2634–2636 Nucleic Acids Research, 1994, Vol. 22, No. 13

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 very similar ~220 base pair (bp) terminal inverted repeats 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 mofif that is shared by MURA and the putative transposases of nine bacterial insertion sequences (ISs), IS*1081* (10), IS*Rm3* (11), IS*6120* (12), IS*256* (13, 14), IS*406* (15, 16), IS*T2* (17), IS*905* (18, 19), IS*1201* (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, IS*256*, IS*Rm3*, IS*T2*, IS*1081* and IS*6120* 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 IS*406*, IS*905*, IS*1201*, the *C.diphtheriae* IS, as well as *Mutator*. Table I lists some characteristics of the ISs in this extended goup. The 8/9 bp target-site duplication of the ISs and *Mutator* could represent a conserved feature of transposase function. Of these ISs five (IS*905*, IS*256*, IS*406*, IS*6120*, and IS*Rm3*) 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 IS*1201* 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 tblastn searches (23) of the non-

* To whom correspondence should be addressed

Table 1. Structural features of the bacterial insertion sequences with amino acid similarity to MURA

Insertion Sequence	Species of Origin	Total Length (bp)	Inverted Repeat (bp)	Target Duplication (bp)	Putative Transposase (aa)	Genbank Accession Number
IS*1081* (10)	*Mycobacterium bovis*	1324	15	?*	415	X61270
IS256 (13, 14)	*Staphylococcus aureus*	1322	26	8	390	M18086
IS*406* (15, 16)	*Pseudomonas cepacia*	1368	41	8	388	J50237
IST2 (17)	*Thiobacillus ferrooxidans*	1408	25	9	296	J03859
IS*6120* (12)	*Mycobacterium smegmatis*	1486	24	9	323	M69182
IS905 (18, 19)	*Lactococcus lactis*	1313	28	?	391	L20851
IS*1201* (20)	*Lactobacillus helveticus*	1387	24	8	369	L26311
IS (21)	*Corynebacterium diphtheriae*	>1500	?	?	343	a07012 (EMBL)
IS*Rm3* (11)	*Rhizobium meliloti*	1298	30	8/9	400	M60971

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

redundant 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-IS*Rm3* comparison) when alignment gaps were allowed, such as by using the MPsrch 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

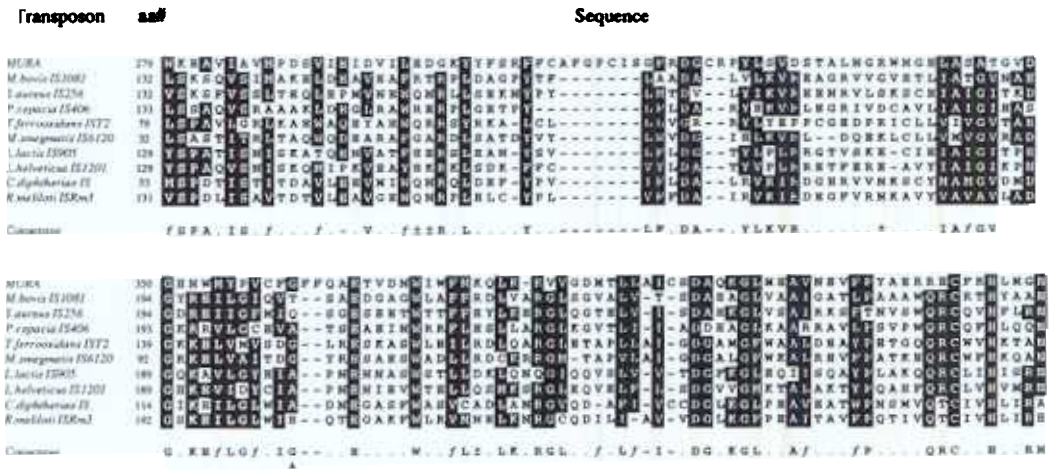

Figure 1. Alignment of a region of the predicted protein sequences of maize MURA and putative transposases from some bacterial insertion sequences. The alignment was generated using the Clustal V program (28) with gap penalties of 20 and a PAM250 matrix. Gaps in the alignment are labeled by an '~'. Groups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

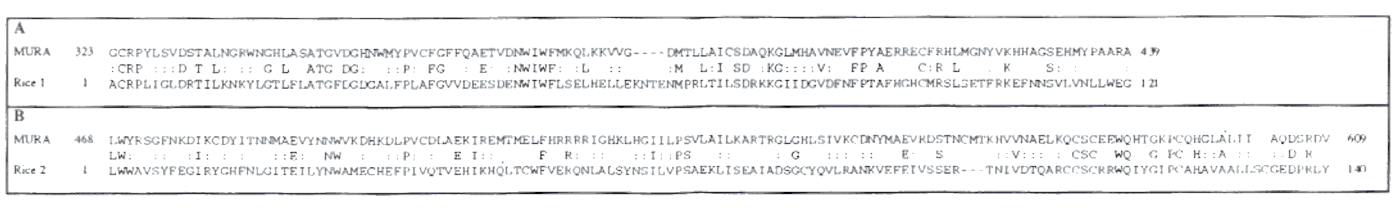

Figure 2. Alignment of regions of MURA with the amino acid sequences encoded by two cDNAs from rice (*Oryza sativa*). A: Rice 1 — cDNA C1069A (Genbank no. d15675). B: Rice 2 — cDNA C2310 2A (Genbank no. d23146). Both cDNAs are partial sequences of genes expressed in a Nipponbare Japonica tissue culture 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

1. Khan,E. *et al.* (1991) *Nucleic Acids Res.* **19**, 851–860.
2. Capy,P. *et al.* (1994) *Trends Genet.* **10**, 7–11.
3. Doak,T.G. *et al.* (1994) *Proc. Natl. Acad. Sci. USA* **91**, 942–946.
4. Chandler,V.L. and Hardeman,K.J. (1992) *Adv. Genet.* **30**, 77–122.
5. Talbert,L.E. *et al.* (1989) *J. Mol. Evol.* **29**, 28–39.
6. Chomet,P. *et al.* (1991) *Genetics* **129**, 261–270.
7. Hershberger,R.J. *et al.* (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10198–10202.
8. Qin,M. *et al.* (1991) *Genetics* **129**, 845–854.
9. Hershberger,R.J. *et al.* (1994) Submitted.
10. Collins,D.M. and Stephens,D.M. (1991) *FEMS Microbiol. Lett.* **83**, 11–16.
11. Wheatcroft,R. and Laberge,S. (1991) *J. Bacteriol.* **173**, 2530–2538.
12. Guilhot,C. *et al.* (1992) *Mol. Microbiol.* **6**, 107–113.
13. Byrne,M.E. *et al.* (1989) *Gene* **81**, 361–367.
14. Byrne,M.E. *et al.* (1990) *Antimicrobial Agents Chemotherapy* **34**, 2106–2113.
15. Wood,M.S. *et al.* (1990) *J. Bacteriol.* **172**, 1719–1724.
16. Wood,M.S. *et al.* (1991) *Gene* **105**, 101–106.
17. Yates,J.R. *et al.* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7284–7287.
18. Dodd,H.M. *et al.* (1992) *Appl. Environ. Microbiol.* **58**, 3683–3693.
19. Dodd,H.M. *et al.* (unpublished) L20851.
20. Tailliez,P. *et al.* (unpublished) L26311.
21. Rappuoli,R. *et al.* (unpublished) a07012.
22. Galas,D.J. and Chandler,M. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington D.C., pp. 109–162.
23. Altschul,S.F. *et al.* (1990) *J. Mol. Biol.* **215**, 403–410.
24. Sturrock,S.S. and Collins,J.F. MPsrch version 1.3. Biocomputing Research Unit, University of Edinburgh, UK.
25. Henikoff,S. (1992) *New Biol.* **4**, 382–388.
26. Plasterk,R.H.A. (1993) *Cell* **74**, 781–786.
27. Bennetzen,J.L. *et al.* (1993) *Crit. Rev. Plant Sci.* **12**, 57–95.
28. Higgins,D. *et al.* (1992) *Comput. Appl. Biosci.* **8**, 189–191.
29. Sasaki,T. and Minobe,Y. (unpublished) D15675, D23146.