Functional Dissection of the *cis*-Acting Sequences of the Arabidopsis Transposable Element *Tag1* Reveals Dissimilar Subterminal Sequence and Minimal Spacing Requirements for Transposition

Dong Liu,[1] Alyson Mack, Rongchen Wang, Mary Galli, Jason Belk,[2] Nan I. Ketpura[3] and Nigel M. Crawford

Section of Cell and Developmental Biology, Division of Biology, University of California, San Diego, California 92093-0116

Manuscript received April 10, 2000
Accepted for publication October 12, 2000

ABSTRACT

The Arabidopsis transposon *Tag1* has an unusual subterminal structure containing four sets of dissimilar repeats: one set near the 5′ end and three near the 3′ end. To determine sequence requirements for efficient and regulated transposition, deletion derivatives of *Tag1* were tested in Arabidopsis plants. These tests showed that a 98-bp 5′ fragment containing the 22-bp inverted repeat and four copies of the AAACC*X* (*X* = C, A, G) 5′ subterminal repeat is sufficient for transposition while a 52-bp 5′ fragment containing only one copy of the subterminal repeat is not. At the 3′ end, a 109-bp fragment containing four copies of the most 3′ repeat TGACCC, but not a 55-bp fragment, which has no copies of the subterminal repeats, is sufficient for transposition. The 5′ and 3′ end fragments are not functionally interchangeable and require an internal spacer DNA of minimal length between 238 and 325 bp to be active. Elements with these minimal requirements show transposition rates and developmental control of excision that are comparable to the autonomous *Tag1* element. Last, a DNA-binding activity that interacts with the 3′ 109-bp fragment but not the 5′ 98-bp fragment of *Tag1* was found in nuclear extracts of Arabidopsis plants devoid of *Tag1*.

A 3.3-kb DNA transposable element of *Arabidopsis thaliana*, *Tag1*, was discovered as an insertion in the nitrate transporter gene *NRT1* (*CHL1*) (Tsay *et al.* 1993). *Tag1* is the only known mobile element native to the Arabidopsis genome to date although sequences or elements with similarity to a large variety of retrotransposon and transposons including *Ac*, *En/Spm*, *MuDR*, MITES, foldbacks and *copia* have been identified (Voytas and Ausubel 1988; Voytas *et al.* 1990; Casacuberta *et al.* 1998; Wright and Voytas 1998; Ade and Belzile 1999; Copenhaver *et al.* 1999; Henk *et al.* 1999; Lin *et al.* 1999; Mayer *et al.* 1999). *Tag1* is an autonomous element (Frank *et al.* 1997) and has the interesting feature that its transposition is controlled during shoot development (Liu and Crawford 1998a). Using 35S-*Tag1*-GUS constructs, it was found that *Tag1* excision is restricted with a few exceptions to late events as manifested by tiny sectors in leaves, flowers, and siliques (Liu and Crawford 1998a), reminiscent of the late excision behavior of the maize *Mutator* element (reviewed in Bennetzen 1996). The late excision of *Tag1* also carries through to the gametes and leads to independent germinal revertants, which appear at frequencies from 0 to 27% (Liu and Crawford 1998a). Late somatic excision has also been observed in rice that has been transformed with *Tag1* (Liu *et al.* 1999).

Analysis of *Tag1* has revealed that it belongs to the *Ac* or hAT superfamily of elements (Warren *et al.* 1994; Essers and Kunze 1995; Liu and Crawford 1998b). *Tag1* produces one major (2.3 kb) and several minor (between 1.0 and 1.2 kb) transcripts in both Arabidopsis and rice (Liu and Crawford 1998b; Liu *et al.* 1999). The major transcript encodes a putative transposase, which contains a signature sequence shared with other members of the hAT family including *Ac* and *Bg* from maize, *Tam3* from snapdragon, *hobo* from Drosophila, *Hermes* from housefly, *Slide* from tobacco, and *restless* from fungus (Liu and Crawford 1998b and references therein). Surrounding the transcribed regions of *Tag1* are 22-bp terminal repeats and subterminal repeat elements (Liu and Crawford 1998b). Unlike other plant transposons, the subterminal repeats of *Tag1* are different at the 5′ and 3′ ends. The 5′ end has 12 copies of a repeat with the consensus AAACCC in direct and inverse orientation. The 3′ end has three nonoverlapping sets of repeats that differ from each other and the 5′ repeat.

For many transposons, the transposase binds to a target sequence at or near the ends of the element. Trans-

Corresponding author: Nigel M. Crawford, Division of Biology, 0116, UCSD, 9500 Gilman Dr., San Diego, CA 92093-0116.
E-mail: ncrawford@ucsd.edu

[1] *Present address:* Dow AgroSciences LLC, 9330 Zionsville Rd., Indianapolis, IN 46268.

[2] *Present address:* School of Pharmacy, University of California, San Francisco, CA 94143-0150.

[3] *Present address:* Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90095-1737.

posase binding results in a synaptic complex that brings the ends of the element together for subsequent DNA cleavage (reviewed in Saedler and Gierl 1996). For the *Caenorhabditis elegans Tc1* and *Tc3* elements, transposase binding occurs within the terminal inverted repeats (Colloms *et al.* 1994; Vos and Plasterk 1994). For *P* elements binding occurs at two sites with the sequence A-T-A/C-C-A-C-T-T-A-A, located very near the 5′ and 3′ ends but outside the inverted terminal repeats (Kaufman *et al.* 1989). For the most-studied plant transposons, sequences in either the terminal inverted repeats or the subterminal repeat region serve as transposase binding sites. For the maize *Mutator* element, a 32-bp sequence in the 210-bp inverted repeat binds the *MURA* transposase protein (Benito and Walbot 1997). The maize *Ac* transposase binds specifically and cooperatively to repetitive ACG and TCG trinucleotides, which are found in >20 copies in both 5′ and 3′ subterminal regions usually within the sequence AAACGG (Kunze and Starlinger 1989; Becker and Kunze 1997). The *Ac* transposase also weakly interacts with the terminal repeats (Becker and Kunze 1997). The *TNPA* transposase protein of the maize *En/Spm* element binds a 12-bp sequence found in multiple copies within the 5′ and 3′ 300-bp subterminal repeat regions (Gierl *et al.* 1988; Trentmann *et al.* 1993). What is similar about all these elements is that the transposase binding sites are identical or very similar at each end. *Tag1*, however, has three sets of repeats within the 3′ subterminal regions that are different from each other and from the repeated sequence within the 5′ subterminal region. This novel organization suggests that *Tag1* excision or transposase binding may involve some unique mechanisms.

This article describes our work to determine what role, if any, the subterminal repeat regions of *Tag1* play in developmentally regulated transposition by establishing the minimal *cis*-acting sequences required for excision and reinsertion of *Tag1* in Arabidopsis. Such studies are typically done by analyzing defective elements native to the host genome or by introducing mutant forms of the element into cells containing a transposase source. Because no defective *Tag1* elements (*dTag1*) have been reported in the Arabidopsis genome, *dTag1* elements were constructed and introduced into plants containing *Tag1* transposase. The initial *dTag1* construct had a deletion of a 1.4-kb internal *Eco*RI fragment (Frank *et al.* 1997). This *dTag1* (2 kb) element was found not to transpose in the commonly used Columbia ecotype of Arabidopsis but did transpose in the Landsberg (*erecta*) ecotype (Frank *et al.* 1997; Liu and Crawford 1998a). Analysis of these ecotypes revealed that Columbia has no *Tag1* elements but Landsberg has two, which are 0.3 cM apart at position 106 cM near the bottom of chromosome 1 (Frank *et al.* 1997, 1998; Bhatt *et al.* 1998). Normally the endogenous elements in Landsberg are dormant and produce no mRNA transcripts (Bhatt *et al.* 1998; Liu and Crawford 1998b). However, transforming this ecotype using Agrobacterium activates the transcription of the native autonomous elements, resulting in the production of transposase, which can mobilize endogenous elements or an introduced defective 2-kb *dTag1* element (Bhatt *et al.* 1998; Liu and Crawford 1998b). This approach of transforming Landsberg plants was then applied to the study of other *dTag1* elements to identify the minimal sequences required for excision. Unfortunately, this approach did not work as smaller *dTag1* elements failed to activate the endogenous elements. Another method for providing transposase was therefore developed and used successfully to reveal which sequences are sufficient and required for efficient and developmentally regulated transposition. During the course of this work, it was discovered that high excision frequencies of *dTag1* are limited by size restrictions on the element.

MATERIALS AND METHODS

Plasmid construction: All *Tag1* deletion derivatives (*dTag1* elements) were generated from plasmid pBT1, which has the entire 3.3-kb *Tag1* element cloned into the *Xba*I and *Bam*HI sites in pBluescript (SK) (Stratagene, San Diego; Liu and Crawford 1998a). This *Tag1* clone contains an additional 8 bp of duplicated target sequences at each end along with 12 and 23 bp of flanking genomic sequences from Arabidopsis (excluding 8-bp target sequence) at the 5′ and 3′ ends, respectively. All constructs were assembled first in the pBluescript (SK) vector, which was modified to delete the *Eco*RI site, and then transferred to the *Xba*I/*Bam*HI sites between CaMV 35S promoter and GUS gene, encoding β-glucuronidase of the plant expression vector pBI121 (CLONTECH, Palo Alto, CA). Details for each *dTag1* element follow.

pTG3: pBTI was digested with *Eco*RI to remove the 1.4-kb internal *Tag1* sequence to produce a 2-kb *dTag1* element (1.1-kb left fragment and 0.9-kb right fragment) in Bluescript to make pTBS33. The 2-kb *dTag1* element was then transferred to pBI121 to make pTG3 (Figure 1).

pTG-A to pTG-E and pTG29 to pTG36: Specific *Tag1* sequences were produced by PCR amplification and then ligated to the 1.1-kb left end fragment or 0.9-kb right fragment of *dTag1* in pTBS33 (Figures 3 and 4). All these *dTag1* elements have the same genomic flanking sequences as found on pBT1 and pTBS33. To generate the 2-kb *dTag1* element without the 8-bp target sequences, primers directed to the 22-bp terminal inverted repeat were used to PCR amplify the left 1.1-kb and right 0.9-kb fragments of *Tag1* element.

pTG0.5, pTG0.9, and pTG5.0: DNA fragments from randomly picked Arabidopsis genomic DNA clones of 0.5, 0.9, and 5 kb in length (corresponding to clones ALC030, ALC119, and ALC093, respectively) were inserted into the *Eco*RI site between the 98-bp 5′ end and 109-bp 3′ end fragments of *Tag1* (see Figure 5).

pTG-5F and pTG-3F: The 5′ and 3′ subterminal fragments (270 and 262 bp, respectively) of *Tag1* were amplified by PCR (see Figure 5). These fragments were used to replace the 1.1- or 0.9-kb end fragments of pTBS33 as shown in Figure 5.

35S-Tag1-SPT (pTS): Tag1 was excised from plasmid pBT1 by *Sac*I/*Xho*I digestion and cloned into the same sites on plasmid pCL0111 (a kind gift from Caroline Dean; Jones *et al.* 1989). In this construct the *Tag1* element is between the 35S promoter and streptomycin phosphotransferase gene

(SPT) in opposite orientation with respect to the 35S-SPT expression cassette (Figure 1).

All DNA constructs were transformed into Agrobacterium strain C58 AGL-0 (Lazo *et al.* 1991), except the construct 35S-*Tag1*-SPT, which was transformed into strain LBA4404.

Plant transformation and DNA analysis: All Arabidopsis plants used in this study are of the Columbia or the Landsberg (*erecta*) ecotype. Plants were grown under continuous light. Vacuum infiltration procedures were used for plant transformation as described (Bechtold *et al.* 1993). After transformation, transgenic plants were selected on GM medium (Valvekens *et al.* 1988) with kanamycin at 30 μg/ml. Isolation of genomic DNA and Southern blot analysis were performed as described (Liu and Crawford 1998a). Flanking sequences were amplified by TAIL PCR as described (Tsugeki *et al.* 1996). PCR products were cloned into pGEM-T PCR vector (Promega, Madison, WI) and sequenced. The degenerate primers used for the PCR were:

AD1: 5′-*N*TCGA*S*T*W*T*S*G*W*GTT-3′
AD2: 5′-*S*TTG*N*TA*S*T*N*CT*N*TGC-3′
AD4: 5′-*N*GTCGA*SW*GA*N*A*W*GAA-3′
AD11: 5′-TCTTTCG*N*ACTT*N*GGA-3′

(*S* = C/G; *W* = A/T; *N* = A/G/C/T). The anchored primers used were:

ALC 119-1: 5′-CGACTGTGATCATCTTCATGC-3′
ALC 119-2: 5′-GGTAATCATATTGGTGTAACCC-3′
ALC 119-3: 5′-GGATATGATCCATACAGTTTGC-3′.

Phenotypic assays for *Tag1* excision frequencies: Histochemical assays for GUS expression were performed as described (Jefferson 1989; Liu and Crawford 1998a). Somatic excisions of *dTag1* elements were visualized as blue sectors on a white background. Germinal excision frequency of *dTag1* elements was scored by counting the number of completely blue staining seedlings from a progeny population of 100 seedlings for each line.

The phenotypic assays for transgenic plants carrying 35S-*Tag1*-SPT construct were performed as described (Dean *et al.* 1992). Seeds were germinated on GM medium containing streptomycin (100 μg/ml). Ten days after germination, germinal revertants were identified as fully green streptomycin-resistant seedlings (Figure 2A). Somatic sectors on cotyledons are observed as green spots on a bleached background (Figure 2B).

Gel mobility-shift assays: Nuclear extracts were prepared from Arabidopsis plant leaves as described (Raghothama *et al.* 1993). Briefly, leaves (25 g) were ground to a fine powder in liquid nitrogen and added to an extraction buffer containing 149 ml Honda buffer (3.3% Ficoll, 6.6% Dextran T40, 33 mm Tris-HCl, pH 8.5, 3.3% Triton X-100, 6.6 mm $MgCl_2$, 44 ml 2 m sucrose, 140 μl 2-mercaptoethanol and 1.89 ml 0.2 m spermine). Material was homogenized in a Polytron three times each 20 sec. Homogenates were filtered through nytex, which was pre-wet with Honda buffer. Filtered homogenate was centrifuged at 5000 rpm for 5 min at 4°. The pellets were resuspended in 15 ml Honda buffer and then centrifuged again. The pellets were resuspended in 10 ml nuclei wash buffer (50 mm Tris-HCl, pH 8.5, 5 mm $MgCl_2$, 1 mm 2-mercaptoethanol, and 20% glycerol) and centrifuged as above. After two washes the pellet was resuspended in 1.5 ml nuclei resuspension buffer (10 mm Hepes, pH 8.0, 50 mm NaCl, 0.5 m sucrose, 0.1 mm EDTA, 0.5% Triton X-100, 1 mm dithiothreitol (DTT), and 5 mm $MgCl_2$). Spermidine and NaCl were added to final concentration of 5 mm and 0.5 m to lyse nuclei on ice for 30 min. After lysis, the sample was centrifuged for 10 min at 4°. The supernatants were dialyzed into 10 mm Hepes, pH 8.0, 1 mm $MgCl_2$, 1 mm DTT, 50% glycerol, 50 mm NaCl, 0.8 mm phenylmethylsulfonyl fluoride overnight with several changes. After dialysis, the nuclei extracts were aliquoted and stored at −70°.

The DNA-protein binding reactions (15 μl) contained 5000 cpm of ^{32}P-labeled target DNA, 2 μg of poly(dI-dC), 5 μg of nuclear protein, 25 mm Hepes, pH 8.0, 40 mm KCl, 5 mm $MgCl_2$, 1 mm DTT, 1 mm EDTA, and 8% glycerol. The binding reaction was incubated at 4° for 30 min. Reaction mixtures were then separated by electrophoresis on 8% polyacrylamide gels in 0.5× TBE buffer. Gels were run at 10 V per cm for 2 hr and exposed to X-ray film at −80° overnight.

RESULTS

Establishing a system for analysis of *Tag1 cis*-acting sequences: We wished to determine which *Tag1* sequences are sufficient for efficient and regulated excision of *Tag1* in Arabidopsis. Our initial strategy was to test a series of *Tag1* deletion derivatives (*dTag1* elements) in Arabidopsis plants of the Landsberg (*erecta*) ecotype (hereafter referred to as Landsberg). Landsberg has two endogenous *Tag1* elements that provide transposase function but only after activation by transformation of plants with *dTag1* DNA (Liu and Crawford 1998b). We introduced *dTag1* elements into Landsberg via Agrobacterium-mediated transformation of whole plants. As described above, the first construct tested was a 2-kb *dTag1* element inserted between a CaMV 35S promoter and GUS marker gene (see Figure 1 for a schematic diagram of this construct, pTG3). This element excises in Landsberg to produce small GUS sectors indicative of late excision (Frank *et al.* 1997; Liu and Crawford 1998a). The size of these sectors showed that the timing of excision for the 2-kb element is late and very similar to that observed for the autonomous *Tag1* element in the same 35S-GUS plasmids (Liu and Crawford 1998a). Thus, the 2-kb *dTag1* element contains sufficient sequence information for developmentally regulated transposition.

We next made a 532-bp *dTag1* element (Figure 1, construct pTG-A) containing only 262-bp sequences from the 5′ end and 270 bp from the 3′ end of *Tag1*, which includes the inverted and subterminal repeats. This *dTag1* was transformed into Landsberg using the 35S-GUS marker construct (construct pTG-A in Figure 1). Eighteen transgenic plants were generated, but none showed any GUS sectors. Northern blot hybridizations failed to detect *Tag1* transcripts in any of the 18 transgenic lines (data not shown). These results suggested that the inability of the 532-bp *dTag1* to excise in these lines was due to a failure to activate the endogenous *Tag1* elements. However, a lack of a required *cis*-acting sequence in the *dTag1* element could also explain the results. To test these possibilities, we developed another method for supplying transposase activity for our *Tag1* deletion mutants.

Instead of relying on the endogenous elements of Landsberg to provide transposase function, we intro-

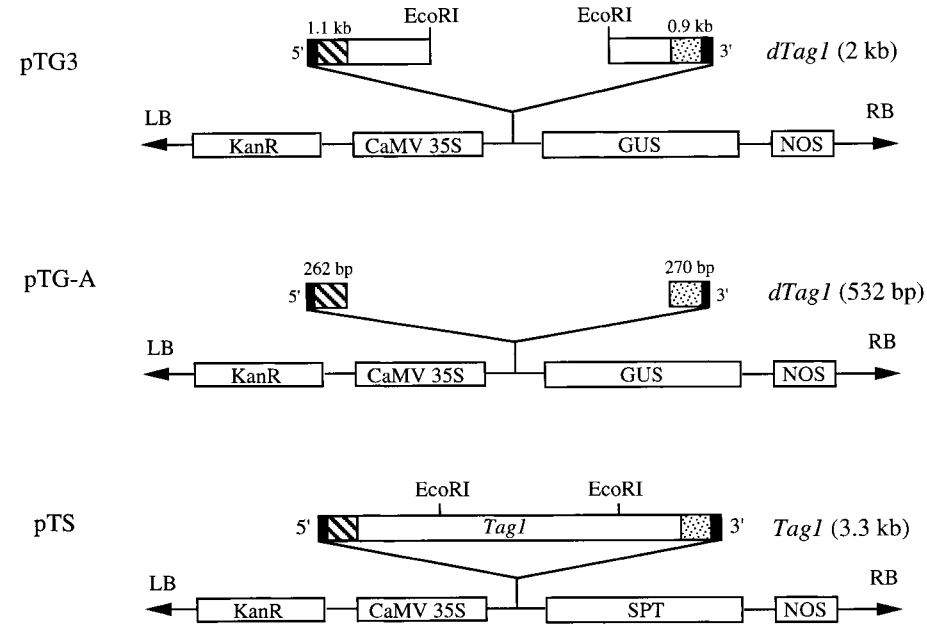

FIGURE 1.—Schematic diagrams of the major DNA constructs used for plant transformation. For each construct, the T-DNA from Agrobacterium is shown along with site of insertion of the *dTag1* construct. CaMV 35S, cauliflower mosaic virus 35S promoter; GUS, β-glucuronidase gene; NOS, nopaline synthase gene terminator; SPT, streptomycin phosphotransferase gene; KanR, kanamycin resistance gene; LB, T-DNA left border; RB, T-DNA right border; (■) terminal inverted repeats; (▧) 5′ subterminal region; (▩) 3′ subterminal region.

duced an intact *Tag1* element into Arabidopsis plants of the Columbia background, which have no endogenous *Tag1* elements (TSAY *et al.* 1993). The *Tag1* element was inserted between a 35S promoter and an SPT (construct pTS in Figure 1) so that excision of the intact *Tag1* could be monitored independently of *dTag1* elements in the GUS vectors. The SPT system has been used to monitor germinal and somatic excision of *Ac* in Arabidopsis (DEAN *et al.* 1992). Germinal excision of *Tag1* should lead to streptomycin-resistant plants appearing as fully green seedlings on agar plates containing streptomycin while untransformed seedlings become completely bleached (Figure 2A). One can also detect green somatic sectors in cotyledons using this system (Figure 2B). Eleven transgenic plants in the Columbia background carrying the 35S-*Tag1*-SPT construct were generated. Evidence for both somatic (Figure 2B) and germinal (Figure 2A) excision events was observed in the progeny of these plants. Northern blot analysis showed that 35S-*Tag1*-SPT transgenic plants produced the typical pattern of *Tag1* mRNA transcripts (data not shown). These results demonstrated that the 35S-*Tag1*-SPT plants could now be used as a source of transposase for testing our *dTag1* constructs.

One line (TS28) was selected that was homozygous for four copies of the 35S-*Tag1*-SPT construct at a single locus (based on analysis of Southern blots and segregation ratios of kanamycin-sensitive to kanamycin-resistant progeny, data not shown). This line has a germinal excision frequency of 5%. This line was crossed with four plants carrying the 2-kb *dTag1* construct pTG-3 in the Columbia background. Recall that the 2-kb *dTag1* element can be mobilized when transformed directly into Landsberg plants, but by itself shows no GUS sectors in Columbia plants (Figure 2C). The F_1 plants from the crosses showed GUS sectors in all plant organs examined (roots, cotyledons, leaves, flowers, and siliques; Figure 2, D–G, and data not shown) indicating that the *Tag1* element from the 35S-*Tag1*-SPT construct had mobilized the *dTag1* element. As evidenced by the tiny sectors, the timing of *dTag1* excision in shoot organs was late in shoot development. The F_1 plants were selfed to produce F_2 seeds, and germinal revertants staining completely blue were identified. These results indicate that *Tag1* in the 35S-*Tag1*-SPT transgenic plants will mobilize a defective element leading to developmentally controlled somatic excision.

The transposase activity in the TS28 35S-*Tag1*-SPT plants was tested next with the 532-bp *dTag1* element (pTG-A, Figure 1, which showed no activity when transformed directly into Landsberg plants). When five plants carrying this *dTag1* element were crossed to TS28, the F_1 progeny from four of the crosses showed GUS sectors that were very small (Figure 2J and data not shown). Thus, the sequences in the 532 *dTag1* element are sufficient for regulated excision when provided with *Tag1* transposase. These results also confirm our hypothesis that the inability of the 532-bp *dTag1* to excise in Landsberg plants is due to its inability to activate the two endogenous *Tag1* elements, rather than to its lacking sufficient *cis*-acting sequences for transposition.

***dTag1* elements less than 445 bp show no *trans*-activation by *Tag1*:** To further dissect the *cis*-acting sequences, four additional deletion derivatives were made of the 532-bp *dTag1* element (Figure 3, constructs pTG-B to pTG-E). These deletions removed some of the repetitive sequences found in the subterminal repeat regions at the 5′ and 3′ ends (shown schematically in Figure 3).

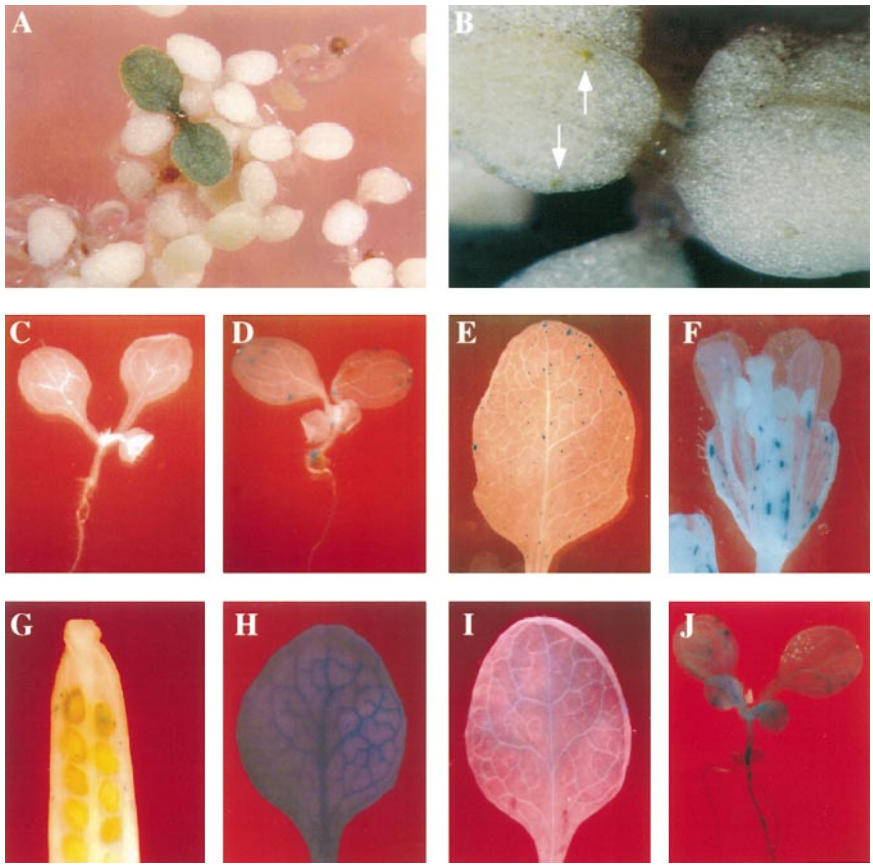

FIGURE 2.—Transposition behavior of *Tag1* and *dTag1* elements in transgenic plants. (A) Seedlings of plant TS28 with 35S-*Tag1*-SPT construct grown on streptomycin-containing medium. *Tag1* germinal revertants are fully green seedlings. Plants retaining the *Tag1* element are bleached. (B) A close-up of a bleached seedling showing somatic sectors (green spots indicated by arrows). (C) Seedling containing pTG3 construct (35S-*dTag1*-GUS) in Columbia stained for GUS activity. (D–G) TS28 plants (containing transposase) were crossed to plants with pTG3 [35S-*dTag1* (2 kb)-GUS], and F_1 progeny were stained for GUS activity to assay for *dTag1* excision. Various organs of F_1 are shown: (D) seedling, (E) leaf, (F) flower, and (G) silique. (H and I) Leaves from two primary transgenic plants containing the construct pTG-A [35S-*dTag1* (532 bp)-GUS] were stained for GUS activity and showed background staining. (J) An F_1 hybrid seedling from a cross between TS28 and the pTG-A containing plant was stained for GUS activity and showed *dTag1* sectors on a blue background.

These constructs were cloned into the 35S-GUS vector and then transformed into Columbia plants. These plants (eight primary transgenic lines per construct) were then crossed to TS28, which provides *Tag1* transposase, and GUS sectors were analyzed in the F_1 plants.

The result from these crosses was that no GUS sectors were observed in any of the F_1 plants (data not shown). To verify that no excision was occurring, primers flanking the *dTag1* insertion sites were used for PCR analysis of DNAs from F_1 hybrid seedlings containing both *Tag1* and *dTag1*. No excision products were detected (data not shown). These results indicate that *dTag1* elements smaller than about 500 bp cannot be mobilized by an autonomous *Tag1* element.

These experiments had the complication that when these smaller *dTag1* elements were tested in the 35S-GUS vectors, significant background or diffuse GUS staining was observed in primary transformants, which

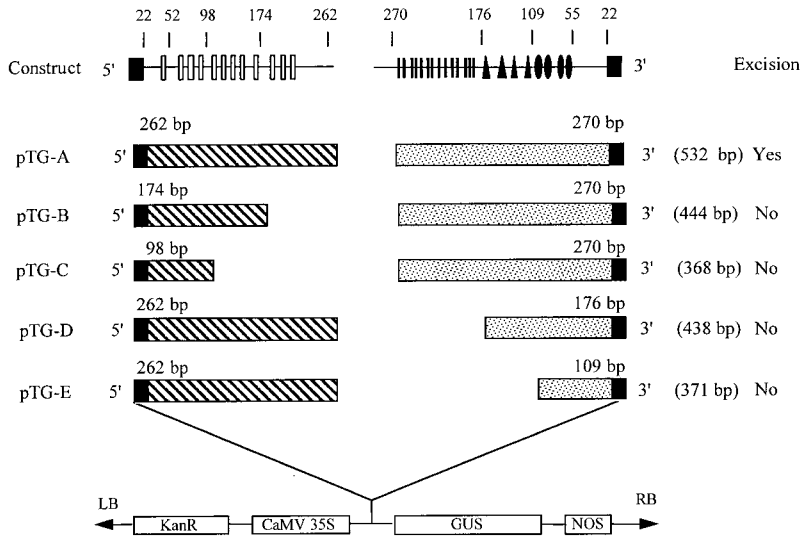

FIGURE 3.—Schematic diagram of small *dTag1* constructs that were inactive. Top line shows the 5′ and 3′ ends (about 300 bp each) of *Tag1* with the position of the subterminal repeats shown using the following symbols: (□) repeat AAACCC, (■) repeat TTATT, (▲) repeat TATATA, (●) repeat TGACCC. The *dTag1* constructs are shown using the same symbols as in Figure 1 and with the size of the element shown in parentheses and its excision activity shown at far right. The bottom shows the T-DNA vector used for plant transformation. Symbols are as described in Figure 1 legend.

TABLE 1

Summary of diffusive GUS staining phenotype of transgenic plants carrying 35S-*dTag1*-GUS constructs

Construct	Size of *dTag1*	No. of lines examined	No. of lines showing diffusive GUS staining
pTG-A	532 bp	14	5 (36%)
pTG-B	444 bp	24	22 (91%)
pTG-C	268 bp	22	19 (81%)
pTG-D	438 bp	24	17 (70%)
pTG-E	371 bp	24	19 (79%)
pTG29	1.28 kb	24	4 (16%)
pTG30	1.21 kb	24	2 (8%)
pTG32	1.07 kb	24	3 (13%)
pTG33	1.00 kb	24	4 (16%)
pTG34	0.95 kb	24	1 (4%)
pTG35	1.12 kb	24	0 (0%)
pTG36	1.16 kb	16	1 (6%)
pTG-0.5	0.70 kb	24	5 (21%)
pTG-0.9	1.10 kb	12	1 (8%)
pTG-5.0	5.20 kb	20	0 (0%)
pTG-5F	1.40 kb	24	5 (21%)
pTG-3F	1.40 kb	24	3 (13%)

contain no transposase, and in F_1 plants containing transposase (see Figure 2, H and I, for examples). In primary transformants, the percentage of plants showing background staining correlates with the size of the element (Table 1). The background staining varied from dark blue to faint blue (Figure 2, H and I) and could be due to a number of factors including "read through" translation. Fortunately, the background staining did not usually obscure visualization of GUS sectors, as can be seen in plants containing the 532-bp *dTag1* element (Figure 2J). In our crosses with the TS28 line, we tested 19 *dTag1*-containing plants with diffusive staining and 13 with no staining. In all cases, no GUS sectors were observed. Therefore, all our data indicate that no excision was occurring for elements <445 bp.

Minimal sequence and spacing requirements for *dTag1* excision: The failure of *Tag1* to mobilize *dTag1* elements smaller than 445 bp could be due to one of two reasons: a necessary sequence is missing in these constructs or all the necessary sequences are present but the distance between the 5′ and 3′ ends is not sufficient. We examined these possibilities by making two additional sets of deletion derivatives that contained a spacer sequence between the two subterminal repeat regions of *Tag1*. Figure 4 shows the various constructs and the relative positions of repetitive sequence motifs. Within the 5′ subterminal region, a motif with a consensus sequence of AAACCC is repeated 12 times in both direct and inverse orientations. Within the 3′ subterminal region, the first motif (TTATT) is repeated 14 times; the second (TATATA) and third (TGACCC) are repeated 4 times each in same orientation. The first set of deletion mutants we made retained the 0.9-kb *Tag1* sequence at the 3′ end fused to various deletion fragments of the 5′ end (Figure 4, construct pTG32, pTG33, and pTG34). Twenty-four independent primary trans-

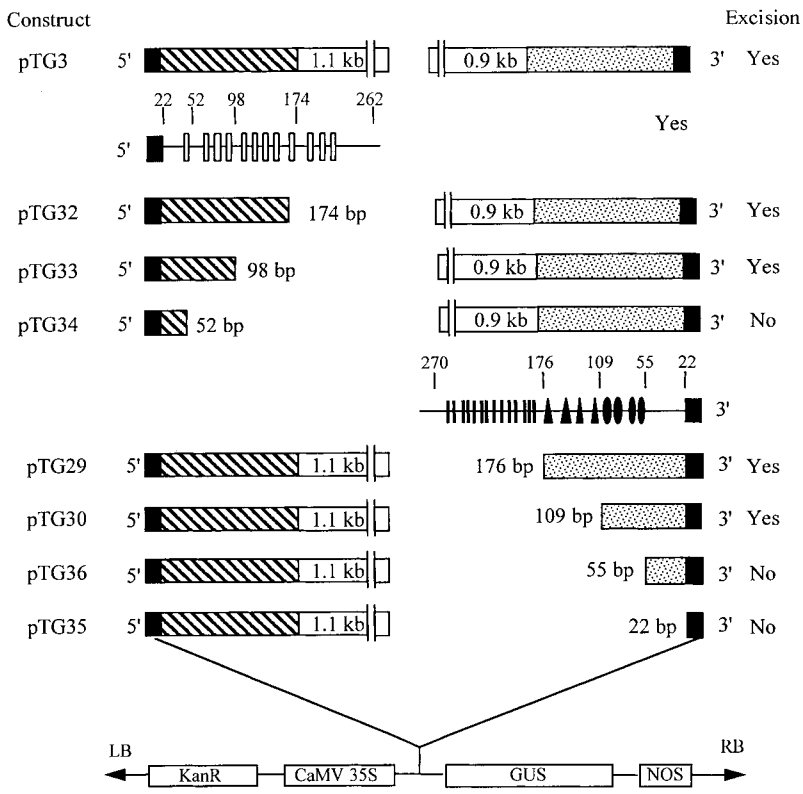

FIGURE 4.—Schematic diagram of larger *dTag1* constructs. A 0.9- to 1.1-kb fragment from either the 5′ or 3′ end of *Tag1* was ligated with various deletion derivatives and inserted into a T-DNA vector and transformed into Columbia. The activity of the *dTag1* elements in these constructs was tested by crossing plants to TS28 (containing transposase). Whether or not excision activity was observed is given on the right of the figure. Positions of subterminal repeats are shown schematically above each set of deletion derivatives. Symbols are as described in Figure 1 legend.

TABLE 2

Somatic and germinal excision frequencies of F_1 hybrids from various 35S-*dTag1*-GUS lines crossed with 35S-*Tag1*-SPT plants

Line	KanR:KanS (selfed progeny)	F_1 hybrid seedling GUS(+)	F_1 hybrid seedling GUS(−)	Excision frequency in leaves	Germinal excision frequency (%)	Average
PTG29-3	3:1	9	5	M, M, M	6, 6, 5	5.6
pTG29-4	50:1	10	18	L, L, L	11, 8, 7	8.6
pTG29-12	3:1	12	21	L, L, L	11, 7, 6	8.0
pTG29-13	3:1	10	11	L, L, L	0, 0, 0	0
pTG29-14	3:1	13	11	M, M, H	1, 4, 3	2.6
pTG29-16	40:1	13	5	H, H, H	13, 11, 11	12.3
pTG29-23	3:1	25	5	H, H, M	5, 7, 8	6.3
pTG29-24	5:1	28	4	L, L, L	0, 0, 0	0
pTG30-3	10:1	14	6	H, H, M	5, 4, 5	4.6
pTG30-5	3:1	19	11	M, M, M	4, 7, 6	5.6
pTG30-9	3:1	7	5	M, M, M	11, 12, 12	11.6
pTG30-12	5:1	0	26	ND	ND	ND
pTG30-13	3:1	6	13	L, L, M	5, 6, 6	5.6
pTG30-17	3:1	9	10	M, M, H	6, 0, 4	3.3
pTG30-21	3:1	5	5	M, M, M	5, 5, 4	4.6
pTG30-23	8:1	9	6	L, M, M	10, 9, 9	9.6
pTG32-4	30:1	36	8	H, H, M	13, 10, 11	11.3
pTG32-7	3:1	5	5	H, M, M	15, 16, 13	14.6
pTG32-10	20:1	11	3	L, L, L	9, 12, 16	12.3
pTG32-15	150:1	23	2	L, M, M	0, 9, 2	3.6
pTG32-19	4:1	15	12	L, L, L	4, 7, 6	5.6
pTG32-23	3:1	18	14	L, L, L	12, 12, 11	11.6
pTG32-27	3:1	12	15	L, L, L	6, 5, 6	5.6
pTG33-2	7:1	20	13	M, M, H	5, 6, 3	4.6
pTG33-3	3:1	15	17	L, L, M	0, 1, 3	1.3
pTG33-4	25:1	15	12	M, L, H	5, 8, 3	5.3
pTG33-6	6:1	16	17	L, L, L	3, 0, 0	1.0
pTG33-14	5:1	9	8	M, M, H	7, 5, 10	7.3
pTG33-18	2:1	15	15	H, H, M	0, 0, 2	0.6
pTG33-19	16:1	17	10	M, M, M	10, 12, 12	11.3

H, >300 sectors/leaf; M, 50–300 sectors/leaf; L, <50 sectors/leaf; ND, none detected.

formants were prepared for each construct, and 7 with no background staining were selected for crossing to TS28 plants. The second set of derivatives retained the 1.1-kb sequence at the 5′ end fused to various fragments of the 3′ end (Figure 4, construct pTG29, pTG30, pTG35, and pTG36). Sixteen to 24 independent transgenic plants were prepared for each construct, and then 6 to 8 with no background staining were selected for crossing to TS28, which contains the *Tag1* transposase. After crossing, the GUS activity in all F_1 seedlings was examined.

The results of this experiment (Table 2) show that the smallest fragments capable of supporting excision are the 98-bp fragment at the 5′ end and the 109-bp fragment at the 3′ end (see Figure 4, constructs pTG33 and pTG30). The next smallest fragments tested, 52 bp at the 5′ end and 55 bp at the 3′ end, did not show any activity *in vivo* (constructs pTG34 and pTG35, Figure 3). The functional 98-bp fragment at the 5′ end contains four copies of the repetitive sequence AAACC*X* (*X* = C, G, A). The nonfunctional 55-bp 5′ end fragment contains only a single copy of the repetitive sequence. At the 3′ end, the functional fragment contained all four copies of the last repeat sequence TGACCC. The two nonfunctional fragments contained no 3′ repeat sequence and one had only the 22-bp inverted repeat sequence (pTG35 and pTG36). These results show that the inverted repeat alone is not sufficient for excision.

The results shown in Figure 4 also indicate that the small, inactive *dTag1* elements shown in Figure 3 do have all the *cis*-acting sequences needed for excision. Therefore, it is likely that a minimal spacing between the two end fragments is required for excision. To test this idea, three constructs were made that had increasing lengths of non-*Tag1* spacer DNA inserted between the 98- and 109-bp end fragments of *Tag1* (Figure 5). Transgenic plants containing these DNAs cloned into the 35S-GUS vector were crossed to the TS28 line. All

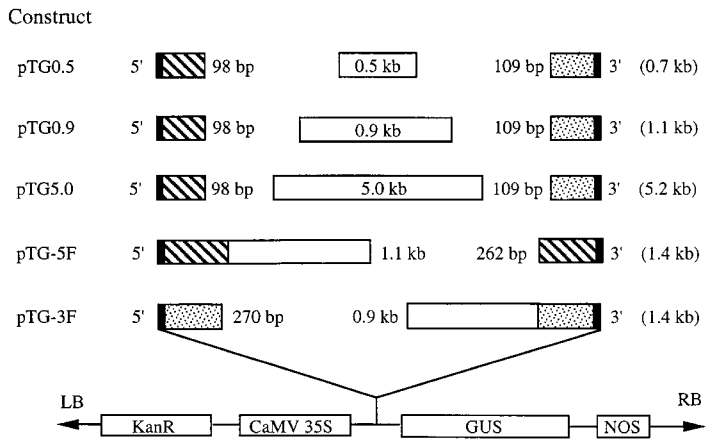

FIGURE 5.—Schematic diagram of minimal *dTag1* constructs containing spacer DNA. The solid, striped, and stippled boxes represent *Tag1* end sequences as detailed in Figure 1 legend. (▨) 5′ subterminal region; (▩) 3′ subterminal region. Open boxes in the top three diagrams indicate randomly selected Arabidopsis genomic DNA of lengths given in box. The bottom two diagrams indicate constructs that have duplicated *Tag1* end sequences. The open boxes represent *Tag1* sequences that are naturally present in that position in *Tag1*. The total length of each *dTag1* element is shown in parentheses. The excision activity for each construct as determined from F_1 progeny between plants containing the *dTag1* construct and line TS28 (containing transposase) is shown at right.

constructs showed excision activity (Figure 5). These results, including those from Figure 3, demonstrate that a minimal spacer of 238–325 bp separating the two end fragments is required for excision. A 5-kb spacer also supports excision (Figure 5). Consistent with results described above, the length of the spacer also affected the percentage of primary transformants showing diffusive, background GUS staining. The larger the spacer was, the lower the percentage of background staining plants (Table 1).

5′ and 3′ minimal sequences of *dTag1* are not interchangeable: Most plant transposable elements, including *Ac*, *Spm*, and *Mutator*, have the same or very similar sequences at their 5′ and 3′ ends; nevertheless, the 5′ and 3′ end fragments are not functionally interchangeable (COUPLAND *et al.* 1989; GIERL 1996). *Tag1* has two different repeated sequences at its 5′ and 3′ ends. We tested if the 5′ and 3′ end fragments of *Tag1* were interchangeable. Two constructs were made (Figure 5). The first (pTG-5F) contained duplicated 5′ ends with the 1.1-kb 5′ fragment fused to the 270-bp 5′ fragment. The second (pTG-3F) had duplicated 3′ ends using the 270-bp 3′ fragment and the 0.9-kb 3′ fragment. Six independent transformants for each construct were crossed to TS28 plants containing the *Tag1* transposase. GUS staining of F_1 plants showed no sectors. These results indicate that 5′ and 3′ sequences are not functionally interchangeable.

Somatic and germinal excision properties of *dTag1* elements: The distinctive behavior of *Tag1* excision from 35S-*Tag1*-GUS constructs during shoot development, which produced tiny somatic sectors and independent germinal revertants, indicates that excision is developmentally regulated in these lines (LIU and CRAWFORD 1998a). In contrast, the frequency of *Tag1* excision was quite variable, with germinal excision rates varying from 0 to 27% (LIU and CRAWFORD 1998a). We wished to determine if the *dTag1* derivatives also displayed developmentally regulated excision and a similar range of excision frequencies to that of *Tag1*. Plants containing each *dTag1* derivative were crossed to TS28 plants containing *Tag1* transposase. To examine somatic excision frequency, GUS sector number was determined on leaf 3 or 4 in three F_1 plants from each TS28 cross and then classified as high (>300 sectors per leaf), medium (50–300 sectors per leaf), and low (<50 sectors per leaf) as described previously (LIU and CRAWFORD 1998a). To determine germinal excision frequency, three F_1 plants showing somatic sectors were allowed to self, and then the percentage of whole blue staining seedlings were counted in the F_2 generation.

Our data (Tables 2 and 3) show that *dTag1* elements with the minimal end sequences and within a specific size range show very similar somatic excision frequencies compared with the autonomous *Tag1* element (LIU and CRAWFORD 1998a). The number of somatic sectors for the simple deletion derivatives is variable from line to line, but the range and average are typical for *Tag1* (Table 2). However, as described above, if the size of the *dTag1* element is 440 bp or less, no excision was observed (Figure 3). For the *dTag1* elements containing spacer DNA, typical ranges and averages of excision frequencies for the pTG0.5 and pTG0.9 elements (0.7 and 1.1 kb in size, respectively) were observed (Table 3). For the 5.2-kb element pTG5.0, however, the somatic excision frequency was much lower (Table 3). This reduction could be due to the size of the spacer or to an inhibitory sequence within the DNA. In conclusion, there appears to be no correlation between somatic excision frequency and the length or sequence of the *dTag1* elements for elements between 0.7 and 1.2 kb. If the autonomous 3.3-kb Tag1 element is included, the range of lengths that accommodate typical excision rates is from 0.7 to 3.3 kb.

When germinal excision of *dTag1* was examined, similar ranges and averages of excision rates were found for *dTag1* elements between 0.7 and 1.2 kb in length (0–16%; Tables 2 and 3), and these values matched those for the autonomous 3.3-kb *Tag1* element (LIU and CRAWFORD 1998a). The average rate for all the active *dTag1* elements tested (minus the pTG5.0 element) was 5.8%. This value is very close to the 5.3%

TABLE 3

Somatic and germinal excision frequencies of F_1 hybrids resulting from plants with 35S-*dTag1*-GUS lines with spacers crossed with 35S-*Tag1*-SPT plants

Line	KanR:KanS (selfed progeny)	F_1 hybrid seedling GUS(+)	F_1 hybrid seedling GUS(−)	Excision frequency in leaves	Germinal excision frequency (%)	Average
pTG0.5-4	25:1	5	8	M, H, M	2, 0, 3	1.6
pTG0.5-6	30:1	17	17	M, M, M	3, 5, 2	3.3
pTG0.5-8	3:1	20	5	H, H, H	10, 11, 12	11.0
pTG0.5-10	3:1	8	2	H, M, H	10, 4, 6	6.6
pTG0.5-21	2:1	12	21	H, M, H	2, 4, 2	2.6
pTG0.9-3	150:1	17	5	L, L, M	6, 4, 5	5.0
pTG0.9-4	40:1	15	13	L, M, M	12, 8, 3	5.6
pTG0.9-6	65:1	21	2	M, M, M	13, 2, 4	6.3
pTG0.9-8	170:1	21	8	M, M, M	9, 0, 0	3.0
pTG0.9-9	3:1	12	7	M, M, M	5, 4, 7	5.3
pTG0.9-11	3:1	13	17	L, M, M	8, 5, 4	5.6
pTG0.9-12	1:1	15	14	M, M, M	1, 3, 5	3.0
pTG5.0-1	3:1	1	24	L, L, L	0, 1, 0	0.3
pTG5.0-2	3:1	0	16	0, 0, 0	0, 1, 2	1.0
pTG5.0-3	3:1	16	26	L, L, L	0, 0, 1	0.3
pTG5.0-4	95:1	9	28	L, L, L	3, 2, 0	1.7
pTG5.0-5	165:1	2	19	0, 0, 0	1, 0, 1	0.7
pTG5.0-6	3:1	1	14	L, L, L	1, 2, 1	1.3

H, >300 sectors/leaf; M, 50–300 sectors/leaf; L, <50 sectors/leaf; ND, none detected.

average value determined for *Tag1*. The latter value is based on the germinal excision rates observed for the original 47 *Tag1* transgenic lines reported previously (Liu and Crawford 1998a) as well as an additional 170 *Tag1* lines constructed more recently (data not shown). As was the case for somatic excision, no correlation was observed between germinal excision frequency and the sequence or length of the *dTag1* element within the size range given above (Tables 2 and 3). The 5.2-kb element pTG5.0 and elements of length 440 bp or less showed no or much reduced germinal excision. We conclude that there is no obvious difference in germinal excision frequencies between the 3.3-kb *Tag1* and the 0.7- to 1.2-kb *dTag1* elements.

The timing of *dTag1* excision in all F_1 plants was examined next. For the simple deletion *dTag1* constructs pTG29, 30, 32, and 33 and for the *dTag1* elements with spacers pTG0.5, 0.9, and 5.0, excision timing was similar to that of *Tag1*. In roots and cotyledons, GUS sector size ranged from small to large as was found for *Tag1* (see Liu and Crawford 1998a). In leaves and siliques, only tiny excision sectors were found for *dTag1*. Occasionally, a few larger sectors were observed in sepals among the many inflorescences examined (one large sepal sector for every five inflorescences, data not shown). These results indicate that 98 bp at the 5′ end and 109 bp at the 3′ end are the only *cis*-acting sequences needed for normal regulation of *Tag1* excision during shoot development.

The 8-bp duplicated target sequence flanking *Tag1* does not affect excision frequency and timing: *Tag1* generates an 8-bp duplication when it inserts into the genome. All the *dTag1* constructs tested in this study were flanked by an 8-bp direct repeat of target sequence. To test if these flanking sequences have any impact on *dTag1* excision frequency and timing, a derivative of the pTG3 2-kb element (Figure 1) was made that lacked the 8-bp repeat. This construct was introduced into Columbia plants and four independent transgenic lines were crossed to TS28 line. The resultant F_1 plants showed excision timing and range of frequencies similar to those of plants containing pTG3 (data not shown). Thus, the target duplication has no apparent impact on *dTag1* excision similar to what was shown for *Ac* (Dooner *et al.* 1988).

Reinsertion of *dTag1* elements into the plant genome: Southern blot and sequence analyses were performed to determine if *dTag1* elements with minimal 5′ and 3′ *cis* sequence still retain their ability to reinsert into the genome after excision. Thirteen germinal revertants were selected from an F_2 population produced by crossing TS28 with plants containing the *dTag1* construct pTG-0.9. The 0.9-kb spacer sequence within this *dTag1* element was used as probe. The 0.9-kb spacer was a randomly selected fragment from the Arabidopsis genome. When the spacer DNA is hybridized with genomic DNA from the TS28 parent digested with *Xba*I, which does not cut the *dTag1* element but does cut once in the T-DNA, five bands appear (Figure 6, lane 2). The top band corresponds to the spacer DNA in its native

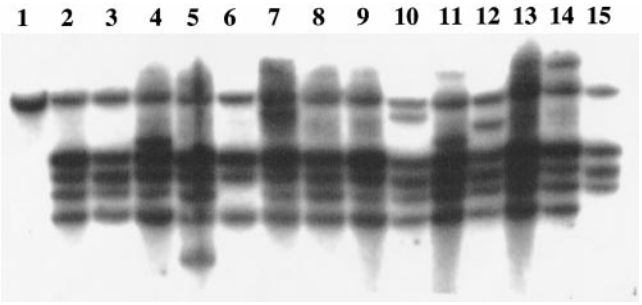

FIGURE 6.—Southern blot analysis of *dTag1* reinsertions into the Arabidopsis genome. Genomic DNA was prepared from (lane 1) untransformed plants of Columbia ecotype (single band in lane is hybridization of probe to endogenous sequence present in Arabidopsis genome); (lane 2) the parental transgenic line (pTG0.9-11 in Table 3) containing *dTag1* elements in the 35S-GUS construct; and (lanes 3–15) germinal revertants of pTG0.9-11 that inherited a *dTag1* excision event (new bands represent reinsertions of *dTag1*). The revertants were identified among a pool of F_2 progeny from a cross between line TS28 (containing *Tag1* transposase) and a line containing the *dTag1* element pTG0.9, which contains 98 bp of 5′ DNA and 109 bp of 3′ DNA of *Tag1* and a 0.9-kb spacer DNA in the middle. Genomic DNA was digested with *Xba*I (does not cut *Tag1*) and analyzed by Southern blot hybridization using the 0.9-kb spacer DNA present in the middle of the *dTag1* element as probe.

position in the Arabidopsis genome (this band is also present in Columbia plants not containing the *dTag1* construct, lane 1), and the bottom four correspond to the *dTag1* elements introduced into TS28. When DNA from the 13 germinal revertants was examined, 8 were found to contain a new band indicative of a reinsertion event (Figure 6, lanes 4, 5, 7, 10–12, 14, and 15; in lane 15 the third "band" is a doublet on the original) and 5 had no new bands (Figure 6, lanes 3, 6, 8, 9, and 13). Based on the unique position of the new bands, most of the reinsertions appear to be independent. Those lines that did not have a new visible band could still have undergone a reinsertion event that was hidden under one of the other bands. To confirm that *dTag1* was indeed reinserting into the genome, DNA flanking *dTag1* reinsertions was amplified by TAIL PCR and then analyzed as described in MATERIALS AND METHODS. DNA from 11 independent germinal revertants was cloned and sequenced. Nine of these 11 sequences had flanking DNAs that had matches to Arabidopsis genomic clones deposited in GenBank (Table 4); the other two gave sequences identical to the T-DNA vector pBI121. The identified genomic clones (usually bacterial artificial chromosome clones) have all been mapped on the Arabidopsis genome (Table 4) and were found to be scattered on all five chromosomes (Figure 7). These results indicate that a *dTag1* element with only 98 bp of 5′ DNA, 109 bp of 3′ DNA, and spacer DNA is capable of reinserting into the genome after its excision at about the same frequency observed for *Tag1*, which is approximately 75% as reported in LIU and CRAWFORD (1998a).

Evidence for a host-encoded DNA-binding activity

TABLE 4

Summary of pTG 0.9 reinsertion sites

Revertant no.	Clone	Chromosome no. (position[a])	Sequence[b]
R1	T24P13	I (10.1 Mb)	5′-**CATTG**CTTGTGCCTTTTTTTGTTCTTGTTCTGGGA AAAGGTTGATGAACTTTTTTTGCCTTAA-3′
R2	T31E10	II (15.6 Mb)	5′-**CATTG**TTTATCACTTATTTTTTTTTGGTGAGAACT ATGCCTTGGCTTTGGCTAGTGCTTTAAA-3′
R4	F2G14	V (7.4 Mb)	5′-**CATTG**GATAAAACCAAAAGTAAAAACGATATTCAT AAACTTGTTTAGAGACGTACCTTTCT-3′
R5	F16L2	III (17.5 Mb)	5′-**CATTG**GTAATATCCTAGATTTTAGCAAAAACGACT ACTATAGAGCTCCATCTAATCTTGAAA-3′
R8	T15K4[c]	I (14.5 Mb)	5′-**CATTG**GTTTTACACCTAAACCCTTCCACTATATAT ATAAACCCCACTTTTGTCTCTATATCTT-3′
R9	AP21	IV (22.3 Mb)	5′-**CCTTG**ATAAATATCTACCCTATACTATTGGCTAAT GTATGTTACATGGTGCAAGAAGCTTTA-3′
R11	F25I18	II (15.1 Mb)	5′-**CATTG**GTTTCAACGTATTTTTATCAGCGAAAATGA ATTTGGATAAGCTAGTTTGAACTTTG-3′
R12	F3I6	I (9.4 Mb)	5′-**CATTG**CTTAAAATACAAAACAATAAAATAACAGTG ATAATTTTTTTTTTTTGGCGAAGTCCTC-3′
R14	F24P17	III (2.7 Mb)	5′-**CATTG**TATTAGAGCTAAAAGTTTTCTTTTATAGAG ACAGTATCATTATTACAACACAAAAT-3′

[a] Clone position determined using TAIR map viewer and AGI maps located at www.arabidopsis.org/servlets/mapper.

[b] Nucleotides in plain text are genomic flanking sequences. Nucleotides in boldface text correspond to the last five nucleotides of Tag1 terminal inverted repeats according to sequencing results.

[c] The R8 insertion is located 80 bp 5′ of the start of the Arabinogalactan-protein (AGP5) mRNA sequence.

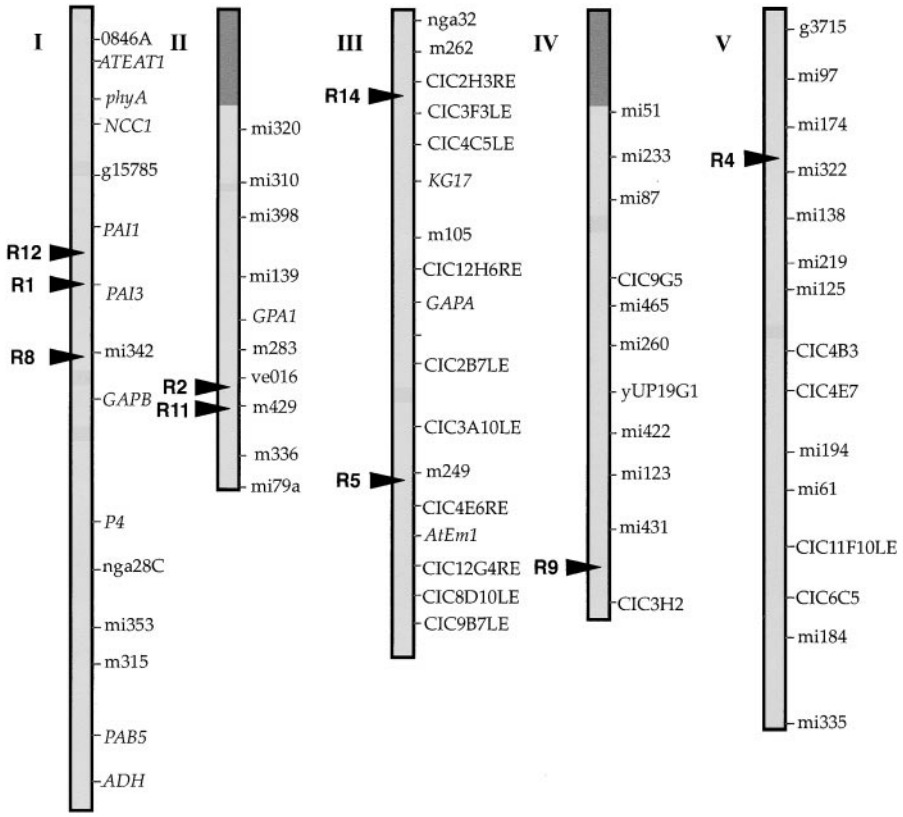

FIGURE 7.—Reinsertion sites for *dTag1* (0.9) on the Arabidopsis chromosomes. The positions of the individual reinsertions were determined as described in Table 4. AGI maps of the Arabidopsis chromosomes were obtained from http://arabidopsis.org/chromosomes/.

that interacts with the 3′ but not 5′ region of *Tag1*: Our above analysis shows that approximately 100 bp at both 5′ and 3′ ends of *Tag1* element is sufficient for *Tag1* transposition in a developmentally regulated manner. To begin examining the mechanism for the regulated excision, DNA-binding activities were assayed in protein extracts from Arabidopsis plants. Nuclear extracts were first prepared from leaves of untransformed Columbia plants, which contain no *Tag1* elements. DNA binding was assayed by gel retardation methods using radiolabeled fragments from each end of *Tag1*. When these nuclear extracts were incubated with radiolabeled 109-bp 3′ end fragment or 98-bp 5′ end fragment, a DNA-protein complex was observed for the 3′ fragment (Figure 8A, lane 2) but not with the 5′ fragment (Figure 8A, lanes 4 and 5). Competition experiments confirmed that this DNA-binding activity is specific (Figure 8B). Next, the DNA-binding activity in the nuclear extracts made from line TS28, which contains active *Tag1* elements, was tested. For the 5′ fragment no binding activity could be detected. For the 3′ fragment a DNA-protein complex with similar mobility to that from *Tag1*-minus nuclear extracts was observed, but its intensity was dramatically increased (Figure 8A, lane 3). Whether this stronger signal indicates another complex or a simple enhancement of the binding activity already existing in *Tag1*-free plants is unknown. These experiments provide evidence that a host-encoded factor(s) binds to the 3′ end of *Tag1*. Because both 5′ and 3′ DNA probes contain the same 22-bp terminal inverted repeat and 8-bp duplicated target sequence, these results suggest that the DNA-binding factor does not simply bind to these sequences but must at least include unique sequence present in the 3′ subterminal fragment.

DISCUSSION

The data from our deletion analysis show that end fragments of 98 bp at the 5′ end and 109 bp at the 3′ end are sufficient and required for excision and reinsertion of *dTag1* in Arabidopsis plants at rates typical for the autonomous *Tag1* element. These end fragments include the 22-bp inverted repeat and four copies of either the 5′ subterminal repeat (AAACC*X*, where *X* = C, A, G in direct or inverse orientation) or the 3′ subterminal repeat (TGACCC). Smaller end fragments of about 50 bp containing either one copy of the 5′ subterminal repeat or no copies of the 3′ subterminal repeat support no excision in Arabidopsis. To be active, the 98- and 109-bp end fragments must be separated by spacer DNA.

These findings shed light on the unusual repeat structure of *Tag1*. At the 3′ end, which contains three sets of unrelated subterminal repeats, the most 3′ repeat region (containing TGACCC) is sufficient for transposition. The other two 3′ repeat regions are dispensable for *dTag1* but may play a role for the intact element, such as serving as signals for processing the transposase

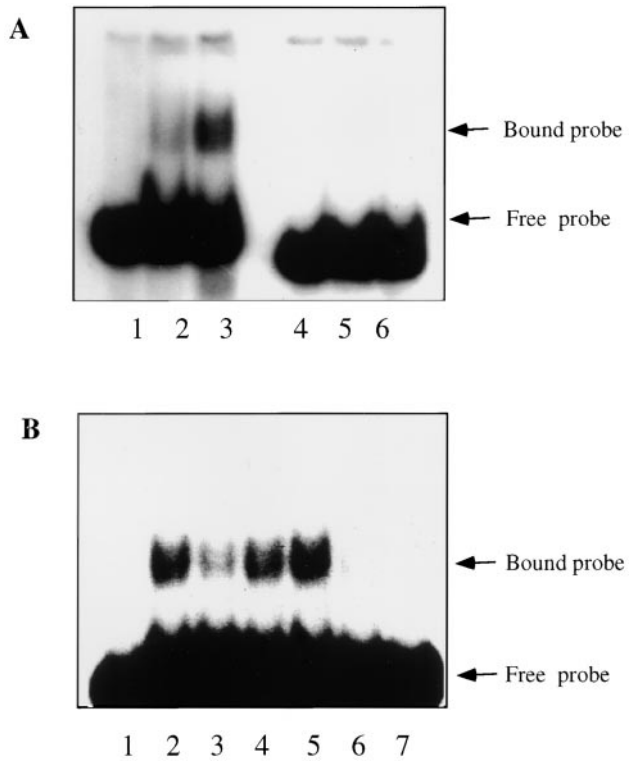

FIGURE 8.—Gel mobility-shift assays of DNA-binding activity to *Tag1* subterminal sequences. Crude nuclear protein extracts were prepared from plants containing or lacking active *Tag1* elements. Extracts were incubated with labeled DNA fragments from the 5′ or 3′ ends of *Tag1*. Complexes were resolved by gel electrophoresis. (A) Extracts from Arabidopsis plants (Columbia ecotype) with or without *Tag1* elements show binding to the 3′ end but not 5′ end fragment. Lanes 1–3, a 3′ *Tag1* end fragment [109 bp of *Tag1* (see right side of pTG30 in Figure 4) with 8-bp target sequence and 23-bp flanking sequence] was radiolabeled and used as target DNA for binding. Lanes 4–6, a 5′ *Tag1* end fragment [98 bp of *Tag1* DNA (see left side of pTG33 in Figure 4) with 8-bp target sequence and 12-bp flanking sequence] was radiolabeled and used as target DNA for binding. Lanes 1 and 4, no proteins added. Lanes 2 and 5, 5 μg nuclear proteins from untransformed Arabidopsis plants (Columbia). Lanes 3 and 6, 5 μg nuclear proteins from transgenic plants pTS28 containing active *Tag1* elements (Columbia). (B) Competition experiments showing the specificity of binding to the 3′ end fragment. Lanes 1–7, labeled DNA fragment was the same 109-kb 3′ end fragment described in A. Lane 1, no nuclear extract added. Lanes 2–7, 5 μg nuclear extract added. Lane 2, no unlabeled competitor DNA added. Lane 3, 1000× poly(dI-dC) (nonspecific competitor). Lane 4, 1000× salmon sperm DNA (nonspecific competitor). Lane 5, 1000× pBluescript plasmid DNA (nonspecific competitor). Lane 6, 25× unlabeled 3′ end fragment DNA (specific competitor). Lane 7, 100× unlabeled 3′ end fragment DNA (specific competitor).

mRNA. At the 5′ end a region with only 4 copies (out of 12) of the AAACCC*X* repeat is required for efficient excision and reinsertion. The 5′ and 3′ end fragments are not interchangeable as duplication of either end leads to an inactive element as is the case for *Ac* (COUPLAND *et al.* 1989) and *Spm* (GIERL 1996). Our functional *dTag1* elements displayed developmentally controlled timing of excision indicating that any *cis*-acting sequences involved in regulation of excision are embedded in the short ends of *Tag1*. This result is similar to what has been found for other elements where key methylation and regulatory sites of *En/Spm* reside in the first 0.55 kb of the element including the terminal 0.2-kb sequence (reviewed in FEDOROFF *et al.* 1995) and binding sites of the KP repressor in *P* elements overlap those of the transposase within the 31-bp inverted repeat and the 11-bp transpositional enhancer sequence (LEE *et al.* 1996, 1998).

Special emphasis is placed here on the 5′ and 3′ subterminal repeats of *Tag1*. On the basis of our *in vivo* data, we cannot state that they are the sequences to which the transposase binds, yet they are the most likely binding sites for the *Tag1* transposase based on what is known about other eukaryotic transposons. As described above, *Ac* and *En/Spm* rely on subterminal repeats for transposase binding (GIERL *et al.* 1988; KUNZE and STARLINGER 1989; TRENTMANN *et al.* 1993; BECKER and KUNZE 1997). We have found for *Tag1*, which is related to *Ac*, that the parts of the subterminal repeat regions are critical for transposition. These regions contain repeats that are not identical, which is unusual. These repeats do, however, share the tetranucleotide sequence ACCC. Studies of *Ac* have shown that a repeated trinucleotide sequence can serve as a binding site (BECKER and KUNZE 1997); therefore, it is certainly possible that the *Tag1* transposase may recognize the ACCC sequence. Alternatively, there may be two distinct recognition sequences for the *Tag1* transposase: AAACCC at the 5′ end and TGACCC at the 3′ end. Further experiments testing the DNA-binding properties of the *Tag1* transposase *in vitro* are required to determine the exact binding sequences of the *Tag1* transposon.

We have shown that *dTag1* elements have a strict minimal size requirement for efficient excision. At least 238–325 bp of DNA must be present between the 98-bp 5′ end and 109-bp 3′ end fragments of *Tag1* for excision to occur. A *dTag1* element of 532 bp showed normal excision rates while an element of 444 bp showed no excision even though it contained all the required *cis*-acting sequences. In contrast, a slightly larger construct, pTG0.5, which had the 98-bp 5′ end and 109-bp 3′ end fragments separated by 0.5-kb spacer DNA, displayed excision frequencies typical of the 3.3-kb element. If *Tag1* excision occurs upon the formation of a synaptic complex held together by oligomers of transposase proteins, our results suggest that *Tag1* has critical length requirements for efficient formation or resolution of such complexes. Perhaps a minimal length is required for bending DNA to form the synaptic complex. We are not aware of any reports describing minimal length requirements for eukaryotic transposons; however, DNA bending for formation of transposase-

DNA complexes has been shown, for example, for *Tc3* (van Pouderoyne *et al.* 1997).

We have also found that there is an activity in crude nuclear extracts from Arabidopsis plants devoid of *Tag1* and *Tag1* mRNA. Binding activity is higher in plants that have active *Tag1* elements. This higher activity could be due to (1) an increase in the plant-encoded activity that was induced by transformation or introduction of *Tag1* or (2) the presence of *Tag1* transposase-DNA complexes, which happen to migrate to the same position. Interestingly, this activity is specific for the 3′ 109-bp fragment and does not show any binding to the 5′ 98-bp fragment, which indicates that it is not binding exclusively to the inverted terminal repeat. In comparison, plant-encoded factors in both maize and tobacco have been found that bind to the subterminal repeat regions of *Ac* but not the transposase binding sites themselves (Becker and Kunze 1996; Levy *et al.* 1996). Two other host-encoded proteins have been described that bind to the ends of transposons but only within the inverted repeats: a Drosophila-encoded inverted repeat binding protein that binds to *P* elements DNA (Rio and Rubin 1988) and a maize-encoded protein that binds the ends of *Mutator* (Zhao and Sundaresan 1991). The function of these proteins has yet to be determined, but they are candidate regulatory factors that might control excision.

This work was supported by a grant from the National Science Foundation (MCB-9808215).

LITERATURE CITED

Ade, J., and F. J. Belzile, 1999 *Hairpin* elements, the first family of foldback transposons (FTs) in *Arabidopsis thaliana*. Plant J. **19:** 591–597.

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.

Becker, H.-A., and R. Kunze, 1996 Binding sites for maize nuclear proteins in the subterminal regions of transposable element *Activator*. Mol. Gen. Genet. **251:** 428–435.

Becker, H.-A., and R. Kunze, 1997 Maize *Activator* transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. Mol. Gen. Genet. **254:** 219–230.

Benito, M.-I., and V. Walbot, 1997 Characterization of the maize *Mutator* transposable element MURA transposase as a DNA-binding protein. Mol. Cell. Biol. **17:** 5165–5175.

Bennetzen, J. L., 1996 The *Mutator* transposable element system of maize, pp. 195–230 in *Transposable Elements*, edited by H. Saedler and A. Gierl. Springer-Verlag, Berlin.

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.

Casacuberta, E., J. M. Casacuberta, P. Puigdomenech and A. Monfort, 1998 Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. Plant J. **16:** 79–85.

Colloms, S. D., H. G. A. M. Van Luenen and R. H. A. Plasterk, 1994 DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase. Nucleic Acids Res. **22:** 5548–5554.

Copenhaver, G. P., K. Nickel, T. Kuromori, M. I. Benito, S. Kaul *et al.*, 1999 Genetic definition and sequence analysis of Arabidopsis centromeres. Science **286:** 2468–2474.

Coupland, G., C. Plum, S. Chatterjee, A. Post and P. Starlinger, 1989 Sequences near the termini are required for transposition of the maize transposon *Ac* in transgenic tobacco plants. Proc. Natl. Acad. Sci. USA **86:** 9385–9388.

Dean, C., C. Sjodin, T. Page, J. Jones and C. Lister, 1992 Behaviour of the maize transposable element *Ac* in *Arabidopsis thaliana*. Plant J. **2:** 69–81.

Dooner, H. K., J. English and E. J. Ralston, 1988 The frequency of transposition of the maize element *Activator* is not affected by an adjacent deletion. Mol. Gen. Genet. **211:** 485–491.

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 Genet. Cooperation Newsl. **69:** 39–41.

Fedoroff, N., M. Schlappi and R. Raina, 1995 Epigenetic regulation of the maize *Spm* transposon. Bioessays **17:** 291–297.

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.

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.

Gierl, A., S. Lutticke and H. Saedler, 1988 *TnpA* product encoded by the transposable element En-1 of *Zea mays* is a DNA-binding protein. EMBO J. **7:** 4045–4053.

Henk, A. D., R. F. Warren and R. W. Innes, 1999 A new *Ac*-like transposon of Arabidopsis is associated with a deletion of the *RPS5* disease resistance gene. Genetics **151:** 1581–1589.

Jefferson, R. A., 1989 The GUS reporter gene system. Nature **342:** 837–838.

Jones, J., F. Carland, P. Maliga and H. Dooner, 1989 Visual detection of transposition of the maize element *Activator* (*Ac*) in tobacco seedlings. Science **244:** 204–207.

Kaufman, P. D., R. F. Doll and D. C. Rio, 1989 Drosophila P element transposase recognizes internal P element DNA sequences. Cell **59:** 359–371.

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.

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

Lee, C. C., Y. M. Mul and D. C. Rio, 1996 The *Drosophila* P-element KP repressor protein dimerizes and interacts with multiple sites on P-element DNA. Mol. Cell. Biol. **16:** 5616–5622.

Lee, C. C., E. L. Beall and D. C. Rio, 1998 DNA binding by the KP repressor protein inhibits P-element transposase activity *in vitro*. EMBO J. **17:** 4166–4174.

Levy, A. A., M. Fridlender, U. Hanania, E. Rubin and Y. Sitrit, 1996 Binding of *Nicotiana* nuclear proteins to the subterminal regions of the *Ac* transposable element. Mol. Gen. Genet. **251:** 436–441.

Lin, X. Y., S. S. Kaul, S. Rounsley, T. P. Shea, M. I. Benito *et al.*, 1999 Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402:** 761–768.

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

Liu, D., and N. M. Crawford, 1998b 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. Genetics **149:** 693–701.

Liu, D., S. Zhang, C. Fauquet and N. M. Crawford, 1999 The *Arabidopsis* transposon *Tag1* is active in rice, undergoing germinal transposition and restricted, late somatic excision. Mol. Gen. Genet. **262:** 413–420.

Mayer, K., C. Schuller, R. Wambutt, G. Murphy, G. Volckaert *et al.*, 1999 Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature **402:** 769–777.

Raghothama, K. G., D. Liu, D. E. Nelson, P. M. Hasegawa and R. A. Bressan, 1993 Analysis of an osmotically regulated patho-

genesis-related osmotin gene promoter. Plant Mol. Biol. **23:** 1117–1128.

Rio, D. C., and G. M. Rubin, 1988 Identification and purification of a Drosophila protein that binds to the terminal 31-base-pair inverted repeats of the P transposable element. Proc. Natl. Acad. Sci. USA **85:** 8929–8933.

Saedler, H., and A. Gierl (Editors), 1996 *Transposable Elements.* Springer-Verlag, Berlin.

Trentmann, S. M., H. Saedler and A. Gierl, 1993 The transposable element En/Spm-encoded TNPA protein contains a DNA binding and a dimerization domain. Mol. Gen. Genet. **238:** 201–208.

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.

Tsugeki, R., E. V. Kochieva and N. V. Fedoroff, 1996 A transposon insertion in the *Arabidopsis SSR16* gene causes an embryo-defective lethal mutation. Plant J. **10:** 479–489.

Valvekens, D., M. Van Montagu and M. Van Lijsebettens, 1988 *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proc. Natl. Acad. Sci. USA **85:** 5536–5540.

van Pouderoyne, G., R. F. Ketting, A. Perrakis, R. H. A. Plasterk and T. K. Sixma, 1997 Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. EMBO J. **16:** 6044–6054.

Vos, J. C., and R. H. Plasterk, 1994 Tc1 transpoase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. EMBO J. **79:** 6125–6132.

Voytas, D. F., and F. M. Ausubel, 1988 A copia-like transposable element family in *Arabidopsis thaliana.* Nature **336:** 242–244.

Voytas, D. F., A. Konieczny, M. P. Cummings and F. M. Ausubel, 1990 The structure, distribution and evolution of the Ta1 retrotransposable element family of *Arabidopsis thaliana.* Genetics **126:** 713–721.

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.

Wright, D. A., and D. F. Voytas, 1998 Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana Ty3/gypsy* retrotransposons that encode envelope-like proteins. Genetics **149:** 703–715.

Zhao, Z.-Y., and V. Sundaresan, 1991 Binding sites for maize nuclear proteins in the terminal inverted repeats of the *Mu1* transposable element. Mol. Gen. Genet. **229:** 17–26.

Communicating editor: V. L. Chandler