Characterization of the Putative Transposase mRNA of *Tag1*, Which Is Ubiquitously Expressed in Arabidopsis and Can Be Induced by Agrobacterium-Mediated Transformation With *dTag1* DNA

Dong Liu and Nigel M. Crawford

Department of Biology, University of California at San Diego, La Jolla, California 92093

Manuscript received December 28, 1997
Accepted for publication March 6, 1998

ABSTRACT

Tag1 is an autonomous transposable element of *Arabidopsis thaliana*. *Tag1* expression was examined in two ecotypes of Arabidopsis (Columbia and No-0) that were transformed with CaMV 35S-*Tag1*-GUS DNA. These ecotypes contain no endogenous *Tag1* elements. A major 2.3-kb and several minor transcripts were detected in all major organs of the plants. The major transcript encoded a putative transposase of 84.2 kD with two nuclear localization signal sequences and a region conserved among transposases of the *Ac* or hAT family of elements. The abundance of *Tag1* transcripts varied among transgenic lines and did not correlate with somatic excision frequency or germinal reversion rates, suggesting that factors other than transcript levels control *Tag1* excision activity. In untransformed plants of the Landsberg ecotype, which contain two endogenous *Tag1* elements, no *Tag1* transcripts were detected. Agrobacterium-mediated transformation of these Landsberg plants with a defective 1.4-kb *Tag1* element resulted in the appearance of full-length *Tag1* transcripts from the endogenous elements. Transformation with control DNA containing no *Tag1* sequences did not activate endogenous *Tag1* expression. These results indicate that Agrobacterium-mediated transformation with *dTag1* can activate the expression of *Tag1*.

An autonomous transposable element of *Arabidopsis thaliana*, *Tag1* undergoes somatic and germinal excision late in shoot development (Tsay *et al.* 1993; Frank *et al.* 1997; Liu and Crawford 1998). *Tag1* was first uncovered as an insertion in the fourth intron of a nitrate transporter gene, *CHL1*, which produced plants resistant to chlorate (Tsay *et al.* 1993). *Tag1* is 3.3 kb in length, has 22-bp terminal inverted repeats, and produces an 8-bp direct repeat upon insertion. Analysis of *Tag1* genomic sequence has revealed that *Tag1* is a member of the *Ac* or hAT family of elements, which includes *Ac* and *Bg* in maize, *Tam3* in snapdragon, *Hobo* in Drosophila, *Hermes* in housefly, *Slide* in tobacco, and *Restless* in the fungus *Tolypocladium inflatum* (Calvi *et al.* 1991; Warren *et al.* 1994; Essers and Kunze 1995; Grappin *et al.* 1996; Kempken and Kuck 1996). All members of this family produce an 8-bp target site duplication and share a signature protein sequence near the C terminus of the transposase.

Although *Tag1* is endogenous to the Arabidopsis genome, it is not found in all ecotypes (geographical races) of *A. thaliana* (Tsay *et al.* 1993; Bhatt *et al.* 1998; Frank *et al.* 1998). For example, two *Tag1* elements are present in the Landsberg *erecta* ecotype but none in Columbia, WS, or certain isolates of No-0. Ecotypes that lack endogenous *Tag1* elements have been used for studying *Tag1* excision from the marker gene CaMV 35S-GUS (Frank *et al.* 1997). Such studies have shown that *Tag1* somatic excision activity is restricted to late stages of vegetative and reproductive development in the shoot (Liu and Crawford 1998). These studies also showed that *Tag1* germinal excision activity is affected by *Tag1* copy number, genetic dosage, and chromosomal location and can be as high as 25–30% (Liu and Crawford 1998). Still unknown are the mechanisms that control *Tag1* excision and the identity of the gene product(s) that are needed for *Tag1* transposition.

To begin elucidating the mechanisms and gene products that control *Tag1* transposition, we have characterized the mRNA expression patterns of *Tag1* and compared them to somatic and germinal excision rates of the element. We have also isolated and sequenced *Tag1* cDNA clones and found that the major *Tag1* transcript encodes a putative transposase protein containing the signature sequence common to transposases of the *Ac* superfamily. We also found that expression of endogenous *Tag1* elements is ubiquitous, encompasses all major organs of the plant, and can be activated by Agrobacterium-mediated transformation with *dTag1* DNA. The results of these experiments are presented below.

MATERIALS AND METHODS

Plant material: DNA constructs were first transformed into *Agrobacterium tumefaciens* strain C58 AGL-0 (Lazo *et al.* 1991) and then into Arabidopsis plants using vacuum infiltration

Corresponding author: Nigel M. Crawford, Department of Biology, 0116, University of California at San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0116. E-mail: ncrawford@ucsd.edu

(Bechtold *et al.* 1993). Tissue from organs other than the root was obtained from Arabidopsis plants grown in peat soil and grown under continuous light at 23–25° for 3 wk before harvesting. For root tissue, plants were grown for 10 days in submerged liquid culture as described (LaBrie and Crawford 1994).

Molecular cloning and sequence analysis: A PCR-based strategy was used for cloning *Tag1* mRNA. Poly(A)$^+$ RNA was isolated from plant leaves of 35S-*Tag1*-GUS transgenic plants (Frank *et al.* 1997) using QuickPrep *Micro* mRNA Purification Kit (Pharmacia Biotech, Piscataway, NJ). cDNA was synthesized using the Ready-To-Go cDNA Synthesis Kit (Pharmacia Biotech). Primers for subsequent reactions are given below and shown in Figure 1A. Primers 1 and 2 were used to generate the middle 1.6-kb fragment of *Tag1* cDNA. Primer 3 was used with oligo(dT) (primer 4) to amplify the 3′ part of *Tag1* cDNA. The 5′ end of the *Tag1* cDNA was generated by the 5′-RACE (5′-rapid amplification of cDNA end) procedure of Frohman *et al.* (1988). Poly(A)$^+$ RNA was reverse transcribed by first heating 0.5 μg RNA in 38 μl of water at 65° for 5 min, mixed with 10 μl of 5× reaction buffer (250 mm Tris-HCl, pH 8.3, 375 mm KCl, 15 mm MgCl, 5 mm DTT, 5 mm of each dNTP) and 1 μl (1 μg) of primer 5, 1 μl (200 units) of Superscript II RNaseH$^-$ reverse transcriptase (GIBCO BRL Life Technologies, Gaithersburg, MD) and incubated at 37° for 30 min. cDNA products were gel-purified to remove the excess primer and dissolved in 22 μl water. cDNA was mixed with 6 μl 5× buffer (500 mm potassium cacodylate, pH 7.2, 10 mm $CoCl_2$, 1 mm DTT) and 1.5 μl 10 mm dCTP, 1.5 μl terminal deoxynucleotidyl transferase (15 units) (GIBCO BRL Life Technologies) and incubated at 37° for 1 hr. One microliter of poly(C)-tailed cDNA was used for PCR reaction with oligo(dG) (primer 7) and primer 6. One PCR band was produced and subsequently cloned. The complete cDNA clone (see Figure 1B) was assembled from the three fragments. The complete *Tag1* cDNA sequence was deposited in GenBank under the accession number AF051562. The sequence of the primers used in RT-PCR (reverse transcription-polymerase chain reaction) and 5′ RACE procedures are listed below where (+) refers to upper strand and (−) to bottom strand. The first nucleotide of the 5′ inverted repeat is designated as position 1 (Figure 1A).

Primer 1: 5′-GAAACACCATCTTGCTGG-3′ (+: 725–742)
Primer 2: 5′-GCTCACATCCAGATGAAG-3′ (−: 2440–2457)
Primer 3: 5′-GGGATGTACCGAGCA-3′ (+: 1959–1973)
Primer 4: Oligo(dT)18
Primer 5: 5′-TGAAGGACCCACATATCC-3′ (−: 1149–1166)
Primer 6: 5′-CCAGCAAGATGGTGTTTC-3′ (−: 725–742)
Primer 7: Oligo(dG)18

PCR reactions were performed at 94°, 1 min; 60°, 2 min; and 72°, 3 min for 35 cycles. All PCR products were cloned into the *Eco*RV site of pBluescript (SK) vector (Stratagene, La Jolla, CA) and sequenced using dideoxy chain termination methods. Sequence analysis was performed by Wisconsin Sequence Analysis Package "GCG" program (Version 8.0).

Northern hybridization: Total RNAs were isolated as described (Crawford *et al.* 1986). Twenty micrograms total RNA was separated on 1.2% agarose gels containing 6% formaldehyde and transferred to nylon membranes. Hybridizations were performed at 42° for 24 hr in a solution containing 50% formamide, 5× SSPE, 5× Denhardt's solution, 0.1% SDS, and 100 μg/ml herring sperm DNA. After hybridization, membranes were washed twice with 2× SSPE, 0.5% SDS for 15 min, then twice with 0.1× SSPE, 0.1% SDS. The first three washes were at room temperature, and the final wash was at 42°.

***Tag1* excision assay in leaves and germinal activity:** *Tag1* excision assays of plant leaves were performed by histochemical staining for GUS (β-glucuronidase) activity as described (Liu and Crawford 1998). *Tag1* germinal reversion rate was determined by counting the number of progeny from primary transformants that stained completely blue.

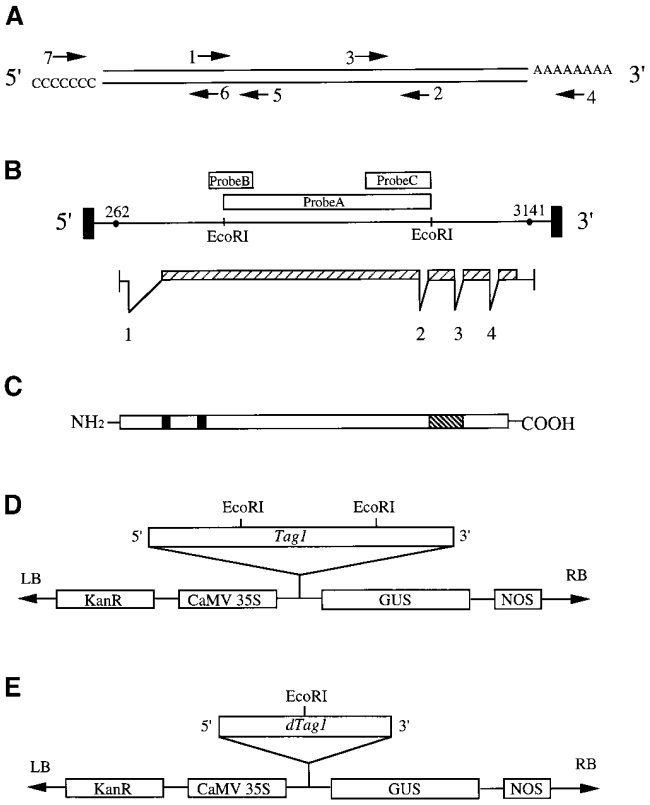

Figure 1.—Diagram of *Tag1*, its products and DNA constructs. DNA cloning steps are described in text. (A) Position of primers used in cloning procedures is shown. The double line represents the cDNA clone of the *Tag1* transcript with poly(C) and poly(G) tails attached at the ends. (B) Diagram of major *Tag1* transcript and introns is shown relative to the *Tag1* genomic DNA. Vertical bars at each end represent terminal inverted repeats. The numbers 262 and 3141 refer to the beginning and end of the *Tag1* transcript. Boxed, hatched line below the diagram refers to the ORF with introns indicated by numbers. Boxes above diagram refer to probes that were a 1.4-kb *Eco*RI fragment (probe A, nucleotides 1096–2424), a 475-bp fragment (probe B, nucleotides 900–1373), and a 662-bp fragment (probe C, nucleotides 1762–2424). (C) Schematic diagram of the protein encoded by major *Tag1* transcript is shown with NLS (■) and *Ac*-transposase homology sequences (▨) indicated. (D) Schematic diagram of *Tag1* in the CaMV 35S-GUS construct used for transformation. (E) Schematic diagram of the defective *dTag1* in the CaMV 35S-GUS construct used for transformation. The *dTag1* is missing the internal 1.4-kb *Eco*RI fragment of *Tag1*.

RESULTS

Expression of *Tag1* element: To examine the mRNA transcripts expressed by *Tag1*, RNA blot hybridization experiments were performed using a 1.4-kb internal *Eco*RI fragment of *Tag1* as probe (probe A in Figure 1B). Total RNA was extracted from leaves of 3-wk-old

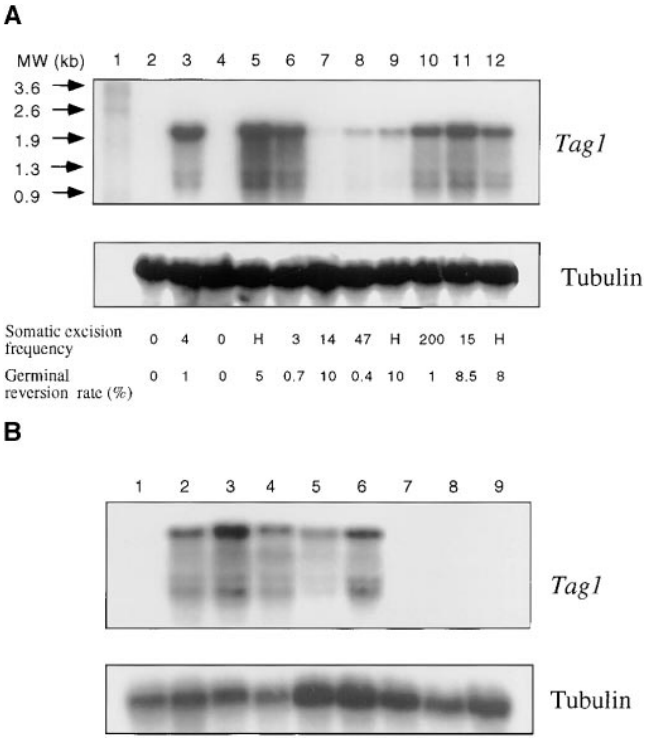

Figure 2.—RNA blot analysis of *Tag1* mRNA in various plant lines. RNA blot conditions are given in materials and methods, and the radiolabeled fragment used for RNA blots was probe A in Figure 1B. (A) RNA blot shows *Tag1* mRNAs along with somatic and germinal reversion frequencies for different transgenic lines of No-0 and Columbia ecotypes carrying 35S-*Tag1*-GUS construct. RNA samples were as follows: lane 1, RNA molecular markers; lane 2, untransformed No-0 plants; lane 3, transformed No-0 plants; lane 4, untransformed Columbia plants; lanes 5–12, transformed Columbia plants. Each transgenic line carries from one to nine copies of *Tag1* at a single locus. Somatic excision frequency is given as the average number of GUS sectors per leaf in leaf 3 or 4 observed in five to 10 plants. Germinal reversion rate was determined by scoring completely blue-staining seedlings among the progeny from primary transformants as described (Liu and Crawford 1998). "H" means high excision frequency, that is, more than 1000 GUS sectors per leaf. (B) RNA blot shows *Tag1* mRNA expression in different Landsberg plant lines. Lane 1, untransformed Landsberg plants; lanes 2–4, Landsberg plants transformed with 35S-*Tag1*-GUS construct; lanes 5 and 6, Landsberg plants transformed with 35S-*dTag1*-GUS construct; lanes 7–9, Landsberg plants transformed with plant expression vector pCGN1578 (McBride and Summerfelt 1990). The defective *Tag1* element (*dTag1*) used for the Landsberg plants is missing the 1.4-*Eco*RI fragment of *Tag1* (Frank *et al.* 1997) that was used as probe A for the RNA blots.

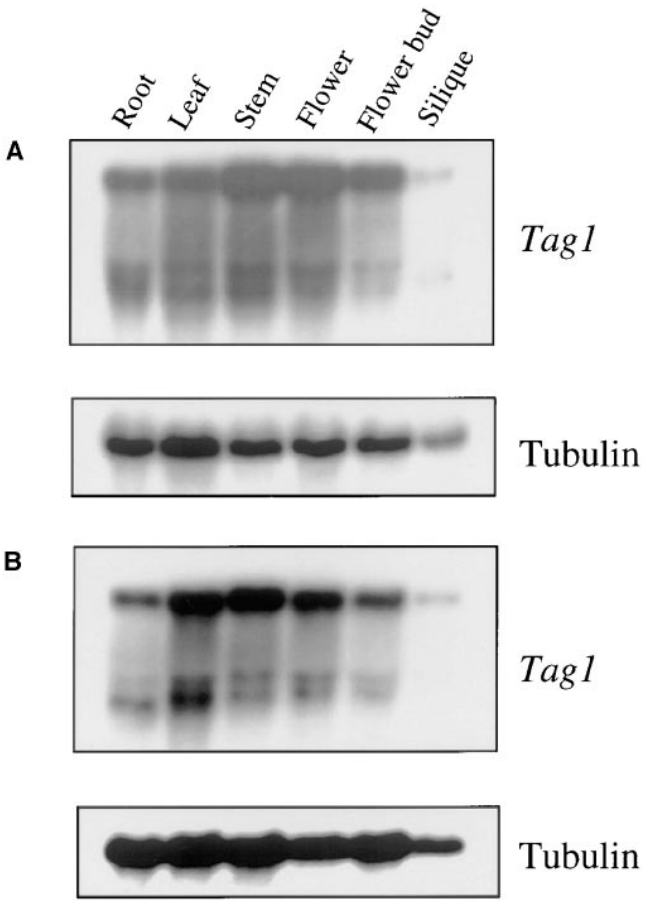

Figure 3.—RNA blot analysis of *Tag1* mRNA expression in various plant organs. Hybridization conditions are given in Figure 2 legend. Blots were hybridized with probe A shown in Figure 1B. (A) RNA samples from a Columbia plant line (see Figure 2A, lane 6) transformed with 35S-*Tag1*-GUS construct. (B) RNA samples from a Landsberg plant line (see Figure 2B, lane 6) transformed with 35S-*dTag1*-GUS construct.

plants that were transformed with a CaMV 35S-*Tag1*-GUS construct (Figure 1D; Frank *et al.* 1997; Liu and Crawford 1998). In untransformed control plants of the No-O or Columbia ecotype, which have no endogenous *Tag1* elements, no *Tag1* transcripts were detected (Figure 2A, lanes 2 and 4). In transformed plants, one major band of ~2.3 kb in size was detected along with two minor bands of 1.0 and 1.2 kb (Figure 2A, lane 3, lanes 5–12).

The transcripts observed in Figure 2A came from *Tag1* elements introduced into plants with the GUS transgene adjacent to the 35S promoter. To determine the transcript pattern produced by endogenous *Tag1* elements, plants of the Landsberg *erecta* ecotype were examined by RNA blot analysis. Plants of this ecotype, hereafter called Landsberg, have two endogenous *Tag1* elements (Frank *et al.* 1997) but show no detectable *Tag1* transcripts by RNA blot analysis (Figure 2B, lane 1). After transformation with the 35S-*Tag1*-GUS construct, the Landsberg plants showed the same *Tag1* transcripts observed for transformed No-0 and Columbia plants as described above (Figure 2B, lanes 2–4). However, one cannot distinguish transcripts of the endogenous elements from those of the transgene in these lines; therefore, transgenic lines were generated that contained a defective *Tag1* element lacking an internal 1.4-kb *Eco*RI fragment (Figure 1E; Frank *et al.* 1997). RNA blots from these transgenic lines probed with the same 1.4-kb *Eco*RI fragment (probe A in Figure 1B)

caatgttttcacgcccgacccgaaccgcccggtccaaccggttaaaccacgacccgagcacttctccggtttgggtttagtgctaaaacccactatctat 100

aaaaccaaaaaaacccataaaaacccgctattaacccacgacccattgaaccggttgaaccgactgggtgtgggttttaaaaactatagttgattttgga 200

acgaacatagggtttttggaacgaacatagggctttttcatagttcataatttcttctcaaACGGCGACTCAATCTCTCGTCTACACTAGCCAAACAAAC 300

CAgttagtttttctatctctcgttttaagtttgaaaatcctatttgtctttatttgtgagcatcatgatcattttatattgtatattgatatagtgattt 400

ctatactctcgtggcctctttgtcttggctccactgaatcaaatataatattcaaagtgtttttgcatcaaaataataattcagcttcatgtttgatgtg 500

catgtctcacagatatttgtttgtttttttgtgtgaatttaatagGTAATGGAAACAGAACACGATGAAAATTATGAAGATATAGCTGCTGCAAACCGTT 600
M E T E H D E N Y E D I A A A N R S

CAATCCGAGGAAAAAGTGATATTGCTTGGTCCTATGTAATTCAATCAAAAGACGAAAAAGGAAAAACGGTTTTGGAATGTGCATTTTGTCATAAAAAAAA 700
I R G K S D I A W S Y V I Q S K D E K G K T V L E C A F C H K K K

AAGAGGAGGTGGAATTAATAGGATGAAACACCATCTTGCTGGCGTGAAAGGGAATACAGATGCATGTTTGAAGATTTCTGCAGATATTAGATTTAAGATA 800
R G G G I N R M K H H L A G V K G N T D A C L K I S A D I R F K I

GTGAACGCATTGAAAGAAGCTGAGAACAAGAAGAAGCAAAATATTGTTTTGGATGATACGAACTTAGACGGTCCAGAAATTGAAGATGTTGATGGAGATG 900
V N A L K E A E N K K K Q N I V L D D T N L D G P E I E D V D G D D

ATATCCGAGTTGATGTTCGGCCAAGTCAAAAGAGGAAGAAACAAGGTATTGATCTACATGATTATTTTAAAAGAGGTGTGCATGATCAGACCCAACCTTC 1000
I R V D V R P S Q K R K K Q G I D L H D Y F K R G V H D Q T Q P S

TATAAAAGCATGCATGCAAAGCAAAGAGAGGATACATGCTGTGGACATGTCTGTTGCATTGTGGTTTTATGACGCTTGTATTCCGATGAATGCGGTGAAT 1100
I K A C M Q S K E R I H A V D M S V A L W F Y D A C I P M N A V N

TCTCCACTTTTTCAGCCTATGATGAGTATGGTAGCAAGTATGGGACATGGATATGTGGGTCCTTCATACCATGCTCTGCGTGTTGGATTGTTGCGTGATG 1200
S P L F Q P M M S M V A S M G H G Y V G P S Y H A L R V G L L R D A

CTAAGTTGCAAGTTTCTTTGATCATTGATAAGTTTAAAAGTAGTTGGGCTAGTACTGGATGTACTCTTATGGCAGATGGATGGAAGGACACAAGACAAAG 1300
K L Q V S L I I D K F K S S W A S T G C T L M A D G W K D T R Q R

ACCACTGATTAACTTCTTAGTTTATTGTCCCAAGGGAATTACATTTCTTAAGTCAGTTGATGCTTCAGATATCTATGCAAGTGCTGAGAATTTATGCAAC 1400
P L I N F L V Y C P K G I T F L K S V D A S D I Y A S A E N L C N

TTGTTTGCTGAACTTGTGGGAATAATTGGTTCAGAAAATATTGTTCATTTTGTGACTGATAGTGCACCCAACTACAAAGCTGCTGGAAAGCTTCTTGTTG 1500
L F A E L V G I I G S E N I V H F V T D S A P N Y K A A G K L L V E

AAAAGTTTCCCACCATTGCTTGGTCTCCTTGTTCAGCTCACTGTATCAATCTTATTTTGGAAGATGTGGCAAAACTACCGCATGTTCACCATATAGTGCG 1600
K F P T I A W S P C S A H C I N L I L E D V A K L P H V H H I V R

CCGTATGTCTAAGGTAACCATTTTTGTTTACAACCATAAGCCGGCTTTAAATTGGGTGAGAAAAAGATCGGGTTGGAGAGAGATCATTCGTCCTGGAGAA 1700
R M S K V T I F V Y N H K P A L N W V R K R S G W R E I I R P G E

ACACGTTTTGCTACAACTTTCATAGCTCTACAAAGTCTCTATCAACATAAAGAAGATCTACAAGCTTTGGTTACAAGTGCAGATCCCGAACTCAAGCAGC 1800
T R F A T T F I A L Q S L Y Q H K E D L Q A L V T S A D P E L K Q L

Figure 4.—*Tag1* nucleotide and protein sequence. Terminal inverted repeat sequences at each end are in boldface. *Tag1* cDNA sequence is in uppercase. Untranscribed regions and intron sequences are in lowercase. Transcription start site is indicated by an arrow. Two putative TATA sequences are boxed. Stop codon is indicated by an asterisk. Four different transcription termination sites are indicated by open circles. Putative polyadenylation signal sequence ATTAAA is double underlined. Two nuclear localization signal sequences are underlined. Protein sequences that show high homology to other transposases are shaded.

should display only transcripts from the endogenous elements. RNA from two such transformed Landsberg lines was found to have the same pattern of major and minor transcripts observed in plants carrying the 35S-*Tag1*-GUS construct (Figure 2B, lanes 5 and 6). No evidence of the probe hybridizing to transcripts from the *dTag1* construct was found. We conclude that, when expressed, *Tag1* primarily produces a 2.3-kb transcript along with several smaller, minor transcripts in leaves and that an adjacent 35S promoter does not affect the transcript pattern produced by *Tag1*.

This RNA analysis also indicated that expression of endogenous *Tag1* elements is much higher in transgenic as compared with untransformed lines, where expression is undetectable in leaves. Perhaps Agrobacterium-mediated transformation activates *Tag1* elements in Landsberg. To test this idea, three Landsberg transgenic lines that had been transformed with control DNA lacking *Tag1* (*i.e.*, with the plant expression vector pCGN1578; McBride and Summerfelt 1990) were examined. No *Tag1* transcripts were detected in the leaves of these lines (Figure 2B, lanes 7–9). Care was taken to

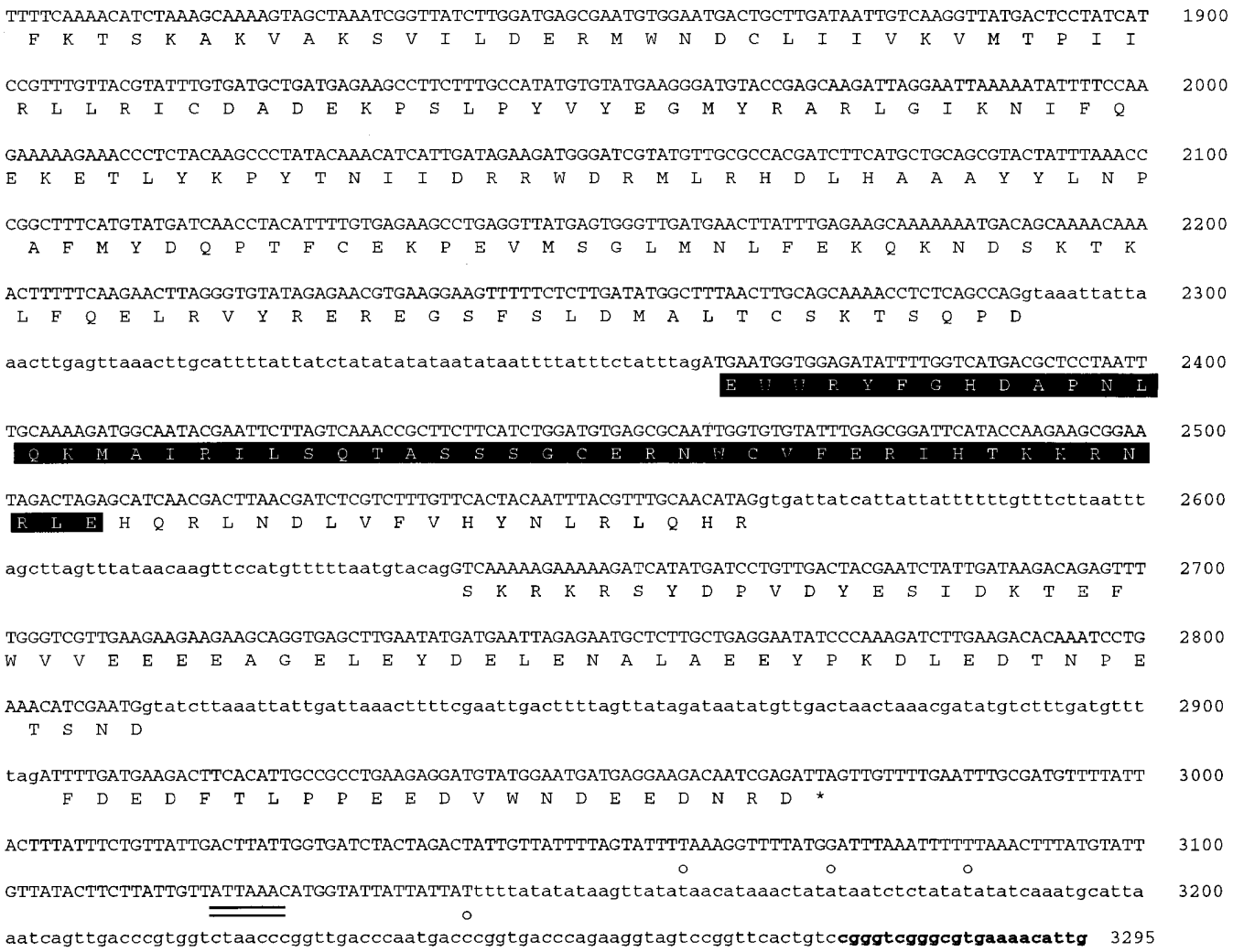

Figure 4.—Continued.

examine plants of the same generation (*i.e.*, T_1 progeny from primary transformants) for both the pCGN and *Tag1*-transformed lines to eliminate propagation effects. These results indicate that transformation with *Tag1* containing DNA, not just transformation itself, activates the expression of the endogenous *Tag1* elements.

The RNA blots in Figure 2A also show the relative abundance of the *Tag1* transcripts in several different lines. Transcript abundance varied from line to line and was compared to the excision activity of *Tag1* in each line. Each transgenic line contained an excision marker (35S-*Tag1*-GUS). Those cells that inherit a 35S-GUS excision allele will stain blue for GUS expression. Somatic excision was assessed by counting the number of GUS sectors in leaf 3 or 4; germinal reversion rate was determined by counting completely blue-staining progeny, as described in Liu and Crawford (1998). No correlation was found between *Tag1* mRNA levels and excision frequency (Figure 2A). Therefore, the *Tag1* mRNA levels in the leaves of these lines do not affect the level of *Tag1* somatic excision activity in leaves and do not correlate with germinal reversion rates.

Next, *Tag1* expression studies were expanded to include other organs of the plant: root, leaf, stem, fully opened flower, young flower bud, and silique. RNA blot analysis was performed using transgenic lines carrying 35S-*Tag1*-GUS or 35S-*dTag1*-GUS constructs. Fairly uniform *Tag1* expression was found throughout the plant (Figure 3). This finding was true both for *Tag1* elements introduced as transgenes in the 35S-*Tag1*-GUS lines (Figure 3A) and for endogenous elements in the 35S-*dTag1*-GUS lines (Figure 3B). This non-organ-specific expression pattern correlates with our finding of *Tag1* excision in all plant organs (Liu and Crawford 1998).

Cloning and sequence analysis of the major *Tag1* transcript: The major *Tag1* transcript was cloned using RT-PCR and 5′ RACE as described in materials and methods. Sequence analysis showed that four introns are removed to produce the final transcript (Figures 1B and 4). All four introns have GT/AG border sequences and contain AT-rich sequences (75–89% AT). All PCR products analyzed had the same 5′ end starting at position 262, but the 3′ end varied to produce spliced products of 2.3–2.4 kb in length (see Figure 4). When we com-

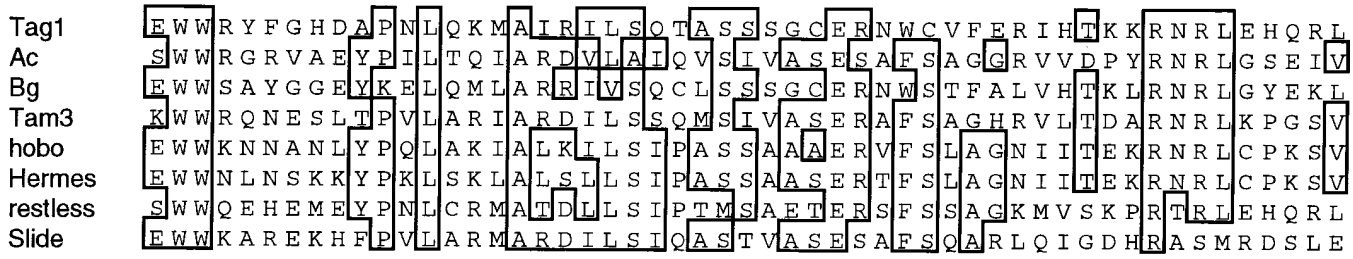

Figure 5.—Conserved transposase sequences in *Ac* or hAT family. Amino acid residues that are identical in at least four sequences are boxed.

pared the cDNA sequence to that published for the *Tag1* genomic clone (accession number L12220; Tsay *et al.* 1993), several discrepancies were found. The original sequence has an extra A, T, T, C, and A at positions 1033, 1630, 1721, 1722, and 3044, respectively, and lacks a T at position 1108. We resequenced the *Tag1* element and found that the cDNA sequence was correct. The correct *Tag1* nucleotide sequence is 3295 bp in length and was provided to GenBank.

Translation of the *Tag1* cDNA reveals a single open reading frame (ORF) that encodes a 729-amino-acid protein with a calculated molecular mass of 84.2 kD and pI of 6.74. Two putative nuclear localization signal (NLS) sequences are located at amino acids 47–51, and 127–144 (Figures 1C and 4). The first NLS sequence consists of five basic amino acids, which is a SV40-like NLS, and the second has a combination of two regions of basic amino acids separated by a space of about 10 residues, which is a bipartite NLS (Hicks *et al.* 1995). The predicted *Tag1* protein also contains the conserved transposase sequence found among members of the *Ac* or hAT family (Figure 1C and 4). The conserved region is also shown in Figure 5, which includes the sequence of the protein encoded by the *Tag1* cDNA. Previous comparisons (Warren *et al.* 1994; Essers and Kunze 1995) used the *Tag1* genomic sequence, which has an extra 4–5 amino acids at the 5′ end of the conserved regions that are not present in the cDNA sequence due to splicing.

Analysis of subterminal regions of *Tag1* element: It has been shown that multiple repetitive sequence motifs in the transposon's subterminal region play an important role for the element's transposition and serve as binding sites for the transposase (reviewed in Saedler and Gierl 1996). We examined the subterminal regions of *Tag1* and found at the 5′ end a motif of AAACCC repeated 12 times in both orientations (Figure 6A). Six of these repeats are perfect, and the other six have one nucleotide change. There is also a 19-bp direct tandem repeat with the putative "TATA" box sequence. Unlike other transposons, the AAACCC motif is not found in the subterminal region at 3′ end. Instead, the 3′ end region has a TTATT sequence motif repeated 14 times, all in the same orientation (Figure 6B). Eight of the repeats are perfect, and six have one nucleotide change. Two other motifs, TATATA and GACCC, are repeated four times each in direct orientation. This 3′ subterminal region is highly AT-rich (80% AT).

DISCUSSION

In this article we present data on the expression of *Tag1*, along with an analysis of the sequence of the putative transposase and its possible binding sites. RNA blot analysis revealed a major 2.3-kb *Tag1* transcript ubiquitously present in all major organs of the plant. The transcript encompasses almost the entire length of *Tag1* and contains an ORF with two NLS sequences and a region that is highly conserved among transposases of the *Ac* or hAT family of transposons. This conserved region is required for transposition of *hobo* elements in Drosophila (Calvi *et al.* 1991). If the major transcript encodes the functional *Tag1* transposase, then *Tag1* would be most similar to *Ac*, which requires only one transposase transcript for transposition (Pohlman *et al.* 1984; Coupland *et al.* 1988; Kunze and Starlinger 1989). If any of the minor transcripts are also required for transposition, then *Tag1* would be analogous to *Spm*, which requires two mRNAs generated by alternative splicing for transposition (Masson *et al.* 1991). *Tag1* transcription does not appear to resemble *Mutator*, which produces two transcripts by convergent, nonoverlapping transcription (Chomet *et al.* 1991; Hershberger *et al.* 1991, 1995; Joanin *et al.* 1997).

Further analysis of the *Tag1* transcripts showed that the major mRNA is produced from the removal of four introns, all with the consensus border sequences GT-AG and AT-rich internal sequences (Brown 1986; Goodall and Filipowicz 1989). The first intron is located in the 5′ untranslated region, a strategic location for regulation (*e.g.*, see Callis *et al.* 1987; Fu *et al.* 1995; Sieburth and Meyerowitz 1997). Because we found two minor transcripts along with the major one in leaves of transgenic plants, we wondered if they might be generated by alternative splicing as is the case for *Spm* (Masson *et al.* 1989). We obtained several additional cDNA clones by the RT-PCR method described above that were smaller than the major transcript, but none of these

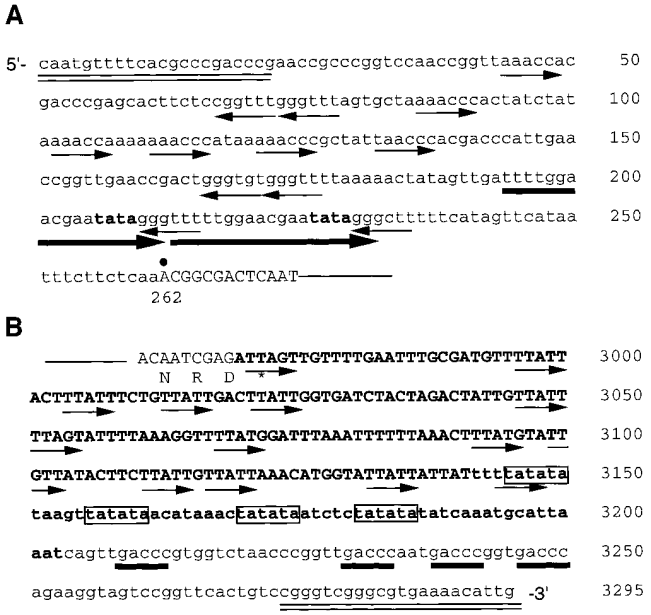

Figure 6.—Analysis of subterminal regions of *Tag1*. (A) Sequence at 5′ end of *Tag1* is shown. Terminal inverted repeat sequence is double underlined. Repeated AAACCC sequences are indicated by thin-lined arrows below. Nine-basepair tandem direct repeats are indicated by arrows with thick line. Putative TATA boxes are in bold. Transcription start site for *Tag1* cDNA is indicated by closed circle at nucleotide 262. Transcribed sequences are capitalized. (B) Sequence of the 3′ end of *Tag1* is shown. Terminal inverted repeat is double underlined. Repeated sequences TTATT are indicated by arrows. TATATA motifs are boxed, and GACCC sequences are indicated by a black bar below. AT-rich sequences are in bold. Transcribed region is capitalized. Stop codon is indicated by an asterisk.

clones corresponded in size to the minor transcripts (data not shown). We also hybridized our RNA blots (as seen in Figure 2) with internal *Tag1* DNA fragments that were missing from the smaller cDNA clones (probes B and C in Figure 1B) and found that all RNA transcripts, major and minor, were labeled (data not shown). The origin of the smaller transcripts remains unknown.

With the borders of the major transcript defined, one can locate and analyze the (100–250 bp) nontranscribed regions of *Tag1*. Typically, sequences adjacent to the inverted repeats (subterminal repeats) of transposons have repeated sequences found at both ends that serve as binding sites for the transposase. For example, *Ac* transposase binds to the AAACGG and related sequence motifs that existed in both 5′ and 3′ subterminal regions (Kunze 1996). For *Spm*, the TNPA component of the transposase binds to 12-bp motifs repeated at both ends (Gierl 1996). A model for transposition is that the transposase binds to the subterminal repeats, bringing together both ends of the element for subsequent cleavage next to the terminal inverted repeats. Sequence analysis of the subterminal repeats in *Tag1* reveals repeated sequences at each end, but the repeated sequences at the 5′ end are different from those at the 3′ end. If these sequence motifs serve as transposase binding sites, instead of some other sequence such as the terminal inverted repeat, then it would appear that *Tag1* does not fit this model for the mechanism of transposition. Further functional tests, including DNA binding assays, will be needed to determine the mechanism of *Tag1* excision.

Tag1 expression is ubiquitous with abundant transcripts in all major organs of Arabidopsis. This finding correlates with the ubiquitous excision activity of *Tag1* in all major organs (Liu and Crawford 1998). However, the abundance of *Tag1* mRNA does not correlate with somatic excision activity, measured as the number of somatic sectors in leaves, in our transgenic lines, nor with germinal reversion rates. These results contrast with those obtained with *Ac* in dicots where a higher level of transposase expression leads to higher frequency of excision up to a limit (Scofield *et al.* 1992, 1993; Swinburne *et al.* 1992). For *Tag1* it appears that some factor other than *Tag1* mRNA abundance as observed on the RNA blots is controlling excision frequency. Perhaps the active transposase levels in these lines are well above the saturating limit in the nucleus. Alternatively, translational or post-translational mechanisms are limiting active transposase levels, regardless of transcript abundance. Last, a component of the transposase not observed on our RNA blots or accounted for by our cDNA clone is limiting or controlling excision frequencies.

Unlike the transgenic lines, untransformed Landsberg plants had no detectable *Tag1* transcripts even though they have two endogenous *Tag1* elements. Transformation of Landsberg plants with a *dTag1* element in the 35S-GUS construct produces plants with high levels of mRNA from the endogenous elements. This apparent activation of *Tag1* expression does not occur when Landsberg plants are transformed with DNA containing no *dTag1* sequences. It is interesting to compare these results with those of Bhatt *et al.* (1998), who showed that *Tag1* transposition, as measured by Southern blot analysis, is activated by Agrobacterium-mediated transformation with any DNA (*i.e.*, *Ac*, *Ds*, and control DNA). A total of 43 new *Tag1* insertion events were observed from 241 transgenic lines. No *Tag1* transposition was observed in untransformed Landsberg lines (188 lines examined), but two out of 118 Landsberg lines crossed to Columbia (*i.e.*, recombinant inbred lines) showed a new *Tag1* insertion. The near absence of transposition in untransformed lines fits with the absence of *Tag1* mRNA that is reported here, but the activation of *Tag1* transposition with non-*Tag1* DNA does not correlate with our finding that transformation with control DNA does not activate *Tag1* expression. The reason for this discrepancy is not clear; however, the transformation protocol was different in these two

sets of experiments: the work of Bhatt *et al.* used root transformation in tissue culture, whereas the work presented here was done using vacuum infiltration of whole plants. In summary, transformation with *dTag1* DNA increases the level of *Tag1* mRNA, possibly by activating the *Tag1* promoter. We do not know if this activation is dependent on a specific sequence within *dTag1*, if it requires transcription of the *dTag1* element to produce some activating protein product, or if it requires excision of the *dTag1*. The *dTag1* element used in these experiments still retains the 5′ upstream regions that include potential binding sites for the transposase and the upstream elements of the promoter. It also can excise from GUS construct (Frank *et al.* 1997). Further analysis of the activation of *Tag1* expression by different *dTag1* sequences will be done to distinguish among these possibilities.

We wish to thank Christian Fankhauser and Joanne Chory for pCGN transgenic Landsberg plants and Mary Frank for No-0 plants transformed with 35S-*Tag1*-GUS. This work was supported by a grant from the National Science Foundation (MCB-9219374).

LITERATURE CITED

Bechtold, N., J. Ellis and G. Pelletier, 1993 In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. Mol. Biol. Genet. **316:** 1194–1199.

Bhatt, A. M., C. Lister, N. Crawford and C. Dean, 1998 The transposition frequency of *Tag1* elements is increased in transgenic Arabidopsis lines. Plant Cell **10:** 427–434.

Brown, J. W. S., 1986 A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res. **14:** 9549–9559.

Callis, J., M. Fromm and V. Walbot, 1987 Introns increase gene expression in cultured maize cells. Genes Dev. **1:** 1183–1200.

Calvi, B. R., T. J. Hong, S. D. Findley and W. M. Gelbart, 1991 Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, Activator and Tam3. Cell **66:** 465–471.

Chomet, P., D. Lisch, K. J. Hardeman, V. L. Chandler and M. Freeling, 1991 Identification of a regulatory transposon that controls the *Mutator* transposable element system in maize. Genetics **129:** 261–270.

Coupland, G., B. Baker, J. Schell and P. Staringer, 1988 Characterization of the maize transposable *Ac* by internal deletions. EMBO J. **7:** 3653–3659.

Crawford, N. M., W. H. Campbell and R. W. Davis, 1986 Nitrate reductase from squash: cDNA cloning and nitrate regulation. Proc. Natl. Acad. Sci. USA **83:** 8073–8076.

Essers, L., and R. Kunze, 1995 Transposable elements *Bg* (*Zea mays*) and *Tag1* (*Arabidopsis thaliana*) encode protein sequences with homology to *Ac*-like transposases. Maize Genetics Cooperation Newsletter **69:** 39–41.

Frank, M. J., D. Liu, Y.-F. Tsay, C. Ustach and N. M. Crawford, 1997 *Tag1* is an autonomous transposable element that shows somatic excision in both Arabidopsis and tobacco. Plant Cell **9:** 1745–1756.

Frank, M. J., D. Preuss, A. Mack, T. C. Kuhlmann and N. M. Crawford, 1998 The *Arabidopsis* transposable element *Tag1* is widely distributed among *Arabidopsis* ecotypes. Mol. Gen. Genet. **257:** 478–484.

Frohman, F. M., M. K. Dush and G. R. Martin, 1988 Rapid production of full-length cDNA from rare transcripts: amplication using a single gene specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:** 8998–9002.

Fu, H., S. Y. Kim and W. D. Park, 1995 High-level tuber expression and sucrose inducibility of a potato *Sus4* sucrose synthase gene require 5′ and 3′ flanking sequences and the leader intron. Plant Cell **7:** 1387–1394.

Gierl, A. 1996 The *En/Spm* transposable element of maize, pp. 145–159 in *Transposable Elements*, edited by H. Saedler and A. Gierl. Springer-Verlag, Berlin.

Goodall, G. J., and W. Filipowicz, 1989 The AU-rich sequence present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell **58:** 473–483.

Grappin, P., C. Audeon, M.-C. Chupeau and M.-A. Grandbastien, 1996 Molecular and function characterization of Slide, an *Ac*-like autonomous transposable element from tobacco. Mol. Gen. Genet. **252:** 386–397.

Hershberger, R. J., C. A. Warren and V. Walbot, 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.

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

Hicks, G. R., H. M. S. Smith and N. V. Raikhel, 1995 Three classes of nuclear import signals bind to plant nuclei. Plant Physiol. **107:** 1055–1058.

Joanin, P., R. J. Hershberger, M.-I. Benito and V. Walbot, 1997 Sense and antisense transcripts of the maize *MuDR* regulatory transposon localized by *in situ* hybridization. Plant Mol. Biol. **33:** 23–36.

Kempken, F., and U. Kuck, 1996 *restless*, an active *Ac*-like transposon from the fungus *Tolypocladium inflatum*: structure, expression, and alternative RNA splicing. Mol. Cell. Biol. **16:** 6563–6572.

Kunze, R. 1996 The maize transposable element *Activator* (*Ac*), pp. 162–194 in *Transposable Elements*, edited by H. Saedler and A. Gierl. Springer-Verlag, Berlin.

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

LaBrie, S. T., and N. M. Crawford, 1994 A glycine to aspartic acid change in the MoCo domain of nitrate reductase reduces both activity and phosphorylation levels in *Arabidopsis*. J. Biol. Chem. **269:** 14497–14501.

Lazo, G. R., P. A. Stein and R. A. Ludwig, 1991 A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotech. **9:** 963–967.

Liu, D., and N. M. Crawford, 1998 Characterization of the germinal and somatic activity of the Arabidopsis transposable element *Tag1*. Genetics **148:** 445–456.

Masson, P., G. Rutherford, J. A. Banks and N. Fedoroff, 1989 Essential large transcripts of the maize *Spm* transposable element are generated by alternative splicing. Cell **58:** 755–765.

Masson, P., M. Strem and N. Fedoroff, 1991 The tnpA and tnpD gene products of the Spm element are required for transposition in tobacco. Plant Cell **33:** 73–75.

McBride, K. E., and K. R. Summerfelt, 1990 Improved binary vectors for *Agrobacterum*-mediated plant transformation. Plant Mol. Biol. **14:** 269–276.

Pohlman, R. F., N. V. Fedoroff and J. Messing, 1984 The nucleotide sequence of the maize controlling element *Activator*. Cell **37:** 635–643.

Saedler, H., and A. Gierl (Editors), 1996 *Transposable Elements. Current Topics in Microbiology and Immunology*. Springer-Verlag, Berlin.

Scofield, S. R., K. Harrison, S. J. Nurrish and J. D. G. Jones, 1992 Promoter fusions to the Activator transposase gene cause distinct patterns of Dissociation excision in tobacco cotyledons. Plant Cell **4:** 573–582.

Scofield, S. R., J. J. English and J. D. G. Jones, 1993 High level expression of the *Activator* transposase gene inhibits the excision of *Dissociation* in tobacco cotyledons. Cell **75:** 507–517.

Sieburth, L. E., and E. M. Meyerowitz, 1997 Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. Plant Cell **9:** 355–365.

Swinburne, J., L. Balcells, S. R. Scofield, J. D. G. Jones and G. Coupland, 1992 Elevated levels of Activator transposase mRNA are associated with high frequencies of Dissociation excision in Arabidopsis. Plant Cell **4:** 583–595.

Tsay, Y.-F., M. J. Frank, T. Page, C. Dean and N. M. Crawford, 1993 Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. Science **260:** 342–344.

Warren, W. D., P. W. Atkinson and D. A. O'Brochta, 1994 The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac* and *Tam3* (*hAT*) element family. Genet. Res. **64:** 87–97.

Communicating editor: D. Preuss