Sequence similarity of putative transposases links the maize Mutator autonomous element and a group of bacterial insertion sequences

Jonathan A. Eisen*, Maria-Ines Benito and Virginia Walbot
Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, USA

Received March 7, 1994; Revised and Accepted June 1, 1994

ABSTRACT

The Mutator transposable element system of maize is the most active transposable element system characterized in higher plants. While Mutator has been used to generate and tag thousands of new maize mutants, the mechanism and regulation of its transposition are poorly understood. The Mutator autonomous element, MuDR, encodes two proteins: MURA and MURB. We have detected an amino acid sequence motif shared by MURA and the putative transposases of a group of bacterial insertion sequences. Based on this similarity we believe that MURA is the transposase of the Mutator system. In addition we have detected two rice cDNAs in genbank with extensive similarity to MURA. This sequence similarity suggests that a Mutator-like element is present in rice. We believe that Mutator, a group of bacterial insertion sequences, and an uncharacterized rice transposon represent members of a family of transposable elements.

Transposons are genetic elements capable of moving within and between continuous segments of genetic material and are likely ubiquitous contributors to genome structure. Examples of transposons include retroviruses like HIV, maturase-encoding introns in mitochondria, insertion sequences in bacteria, P elements in Drosophila, LINEs in humans, and Ac and Mutator in maize. Closely related elements are classified into transposon families. Within a family elements can be divided into two functional classes, autonomous and non-autonomous. Autonomous elements are capable of directing their own transposition as well as transactivating the transposition of non-autonomous elements by producing the factors (transposases) that are required along with host factors for transposition. Sequence comparison of transposases from different families has led to the creation of transposase superfamilies. This broad grouping of transposases has provided insight into the evolution of transposon families and their respective mechanisms of transposition (1–3).

We are interested in the characterization of the transposase of the Mutator transposable element system of maize (Zea mays).

The Mutator family is composed of diverse classes of elements whose transposition and excision activities are associated with mutation frequencies of $10^{-4}$ per locus per generation (4). These elements share a similar 220 base pair (bp) terminal inverted repeat and their insertion sites are flanked by a 9 bp direct duplication of host sequence (4). Different classes of elements within the Mutator family are defined by internal sequence similarity (4, 5). Mutator is the most active transposable element system characterized in higher plants, and has been used to generate and tag thousands of new mutants in maize. Only recently has an element been isolated which meets the genetic criteria for an autonomous Mutator element (6–8). This 4.9 kb element, MuDR, produces at least two primary transcripts: mudra and mudrB (7), which are predicted to encode proteins of 823 amino acids (MURA) and 207 amino acids (MURB) (9). The functions of these transcripts and their putative protein products remain unknown.

We have detected a sequence motif that is shared by MURA and the putative transposases of nine bacterial insertion sequences (ISs), IS1081 (10), ISRm3 (11), IS6120 (12), IS256 (13, 14), IS406 (15, 16), IST2 (17), IS905 (18, 19), IS1201 (20), and an unnamed IS isolated from Corynebacterium diphtheriae (21) (Figure 1). Like other transposons, ISs have been classified based on conserved structural features (22). Previously, IS256, ISRm3, IST2, IS1081 and IS6120 have been variously linked to each other based on sequence similarity of putative transposases (10–12). We believe that this group should be extended to include IS406, IS905, IS1201, the C. diphtheriae IS, as well as Mutator. Table I lists some characteristics of the ISs in this extended group. The 8/9 bp target-site duplication of the ISs and Mutator could represent a conserved feature of transposase function. Of these ISs five (IS905, IS256, IS406, IS6120, and ISRm3) have been shown to actively transpose (11–18, 20). All are known to be present in multiple copies in their respective genomes (10–18) except IS1201 and the C. diphtheriae IS for which this information is unavailable.

The cohesiveness of this extended group is supported by a variety of analyses. First, none of the previously characterized motifs from other transposase groups could be detected in these proteins. In addition, blastp and tblast searches (23) of the non-
Table 1. Structural features of the bacterial insertion sequences with amino acid similarity to MURA

<table>
<thead>
<tr>
<th>Insertion Sequence</th>
<th>Species of Origin</th>
<th>Total Length (bp)</th>
<th>Inverted Repeat (bp)</th>
<th>Target Duplication (bp)</th>
<th>Putative Transposase (aa)</th>
<th>Genbank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1082 (10)</td>
<td>Mycobacterium bono</td>
<td>1324</td>
<td>15</td>
<td>?</td>
<td>415</td>
<td>X61270</td>
</tr>
<tr>
<td>IS256 (12, 14)</td>
<td>Staphylococcus epidermidis</td>
<td>1322</td>
<td>26</td>
<td>8</td>
<td>390</td>
<td>M18056</td>
</tr>
<tr>
<td>IS406 (15, 16)</td>
<td>Paracoccus denitrificans</td>
<td>1368</td>
<td>41</td>
<td>8</td>
<td>388</td>
<td>IS5223</td>
</tr>
<tr>
<td>IS72 (17)</td>
<td>Thiothrix ferrooxidans</td>
<td>1408</td>
<td>25</td>
<td>9</td>
<td>296</td>
<td>J03859</td>
</tr>
<tr>
<td>IS6120 (12)</td>
<td>Mycobacterium smegmatis</td>
<td>1486</td>
<td>24</td>
<td>9</td>
<td>332</td>
<td>M69182</td>
</tr>
<tr>
<td>IS905 (18, 19)</td>
<td>Lactobacillus acidophilus</td>
<td>1313</td>
<td>28</td>
<td>?</td>
<td>391</td>
<td>L38851</td>
</tr>
<tr>
<td>IS1201 (20)</td>
<td>Lactobacillus helveticus</td>
<td>1387</td>
<td>24</td>
<td>8</td>
<td>369</td>
<td>L.28331</td>
</tr>
<tr>
<td>IS (21)</td>
<td>Corynebacterium diphtheriae</td>
<td>&gt;1500</td>
<td>?</td>
<td>?</td>
<td>343</td>
<td>a07012 (EMBL)</td>
</tr>
<tr>
<td>ISR01 (11)</td>
<td>Rhizobium meliloti</td>
<td>1298</td>
<td>30</td>
<td>8/9</td>
<td>400</td>
<td>M60971</td>
</tr>
</tbody>
</table>

*A '?' indicates that the information for that insertion sequence is unknown.

 redundantly database at NCBI using each of the proteins in this group as well as the consensus sequence (Figure 1) did not yield any significant alignments over the region of the motif with any proteins other than those in the group. Blastp generated pairwise alignments were significant for all comparisons among the ISs (most P values < 10^-5). Alignments between MURA and the ISs were most significant (e.g. P = 10^-5 for a MURA-ISRm3 comparison) when alignment gaps were allowed, such as by using the MPscratch program (24).

The grouping of Mutator and bacterial ISs is not unreasonable; other transposable elements have been grouped across similar evolutionary distances (25, 26). Why is Mutator the only eukaryotic transposon in this group? We believe that other transposons that would fit into this group do exist in other eukaryotes but have not yet been characterized. Southern analysis with Mutator probes suggests that Mutator-like elements are present in other monocot species (27). In addition, we have discovered similarity between MURA and putative proteins from two partially sequenced cDNAs from rice (Figure 2). Although these cDNAs and the proteins they may encode are uncharacterized, the high percentage of similarity with MURA suggests that Mutator-like transposons exist in rice. It remains to be determined if such elements are actively transposing or possibly transactivating the transposition of other Mutator-like

Figure 1. Alignment of the predicted protein sequences of maize MURA and putative transposases from some bacterial insertion sequences. The alignment was generated using the Clustal V program (28) with gap penalties of 20 and a PAM250 matrix. Gaps in the alignment are labeled by an ‘-’. Groups of residues were shaded at an alignment position if they included ≥ 50% identical amino acids, ≥ 70% in the groups DENQ, KR, MVIL, or WY, or ≥ 90% in groups DENQST or WYILMY. Consensus abbreviations are f (hydrophobic), + (basic), − (acidic), and ± (charged). Complete species names, citations, and accession numbers are given in Table 1. Numbers after transposase name refer to the amino acid residue of the predicted protein.

Figure 2. Alignment of regions of MURA with the amino acid sequences encoded by two cDNAs from rice (Oryza sativa). A: Rice 1 — cDNA C1069A (Genbank no. d15075). B: Rice 2 — cDNA C25102A (Genbank no. d23146). Both cDNAs are partial sequences of genes expressed in a Nipponbare Japonica tissue culture line (29).
elements. Perhaps Mutator-like elements have not been detected in other eukaryotes because they are repressed or inactivated as they seem to be in many lines of maize (27).

ACKNOWLEDGEMENTS
We would like to thank M. Feldman for the use of computer equipment, S. Smith for the GDE program, the NCBI and EMBL for free use of sequence analysis email servers, and two anonymous reviewers for very helpful suggestions. M.-I. B. and J. A. E. are supported by predoctoral fellowships from HHMI and NSF, respectively. This work was also supported by NIH grant GM49681. This paper is dedicated to the memory of Benjamin Post.

REFERENCES