, Palo Alto, California.

Software

DNA sequence information was studied by use of the Genetics Computer Group (GCG) package (Devereux et al. 1984), specifically programs COMPARE, DOTPLOT, and PILEUP. Access to the databases and the GCG package was provided by the Purdue AIDS Center Laboratory for Computational Biochemistry (ACLCB).

Direct or terminal inverted repeats were arbitrarily chosen as significant only if they were greater than 80% identical over at least 11 contiguous base pairs.

Results

Internal repeats are unusually abundant in non-autonomous plant transposable elements

Two of the many oddities of the *Mutator* system, when compared to other eukaryotic transposable element systems, are the relatively long (and nearly identical) terminal inverted repeats and the additional direct and inverted repeats associated with the *Mu* elements. Hence, in our sequence analyses, we not only compared each element to each other element (and their reverse sequence) but also compared each element to itself and the reverse of itself.

Our sequence analyses confirmed and extended the observations that both terminal and internal repeats are common in *Mutator* elements (Fig. 1, Table 1). The putative autonomous element *Mu9* is relatively deficient in extra repeats, containing three different types of repeats ("a", "i", and "r") located in or near the TIRs (Fig. 1, Table 1). Similar analyses of the maize *Ac* and

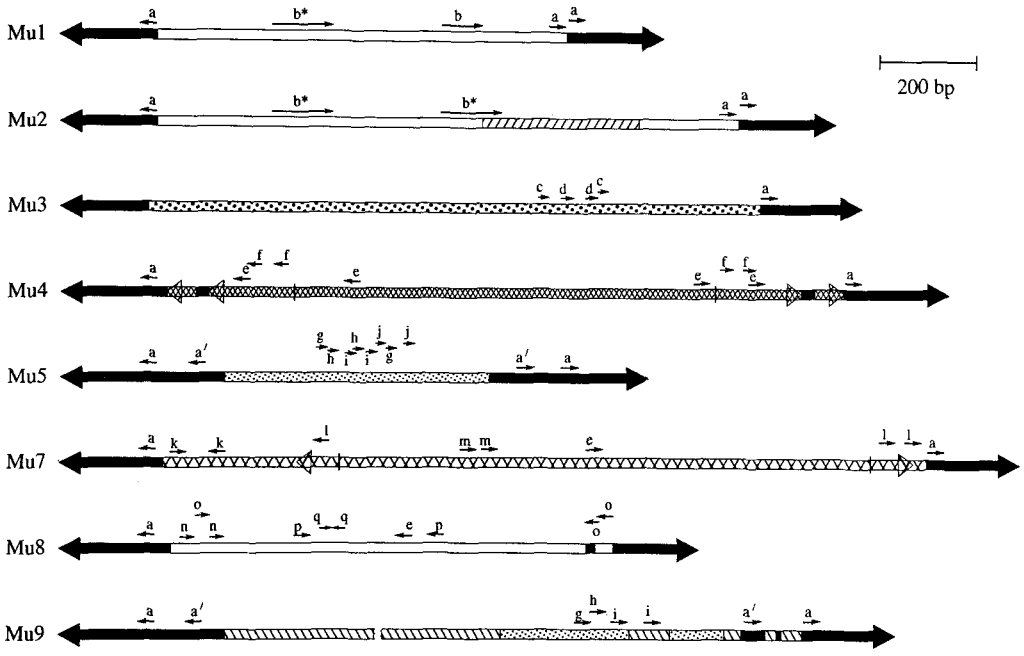


Fig. 1. Maps of the sequenced *Mutator* transposable elements. *Arrows within the bars* indicate the terminal inverted repeats (TIRs) of each element, while *arrows above the bars* indicate additional repeats. The *fill of the bar* indicates homology between sequences; the *solid fill* indicates homology to the TIRs of *Mu5*, which were defined as the reference *Mu* element TIRs. The orientation of other *Mu* element sequences was then chosen to optimize this homology. Hence, in the case of *Mu9*, the element is presented in reverse orientation relative to the published sequence (Hershberger et al. 1991). Only repeats within an element that exhibited 80% or greater identity over 11 or more contiguous basepairs are shown. The sequences of all of these repeats except "*a'*", "*b*" and "*b**" are shown in Table 1. The "*a*" repeats found in *Mu5* and *Mu9* TIRs are relatively degenerate versions of the "*a*" repeat, with only 12–18/24-bp homology with the consensus sequence shown in Table 1. The inverted repeats labelled "*r*" (see Table 1) are found in each TIR, but were left out in order to simplify the figure somewhat. Repeat "*s*" (see text and Table 1) is also not shown, since it was not repeated within any single element

Spm(En) transposable elements (Muller-Neumann et al. 1984; Pereira et al. 1986; Pohlman et al. 1984) and the maize *Adh1* gene (Dennis et al. 1984) did not indicate any extensive inverted or direct repeats other than the interrupted terminal inverted repeats of *Ac* and *Spm(En)* (data not shown). However, *Ds(wx-m5)*, *Ds(bz-m4)*, *Ds2*, and *dSpm(I)* elements (Gierl et al. 1985; Klein et al. 1988; Merckelbach et al. 1986; Weil et al. 1992) from maize did exhibit many additional repeats and, as with the *Mutator* elements, these extra repeats

Table 1. Nucleotide sequences of *Mu* element repeats

Repeat	Sequence
a	5'-A/TGAC/GA/GAGAA/GGAGTACGCC/GA/TGACGG-3'
c	5'-TGAACAGCTGGAGGA/TCAA-3'
d	5'-CGCAATATCCTGGACTGGGATACTCGTGA-3'
e	5'-AACCAGCTCATCGG-3'
f	5'-CTGTCG/CTCNAACG/CAGCT-3'
g	5'-CATCAATCCTCTGAACTTCCTATAT-3'
h	5'-ATTACAGACAATGCAGGAGCCTCCAA-3'
i	5'-ACTATTTTAAAAATTG/AAAGA-3'
j	5'-AGATG/AAACTTGGTG/ATACCAT-3'
k	5'-CGCTGTGGAGAAGAAGGAGCAGAGCGCC/GCTGG-3'
l	5'-GGCGCAGAGCCGAACCAGGGGCGAGCTCAAG-3'
m	5'-CTCAAAGCTTCTTTC-3'
n	5'-GCGCAGGTTGGCCA/TCG-3'
o	5'-CGTGTTCATGCTCTCG-3'
p	5'-CAG/TCA/CTCAGCGCCG-3'
q	5'-CCAGCGCGA/CGCG/CTGGC-3'
r	5'-TCT/CGTTTTGGAG-3'
s	5'-CGTCTGCA/GCC/AGACGCTGCG/CG/CG/TAGCCG-3'

were most abundant within and adjacent to the terminal inverted repeats (data not shown and Fig. 1). Hence, extensive sequence repetition beyond the inverted termini appears to be a common feature of non-autonomous maize elements, but not of autonomous elements of the same families. Of course, non-autonomous elements that have been simply derived from *Ac* or *Spm* by a single mutational (commonly, deletional) event (Fedoroff et al. 1983) have not accumulated significantly more duplications than the parental autonomous element.

Many of the additional duplications of sequences in *Mutator* elements are tightly associated with the TIRs, particularly with the inward-most ends of the TIRs. For instance, *Mu2* exhibits two copies of a 24-bp direct duplication ("a") separated by 9 bp; one of the repeats is within the right TIR and one is just outside it (Fig. 1). Similarly, *Mu7* has 31-bp direct repeats ("I"), separated by 46 bp, that are within its extended right TIR, while a pair of 32-bp inverted repeats ("k") is separated by 63 bp and is totally contained within an interruption in the long *Mu7* left TIR (Fig. 1 and Schnable et al. 1989). Hence, the mechanism(s) associated with generation and/or retention of additional repeats inside evolving *Mu* elements appear(s) to be most effective near the inner-most ends of the TIRs in the *Mutator* system. In many cases, the sequences found between two direct repeats appear to be one or more degenerate versions of the duplicated sequences; this is true for the 24-bp repeats in *Mu1/Mu2* ("a"), the 17-bp repeats in *Mu4* ("f"), and the 31-bp repeats in *Mu7* ("I"). Some of the direct repeats in *Mu3* and *Mu5* have a more complicated repeat-within-repeat structure (Fig. 1). Many of these cases of adjacent multiplications of short sequences appear to be similar to the "simple" repetitive DNAs found in all eukaryotes (Blaisdell 1983; Tautz et al. 1986). Some direct repeats (e.g., the 138-bp duplications in *Mu2*) appear to have totally novel sequences between the duplications, however.

One specific repeat, "a", is particularly interesting. This element has the sequence 5′A/T G A C/G A/G A G A A/G G A G T A C G C C/G A/T G A C G G-3′ and is found in all *Mu* element TIRs except the left end of *Mu3* and the right end of *Mu8* (Fig. 1). In both of these exceptional cases, however, part of the "a" repeat is present at the internal end of the TIR and is then followed directly by internal sequences that are novel to this element subfamily. The "a" repeat is usually preceded, within zero to 1 bp, by the sequence 5′-GACGG-3′. The "a" repeat is commonly found at or very near to the point where *Mu* element ends differ due to some apparent structural rearrangement. The 5′-GACGG-3′ sequence just upstream from each "a" repeat has the potential to form an unequal and illegitimate pairing with the complementary nucleotides on the other strand of the "a" repeat. Such a structure could give rise to unequal or "slipped strand" recombination events (Ehrlich 1989; Levinson and Gutman 1987; Meuth 1989) that would duplicate the "a" sequence or might initiate other rearrangements. For instance, the TIRs of *Mu4* and *Mu8* share a unique repetitive element, "s", that appears to be derived from a tripartite interaction of the "a" repeat with the "b" or "b*" repeat of *Mu1/Mu2* and with *Mu5/Mu9* TIR sequences (Table 1). The structure of the 26 nucleotide "s" repeat is the following: 6 bp of perfect complementarity with the 5′-C/G A/T GAC G-3′ sequence near the end of "a", followed by 15 bp with 67-80% homology to a sequence in "b" and "b*" that is positioned exactly at the site where the "b*" deletion occurred to give rise to "b" (Taylor and Walbot 1987), followed by 11 bp with 73–82% homology to portions of the *Mu5/Mu9* TIRs that would not be otherwise represented in *Mu4* or *Mu8*. These last two components of the "s" repeat overlap for 6 bp. This tripartite element is present only in the right end of *Mu4* and the right end of *Mu8* (in opposite orientations relative to each other), and both are found at the sites where the TIRs of these elements begin to show wholesale divergence with other *Mu* element ends (see below).

Sequence relatedness among Mu element subfamilies

Figure 1 presents a pictorial representation of previously sequenced *Mu* elements. The *Mu2*(*Mu1.7*) and *Mu1* elements are members of the same subfamily, since they share internal sequences and differ primarily by the absence of 385 bp of *Mu2* sequence within *Mu1*. This 385 bp of sequence information extends the long direct repeats within *Mu1* from 104 bp to 138 bp in *Mu2* and suggests that *Mu1* was generated primarily via deletion of a portion of *Mu2* (Taylor and Walbot 1987). None of the other *Mu* elements presented here share extensive internal homology, with the exception of *Mu9* and *Mu5*. Most of the sequences found in the 1320-bp *Mu5* element are also found in the 4942-bp *Mu9* element (Hershberger et al. 1991; Talbert et al. 1989) (Fig. 1).

Since *Mu1* was the first *Mutator* element cloned, its 215-bp (and 95% identical) inverted termini have been used to define the termini of all *Mu* elements subsequently identified. However, two *Mu* subfamiles were found to have even longer TIRs than *Mu1*. *Mu4* has 516-bp inverted repeats that are 94% identical. Only the most terminal 215 bp of this repeat are related (ca. 85% identity) to the *Mu1* TIRs (Talbert et al. 1989). Similarly, *Mu5* has 359-bp TIRs of which only the most terminal 215 bp are related (ca. 88% identity) to the *Mu1* TIRs (Talbert et al. 1989). Our sequence comparisons uncovered the fact that all of the *Mu* elements that have been sequenced contain more extensive homologies to the *Mu5* TIRs than they do to the *Mu1* TIRs

(Fig. 1). For instance, although *Mu8* has been reported to have an 185-bp TIR homologous to the outward-most 185 bp of the *Mu1* TIR (Fleenor et al. 1990), the element contain 185 colinear bp of the *Mu1*/*Mu5* right TIR on its right end plus the most terminal 249 colinear bp of the larger *Mu5* left TIR as its left TIR. In addition, directly adjacent to the 185-bp right TIR of *Mu8* are 40 bp of DNA that does not pair with the left TIR of *Mu8*, followed by 14 bp of *Mu5* TIR-related sequence that pairs exactly with nucleotides 226–239 of the *Mu8* left end (Fig. 1).

These analyses indicated a general pattern of extended *Mu5*-related TIR at one end or the other of each *Mu* element subfamily member. Hence, these data demonstrate that the progenitor element(s) that gave rise to most or all of these subfamily members had TIRs much like at least the most terminal 250 bp of *Mu5*. Relative to the *Mu5* TIRs, the current TIRs of these other subfamilies of elements have been truncated to varying degrees off one or both ends. In the cases of *Mu4*, *Mu7*, *Mu8*, and *Mu9*, additional sequences are found within one or both TIRs. With *Mu7*, *Mu8* and *Mu9*, these additional sequences lead to interruptions in the TIRs. For *Mu4*, the additional sequences are found in both TIRs, thereby leading to a longer pair of TIRs (Fig. 1).

Generation of the extended Mu4 and Mu7 terminal inverted repeats

The uniquely placed and composed interruptions in *Mu4*, *Mu7*, and *Mu8* suggest that insertion of sequences into a *Mu* end is an evolutionarily frequent process. Some of these events could be due to the insertion of other mobile elements, as appears to have been the case for the inverted repeat-flanked insertion found in the left end of *Mu7*. This type of small mobile element is frequently observed in maize DNA and is often associated with restriction fragment length polymorphisms of no detected biological effect (Ralston et al. 1988; Sachs et al. 1986).

Chandler and coworkers (Talbert et al. 1989) described a simple process wherein a double-stranded break within a *Mu* element, followed by gap repair between the TIRs of a progenitor element, could generate newly extended TIRs. Our related model, employing repair/conversion of a single-stranded cruciform structure, provides an alternative process to transform a TIR-adjacent or TIR-interrupting sequence into a component of an extended TIR. Figure 2 describes the steps and outcome of a process wherein (1) an insertion occurs in one TIR, (2) the terminal inverted repeats of the *Mu* element form a stem-loop structure from either the Watson or Crick strand, and (3) DNA repair or replication occurs across the interruption within the TIR, thereby converting the TIR insertion into an

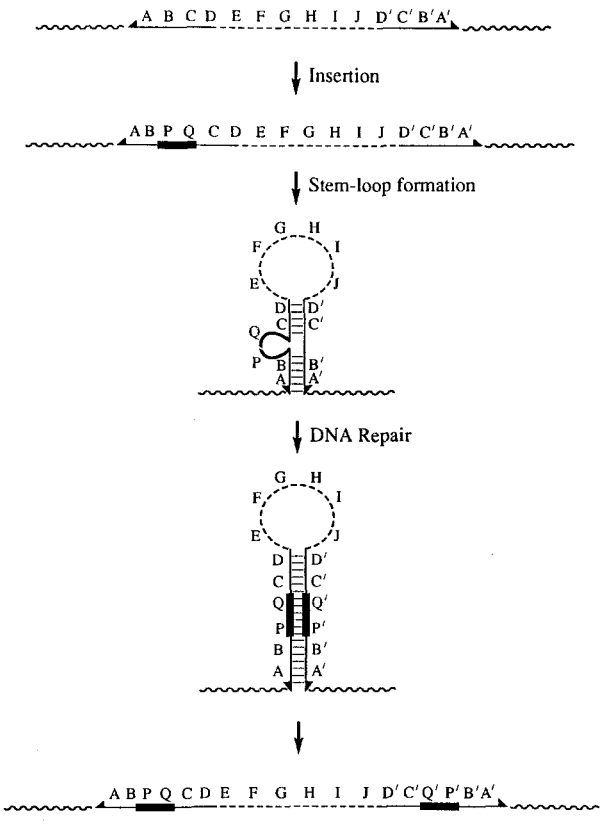


Fig. 2. Generation of an extended *Mu* element TIR by a repair/conversion process acting on a single stranded stem-loop structure. The *bold line designated P-Q* represents an insertion in one TIR that is eventually converted to part of a longer, shared TIR through a repair/conversion process

inversely repeated component of a longer TIR. After this, the disparity, between the newly repaired/converted strand and the other strand would need to be resolved by a second repair process acting on the duplex element. This simple model could explain the origin of the extra inverted repeat sequences within the TIRs of *Mu4*.

Extension of the TIR into previously TIR-adjacent sequences could be due to a co-conversion extending from the repair event described above, or could be initiated by DNAse attack of the single-stranded loop next to the double-stranded stem. Standard DNA gap repair of sequences deleted from the loop of such a stem-loop structure, using the free 3′ hydroxyl end of the left TIR as a primer and the sequences 5′ to the right TIR as template, would allow extension of a conversion process into the TIR-adjacent regions within the *Mu* element (Fig. 3A). This process could explain the extended internal ends of the TIRs of *Mu4* and *Mu7*.

Two types of evidence support this model. First, the unusual length and degree of sequence identity of the *Mu* TIRs provides them with a relatively high positive

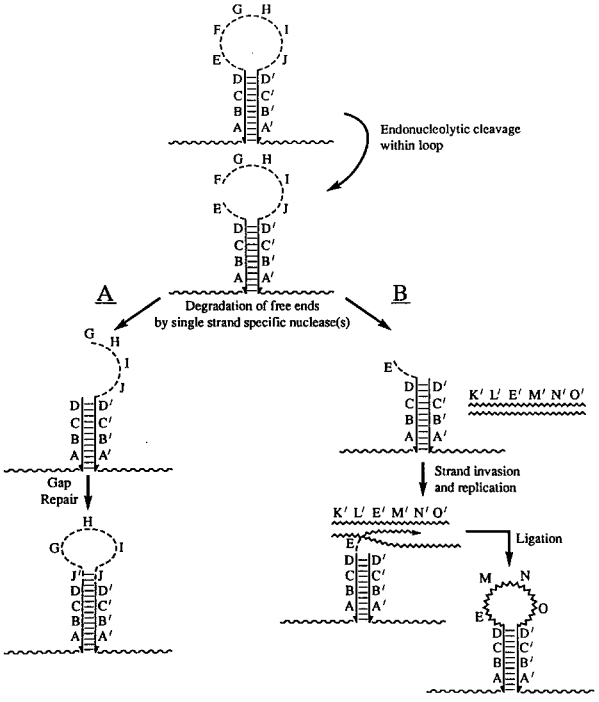


Fig. 3A, B. Alternative outcomes of DNA repair/conversion of a degraded loop in a *Mu* stem-loop structure. **A** indicates an intramolecular repair/conversion event where sequences within the original *Mu* element are lost and others are duplicated to create an element with an extended TIR. **B** indicates an intermolecular event where the free loop end generated by the degradation process invades a related sequence on another DNA molecule. One outcome of this process would be the generation of a new *Mu* element subfamily

energy for stem-loop formation. The inability of the polymerase chain reaction (PCR) to run through a *Mu* end, unless the two ends have been separated by restriction enzyme digestion (Britt and Walbot 1991), gives some idea of the high stability of this stem-loop structure in vitro. These failures have occurred despite extensive efforts and the use of PCR conditions that work routinely for the interrupted TIRs of the *Ac* element (our unpublished observations). Second, the high level of identity of the *Mu* element TIRs (94–99%, between any pair of ends in the same element) could be explained by mismatch repair within the stem formed in a stem-loop structure. In addition, the higher level of identity observed in the inner-most, extended portion of the *Mu4*, *Mu5*, and *Mu7* TIRs than in the outer-most sequences that are in common to all *Mu* elements, noted by Talbert et al. (1989), agrees with the idea that these internal repeats were generated more recently and by a repair/conversion process in which only one strand of DNA provided information. In this regard, the invasion/repair mechanism previously proposed by Chandler and coworkers (Talbert et al. 1989) to explain the extended *Mu4* and *Mu5* TIRs has many similarities with our model.

Generation of new Mu element subfamilies

One aspect of our model, that progressive degradation of the loop DNA from a *Mu* stem-loop structure could initiate intramolecular repair/conversion, leads directly into a model that would explain the generation of new *Mu* element subfamilies by a similar intermolecular repair/conversion phenomenon (Fig. 3B). The only requirement for such an intermolecular event is the formation of a single-stranded region in the *Mu* element with homology to some sequence elsewhere in the genome. This homology need not be extensive, since "illegitimate" recombination (and subsequent recombinational repair) can be initiated between sequences with as little as 4 bp of colinear identity (reviewed in Ehrlich 1989 and in Meuth 1989).

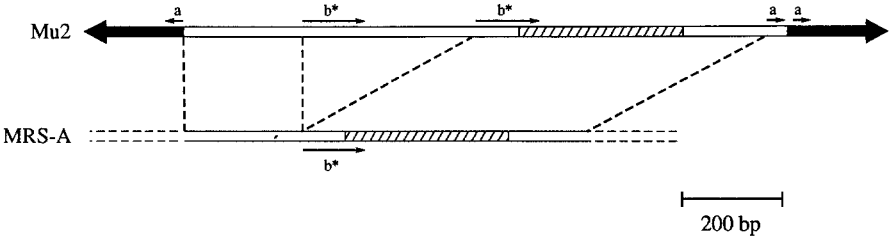
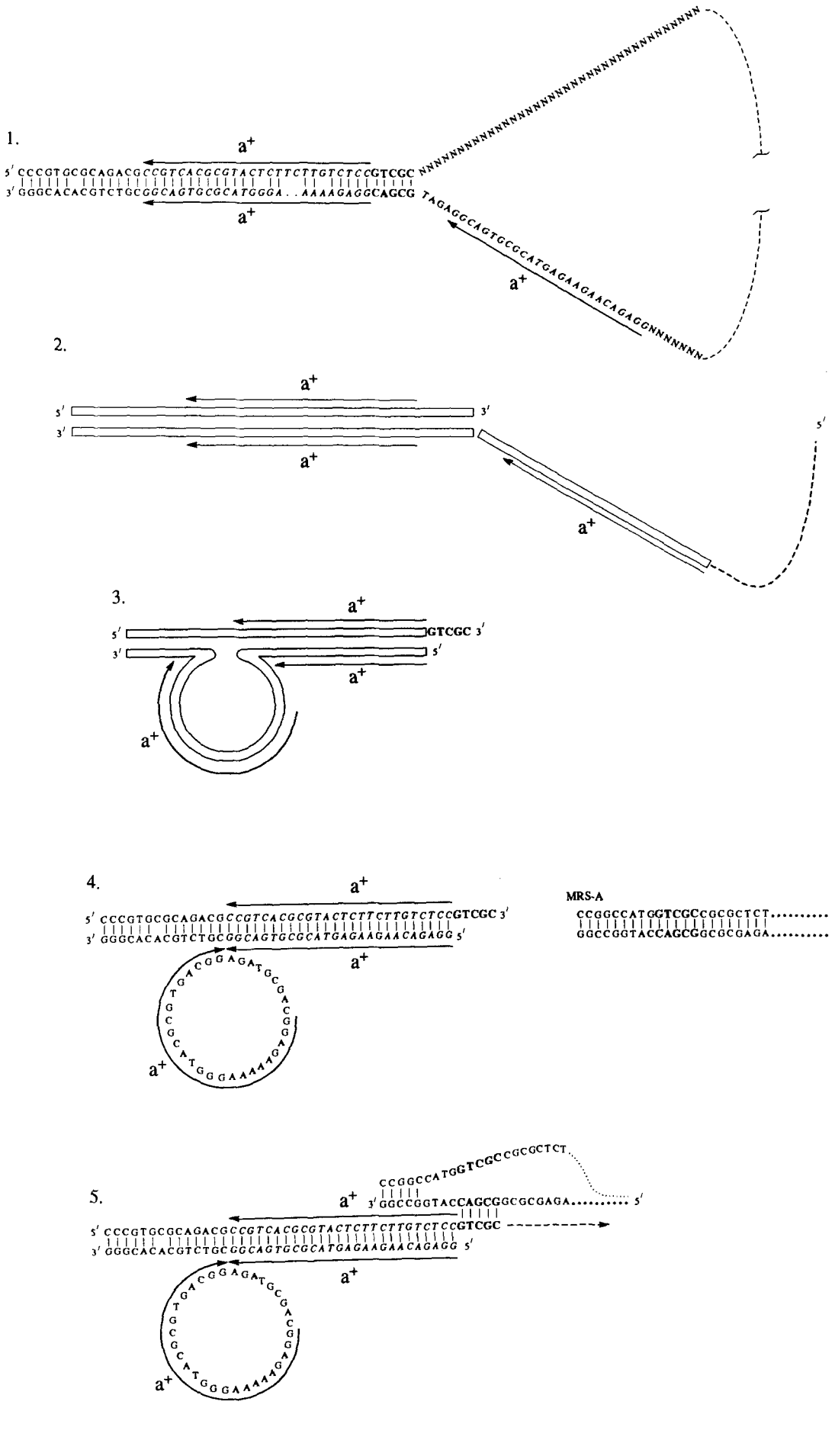


Fig. 4. Comparison of the structure of MRS-A and *Mu2*. *Dotted lines* between *Mu2* and MRS-A indicate the boundaries of the two adjacent regions of MRS-A that are found in *Mu2*. All other symbols are as indicated in the legend to Fig. 1

Fig. 5. Possible steps in the generation of a *Mu2*-related element through invasion of MRS-A by a partially degraded *Mu* element. The 26-bp repeat shown is composed of the 24-bp "*a*" sequence shown in Table 1 and Figs. 1 and 2, plus two additional base pairs. *1* A portion of the stem-loop structure that could be formed by the progenitor *Mu* element. *2* Degradation of the loop by exonucleolytic action that follows an endonucleolytic cleavage at the stem/loop junction. *3* Ectopic pairing of the extra 26-bp repeat adjacent to the progenitor *Mu* element's right TIR with the single 26-bp repeat in the left *Mu* TIR, followed by completion of the exonucleolytic removal of loop sequences in a 5′ to 3′ digestion. *4* Sequences of the resultant *Mu* fragment and of MRS-A, with a short stretch of homology that would allow an illegitimate recombinational interaction. *5* Invasion of MRS-A by the unpaired 5′-GTCGC-3′ free end to initiate a conversion process


1.
a+
5′ CCCGTGCGCAGACGCCGTCACGCGTACTCTTCTTGTCTCCGTCGC NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
3′ GGGCACACGTCTGCGGCAGTGCGCATGGGA..AAAAGAGGCAGCG TAGAGGCAGTGCGCATGAGAAGAACAGAGGNNNNNNN
a+
a+
2.
a+
5′
3′
3′
a+
5′
a+
3.
a+
5′
GTCGC 3′
3′
5′
a+
a+
4.
a+
5′ CCCGTGCGCAGACGCCGTCACGCGTACTCTTCTTGTCTCCGTCGC 3′
3′ GGGCACACGTCTGCGGCAGTGCGCATGAGAAGAACAGAGG 5′
a+
a+
MRS-A
CCGGCCATGGTCGCCGCGCTCT.....
GGCCGGTACCAGCGGCGCGAGA.....
5.
CCGGCCATGGTCGCCGCGCTCT
a+ 3′ GGCCGGTACCAGCGGCGCGAGA..... 5′
5′ CCCGTGCGCAGACGCCGTCACGCGTACTCTTCTTGTCTCCGTCGC ---------->
3′ GGGCACACGTCTGCGGCAGTGCGCATGAGAAGAACAGAGG 5′
a+
a+

The maize DNA fragment MRS-A, cloned and sequenced by Talbert and Chandler (1988), was originally isolated due to its hybridization with the internal regions of the *Mu1/Mu2* element subfamily. MRS-A contains no sequences related to the *Mu* ends, but it is part of an apparently *Mutator*-unrelated gene that has been truncated within the *Mu1/Mu2* elements (Fig. 4). Hence, the *Mu1/Mu2* subfamily of elements was probably generated through the acquisition of a portion of MRS-A by a pair of *Mu* TIRs. Thus, according to our model of *Mu* element subfamily generation, MRS-A should have significant sequence identity with the parental element that was converted into a *Mu2*-like element by the acquisition of MRS-A sequences. Moreover, this sequence identity should be situated at the site where the repair/conversion event was directionally initiated.

Our sequence analyses indicate that just such a short sequence of identical colinear nucleotides is shared by MRS-A and the left end of the *Mu1/Mu2* TIR, and that this sequence identity occurs exactly at the point where the homology between MRS-A and the *Mu1/Mu2* elements begins. This short sequence found in MRS-A, 5′-GTCGC-3′, must have been present in the parental *Mu* element since it is located in the part of the TIR (bp 211–215) that is conserved between *Mu1/Mu2*, *Mu3*, *Mu4*, *Mu5*, *Mu7*, *Mu8*, and *Mu9*. Hence, a simple nucleolytic event in the parental element that would free a single-stranded 5′-GTCGC-3′ sequence with a 3′ hydroxyl at the left TIR would allow this sequence to invade MRS-A and to serve as primer for the repair/conversion event.

The sequence of the *Mu1/Mu2* elements provides further elaboration and support of this mechanism for *Mu* subfamily generation. The generation of a single-stranded 5′-GTCGC-3′ sequence terminating with a free 3′ hydroxyl in the parental *Mu* element could have occurred by any of a number of nucleolytic events. However, the structure of the right end of *Mu1/Mu2* suggests that the specific nucleolytic events responsible for removal of most of the parental *Mu* element loop were an endonucleolytic event at the 3′ end of the left TIR, followed by progressive 5′ to 3′ exonucleolytic removal of loop nucleotides. A direct duplication of 26 bp at the internal end of the *Mu1/Mu2* right TIR surrounds the 5-bp sequence that could initiate the intermolecular repair/conversion event (Fig. 4). The 5-bp invasion/primer sequence is at the very end of the identity between the *Mu1/Mu2* left and right TIRs, and the 26-bp direct repeat is adjacent to this 5-bp sequence. As shown in Fig. 5, a progressive removal of the single-stranded sequences within the loop that began at the 3′ end of the left TIR would leave a stem with a potential for unequal pairing at this 26-bp sequence. When this unequal pairing occurred, the 5′-GTCGC-3′ end would now be single-stranded and available to invade the duplex of MRS-A. The stability of the unequal pairing event would be relatively high, given that the unequal pairing gives 26/26 paired nucleotides and the standard pairing yields only 22/26 paired nucleotides (Fig. 5). Further single-strand-specific, 5′ to 3′ exonuclease action would eventually remove all loop sequences except those within and 3′ to the 26-bp repeat. After invasion of MRS-A by this molecule, gap repair would simply insert MRS-A sequences to replace the degraded loop sequences previously present. The expected outcome of such a process exactly describes the current *Mu1/Mu2* subfamily: a 215-bp left TIR ending in 5′-GTCGC-3′, followed by MRS-A sequences, followed by the 26-bp duplication, 4 bp of unpaired sequence, and the right TIR.

Discussion

Compared to other eukaryotic transposable elements, the *Mu* elements are exceptional both in structure and diversity. The terminal inverted repeats (TIRs) of *Mu* elements are unusually long (185–359 bp), very precise (94–99% identity), and quite variable between *Mu* subfamilies. By comparison, the TIRs of the *Ac/Ds* and *Spm(En)* elements of maize and the P and Tc1 elements of animals are 11 bp, 13 bp, 31 bp, and 54 bp, respectively (Muller-Neumann et al. 1984; O'Hare and Rubin 1983; Pereira et al. 1986; Pohlman et al. 1984; Rosenzweig et al. 1983), and are nearly identical for all autonomous and non-autonomous elements of the same family. Sequences that effect the response of the elements to transposase can be found over 100 bp from the exact termini of the *Spm(En)* or P elements (Karess and Rubin 1984; Schiefelbein et al. 1985), and the simple repeats within 200 bp of the exact *Spm(En)* ends can be theoretically assembled into a stem of over 65 bp, with multiple interruptions (Gierl et al. 1985). Still, these elements do not seem to "repair" the extensive non-pairing regions of their greater TIRs. Most eukaryotic transposable elements with long and well-matched TIRs, like the fold-back elements FB of *Drosophila melanogaster* (reviewed in Bingham and Zachar 1989) and TU of sea urchin (*Strongylocentrotus purpuratus*) (reviewed in Hoffman-Liebermann et al. 1989), have TIRs composed primarily of short direct repeats and may have little or no sequence internal to the TIRs. The recently discovered element Tc4, from *C. elegans*, seems to be intermediate between the *Mu* elements and the fold-back elements; it has long (774 bp, > 99% identical) TIRs that are not composed of short direct repeats, and it has a unique internal sequence of 57 bp (Yuan et al. 1991). Other direct and inverted repeats, in addition to the TIRs, were also found within the first *Mu* elements studied (Barker et al. 1984; Schnable et al. 1989; Taylor and Walbot 1987).

Our studies of the full set of sequenced *Mutator* elements and of members of the *Ac/Ds* and *Spm(En)* families have indicated that internal duplications are particularly frequent in *Mu* elements, but are also common to the non-autonomous members of other element systems. These duplications are primarily short direct repeats interrupted by a few base pairs of DNA. The largest such internal direct duplication observed in a *Mu* element is the 138-bp direct repeat found in *Mu2* (Taylor and Walbot 1987). Only one copy of this 138-bp sequence is found in MRS-A, and the 263 bp of DNA between the 138-bp repeats has no sequence relatedness with MRS-A (Talbert and Chandler 1988) or with the DNA of any other subfamily of *Mu* elements. Hence, it is possible that these 263 bp of DNA were "picked up" in the same sense (but, perhaps, via a different mechanism) that MRS-A sequences were picked up by *Mu* ends to form the original founder of the *Mu1/Mu2* subfamily. Conversely, the "ancestral" MRS-A region may have contained this insert flanked by the "b*" duplication and subsequently lost it by an unequal recombinational event. Our ability to propose a mechanism for the acquisition of MRS-A sequences by *Mu* TIRs was dependent upon the isolation and sequencing of MRS-A DNA (Talbert and Chandler 1988). A similar search for and analysis of the original source of this 263-bp sequence would provide some of the raw information needed to identify the processes that created this additional variation.

Many of the short direct and inverted repeats found in *Mu*, *Ac/Ds*, and *Spm(En)* elements are tightly associated with the internal ends of these elements' TIRs. This suggests that internal ends of the TIRs are very active in the generation of duplications. Alternatively, the sites of generation may be random, but their chance of survival into subsequent generations could be greater if the duplications were end-adjacent. This is particularly likely for the autonomous elements, since internal duplications would commonly inactivate the coding potential for transposase or other essential trans functions. In this regard, the rare duplications that we found within the autonomous (or suspected autonomous) members of the *Ac/Ds*, *Spm(En)*, and *Mu* families were all quite small and directly adjacent to the extended TIRs where they would not be likely to interrupt any coding frame. Selection for other internal structures, probably not related to gene expression, is also suggested by the observation that different *Mu* elements transpose at different frequencies (e.g., *Mu2* transposes about 3 times more frequently than *Mu3*, despite the fact that both have nearly identical TIRs) (Bennetzen et al. 1993). Since *Mu* element transposition can lead to *Mu* element amplification (Alleman and Freeling 1986; Bennetzen et al. 1987; Hardeman and Chandler 1989), internal alterations that increase or decrease transposition frequency would influence the likelihood that any given *Mu* element would pass into the next generation.

The internal ends of *Mu* TIRs appear to be exceptionally active in various aspects of *Mu* element instability. The cruciform or stem-loop structures feasible with a *Mu* element would generate two sites at which double-stranded DNA is juxtaposed with single-stranded DNA: the internal ends of the TIRs. It seems likely that specific enzymes associated with DNA repair, for instance, could be attracted to such an unusual structure. Studies in bacterial and animal systems have shown that palindromic sequences (like TIRs) can greatly enhance the frequency of illegitmate recombination events nearby, including transposon excision (reviewed in Ehrlich 1989 and in Meuth 1989). Whatever the mechanism, it seems likely that extra repeat association with the internal ends of the TIRs is not coincidental and that repair associated with a cleaved stem-loop structure is a component of the process that generates these duplications.

The "a" repeat is found at or very near the break points for many of the rearrangements that differentiate *Mu* element subfamilies. The close 5′ association of "a" with a short direct repeat of its 3′ sequences (5′-GACGG-3′) is a conserved and compelling aspect of this element's structure. Illegitimate recombination or "slipped strand" recombination (Levinson and Gutman 1987) between these short repeats may explain the involvement of "a" in many rearrangement events.

The exceptional length and within-element similarity of the *Mu* TIRs may be self-reinforcing; long and well-paired TIRs would be most likely to form a stem and thereby allow the repair/conversion events that we have proposed as the mechanism accounting for the unusual length, diversity, and degree of identity of the *Mu* TIRs. Hence, the less frequent internal duplications, shorter (and less variable) TIRs, and less diverse internal sequences of other element families (e.g., *Ac/Ds* or *Spm(En)*) may largely be an outcome of their shorter TIRs. Alternatively, differences between transposable element systems in the interaction of transposases and other factors with the elements' TIRs might explain a higher frequency of stem-loop formation or subsequent variability generation with *Mutator* than with other transposable elements.

The *Mu* element stem-loop formation, or its involvement in the generation of variability, cannot be a particularly frequent event since some sequence differences do exist between a pair of *Mu* TIRs. Similarly, the generation of new *Mu* element subfamilies appears to occur on an evolutionary timescale; we have no evidence suggesting that any new *Mu* element subfamily has been generated in the 20 plus years that *Mutator* studies have been actively pursued. In this regard, the transpositional behavior, high level of homogeneity, and relatively limited dispersal of

Mu1/Mu2 subfamily members have suggested that this element subfamily is the most recently generated of those that have been detected (Bennetzen et al. 1993). Yet, *Mu1* and *Mu2* elements are found in low genomic copy numbers across a wide array of *Zea* species (Bennetzen 1984; Chandler et al. 1986, 1988). Hence, all of the events that we have modeled, and the mechanism(s) that could give rise to them, act very infrequently and require detection in an evolutionary time frame.

The variability of *Mu* element subfamilies is, in many ways, the most unusual feature of the *Mutator* transposable element system. Retrotransposons can acquire adjacent DNA sequences through reverse transcription of "fused" or "readthrough" transcripts (Swain and Coffin 1992), but the recruitment of novel internal sequences by other eukaryotic transposable elements has been rarely documented (Heslip et al. 1992; Klein et al. 1988; Tsubota and Dang-Vu 1991). In the *Mutator* system, novel sequence acquisition appears to be a relatively common phenomenon. Moreover, as with the acquisition of MRS-A sequences to assemble the *Mu1/Mu2* subfamily progenitor, this insertion of new DNA is exceptional in its ability to proceed via a mechanism that replaces (i.e., removes) the DNA previously found between two *Mu* TIRs. Of course, we have no evidence that acquisition of sequences by *Mu* TIRs always requires complete replacement. This issue will only be resolved by further analyses of other *Mu* element subfamilies and the chromosomal sequences that they have captured. The cloning and ongoing analysis of *Mu3*-related sequences that lack *Mu* ends by Oishi and coworkers (K. Oishi, personal communication) will provide the first test of this question and of our models. Further isolation and analysis of TIR-free sequences that hybridize to the other *Mu* subfamily internal segments are also warranted.

At the moment, we have no idea of the parental element(s) that may have given rise to any new *Mu* element subfamily. In cases like the acquisition of MRS-A sequences to form the *Mu1/Mu2* family, very few legacies of the parental *Mu* element have been left behind. This may not always be the case, however, and some elements may contain internal segments that are homologous to one or more parental elements. Given the variable distribution of the known *Mu* element subfamilies, it is likely that some additional subfamilies will only be present in a few maize lines or maize relatives, or that some *Mu* subfamilies may have become extinct. Hence, it is possible that the parental elements for some current *Mu* element subfamilies will not be found even with extensive efforts.

On the basis of structural and transcript hybridization criteria, *Mu9* has been proposed to be the autonomous *Mutator* element (Hershberger et al. 1991) and, by analogy to other systems, could be viewed as the most likely "original" *Mu* element. However, our sequence comparisons indicate that *Mu9* could be a rearranged version of the "model" *Mutator* element; the internal-most 145 bp of the right end of *Mu9* appears to have been extensively rearranged relative to its left TIR and the *Mu5* TIRs (Bennetzen et al. 1993). There is also a 110-bp region in *Mu9* primarily composed of scrambled sequences from a similarly placed 155-bp region in *Mu5* (Bennetzen et al. 1993). The TIR rearrangement might have occurred without inactivation of the cis-or trans-activational components of an autonomous element. Since only about 200 bp of the most external portion of the *Mu* TIR is shared by most *Mu* element subfamilies, it is very likely that the interrupted *Mu9* TIRs are actually the canonical *Mu* inverted repeats. If this is true, then the longer and more perfect TIRs of *Mu5* could have been generated by the conversion/repair processes discussed above. The fact that the innermost 138 bp of the *Mu5* TIRs are 99.5% identical, while the outermost 220 bp of these same TIRs are only 90% identical (Talbert et al. 1989), suggests just such a relatively recent origin of the internal portions of the *Mu5* TIRs. In this regard, it will be interesting to see the primary structure of the cloned *Mu* element that has been proven to be autonomous, *MuR1* (Chomet et al. 1991).

Acknowledgements. We thank K. Oishi and V. Walbot for providing unpublished information and P. SanMiguel for technical assistance. This research was supported by USDA/CRGO grant no. 89-37262-4362 and by NIH grant AI27713 (in support of Purdue's ACLCB).

References

Alleman M, Freeling M (1986) The *Mu* transposable elements of maize; evidence for transposition and copy number regulation during development. Genetics 112:107–119

Barker RF, Thompson DV, Talbot DR, Swanson J, Bennetzen JL (1984) Nucleotide sequence of the maize transposable element *Mu1*. Nucleic Acids Res 12:5955–5967

Bennetzen JL (1984) Transposable element *Mu1* is found in multiple copies only in Robertson's *Mutator* maize lines. J Mol Appl Genet 2:519–524

Bennetzen JL, Swanson J, Taylor WC, Freeling M (1984) DNA insertion in the first intron of maize *Adh1* affects message levels: cloning of progenitor and mutant *Adh1* alleles. Proc Natl Acad Sci USA 81:4125–4128

Bennetzen JL, Fracasso RP, Morris DW, Robertson DS, Skogen-Hagenson MJ (1987) Concomitant regulation of *Mu1* transposition and *Mutator* activity in maize. Mol Gen Genet 208:57–62

Bennetzen JL, Springer PS, Cresse AD, Hendrickx M (1993) Specificity and regulation of the *Mutator* transposable element system in maize. Crit Rev Plant Sci 12:57–95

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

Blaisdell BE (1983) A prevalent persistent global nonrandomness that distinguishes coding and non-coding eukaryotic nuclear DNA sequences. J Mol Evol 19:122–133

Britt AB, Walbot V (1991) Germinal and somatic products of *Mu1* excision from the *Bronze-1* gene of *Zea mays*. Mol Gen Genet 227:267–276

Brown WE, Robertson DS, Bennetzen JL (1989) Molecular analysis of multiple *Mutator*-derived alleles of the *Bronze* locus of maize. Genetics 122:439–445

Chandler VL, Hardeman KJ (1992) The *Mu* elements of *Zea mays*. Adv Genet 30:77–122

Chandler VL, Rivin CJ, Walbot V (1986) Stable non-*Mutator* stocks of maize have sequences homologous to the *Mu1* transposable element. Genetics 114:1007–1021

Chandler VL, Talbert LE, Raymond F (1988) Sequence, genomic distibution and DNA modification of a *Mu1* element from non-*Mutator* maize stocks. Genetics 119:951–958

Chomet P, Lisch D, Hardeman KJ, Chandler VL, Freeling M (1991) Identification of a regulatory transposon that controls the *Mutator* transposable element system in maize. Genetics 129:261–270

Dennis ES, Gerlach WL, Pryor AJ, Bennetzen JL, Inglis A, Llewellyn D, Sachs M, Ferl RJ, Peacock WJ (1984) The *Adh1* gene of maize. Nucleic Acids Res 12:3983–4000

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Ehrlich SD (1989) Illegitimate recombination in bacteria. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, pp 799–832

Fedoroff N, Wessler S, Shure M (1983) Isolation of the transposable maize controlling elements *Ac* and *Ds*. Cell 35:235–242

Fleenor D, Spell M, Robertson D, Wessler S (1990) Nucleotide sequence of the *Mutator* element, *Mu8*. Nucleic Acids Res 18:6725

Gierl A, Schwarz-Sommer Z, Saedler H (1985) Molecular interactions between the components of the *En-I* transposable element system of *Zea mays*. EMBO J 4:579–583

Hardeman KJ, Chandler VL (1989) Characterization of *bz1* mutants isolated from *Mutator* stocks with high and low numbers of *Mu1* elements. Dev Genet 10:460–472

Hershberger RJ, Warren CA, Walbot V (1991) *Mutator* activity in maize correlates with the presence and expression of the *Mu* transposable element *Mu9*. Proc Natl Acad Sci USA 88:10198–10202

Heslip TR, Williams JA, Bell JB, Hodgetts RB (1992) A P element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics 131:917–927

Hoffman-Liebermann B, Liebermann D, Cohen SN (1989) TU elements and Puppy sequences. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, pp 575–592

Karess RE, Rubin GM (1984) Analysis of P transposable element functions in Drosophila. Cell 38:35–146

Klein AS, Clancy M, Paje-Manalo L, Furtek DB, Hannah LC, Nelson OE Jr (1988) The mutation *bronze-mutable 4 derivative 6856* in maize is caused by the insertion of a novel 6.7-kilobase pair transposon in the untranslated leader region of the *Bronze-1* gene. Genetics 120:779–790

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Merckelbach A, Doering H-P, Starlinger P (1986) The aberrant *Ds* element in the *adh1-2F11::Ds2* allele. Maydica 31:109–122

Meuth M (1989) Illegitimate recombination in mammalian cells. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, pp 833–860

Muller-Neumann M, Yoder JI, Starlinger P (1984) The DNA sequence of the transposable element *Ac* of *Zea mays* L. Mol Gen Genet 198:19–24

O'Hare K, Rubin GM (1983) Structure of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34:25–35

Pereira A, Cuypers H, Gierl A, Schwarz-Sommer Z, Saedler H (1986) Molecular analysis of the *En/Spm* transposable elements of *Zea mays*. EMBO J 5:835–841

Pohlman RF, Fedoroff NV, Messing J (1984) The nucleotide sequence of the maize controlling element *Activator*. Cell 37:635–643

Qin M, Ellingboe AH (1990) A transcript identified by *MuA* of maize is associated with *Mutator* activity. Mol Gen Genet 224:357–363

Qin M, Robertson DS, Ellingboe AH (1991) Cloning of the *Mutator* transposable element *MuA2*, a putative regulator of somatic mutability of the *a1-Mum2* allele in maize. Genetics 129:845–854

Ralston EJ, English JJ, Dooner HK (1988) Sequence of three bronze alleles of maize and correlation with the genetic fine structure. Genetics 119:185–197

Robertson DS (1978) Characterization of a *Mutator* system in maize. Mutat Res 51:21–28

Robertson DS, Stinard PS (1989) Genetic analysis of putative two-element systems regulating somatic mutability in *Mutator*-induced aleurone mutants in maize. Dev Genet 10:482–506

Rosenzweig B, Liao L, Hirsh D (1983) Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res 11:7137–7140

Sachs MM, Dennis ES, Gerlach WL, Peacock WJ (1986) Two alleles of maize alcohol dehydrogenase 1 have 3′ structural and poly(A) addition polymorphisms. Genetics 113:449–467

Schiefelbein JW, Raboy V, Fedoroff NV, Nelson OE Jr (1985) Deletions within a defective *Suppressor-mutator* element in maize affect the frequency and developmental timing of its excision from the *bronze* locus. Proc Natl Acad Sci USA 82:4783–4787

Schnable PS, Peterson PA, Saedler H (1989) The *bz-rcy* allele of the *Cy* transposable element system of *Zea mays* contains a *Mu*-like element insertion. Mol Gen Genet 217:459–463

Swain A, Coffin JM (1992) Mechanism of transduction by retroviruses. Science 255:841–845

Talbert LE, Chandler VL (1988) Characterization of a highly conserved sequence related to *Mutator* transposable elements in maize. Mol Biol Evol 5:519–529

Talbert LE, Patterson GI, Chandler VL (1989) *Mu* transposable elements are structurally diverse and distributed throughout the genus *Zea*. J Mol Evol 29:28–39

Tautz D, Martin T, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652–656

Taylor LP, Walbot V (1987) Isolation and characterization of a 1.7-kb transposable element from a *Mutator* line of maize. Genetics 117:297–307

Tsubota SI, Dang-Vu H (1991) Capture of flanking DNA by a P element in *Drosophila melanogaster:* creation of a transposable element. Proc Natl Acad Sci USA 88:693–697

Weil CF, Marillonnet S, Burr B, Wessler SR (1992) Changes in state of the *Wx-m5* allele of maize are due to intragenic transposition of *Ds*. Genetics 130:175–185

Yuan J, Finney M, Tsung N, Horvitz HR (1991) Tc4, a *Caenorhabditis elegans* transposable element with an unusual fold-back structure. Proc Natl Acad Sci USA 88:3334–3338