**Tnat1 and Tnat2 from Arabidopsis thaliana: Novel Transposable Elements with Tandem Repeat Sequences**

Kenichi NOMA and Eiichi OHTSUBO*

Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan

(Received 26 October 1999; revised 15 December 1999)

Abstract

A computer-aided homology search of databases found that the nucleotide sequences flanking ATLN44, a non-LTR retrotransposon (LINE) from Arabidopsis thaliana, are repeated in the A. thaliana genome. These sequences are homologous to flanking sequences of 664 bp with terminal inverted repeat sequences of about 70 bp. The 664-bp sequence and most of the 14 homologues identified were flanked by direct repeat sequences of 9 bp. These findings indicate that the repeated sequence, named Tnat1, is a transposable element that duplicates a 9-bp sequence at the target site on transposition and that ATLN44 is inserted in one Tnat1 member. Interestingly, all of the Tnat1 members had tandem repeats comprised of several units of a 60-bp sequence, the number of repeats differing among Tnat1 members. Of the Tnat1 members identified, one was inserted into another sequence repeated in the A. thaliana genome: that sequence is about 770 bp long and has terminal inverted repeat sequences of about 110 bp. The sequence is flanked by direct repeats of a 9-bp sequence, indicating that it is another transposable element, named Tnat2, from A. thaliana. Moreover, Tnat2 members had a tandem repeat about 240 bp long. Tnat1 and Tnat2 with tandem repeats in their internal regions show no homology to each other or to any of the elements identified previously; therefore they appear to be novel transposable elements.

**Key words:** Arabidopsis thaliana; transposable element; tandem repeats; target site duplication

1. Introduction

Eukaryotic genomes have various types of interspersed DNA sequences, many of which have been characterized as mobile genetic elements.1 These elements include transposable DNA elements which move from one site to another by excision and reintegration. Many types of transposable DNA elements in plants have been identified and classified into three major families;2 Ac, En/Spm and MuDR. These elements have terminal inverted repeat sequences (TIRs) characteristic of individual elements and a gene(s) encoding transposase which promotes transposition. By the action of the transposase, a target sequence several bp long is duplicated during transposition. Small non-autonomous elements are produced from these elements by deletion of a transposase gene. These non-autonomous elements transpose when transposase is provided by an autonomous element.3,4 In addition, plant genomes contain other elements, such as Tn1/Stowaway7 and Tourist,8,9 which are too small to have the transposase gene. These elements, however, do have TIRs and appear to generate duplication of a sequence of a few base pairs at the target site. These and some other elements with TIRs often are collectively called MITEs (miniature inverted-repeat transposons).10 Retroelements which transpose via cDNA synthesis from an RNA intermediate comprise another category of mobile genetic elements which consist of retrotransposons with long terminal repeats (LTRs) and non-LTR retrotransposons (also called LINEs).11,12 We recently reported that LINEs are ubiquitous components in the plant kingdom as are LTR retrotransposons.13 During the course of our work, many LINEs present in A. thaliana (named ATLN) were identified by a computer-aided homology search of nucleotide sequences in databases. We report here that one of the ATLN elements, ATLN44, is inserted in a novel transposable element that has tandem repeats of a sequence about 60 bp long and that one of the Tnat1 members in the A. thaliana genome is inserted into Tnat2, another novel transposable element with a tandem repeat about 240 bp long. Although these elements have no sequence homology, they may constitute a new class of transposable elements.
2. Materials and Methods

2.1. Plant materials

Total DNAs from three ecotype strains of *A. thaliana*, Columbia, Landsberg and Wassilewskija, were used. The DNA samples were provided by Dr. M. Umeda.

2.2. Computer analyses

The nucleotide sequence homology search was done with the programs FASTA, BLAST and MP search against the sequences in the DDBJ/GenBank/EMBL DNA databases. Multiple sequences were aligned using the program CLUSTAL W, version 1.7. Primary nucleotide sequences were analyzed with the programs HarrPlot 2.0 and GENETYX-Mac 10.1 system (Software Development).

2.3. PCR

The polymerase chain reaction (PCR) was done with a reaction mixture (50 μl) containing 200 to 300 ng plant DNA, 10 nmol of each dNTP, 25 pmol of each primer and 1.25 units of EX Taq DNA polymerase (Takara) in buffer provided by the supplier of the enzyme. Thirty cycles of amplification were done under the following conditions: denaturation for 30 sec at 94°C, annealing for 30 sec at 55°C, and DNA synthesis for 2 min at 72°C. A Gene Amp PCR System 9700 (Perkin Elmer) was used for the PCR. PCR products were separated in a 1.8% agarose gel.

2.4. Nucleotide sequence accession numbers

The consensus nucleotide sequences of *Tnatl* and *Tnat2* will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases under accession numbers AB033593 and AB033594.

3. Results

3.1. Identification of a novel transposable element, Tnat1

Many LINE homologues from *A. thaliana* (Columbia) were identified by a computer-aided homology search of databases using the nucleotide sequence of the LINE member, *ATLN4*. A further homology search of the sequences of the flanking regions of each homologue showed that sequences that flank the homologue *ATLN4* were present as a repeated sequence in the *A. thaliana* genome (Table 1). The length of the repeated sequence (without the inserted *ATLN4* sequence) was 664 bp. This sequence had terminal inverted repeats (TIRs) of about 70 bp. Direct repeats of a 9-bp sequence were present in the regions adjacent to the 664-bp sequence (Fig. 1), indicating that this sequence is a transposable element, designated *Tnat1* (Transposable element of *A. thaliana* 1), that generates the target site duplication (TSD) with the 9-bp sequence. Interestingly, this element had a sequence about 60 bp long which is repeated 7 times in tandem (Fig. 1). Dot matrix analysis showed a box-like structure with parallel lines at intervals of about 60 bp (Fig. 2).

Of the 14 *Tnat1* members (called *Tnat1-1* to *Tnat1-14*) identified, most had TIRs of 70 bp, but a few showed truncation for one of the TIRs of the *Tnat1* sequence (Table 1). Most *Tnat1* members with TIRs appeared to generate a TSD with a 9-bp sequence (Fig. 1, Table 1). The target sequences of *Tnat1* differed but were very rich in AT content, 88.9% on average (Fig. 1, Table 1). All of the *Tnat1* members had tandem repeats of a unit sequence about 60 bp long (Fig. 1). The *Tnat1* members had different unit numbers of tandem repeats (Fig. 1), as shown by the numbers of parallel lines in the dot matrices (Fig. 2). These tandem repeat sequences of the *Tnat1* members were homologous to one another (Fig. 1).

3.2. Presence of Tnat1 members in three ecotype strains of *A. thaliana*

PCR was done with a pair of primers that hybridized with the flanking sequences of *Tnat1-2* to *Tnat1-6* of the total DNAs from three ecotype strains of *A. thaliana* as templates: Columbia, Landsberg and Wassilewskija. All three strains generated fragments (852 bp long) with *Tnat1-4* (Fig. 3). They also generated fragments with *Tnat1-5* or *Tnat1-6* when a relevant pair of primers was used (data not shown). These results show that *Tnat1-4* to *Tnat1-6* are present in the three *A. thaliana* strains and suggest that these *Tnat1* members were present before the divergence of the *A. thaliana* strains. Two *A. thaliana* strains, Columbia and Landsberg, generated fragments (780 bp long) with *Tnat1-3* (Fig. 3), evidence that *Tnat1-3* is present in those strains. Another *A. thaliana* strain, Wassilewskija, did not generate any fragments (Fig. 3), indicating that Wassilewskija genomic DNA has had a base substitution(s) or deletion(s) which inhibits the ability of a primer(s) to hybridize to the genomic DNA, resulting in PCR amplification failure.

Interestingly, the three *A. thaliana* strains generated two fragments, one (1011 bp long) with another *Tnat1* member, *Tnat1-2*, and the other (260 bp long) without *Tnat1-2* (Fig. 3), evidence that the sequences that hybridize with the primers used are repeated in the *A. thaliana* genome and that *Tnat1-2* is inserted in one of the repeating units.

3.3. Characterization of the target sequence for Tnat1-2: Finding of a second transposable element, Tnat2

A homology search of the databases using the flanking sequences of *Tnat1-2* found homologous sequences in many different loci (Table 1), which supports the indication that the sequence used as the target for insertion of *Tnat1-2* is repeated in the *A. thaliana* genome.
Figure 1. (A) Schematic representation of the structures of four Tnati members. The open arrow represents the sequence and orientation of ATLN44 (5.7 kb) inserted in Tnati-1. Boxes with a solid triangle represent TIRs (IRL and IRR) at the left and right ends of each member, and open triangles the TSD with a 9-bp sequence. Short solid arrows indicate tandem repeat sequences. Thick and thin lines respectively indicate the Tnati and flanking sequences of the element. (B) Alignment of the nucleotide sequences of the end regions of Tnati members. A consensus sequence derived from aligned sequences of the Tnati members is shown by uppercase letters, in which boldface letters indicate TIRs (IRL and IRR). In the sequences of Tnati members, nucleotides identical to those of the consensus sequence are shown by dashes, and deleted nucleotides are shown by slashes. A tandemly duplicated sequence present within IRR of Tnati-1 is underlined. Lowercase letters indicate the sequences flanking Tnati members, in which the possible target site sequences are underlined. Numbers in parentheses show the abbreviated nucleotide length of the middle sequence. (C) Alignment of tandem repeat sequences (TR1-TR7) in the consensus sequence of Tnati. These sequences are aligned with a representative sequence for comparison.

Most of the homologous sequences had TIRs of about 110 bp and were flanked by direct repeats of a 9-bp sequence (Fig. 4, Table 1), evidence that the sequences are members of another transposable element which was designated Tnat2 (Transposable element of A. thaliana 2). The 9-bp sequences duplicated differed from one another, but all were rich in AT content, 88.9% on average (Fig. 4, Table 1). Interestingly, all of the Tnat2 members had tandem repeats of a sequence of about 240 bp (Fig. 2, Fig. 4).

PCR with a relevant pair of primers that hybridized with the sequences of the flanking regions of Tnat2-2, Tnat2-3 or Tnat2-4 was done to see whether these elements are present in three ecotype strains of A. thaliana: Columbia, Landsberg and Wassilewskija. All the strains generated fragments with Tnat2 (data not shown), showing that Tnat2 members are present at the respective loci in the A. thaliana genomes.

4. Discussion

We have shown here that Tnati with TIRs of about 70 bp is a transposable element which appears to generate duplication of a 9-bp target sequence and that Tnat2 with TIRs 110-bp long is another transposable element that generates duplication of a 9-bp target sequence. We have identified 14 Tnati members and 10 Tnat2 members in the A. thaliana genome by a homology search of databases (see Table 1). Considering that about 60% of the genome sequence has been determined as of this search, the A. thaliana genome is assumed to contain about 24 members of Tnati and about 17 members of Tnat2. These Tnati and Tnat2 members appear to be distributed in all the chromosomes of A. thaliana, as judged from the locations of BAC clones used for genome sequencing (Table 1). All the members of Tnati and Tnat2, except one, were less than 1000 bp long (Table 1), indicating that these are not autonomous transposable elements and that the transposase gene is
Novel Transposons \textit{Tnat1} and \textit{Tnat2} from \textit{A. thaliana}

Table 1. Members of \textit{Tnat1} and \textit{Tnat2} in the \textit{A. thaliana} genome.

| Member \(^{a)} \) | Length (bp) | Target site \(^{b)} (5' \rightarrow 3') | Accession | Location \(^{c)} (\text{bp}) | Chromosome \(^{d)} |
|-----------------|-------------|-----------------------------|-----------|--------------------------|---------|
| \textit{Tnat1}  | 1           | ttaatatat                  | AC005957  | 19796-26097              | 2       |
|                 | 2           | tatccgat                   | AB018121  | 34233-34974              | 3       |
|                 | 3           | ttatcaat                   | AC005561  | 30669-31251              | 2       |
|                 | 4           | atatatag                   | AC006267  | 17990-17325              | 4       |
|                 | 5           | taaaaagaa                  | AC006216  | 21689-22265              | 1       |
|                 | 6           | aaaaaaaaaa                 | AC006194  | 57357-56891              | 2       |
|                 | 7           | aaagaattt                  | AC04135   | 12228-12756              | 1       |
|                 | 8           | tttaagatt                  | AF18223   | 36967-37514              | 4       |
|                 | 9           | (tttttttta)                | AC004136  | 8872-8306               | 2       |
|                 | 10*         | (ttttgatcaa)               | 29733    | 150385-149991            | 4       |
|                 | 11*         | (ttttctatct)               | AB015477  | 39933-39452              | 5       |
|                 | 12*         | (ttatataaaa)               | AF05914   | 107152-107700            | 5       |
|                 | 13*         | (ttattttatc)               | AC006283  | 19808-19450              | 2       |
|                 | 14*         | —                          | AC006702  | 27405-27509              | 5       |
| \textit{Tnat2}  | 1*          | (tttttaaaaaa)             | AB018121  | 33956-35231              | 3       |
|                 | 2           | aagaaacaa                  | AC006424  | 96502-95732              | 1       |
|                 | 3           | attttttta                  | AC005170  | 14759-15519              | 2       |
|                 | 4           | ttatatatta                 | AB009050  | 25911-26678              | 5       |
|                 | 5           | ataaacaa                   | AC006233  | 5075-5927               | 2       |
|                 | 6           | (gatttttaa)                | AC006216  | 95422-96199              | 1       |
|                 | 7           | (tttattatat)               | AF075598  | 42386-44119              | 4       |
|                 | 8           | (ttttttttat)               | —         | —                        | —       |
|                 | 9           | ttgaacata                  | AC002329  | 4728-3837               | 4       |
|                 | 10*         | ttttttatc                  | AC004392  | 22582-21639              | 1       |
|                 | 11         | —                          | AB013394  | 82772-82914              | 5       |

a) Asterisks indicate \textit{Tnat1} and \textit{Tnat2} members that appear to truncate for one of the TIRs. b) Target site sequences identified in \textit{Tnat} members with TIRs are shown without parentheses. Sequences in parentheses are possible target sequences, each of which is connected with a TIR of a \textit{Tnat} member. — indicates target site sequences not identified in \textit{Tnat} members that appear to truncate for both TIRs. c) Locations of \textit{Tnat1} and \textit{Tnat2} sequences in the registered sequence at the accession number. Numbers underlined are tentative because all the nucleotide sequences of the BAC clones, registered at the accession number, have yet to be determined. d) Chromosome numbers assigned in http://www.kazusa.or.jp/arabi/displayer/.

required for transposition. To identify autonomous elements, we made a homology search of the databases using the sequences of \textit{Tnat1} and \textit{Tnat2} members, but could not identify them.

The TIRs of \textit{Tnat1} and \textit{Tnat2} were not homologous to those of \textit{Ac}/\textit{Ds} and \textit{Spm}/\textit{dSpm} nor to those of \textit{Tourist} and \textit{Tnrl/Stowaway}, small elements identified in plants.\textsuperscript{5-9} The respective TSD lengths are 8 and 3 bp in the \textit{Ac}/\textit{Ds} and \textit{Spm}/\textit{dSpm} elements and 2 and 3 bp in \textit{Tourist} and \textit{Tnrl/Stowaway}. The lengths of the TSDs of \textit{Tnat1} and \textit{Tnat2}, however, differed by 9 bp from those of the other elements. Interestingly, both \textit{Tnat1} and \textit{Tnat2} had tandem repeat sequences in their internal regions. Some transposable elements have short repeated sequences in their internal regions. For example, \textit{Ac} and \textit{En}/\textit{Spm} have repeat sequences with core sequences of 6-12 bp in their subterminal regions.\textsuperscript{18,19} These subterminal repeats however are present at irregular intervals, and some are present in reverse orientation. No transposable elements with simple tandem repeats like \textit{Tnat1} and \textit{Tnat2} have been reported. This and the fact that the lengths of the TSDs of \textit{Tnat1} and \textit{Tnat2} differ from
those of the other elements indicate that these two elements are novel transposable elements. The transposable element \textit{MuDR} (4942 bp) from maize\textsuperscript{20} has TIRs of large size and generates duplication of a 9-bp sequence at the target site, as do \textit{Tnat1} and \textit{Tnat2}. \textit{MuDR} has five different sets of direct repeats 11–27 bp long, each repeated three to five times in a central region between two \textit{MuDR} genes, \textit{muadrA} and \textit{muadrB}. Therefore it is likely that \textit{Tnat1} and \textit{Tnat2} are related to \textit{MuDR}. The tandem repeat sequences in the \textit{Tnat} elements, however, were not homologous to the direct repeat sequences in \textit{MuDR}, and the TIRs of \textit{Tnat1} and \textit{Tnat2} were not homologous to those of \textit{MuDR}. A recently identified MITE element, \textit{Bigfoot},\textsuperscript{23} also generates duplication of a 9-bp sequence at the target site, as do \textit{Tnat1} and \textit{Tnat2}. This element, however, has no tandem repeat sequences within it, and its TIRs were not homologous to those of \textit{Tnat1} and \textit{Tnat2}.

As stated, \textit{Ac} and \textit{En/Spm} have repetitive sequences in their subterminal regions. These repeat regions, which are required for transposition, are bound by the transposase encoded in each element.\textsuperscript{21,22} The tandem repeat sequences of \textit{Tnat1} and \textit{Tnat2} therefore are thought to be required for the transposition of these elements. In the subterminal regions of \textit{Ac}, a common AAACGG motif sequence is present, and methylation of the CGG sequence is considered to be involved in the regulation of transposition.\textsuperscript{22} The elements in the \textit{En/Spm} family have very different repetitive sequences in their subterminal regions but have the CGG sequence, whose methylation regulates their transposition.\textsuperscript{22} The tandem repeat sequences of \textit{Tnat1} have two AAACGA sequences, similar to the subterminal motif sequence of \textit{Ac} (AAACGG), as well as a TATCGG sequence, which suggests that this element also may be regulated by methylation. The tandem repeat sequences of \textit{Tnat2} do not have the CGG sequence but do have several CG and CNG sequences, suggesting that methylation at these sequences in \textit{Tnat2} may regulate transposition.

Small elements from \textit{Oryza sativa}, called \textit{Tnr2}, \textit{Tnr4}, and \textit{Tnr8}, that have tandem repeats in the internal regions, have recently been identified in our laboratory (unpublished results). \textit{Tnr2} (157 bp) with TIRs of 56 bp has a tandem repeat of 20 bp in its internal region. \textit{Tnr4} (1767 bp) with TIRs of 64 bp has two kinds of tandem repeats, 84 and 93 bp, in its internal region. \textit{Tnr8} (418 bp) with TIRs of 187 bp has tandem repeats of 30 bp. These elements do not show sequence homology to one another, rather they appear to generate a TSD with a 9-bp sequence like that of the \textit{Tnat1} and \textit{Tnat2} described in this paper. The dot matrices made by the same sequences of \textit{Tnat1} and \textit{Tnat2} members showed a box-like structure with parallel lines (Fig. 2). The same structure was present in \textit{Tnr2}, \textit{Tnr4} and \textit{Tnr8}. Therefore when a sequence with tandem repeats occurs in a genome, it may be a critical portion of a transposable element. Another interesting speculation is that transposable elements, including \textit{Ac} and \textit{En/Spm}, may be derived from elements with simple tandem repeat sequences, converting them into elements with complicated subterminal repeats by genetic recombinations in the tandem repeat sequences.

Acknowledgments: We thank Dr. M. Umeda for providing the DNA samples used in this study. This work was supported by a Grant-in-Aid for Scientific Research
from the Ministry of Education, Science, Sports and Culture of Japan. K. N. is the recipient of JSPS Fellowships for Japanese Junior Scientists.

References

1. Berg, D. E. and Howe, M. M. 1989, Mobile DNA, American Society for Microbiology, Washington, DC.
2. Saedler, H. and Gierl, A. 1996, Transposable Elements, Springer-Verlag, Berlin, Heidelberg.
3. Fedoroff, N. V. 1989, About maize transposable elements and development, Cell, 66, 181–191.
4. Benetzen, J. L. 1996, The Mutator transposable element system of maize. In: Saedler, H. and Gierl, A. (eds) Transposable elements, Springer-Verlag, Berlin, Heidelberg, pp. 195–229.
5. Umeda, M., Ohtsubo, H., and Ohtsubo, E. 1991, Diversification of the rice Waxy gene by insertion of mobile DNA elements into introns, Jpn. J. Genet., 66, 569–586.
6. Tenzen, T., Matsuda, Y., Ohtsubo, H., and Ohtsubo, E. 1994, Transposition of Tnrl in rice genomes to 5′-PuTAp-3′ sites, duplicating the TA sequence, Mol. Gen. Genet., 245, 441–448.
7. Bureau, T. E. and Wessler, S. R. 1994, Stowaway: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants, Plant Cell, 6, 907–916.
8. Bureau, T. E. and Wessler, S. R. 1992, Tourist: A large family of small inverted repeat elements frequently associated with maize genes, Plant Cell, 4, 1283–1294.
9. Bureau, T. E. and Wessler, S. R. 1994, Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses, Proc. Natl. Acad. Sci. USA, 91, 1411–1415.
10. Bureau, T. E., Ronald, P. C., and Wessler, S. R. 1996, A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice plants, Proc. Natl. Acad. Sci. USA, 93, 8524–8529.
11. Gabriel, A. and Boeke, J. D. 1993, Retrotransposon reverse transcription, In: Skalka, A. M. and Goff, S. P. (eds), Reverse Transcriptase, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 275–328.
12. Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. 1989, LINEs and related retrotransposons: long interspersed repeated elements in the eucaryotic genome, In: Berg, D. E. and Howe, M. M. (eds) Mobile DNA, American Society for Microbiology, Washington, DC, pp 593–617.
13. Noma, K., Ohtsubo, E., and Ohtsubo, H. 1999, Non-LTR retrotransposons (LINEs) as ubiquitous components of the plant genomes, Mol. Gen. Genet., 261, 71–79.
14. Pearson, W. R. and Lipman, D. J. 1988, Improved tools for biological sequence comparison, Proc. Natl. Acad. Sci.
15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

16. Smith, T. F. and Waterman, M. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197.

17. Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* **22**, 4673–4680.

18. Gierl, A. 1996. The *En/Spm* transposable element of maize, In: Saedler, H. and Gierl, A. (eds) Transposable elements, Springer-Verlag, Berlin, Heidelberg, pp. 145–159.

19. Kunze, R. 1996. The maize transposable element of *Activator* (*Ac*), In: Saedler, H. and Gierl, A. (eds) Transposable elements. Springer-Verlag, Berlin, Heidelberg, pp. 161–194.

20. Hershberger, R. J., Benito, M. I., Hardeman, K. J., Warren, C., Chandler, V. L., and Walbot, V. 1995. Characterization of the major transcripts encoded by the regulatory *MuDR* transposable element of maize, *Genetics* **140**, 1087–1098.

21. Gierl, A., Litticke, S., and Saedler, H. 1988, Tnp4 product encoded by the transposable element En-l of *Zea mays* is a DNA binding protein, *EMBO J.* **7**, 4045–4053.

22. Kunze, R. and Starlinger, P. 1989. The putative transposase of transposable element *Ac* from *Zea mays* L. interacts with subterminal sequences of *Ac*, *EMBO J.* **8**, 3177–3185.

23. Charrier, B., Foucher, F., Kondorosi, E. et al. 1999, *Bigfoot*: a new family of MITE elements characterized from the *Medicago* genus, *Plant J.* **18**, 431–441.
