



## A collection of sequenced and mapped *Ds* transposon insertion sites in *Arabidopsis thaliana*

Surabhi Raina, Ramamurthy Mahalingam, Fuqiang Chen<sup>1</sup> and Nina Fedoroff\*

Life Sciences Consortium and Biotechnology Institute, Pennsylvania State University, University Park, PA 16802, USA (\*author for correspondence; e-mail [nvf1@email.psu.edu](mailto:nvf1@email.psu.edu)); <sup>1</sup>Present address: Sigma-Aldrich Chemical Company, St. Louis, USA

Received 19 June 2001; accepted in revised form 3 January 2002

**Key words:** Activator (*Ac*), *Arabidopsis thaliana*, Dissociation (*Ds*), mapping, searchable database, TAIL PCR, transposon insertion lines

### Abstract

Insertional mutagenesis is a powerful tool for generating knockout mutations that facilitate associating biological functions with as yet uncharacterized open reading frames (ORFs) identified by genomic sequencing or represented in EST databases. We have generated a collection of *Dissociation* (*Ds*) transposon lines with insertions on all 5 *Arabidopsis* chromosomes. Here we report the insertion sites in 260 independent single-transposon lines, derived from four different *Ds* donor sites. We amplified and determined the genomic sequence flanking each transposon, then mapped its insertion site by identity of the flanking sequences to the corresponding sequence in the *Arabidopsis* genome database. This constitutes the largest collection of sequence-mapped *Ds* insertion sites unbiased by selection against the donor site. Insertion site clusters have been identified around three of the four donor sites on chromosomes 1 and 5, as well as near the nucleolus organizers on chromosomes 2 and 4. The distribution of insertions between ORFs and intergenic sequences is roughly proportional to the ratio of genic to intergenic sequence. Within ORFs, insertions cluster near the translational start codon, although we have not detected insertion site selectivity at the nucleotide sequence level. A searchable database of insertion site sequences for the 260 transposon insertion sites is available at <http://sgio2.biotech.psu.edu/sr>. This and other collections of *Arabidopsis* lines with sequence-identified transposon insertion sites are a valuable genetic resource for functional genomics studies because the transposon location is precisely known, the transposon can be remobilized to generate revertants, and the *Ds* insertion can be used to initiate further local mutagenesis.

**Abbreviations:** *Ac*, Activator; *ALS*, acetolactate synthase; *Ds*, Dissociation; EDS, empty donor site; *NPTII*, neomycin phosphotransferase; ORF, open reading frame; TAIL PCR, thermal asymmetric interlaced PCR; *tms2*, indoleacetamide hydrolase

### Introduction

With the completion of the *Arabidopsis thaliana* genome sequence, attention is turning increasingly to the challenges of identifying genes and determining their functions. Insertional mutagenesis with transposons and *Agrobacterium* T-DNAs has become the most widely used approach in reverse genetics. Several recent reports describe collections of T-DNA and transposon insertion lines (Mathur *et al.*, 1998; Wis-

man *et al.*, 1998; Krysan *et al.*, 1999; Parinov *et al.*, 1999; Speulman *et al.*, 1999; Tissier *et al.*, 1999). Each of the insertional systems in use today has both advantages and disadvantages. For example, T-DNA insertion lines are readily generated in large numbers (Bechtold and Pelletier, 1998; Krysan *et al.*, 1999). However, it is generally acknowledged that T-DNA insertional mutagenesis often results in the production of mutations that are not associated with T-DNA insertions. There is also a growing awareness that ma-

jor, sometimes multiple, chromosomal rearrangements can accompany integration events (Van Lijsebettens *et al.*, 1986; Gheysen *et al.*, 1991; Mayerhofer *et al.*, 1991; Castle *et al.*, 1993; Ohba *et al.*, 1995; Takano *et al.*, 1997; Nacry *et al.*, 1998; Fladung, 1999; Kumar and Fladung, 2000).

Multi-element transposon tagging systems based on the maize *Suppressor-mutator (Spm)* transposon, also called *Enhancer*, have recently been described for *Arabidopsis* (Bensen *et al.*, 1995; Das and Martienssen, 1995; Chuck *et al.*, 1998; Hu *et al.*, 1998; Wisman *et al.*, 1998; Speulman *et al.*, 1999). These offer the advantage of high-frequency transposition, with several to many new insertions in each progeny plant. However, the downside, as with plants containing multiple T-DNA insertions, is that additional genetic analysis is required to correlate a phenotype with the insertion causing it. Several groups have developed more elaborate insertional mutagenesis systems for *Arabidopsis* based on both the *Activator (Ac)* and *Spm* transposons of maize (Bancroft *et al.*, 1992; Fedoroff and Smith, 1993; Osborne *et al.*, 1995; Sundaresan *et al.*, 1995; Smith *et al.*, 1996; Long *et al.*, 1997; Tissier *et al.*, 1999). Transposant lines obtained using these systems generally contain a single transposed element (Parinov *et al.*, 1996; Smith *et al.*, 1996; Tissier *et al.*, 1999).

There are several strategies for recovering the unknown sequences flanking insertions of known sequences, such as T-DNAs and transposons (Hui *et al.*, 1998). Among them, thermal asymmetric interlaced PCR (TAIL PCR) has the advantage that it minimizes DNA manipulations and is therefore readily adaptable to multiplexing (Liu *et al.*, 1995; Liu and Whittier, 1995). We have used it for the isolation of flanking sequences from individual transposon insertion lines and it is an increasingly popular method for characterizing both T-DNA and transposon insertion lines (Liu *et al.*, 1995; Tsugeki *et al.*, 1996; Parinov *et al.*, 1999; Tissier *et al.*, 1999; Tsugeki and Fedoroff, 1999). Amplified flanking TAIL-PCR fragments can be directly sequenced (Liu *et al.*, 1993) and the sequence used to search genomic and EST databases for identities and similarities, in some cases leading to the immediate identification of the gene containing the insertion (Tsugeki *et al.*, 1996). With the completion of the *Arabidopsis* genome sequence, essentially all insertion site sequences can be linked to an exact chromosomal position, enhancing the value of insertion site sequencing. Databases of transposon insertion site sequences

are beginning to be established (Parinov *et al.*, 1999; Tissier *et al.*, 1999).

We have generated many independent *Arabidopsis Ds* insertion lines using a previously described *Ac*-based tagging system (Fedoroff and Smith, 1993; Smith *et al.*, 1996). In the present study we used TAIL PCR to amplify and determine the sequences flanking the *Ds* insertion sites. We were able to identify the precise chromosomal location of the insertion site for 260 lines by matching the flanking sequence with that of a mapped genomic clone or EST sequence. These *Ds* transposons can be remobilized to generate revertants, as well as insertions into the same or nearby genes of interest.

## Materials and methods

### *Plant material*

Transgenic plants of *Arabidopsis thaliana* ecotype Nossen (No-0) were used in all experiments. The *Ds* transposon was mobilized by genetic crosses between plants that were homozygous for a single *Ds*-T-DNA and plants with a stabilized *Ac* as a source of the transposase as previously described (Fedoroff and Smith, 1993; Smith *et al.*, 1996). The *Ac*-T-DNA carries an agrobacterial indole acetamide hydrolase gene (*tms2*) as a negative selectable marker (Fedoroff and Smith, 1993). For simplicity, the *Ds*-T-DNA insertion sites have been redesignated from those previously published (Smith *et al.*, 1996) as follows: *Ds* 390-1 is donor site 1 (D1), *Ds* 389-14 is donor site 2 (D2), *Ds* 392-13 is donor site 3 (D3) and *Ds* 393-9 is donor site 4 (D4). It should be noted that these and other *Ds*-T-DNA lines are available from the *Arabidopsis* Biological Resource Center (<http://aims.cps.msu.edu/aims>).

In most of the crosses, the *Ds*-T-DNA plant was the pollen parent and *Ac*-T-DNA plant was the egg parent. The success of the cross was determined by selecting F<sub>1</sub> progeny on MS medium containing hygromycin, resistance to which is carried on the *Ds* transposon (*aph4* gene). In reciprocal crosses in which the *Ds*-T-DNA plant was the female parent, F<sub>1</sub> plants were selfed to generate F<sub>2</sub> progeny. To identify plants in which transposition had occurred, but which did not carry the transposase gene, the F<sub>2</sub> seeds were surface-sterilized and germinated on solid MS medium (Sigma, St. Louis, MO) supplemented with hygromycin (20 mg/l; Gibco-BRL), chlorosulfuron (10 mg/l; Chem Service, West Chester, PA) and

naphthalene acetamide (NAM, 10  $\mu$ M; Sigma, MO), as described previously (Fedoroff and Smith, 1993). About 2 weeks after germination, seedlings resistant to all three compounds were transplanted to Metro-Mix 200 (Geiger, Harleysville, PA) and placed in growth chambers at 22 °C with a 16 h light/8 h dark photoperiod.

#### DNA isolation

DNA was isolated from young rosette leaves by a modification of the hexadecyltrimethylammonium bromide (CTAB) method (Stewart and Via, 1993). Frozen leaves (50–100 mg) were ground to a fine powder with a plastic pestle in liquid nitrogen in 1.5 ml microcentrifuge tube, resuspended in 750  $\mu$ l warm (70 °C) CTAB buffer (2% w/v CTAB, 1.42 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0, 2% w/v polyvinylpyrrolidone (PVP-40; Sigma), 5 mM ascorbic acid, 4.0 mM diethyldithiocarbamic acid (DIECA; Sigma)) containing 75.5 mM 2-mercaptoethanol and incubated at 70 °C for 30 min. Samples were extracted with an equal volume of chloroform/isoamyl alcohol (24:1, v/v) and centrifuged for 5 min at 15 000 rpm at room temperature. The upper aqueous phase was transferred to a new tube, and DNA was precipitated with 0.7 volumes of isopropanol at room temperature for 5 min. The tubes were centrifuged for 25 min at 15 000 rpm. The DNA pellet was washed with 100  $\mu$ l of 70% ethanol, dried and resuspended in 30  $\mu$ l of (10–20 ng/ $\mu$ l) Tris-EDTA (10 mM Tris, 1 mM EDTA) buffer. About 10–20 ng of DNA was used for TAIL-PCR analysis.

#### Eliminating lines lacking transposed elements

Plants heterozygous for an empty donor site (EDS) and a non-rearranged T-DNA and plants that have an EDS and a transposed element are both resistant to both hygromycin and chlorsulfuron. Plants that have an EDS and a non-rearranged T-DNA were identified by a preliminary screen and discarded. We tested for the ability to amplify the EDS using primers A (5'-GATCGAGCGTATCCTTCTGCTGCG-3') and B (5'-CAGCTATCTGTCACTTCATC-3'), which are homologous to the ALS gene and 35S promoter, respectively, and the ability to amplify an ALS-*Ds* junction fragment characteristic of the original *Ds*-T-DNA using primer A and a primer homologous to the 5' end of the *Ds* transposon (primer 5'-3: 5'-TACCTCGGGTTCGAAACGAT-3') (Figure 1). The PCR reaction contained the three primers (0.1  $\mu$ M

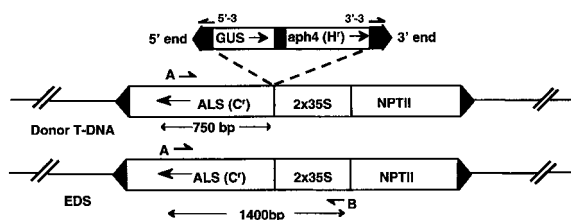


Figure 1. A schematic representation of the *Ds*-T-DNA. The *Ds* element is inserted in the leader region of a chlorsulfuron-resistant acetolactate synthase gene (*ALS*). The transposon contains both a promoterless  $\beta$ -glucuronidase (*GUS*) gene and a bacterial *aph4* gene expressed from a CaMV 19S promoter, which confers resistance to hygromycin and serves as a selection marker for the presence of the *Ds* transposon. The *GUS* gene may be activated upon insertion of the transposon in or near a promoter in the correct orientation. The location and orientation of the *Ds* 5'- and 3'-specific primers are shown by half arrows.

each), 10–20 ng of genomic DNA, 0.2 mM dNTPs, 1 $\times$  Red *Taq* PCR buffer and one unit of Red *Taq* polymerase (Sigma). Amplifications were performed at 94 °C for 3 min, 94 °C for 30 s, 58 °C for 1 min, 72 °C for 2 min for 30 cycles, followed by final extension for 7 min at 72 °C. The products were analyzed on a 1.5% agarose gel.

#### TAIL PCR amplification

The transposon-specific primers and the arbitrary degenerate primers were obtained from Integrated DNA Technologies (Coralville, IW). The two sets of three nested primers homologous to the 5' end of *Ds* (Ds5'-1, Ds5'-2 and Ds5'-3) and the 3' end of *Ds* (Ds3'-1, Ds3'-2 and Ds3'-3) respectively, as well as the 8 arbitrary degenerate primers (AD1–AD8), have been described previously (Liu *et al.*, 1995; Liu and Whittier, 1995; Tsugeki *et al.*, 1996). The PCR reactions were carried out as described previously (Tsugeki *et al.*, 1996) with the following modifications: 0.5 unit of RedTaq (Sigma) was used in the primary, secondary and tertiary PCR reactions and the reaction volume was 10  $\mu$ l. The secondary and tertiary PCR reactions were analyzed side by side on 1.5% agarose gels.

#### Band stab amplification of TAIL PCR products

Band stab amplification of TAIL PCR products was carried out as described by Wilton *et al.* (1997). The secondary PCR fragments on the gel that were slightly larger than the corresponding tertiary PCR fragments were 'stabbed' with sterile pipette tips. The tip was gently wiped and placed in the tube containing the PCR reaction mixture on ice. The tube contents were

mixed well using the same tip and subjected to the tertiary TAIL PCR procedure as described previously (Tsugeki *et al.*, 1996). The amplified DNA was purified using the QIAquick PCR purification kit (Qiagen, Chatsworth, CA) and eluted in 30  $\mu$ l of water. About 80 ng of the purified PCR-amplified DNA was used for cycle sequencing with the Ds5'-3 or Ds3'-3 primer under the conditions specified by manufacturer (Applied Biosystems, Foster City, CA). The sequencing reaction products were purified with Centriflex gel filtration cartridges (EDGE Biosystems, Gaithersburg, MD) and resolved on an ABI 373A sequencer (Nucleic Acid Facility, Pennsylvania State University).

#### *Sequence analysis and physical mapping*

Sequences were edited to correct nucleotide ambiguities, then used to query the GenBank nucleotide sequence database by using BLASTN and BLASTX programs (Altschul *et al.*, 1990, 1997). *Arabidopsis* genomic clones matching the sequences flanking the insertion site were identified. Anchored RFLP markers closest to the matching sequence were used to determine the precise chromosomal locations of the insertion sites.

## **Results**

#### *Generation of Arabidopsis lines containing transposed Ds elements*

The structure of the *Ds*-T-DNA used in the present study is shown in Figure 1 (Fedoroff and Smith, 1993). *Arabidopsis* plants containing a transposed *Ds* element were produced by crossing plants containing a *Ds*-T-DNA insertion at one of four chromosomal sites, designated D1–D4, and plants carrying an *Ac*-T-DNA as described previously (Fedoroff and Smith, 1993; Smith *et al.*, 1996). The donor site D1 is located on chromosome 1 at 9.31 cM, and D2 and D3 are located close together on chromosome 5, at 111.89 and 112.91 cM, respectively. To determine the map position of the *Ds*-T-DNA precisely, sequences flanking the *Ds*-T-DNAs were amplified by TAIL PCR, sequenced and mapped. Donor sites D2 and D3 mapped to BAC clones MXC20 and MGO3, respectively, on chromosome 5, and D1 and D4 mapped to BAC clones T28P6 and F23A5, respectively, on chromosome 1. While close together, the D2 and D3 donor sites are 2.8 Mb apart. The chromosomal location of the four donor sites is shown in Figure 2.

Because we screened self-fertilized progeny for transposed elements and transposition occurs throughout the plant's developmental cycle, some plants that exhibit resistance to chlorsulfuron and hygromycin do not have a reinserted element, but are instead heterozygous for an empty donor site and a non-rearranged donor T-DNA. This situation arises because developmentally late excision events are transmitted through only one of the two gametes, while the other gamete has a T-DNA from which the *Ds* has not transposed. While a fraction of such plants should have a reinserted element elsewhere in the genome, in practice we have found that most do not (data not shown). We therefore routinely carried out a preliminary screen for the presence of both an untransposed *Ds*-T-DNA and a EDS. About 40% (319 of 771 lines screened) of the hygromycin- and chlorsulfuron-resistant lines contained a non-rearranged *Ds*-T-DNA and an EDS.

#### *TAIL PCR amplification and sequencing of Ds insertion sites*

Sequences flanking the *Ds* transposons were amplified from total genomic DNA by thermal asymmetric interlaced (TAIL) PCR as described in Materials and methods (Liu *et al.*, 1995; Liu and Whittier, 1995; Tsugeki *et al.*, 1996), using arbitrary degenerate (AD) primers previously designed for *Arabidopsis* (Liu *et al.*, 1995; Liu and Whittier, 1995). Initially we used all eight AD primers (Liu *et al.*, 1995). However, because we had a more than 70% success rate in TAIL-PCR amplification with just primers 7 and 8, we used these primers routinely, repeating the amplification with additional primers only when these did not yield *Ds*-flanking amplification products.

Flanking sequences were successfully amplified from 83% of the hygromycin- and chlorsulfuron-resistant lines (375 of 452 lines tested) after elimination of those that contained an EDS and an untransposed element (see Materials and methods). The fidelity of TAIL PCR amplification was judged by the appearance on gels of secondary and tertiary TAIL PCR amplification products differing in size by the 25–30 nucleotides expected from the location of the nested *Ds*-specific primers (Liu and Whittier, 1995; Tsugeki *et al.*, 1996). The sizes of the secondary TAIL PCR products ranged from 0.1–2.0 kb in length, the majority being in the 0.2–0.6 kb range. Amplified flanking sequences were re-amplified from 'band stabs' as described in Materials and methods and directly sequenced using the 5' *Ds*-specific primer 2

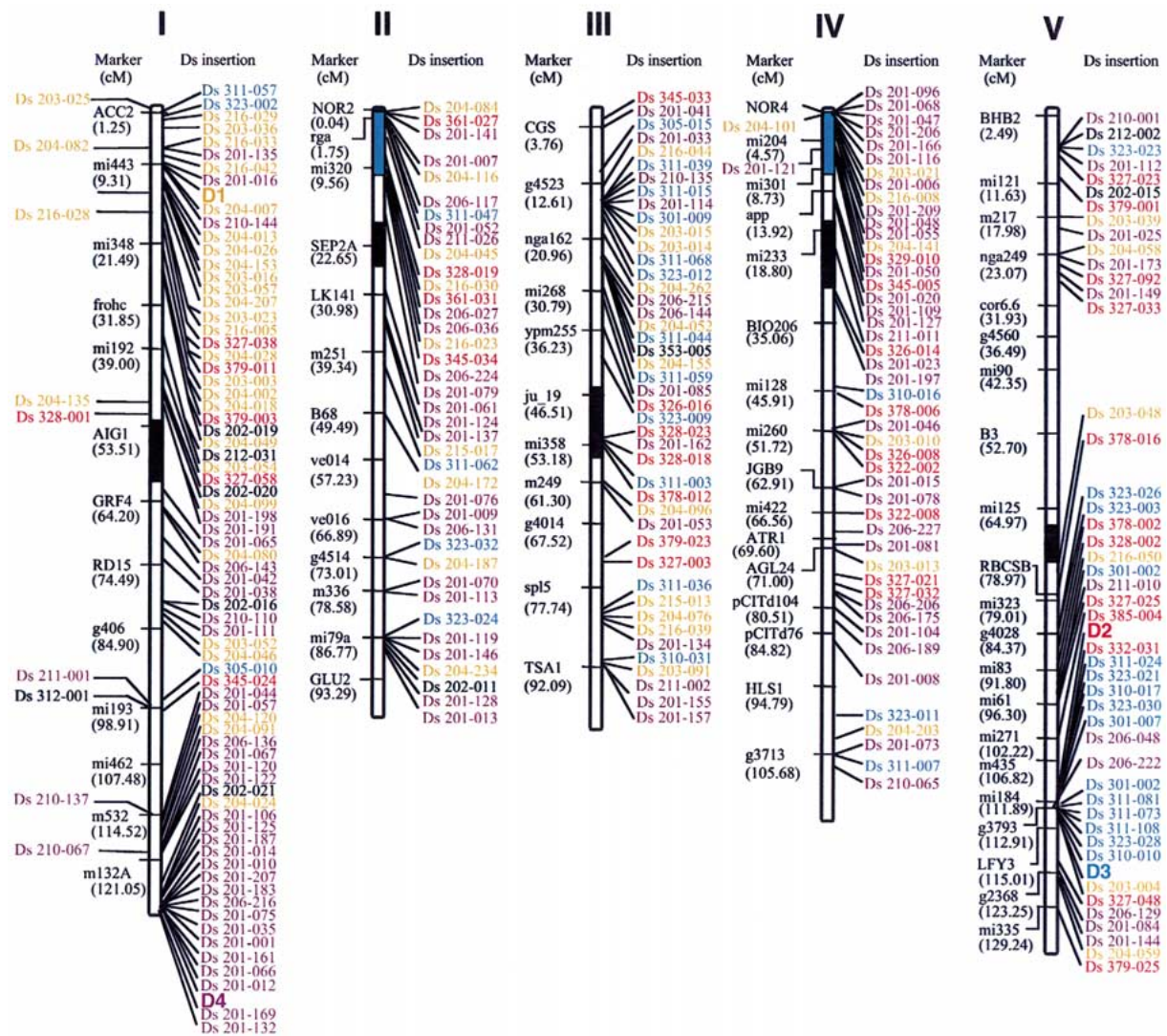


Figure 2. The chromosomal locations of *Ds* insertion sites. Insertion sites are represented in the same color as the donor site from which each originated. Transposons originating from an unmapped donor locus are shown in black.

and/or the 3' *Ds*-specific primer 3 (Figure 1). In cases in which the band stab amplification was unsuccessful (ca. 14%), the size of the fragment was often more than 1 kb, suggesting that the amount of DNA obtained by stabbing was insufficient for amplification. One or more of the following three tests was used to confirm the authenticity of the TAIL PCR amplified fragments: (1) fragments flanking both 5' and 3' ends of the *Ds* were amplified and sequenced and the 8 bp duplication caused by insertion of *Ds* element was found; (2) several TAIL PCR products amplified from a single end were sequenced and found to be identical; and (3) primers were designed from the sequence of a

single amplified fragment and used together with the appropriate *Ds* primer to amplify a precise junction fragment. Sequences that did not meet one or more of these criteria were considered unconfirmed and have not been included in this paper.

#### Genomic location of insertion sites

*Ds* flanking sequences in 260 lines showed extensive homology (greater than 95%) or identity to either a mapped and sequenced *Arabidopsis* genomic clone (BAC, YAC, TAC, and P1) or an EST as identified using the BLASTN and BLASTX algorithms (Altschul *et al.*, 1990, 1997). The minor differences in the

sequences are likely to be attributable to either our sequencing errors, errors in the sequence deposited in the database, or polymorphisms between ecotypes. In less than 2% of the lines flanking sequences showed homology with donor T-DNA.

There was sufficient information about the location of the genomic clone to determine the precise chromosomal location of the insertion site in 259 of the 260 lines with good flanking sequence homology. In one line, the insertion is in a BAC that has been sequenced, but whose chromosomal location has not yet been determined. The chromosomal locations for all of the genomic clones with *Ds* insertions was concomitantly determined by searching the *Arabidopsis* Genome Initiative (AGI) sequencing database (<http://www.arabidopsis.org/agi.html>) and the RI linkage map database (Lister and Dean, 1993; <http://www.arabidopsis.org/cgi-bin/maps/RIintromap>). The positions of the mapped insertion sites are shown in Figure 2.

#### *The distribution of reinsertion sites around ORFs*

A compilation of the genomic clones and precise locations within them of *Ds* insertion sites for all 260 lines in which the flanking sequence homology was greater than 95% is presented in Table 1, [http://sgio2.biotec.psu.edu/sr/pmb\\_table.htm](http://sgio2.biotec.psu.edu/sr/pmb_table.htm). The table also lists the location of the insertion site with respect to the nearest adjacent ORF (or predicted ORF). Insertions were identified in all major categories of genes, including genes encoding transcription factors, signaling proteins and proteins involved in metabolism as well as in tRNA and rRNA genes ([http://sgio2.biotec.psu.edu/sr/pmb\\_table.htm](http://sgio2.biotec.psu.edu/sr/pmb_table.htm)). More than half (57%) of the *Ds* insertions are within ORFs (known or putative) or introns (Figure 3). This is consistent with the estimate that half of the *Arabidopsis* genome comprises coding sequences (Lin *et al.*, 1999; Mayer *et al.*, 1999) and suggests that transposons of this family do not have a marked preference for either genic or intergenic sequences in *Arabidopsis*.

A finer analysis of insertion sites within ORFs revealed a pattern of preferential insertion around the known or putative ATG translational initiation codon, peaking within the first 200 bp of the ORF (Figure 4). A similar distribution has been reported for another *Ds* collection (Parinov *et al.*, 1999). No significant sequence preference for insertion was detected at a finer level of resolution. We determined the frequencies of individual nucleotides, triplets and hexamers

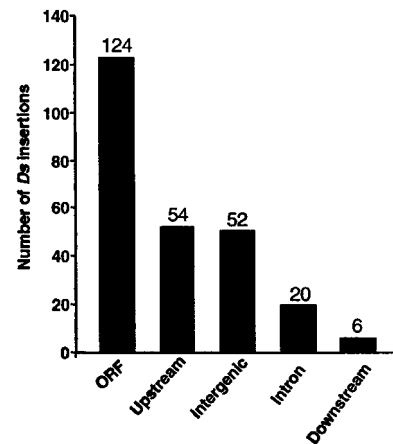


Figure 3. The distribution of the *Ds* insertions in and near genes. ORF represents insertions within open reading frame of known or putative genes, exclusive of introns; upstream designates insertions within 1 kb upstream of the translational start codon; downstream designates insertions within 0.2 kb downstream of the translational stop codon; intergenic refers to insertions at a distance of more than 1 kb from an ORF and an orphan exon predicted by the GRAIL program. Four lines with insertions in unannotated BACs are not included.

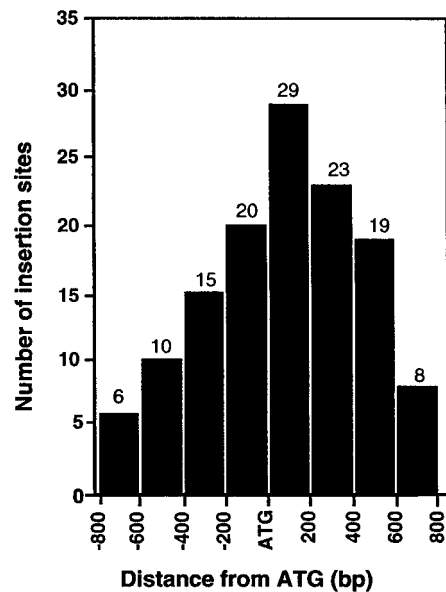


Figure 4. Distribution of the *Ds* insertions around the translational start codon. The figure shows the number of insertions in each 200 bp interval for 800 bp upstream and downstream from the known or putative ATG translational start codon.

Table 1. Analysis of *Ds* insertion sites in genes.

	<i>Ds</i> lines	Genomic clone	Accession number	Position of the insertion within the clone	Chr	Relationship of insertion site to ORF	Gene identity or homology
1	<u>201-001</u>	<u>F23A5</u>	<u>AC011713</u>	54620	I	Upstream	Cinnamoyl-CoA reductase from <i>Zea mays</i>
2	<u>201-006</u>	<u>F6N15</u>	<u>AF069299</u>	57305	IV	Upstream	Putative protein *
3	<u>201-007</u>	<u>F10A8</u>	<u>AC006200</u>	8183	II	Upstream	Putative fructose-bisphosphate aldolase
4	<u>201-008</u>	<u>F23E13</u>	<u>AL022141</u>	52179	IV	ORF	NTL1-like
5	<u>201-009</u>	<u>T4C15</u>	<u>AC004667</u>	69080	II	ORF	Unknown protein **
6	<u>201-010</u>	<u>T21F11</u>	<u>AC018849</u>	5529	I	ORF	Putative protein kinase
7	<u>201-012</u>	<u>F23A5</u>	<u>AC011713</u>	82857	I	ORF	CorA-like Mg <sup>2+</sup> transporter protein family
8	<u>201-013</u>	<u>T9J23</u>	<u>AC006072</u>	30399	II	ORF	Putative protein
9	<u>201-014</u>	<u>F18B13</u>	<u>AC009322</u>	68293	I	ORF	Putative protein
10	<u>201-015</u>	<u>T8O5</u>	<u>AL021890</u>	77213	IV	Intergenic	
11	<u>201-016</u>	<u>F14J9</u>	<u>AC003970</u>	76629	I	Downstream	Germin-like protein
12	<u>201-020</u>	<u>F11O4</u>	<u>AF096370</u>	40836	IV	ORF	Putative AP2 homologue
13	<u>201-023</u>	<u>F17A8</u>	<u>AL049482</u>	44262	IV	Predicted exon	
14	<u>201-025</u>	<u>MHF15</u>	<u>AB006700</u>	78849	V	ORF	Ubiquitin-activating enzyme UBA2
15	<u>201-033</u>	<u>T6K12</u>	<u>AC016829</u>	72609	III	ORF	Putative protein
16	<u>201-035</u>	<u>F23A5</u>	<u>AC011713</u>	34441	I	Downstream	Similar to ribosomal protein L7
18	<u>201-038</u>	<u>F13F21</u>	<u>AC007504</u>	47320	I	Downstream	Similar to acyl CoA synthetase
19	<u>201-041</u>	<u>F28J7</u>	<u>AC010797</u>	70999	III	Downstream	Similar to WARKY-like transcriptional activator
20	<u>201-042</u>	<u>F16N3</u>	<u>AC007519</u>	70849	I	Down stream	Similar to thioglucosidase
21	<u>201-044</u>	<u>F3N23</u>	<u>AC008017</u>	64788	I	Upstream	Acid $\beta$ -galactosidase
22	<u>201-046</u>	<u>ESSA1</u>	<u>Z99707</u>	435	IV	ORF	Cytochrome-like protein
23	<u>201-047</u>	<u>F6N15</u>	<u>AF069299</u>	13299	IV	ORF	Similar to transcriptional activator
24	<u>201-048</u>	<u>F5I10</u>	<u>AF013293</u>	108634	IV	ORF	Putative protein
25	<u>201-050</u>	<u>F6N23</u>	<u>AF058919</u>	61782	IV	Upstream	ATP-dependent RNA helicase-like
26	<u>201-052</u>	<u>F10A8</u>	<u>AC006200</u>	19703	II	ORF	Putative phosphatidic acid phosphatase
27	<u>201-053</u>	<u>T15B3</u>	<u>AL163975</u>	103099	III	Intron	Putative protein
28	<u>201-055</u>	<u>F5I10</u>	<u>AF013293</u>	53675	IV	Intron	Putative protein
29	<u>201-057</u>	<u>F10A5</u>	<u>AC006434</u>	17501	I	ORF	GAST like
30	<u>201-061</u>	<u>F1P15</u>	<u>AC004708</u>	46044	II	ORF	Putative protein
31	<u>201-065</u>	<u>F28J9</u>	<u>AC007918</u>	23705	I	Intergenic	
32	<u>201-066</u>	<u>F23A5</u>	<u>AC011713</u>	63157	I	Intergenic	
33	<u>201-067</u>	<u>T11I11</u>	<u>AC012680</u>	27084	I	ORF	Putative protein
34	<u>201-068</u>	<u>F6N15</u>	<u>AF069299</u>	4321	IV	Upstream	BRCA2-like
35	<u>201-070</u>	<u>T11A7</u>	<u>AC002339</u>	40784	II	ORF	Putative protein
36	<u>201-073</u>	<u>T19P19</u>	<u>AL022605</u>	17093	IV	Downstream	Unknown protein
37	<u>201-075</u>	<u>T21F11</u>	<u>AC018849</u>	69589	I	ORF	Putative glycerol kinase
38	<u>201-076</u>	<u>T31E10</u>	<u>AC004077</u>	58906	II	ORF	Putative pyruvate dehydrogenase complex E1 $\beta$ subunit
39	<u>201-078</u>	<u>T10I14</u>	<u>AL021712</u>	38264	IV	Downstream	Unknown protein
40	<u>201-079</u>	<u>T17H1</u>	<u>AC007730</u>	88228	II	ORF	Mitochondrial genome
41	<u>201-081</u>	<u>M7J2</u>	<u>AL022197</u>	13372	IV	ORF	Actin-depolymerizing factor-like
42	<u>201-084</u>	<u>K19B1</u>	<u>AB015469</u>	8618	V	Downstream	H-protein promoter binding factor-like protein
43	<u>201-085</u>	<u>F5N5</u>	<u>AP001300</u>	32696	III	-	Similar to mitochondrial genome
44	<u>201-096</u>	<u>T15P10</u>	<u>AF167571</u>	1544	IV	-	
45	<u>201-104</u>	<u>F28A23</u>	<u>AL021961</u>	48370	IV	ORF	Unknown protein
46	<u>201-106</u>	<u>F19K16</u>	<u>AC011717</u>	62891	I	Downstream	Chloroplast ribosomal protein S17
47	<u>201-109</u>	<u>T2H3</u>	<u>AF075597</u>	1895	IV	ORF	Putative protein
48	<u>201-111</u>	<u>T24C10</u>	<u>AC064840</u>	20041	I	Upstream	Putative protein

Table 1 continued.

	<i>Ds</i> lines	Genomic clone	Accession number	Position of the insertion within the clone	Chr	Relationship of insertion site to ORF	Gene identity or homology
49	<u>201-112</u>	T7H20	AL162508	59754	V	ORF	Similar to amino acid transport protein
50	<u>201-113</u>	T24P15	AC002561	15378	II	ORF	Unknown protein
51	<u>201-114</u>	T16O11	AC010871	18909	III	ORF	DNAj-like protein
52	201-116	F6N15	AF069299	54865	IV	Intergenic	
53	<u>201-119</u>	F14M4	AC004411	53109	II	Upstream	Putative glucoamylase
54	<u>201-120</u>	F3F9	AC013430	41369	I	ORF	Glutathione transferase
55	201-121	F15P23	AF128392	865	IV	Intron	
56	<u>201-122</u>	F9K20	AC005679	78581	I	Upstream	Unknown protein
57	<u>201-124</u>	F6P23	AC002354	31273	II	ORF	NAM (no apical meristem)-like protein
58	<u>201-125</u>	F19K16	AC011717	76325	I	Intergenic	
59	<u>201-127</u>	T14P8	AF069298	49353	IV	ORF	Putative receptor-like protein kinase
60	<u>201-128</u>	F17A22	AC005309	71777	II	ORF	Putative glutaredoxin
61	<u>201-132</u>	F23A5	AC011713	102282	I	Downstream	Unknown protein with F box domain
62	<u>201-134</u>	T17J13	AL138651	67166	III	ORF	Similar to phosphoprotein phosphatase
63	201-137	T23A1	AC002329	38280	II	ORF	Putative protein
64	<u>201-135</u>	T2D23	AC068143	9112	I	Upstream	Similar to copper homeostasis factor
65	<u>201-141</u>	F23H14	AC006837	84378	II	Intron	Putative sec-independent protein translocase
66	<u>201-144</u>	MSJ1	AB008268	4671	V	ORF	Putative protein
67	<u>201-146</u>	T30B22	AC005309	291	II	ORF	Putative SWI/SNF family transcription activator
68	<u>201-149</u>	F15N18	AL163815	55350	V	Intergenic	
69	<u>201-155</u>	MAA21	AL163818	52211	III	Upstream	Putative protein
70	<u>201-157</u>	MAA21	AL163818	236140	III	Upstream	Putative protein
71	<u>201-161</u>	F23A5	AC011713	29588	I	Intron	Chromomethylase (CMT1) gene
72	201-162	T25F15	AC009529	130	III	Not annotated	
73	<u>201-166</u>	F6N15	AF069299	27611	IV	ORF	TRNA
74	<u>201-169</u>	F23A5	AC011713	100852	I	Intron	F box domain
75	<u>201-173</u>	F17I14	AL353994	32552	V	ORF	$\beta$ -xylosidase
76	<u>201-174</u>	IG005110	AF013293	40559	IV	ORF	Putative protein
77	201-183	T22E19	AC016447	90278	I	ORF	similar to MtN21
78	<u>201-187</u>	F19K16	AC011717	108706	I	ORF	Putative protein
79	<u>201-191</u>	F7P12	AC023913	62678	I	ORF	Putative transcription factor
80	<u>201-197</u>	T7M24	AF077408	24204	IV	Intergenic	
81	<u>201-198</u>	T2H7	AC074176	13376	I	Repeat region	LTR retrotransposon
82	<u>201-206</u>	F6N15	AF069299	14660	IV	Upstream	Putative protein
83	<u>201-207</u>	T21F11	AC018849	4156	I	Upstream	Putative protein
84	<u>201-209</u>	F6N15	AF069299	88543	IV	ORF	Subtilase family serine protease-like
85	<u>202-011</u>	F11L15	AC006072	46094	II	Downstream	Putative protein
86	<u>202-015</u>	F15A17	AL163002	4012	V	ORF	Putative protein
87	<u>202-016</u>	F14G24	AC019018	69612	I	Intergenic	
88	202-019	F2D10	AC069251	120626	I	ORF	Putative protein
89	<u>202-020</u>	T24P13	AC006535	73181	I	Downstream	H-protein promoter binding factor
90	202-021	F20B17	AC010793	31705	I	ORF	Putative protein
91	<u>203-003</u>	T24D18	AC010924	65651	I	ORF	Wall-associated kinase
92	<u>203-004</u>	F15N18	AL163815	49242	V	Intron	Aspartate aminotransferase mRNA
93	<u>203-010</u>	ESSA1	Z99707	119490	IV	Upstream	Unknown protein
94	<u>203-013</u>	M3E9	AL022223	39449	IV	ORF	Putative protein
95	<u>203-014</u>	F24K9	AC008153	17597	III	ORF	Unknown protein
96	<u>203-015</u>	F13M14	AC011560	27043	III	ORF	Diadenosine tetraphosphate hydrolase-like
97	<u>203-016</u>	T28P6	AC007259	68987	I	Downstream	Putative protein

Table 1 continued.

	<i>Ds</i> lines	Genomic clone	Accession number	Position of the insertion within the clone	Chr	Relationship of insertion site to ORF	Gene identity or homology
98	<u>203-021</u>	<u>F6N15</u>	<u>AF069299</u>	34746	IV	Upstream	Similar to nucleotide sugar epimerases
99	203-023	F12F1	AC002131	52211	I	Upstream	Putative protein
100	203-025	T25K16	AC007323	78876	I	Intergenic	
101	<u>203-036</u>	<u>F3F20</u>	<u>AC007153</u>	45339	I	ORF	Putative protein
102	203-039	MOP10	<u>AB005241</u>	16844	V	Intron	Unknown protein
103	<u>203-052</u>	<u>F14C21</u>	<u>AC069144</u>	76791	I	–	Unknown protein
104	<u>203-054</u>	<u>F5A9</u>	<u>AC004133</u>	62241	I	ORF	Unknown protein
105	<u>203-057</u>	<u>F25C20</u>	<u>AC007296</u>	3791	I	Upstream	Unknown protein
106	<u>203-091</u>	<u>T20O10</u>	<u>AL163816</u>	11760	III	ORF	Putative glutaredoxin
107	<u>204-002</u>	<u>T16N11</u>	<u>AC013453</u>	15624	I	ORF	Anion exchanger family
108	<u>204-007</u>	<u>F21M12</u>	<u>AC000132</u>	26477	I	ORF	Double-stranded RNA-binding protein
109	<u>204-013</u>	<u>T10O24</u>	<u>AC007067</u>	11353	I	ORF	Protein phosphatase
110	<u>204-018</u>	<u>T16N11</u>	<u>AC013453</u>	45287	I	ORF	Putative serine/threonine-specific protein kinase
111	<u>204-024</u>	<u>F19K16</u>	<u>AC011717</u>	9385	I	ORF	Putative protein
112	<u>204-026</u>	<u>T28P6</u>	<u>AC007259</u>	88333	I	Intergenic	
113	<u>204-028</u>	<u>F14L17</u>	<u>AC012188</u>	79222	I	ORF	Similarity to an extensin precursor
114	<u>204-045</u>	<u>F2I9</u>	<u>AC005560</u>	19704	II	Downstream	WRKY family DNA-binding protein
115	<u>204-046</u>	<u>F20N2</u>	<u>AC002328</u>	20697	I	ORF	Putative protein
116	<u>204-049</u>	<u>F2D10</u>	<u>AC069251</u>	122778	I	ORF	Putative protein
117	204-052	MLN21	AB022220	63638	III	ORF	Protein kinase
118	<u>204-058</u>	<u>F15M7</u>	<u>AP002543</u>	57245	V	Intron	Putative protein
119	<u>204-059</u>	<u>MQN23</u>	<u>AB013395</u>	15647	V	ORF	RAP2.4 homologue
120	<u>204-076</u>	<u>T4C21</u>	<u>AL162295</u>	81812	III	Upstream	Putative protein
121	204-080	F5M6	AC079041	18976	I	ORF	Unknown protein
122	204-082	F9P14	AC025290	36666	I	Intron	
123	<u>204-084</u>	<u>F23H14</u>	<u>AC006837</u>	73865	II	ORF	Transfactor-like protein
124	<u>204-091</u>	<u>T5M16</u>	<u>AC010704</u>	100832	I	Intron	Putative aminotransferase
125	<u>204-096</u>	<u>T28A8</u>	<u>AL162691</u>	43388	III	Intron	Transporter-like protein
126	204-099	<u>T22C5</u>	<u>AC012375</u>	46246	I	ORF	ADP-glucose pyrophosphorylase
127	204-101	F10A8	AC006200	8368	II	Upstream	Putative fructose-bisphosphate aldolase
128	<u>204-116</u>	<u>F10A8</u>	<u>AC006200</u>	8229	II	Upstream	Putative fructose-bisphosphate aldolase
129	<u>204-120</u>	<u>F22K20</u>	<u>AC002291</u>	11882	I	ORF	Unknown protein
130	204-135	F7P12	AC023913	62564	I	Intergenic	
131	<u>204-141</u>	<u>F5I10</u>	<u>AF013293</u>	79196	IV	Upstream	Putative protein
132	<u>204-153</u>	<u>T28P6</u>	<u>AC007259</u>	39132	I	ORF	Glucose transporter
133	<u>204-155</u>	<u>MOE17</u>	<u>AB025629</u>	76461	III	Upstream	Zinc-finger-like protein
134	<u>204-172</u>	<u>F24D13</u>	<u>AC005851</u>	51918	II	Intergenic	
135	<u>204-187</u>	<u>T2P4</u>	<u>AC002336</u>	2697	II	ORF	Putative protein
136	<u>204-203</u>	<u>T9A14</u>	<u>AL035356</u>	28875	IV	Downstream	Cyclophilin-type peptidyl-prolyl <i>cis</i> -transisomerase
137	<u>204-207</u>	<u>F25C20</u>	<u>AC007296</u>	27659	I	Intron	Unknown protein
138	<u>204-234</u>	<u>T30B22</u>	<u>AC002535</u>	36670	II	Upstream	Putative protein
139	<u>204-262</u>	<u>MJK13</u>	<u>AB022218</u>	45763	III	–	No homology
140	<u>206-027</u>	<u>F23I14</u>	<u>AC006532</u>	66338	II	Intergenic	
141	206-036	F5O4	AC005936	92848	II	Predicted exon	
142	<u>206-048</u>	<u>MNC17</u>	<u>AB016890</u>	13507	V	Downstream	Similar to elongin

Table 1 continued.

	<i>Ds</i> lines	Genomic clone	Accession number	Position of the insertion within the clone	Chr	Relationship of insertion site to ORF	Gene identity or homology
143	<u>206-117</u>	<u>F10A8</u>	<u>AC006200</u>	52222	II	ORF	Putative protein
144	<u>206-129</u>	<u>MFC19</u>	<u>AB018113</u>	76437	V	ORF	Putative protein
145	<u>206-131</u>	<u>T4C15</u>	<u>AC004667</u>	12555	II	Upstream	Unknown protein
146	<u>206-136</u>	<u>F28K19</u>	<u>AC009243</u>	6303	I	ORF	Putative reverse transcriptase
147	206-143	F9L11	AC006424	74730	I	Upstream	Putative protein
148	206-144	MKP6	AB022219	6662	III	ORF	Unknown protein
149	<u>206-175</u>	<u>F17I5</u>	<u>AL031032</u>	12024	IV	Intron	Putative protein
150	<u>206-189</u>	<u>F9N11</u>	<u>AL109796</u>	85401	IV	ORF	Putative YIP1 protein
151	<u>206-206</u>	<u>T27E11</u>	<u>AL078579</u>	16510	IV	Intron	Putative $\beta$ -glucosidase
152	<u>206-215</u>	<u>MLD14</u>	<u>AB025624</u>	6600	III	Upstream	Putative protein
153	<u>206-216</u>	<u>T21F11</u>	<u>AC018849</u>	45272	I	ORF	Putative protein
154	<u>206-222</u>	<u>MCD7</u>	<u>AB009049</u>	53691	V	Upstream	Putative protein
155	<u>206-224</u>	<u>T18E12</u>	<u>AC006284</u>	74624	II	ORF	Putative protein
156	<u>206-227</u>	<u>F16G20</u>	<u>AL031326</u>	1526	IV	Intron	Serine/threonine kinase-like protein
157	<u>210-001</u>	<u>F7J8</u>	<u>AL137189</u>	71241	V	Intergenic	
158	210-065	MUL3***	AB023042	47466	IV	Intergenic	
159	210-067	F9K20	AC005679	36376	I	ORF	Glycoprotein family protein
160	<u>210-110</u>	<u>T20N10</u>	<u>AL353032</u>	73272	I	ORF	Putative protein
161	210-135	T1B9	AC012395	77055	III	Intron	Purple acid phosphatase
162	210-137	T18K17	AC010556	132379	I	Repeat region	
163	<u>210-144</u>	<u>F21M12</u>	<u>AC000132</u>	44345	I	Upstream	Unknown protein
164	211-001	F2K11	AF372939	15752	I	ORF	Unknown protein
165	211-002	T20O10	AL163816	61818	III	ORF	tRNA isopentenyl transferase-like protein
166	211-010	MWF20	AB025638	50644	V	Intergenic	
167	<u>211-011</u>	<u>T10P11</u>	<u>AC002330</u>	36895	IV	Upstream	Acetic acid transporter
168	<u>211-026</u>	<u>F23I14</u>	<u>AC007265</u>	29944	II	Upstream	putative Na <sup>+</sup> /H <sup>+</sup> antiporter
169	<u>212-002</u>	<u>F7J8</u>	<u>AL137189</u>	13055	V	Upstream	Short-chain dehydrogenase/reductase family-like
170	<u>212-031</u>	<u>F21J9</u>	<u>AC000103</u>	3775	I	ORF	Similar to 3-oxoacyl-reductase precursor
171	<u>215-013</u>	<u>F9D24</u>	<u>AL137081</u>	89636	III	Downstream	Putative protein
172	<u>215-017</u>	<u>F26H11</u>	<u>AC006264</u>	25049	II	Intron	Hypothetical protein with C2-domain profile
173	<u>216-005</u>	<u>F12F1</u>	<u>AC002131</u>	103953	I	Upstream	Similar to vesicle trafficking protein
174	<u>216-008</u>	<u>F6N15</u>	<u>AF069299</u>	58702	IV	ORF	Putative protein
175	<u>216-023</u>	<u>T20F6</u>	<u>AC002521</u>	16063	II	ORF	Putative receptor-like protein kinase
176	216-028	F13B4	AC027134	63362	I	ORF	Putative protein
177	<u>216-029</u>	<u>F21B7</u>	<u>AC002560</u>	70184	I	Upstream	Putative protein
178	<u>216-030</u>	<u>T23K3</u>	<u>AC007069</u>	44279	II	ORF	Putative purple acid phosphatase
179	<u>216-033</u>	<u>F22G5</u>	<u>AC022464</u>	5701	I	ORF	Metallothionein gene/hypothetical protein
180	<u>216-039</u>	<u>F2A19</u>	<u>AL132962</u>	29459	III	Upstream	Cysteine synthase AtcysC1
181	<u>216-042</u>	<u>F24B9</u>	<u>AC007583</u>	28964	I	Upstream	Similar to ER-type calcium pump protein
182	<u>216-044</u>	<u>F24P17</u>	<u>AC011623</u>	52027	III	Upstream	Putative protein
183	<u>216-050</u>	<u>K21C13</u>	<u>AB010693</u>	32419	V	ORF	Similar to disease resistance protein
184	<u>301-002</u>	<u>K15I22</u>	<u>AB016870</u>	13957	V	ORF	Receptor protein kinase-like protein
185	<u>301-007</u>	<u>K19M22</u>	<u>AB016885</u>	5140	V	ORF	Subtilisin-like serine protease
186	<u>301-009</u>	<u>T22K18</u>	<u>AC010927</u>	51615	III	Upstream	Unknown protein
187	<u>305-010</u>	<u>F1N21</u>	<u>AC002130</u>	9340	I	ORF	Similar to CREB-binding protein
188	<u>305-015</u>	<u>T12J13</u>	<u>AC009327</u>	34433	III	Intron	Putative protein
189	<u>310-010</u>	<u>MGO3</u>	<u>AB019231</u>	22639	V	Intron	Protein kinase APK1A-like

Table 1 continued.

<i>Ds</i> lines	Genomic clone	Accession number	Position of the insertion within the clone	Chr	Relationship of insertion site to ORF	Gene identity or homology	
190	310-016	F23K16	<a href="#">AL078620</a>	61255	IV	ORF	Cytochrome P450-like protein
191	<a href="#">310-017</a>	MZN1	<a href="#">AB020755</a>	73216	V	ORF	Putative protein
192	310-024	T23J18	AC011661	31604	I	Upstream	Putative protein
193	310-031	T12C14	<a href="#">AL162507</a>	58711	III	ORF	Unknown protein
194	311-003	T21C14	AL138639	75049	III	Predicted exon	
195	<a href="#">311-007</a>	<a href="#">T19P19</a>	<a href="#">AL022605</a>	41140	IV	ORF	Putative protein
196	<a href="#">311-015</a>	F21O3	AC009853	15961	III	ORF	Similar to helix-loop-helix DNA binding motif
197	<a href="#">311-024</a>	MSF19	<a href="#">AB016891</a>	1	V	ORF	ATPase
198	<a href="#">311-036</a>	F24I3	AL138655	76218	III	ORF	Unknown protein
199	<a href="#">311-039</a>	F24P17	AC011623	45396	III	Downstream	Unknown protein
200	<a href="#">311-044</a>	<a href="#">MLD14</a>	<a href="#">AB025624</a>	37541	III	ORF	Cysteine protease
201	<a href="#">311-047</a>	F10A8	<a href="#">AC006200</a>	78859	II	Upstream	Putative nicotinate-nucleotide
202	<a href="#">311-057</a>	F22L4	AC061957	11618	I	ORF	ACS synthase
203	311-059	<a href="#">MCB17</a>	<a href="#">AB022215</a>	15238	III	ORF	Putative protein
204	<a href="#">311-062</a>	F27C12	<a href="#">AC006585</a>	67248	II	ORF	Unknown protein
205	<a href="#">311-068</a>	F24K9	<a href="#">AC008153</a>	15203	III	Downstream	Putative protein
206	<a href="#">311-073</a>	MTH12	<a href="#">AB006705</a>	44994	V	ORF	Receptor-like protein kinase
207	<a href="#">311-081</a>	MTH12	<a href="#">AB006705</a>	36581	V	ORF	Receptor-like protein kinase
208	<a href="#">311-108</a>	MMN10	<a href="#">AB015475</a>	22775	V	Upstream	Small GTP-binding protein-like
209	312-001	T12P18	AC010852	33324	I	ORF	Putative disease resistance protein
210	<a href="#">322-002</a>	ESSAI	<a href="#">Z97337</a>	168639	IV	ORF	Similar to endoprotease
211	<a href="#">322-008</a>	F7H19	<a href="#">AL031018</a>	3834	IV	Downstream	Similar to Na(+)-dependent transporter
212	<a href="#">323-002</a>	<a href="#">F10O3</a>	<a href="#">AC006550</a>	4670	I	Upstream	Unknown protein
213	323-003	T23E23	<a href="#">AC002423</a>	60147	V	ORF	Ring H2 finger protein-like
214	323-009	F20C19	AP001298	67428	III	Upstream	Putative protein
215	<a href="#">323-011</a>	T10C21	<a href="#">AL109787</a>	18958	IV	ORF	Putative protein
216	<a href="#">323-012</a>	<a href="#">MBK21</a>	<a href="#">AB024033</a>	72614	III	Downstream	FUSCA protein FUS6-like protein
217	<a href="#">323-021</a>	MZN1	<a href="#">AB020755</a>	56581	V	ORF	PRL1-associated protein
218	323-023	T20L15	AL162351	30428	V	Upstream	Putative protein
219	<a href="#">323-024</a>	<a href="#">F4I1</a>	<a href="#">AC004521</a>	53863	II	Repeat region	
220	<a href="#">323-026</a>	MEE6	AB010072	80012	V	Upstream	Unknown protein
221	<a href="#">323-028</a>	MMN10	<a href="#">AB015475</a>	63825	V	Downstream	Similar to RNaseP protein
222	323-030	MZN1***	<a href="#">AB020755</a>	75973	V	Upstream	Putative protein
223	<a href="#">323-032</a>	<a href="#">T2P4</a>	<a href="#">AC002336</a>	10910	II	Intergenic	
224	326-008	ESSAI	<a href="#">Z97338</a>	128028	IV	Upstream	Putative protein
225	<a href="#">326-014</a>	F9H3	<a href="#">AF071527</a>	54238	IV	ORF	Mammalian ankyrin-like
226	<a href="#">326-016</a>	<a href="#">K14B15</a>	<a href="#">AB025608</a>	9553	III	ORF	Unknown protein
227	<a href="#">327-003</a>	F4F15	<a href="#">AL049711</a>	92929	III	ORF	Similar to RNA-binding protein
228	<a href="#">327-021</a>	F10M23	<a href="#">AL035440</a>	125844	IV	Upstream	Putative protein
229	<a href="#">327-023</a>	T7H20	AL162508	34170	V	ORF	Unknown protein
230	<a href="#">327-025</a>	MCA23	<a href="#">AB016886</a>	16510	V	ORF	Putative protein
231	327-032	F6G3	AL078464	67932	IV	ORF	Cadmium-transporting ATPase-like protein
232	<a href="#">327-033</a>	T9L3	AL391149	59922	V	ORF	Unknown protein
233	327-038	F13B4	AC027134	67824	I	ORF	Putative protein
234	<a href="#">327-048</a>	MAC9	AB010069	50880	V	ORF	Leafy

Table 1 continued.

	<i>Ds</i> lines	Genomic clone	Accession number	Position of the insertion within the clone	Chr	Relationship of insertion site to ORF	Gene identity or homology
235	327-058	F3H9	AC021044	44835	I	Upstream	Putative protein
236	<u>327-092</u>	MXC9	<u>AB007727</u>	48635	V	Downstream	Similar to RNA-binding protein
237	328-001	T18N24	AC074111	56372	I	Not annotated	
238	<u>328-002</u>	K21C13	<u>AB010693</u>	37171	V	ORF	Unknown protein
239	<u>328-018</u>	MIL15	AB028615	10846	III	Intergenic	
240	<u>328-019</u>	<u>T8O11</u>	<u>AC006069</u>	10257	II	ORF	Putative protein
241	<u>328-023</u>	<u>K5K13</u>	<u>AB025615</u>	14672	III	Intergenic	
242	<u>329-010</u>	F6N23	<u>AF058919</u>	48755	IV	Upstream	Putative dehydrogenase/cyclohydrolase
243	<u>332-031</u>	K19E1	<u>AB013388</u>	46429	V	Downstream	Ubiquitin-conjugating enzyme
244	<u>345-005</u>	F6N23	<u>AF058919</u>	77233	IV	ORF	Putative protein
245	<u>345-024</u>	F2K11	<u>AC008047</u>	15621	I	Upstream	Similar to receptor-like-protein kinase
246	<u>345-033</u>	<u>F4P13</u>	<u>AC009325</u>	51706	III	ORF	Putative protein
247	<u>345-034</u>	F28I8	<u>AC006955</u>	17698	II	Intergenic	
248	<u>353-005</u>	<u>MLD14</u>	<u>AB025624</u>	33369	III	Intergenic	
249	<u>361-027</u>	F23H14	AC006837	75025	II	ORF	Putative protein
250	<u>361-031</u>	F23I4	<u>AC007265</u>	133	II	Predicted exon	
251	<u>378-002</u>	MBK23	<u>AB005233</u>	60096	V	Downstream	Putative protein
252	<u>378-006</u>	T1P17	<u>AL049730</u>	113474	IV	ORF	Similar to polyubiquitin 6
253	<u>378-012</u>	MVA11	AP001311	40482	III	ORF	Similar to reverse transcriptase
254	<u>378-016</u>	MNJ8	<u>AB017069</u>	19587	V	Upstream	Unknown protein
255	<u>379-001</u>	F12E4	AL162751	54443	V	ORF	Histidyl transfer RNA synthase
256	<u>379-003</u>	F2E2	AC069252	18196	I	Downstream	Putative protein
257	<u>379-011</u>	<u>F14L17</u>	<u>AC012188</u>	5176	I	ORF	Similar to PIT1
258	<u>379-023</u>	T6H20	<u>AL096859</u>	58142	III	ORF	Putative protein
259	<u>379-025</u>	K2A18	<u>AB011474</u>	54076	V	ORF	Proteasome $\alpha$ subunit
260	385-004	K19E20	<u>AB017061</u>	15832	V	Upstream	Receptor protein kinase-like protein

Predicted protein \*: Protein predicted by computer algorithm.

Unknown protein \*\*: Predicted protein with EST match.

\*\*\*Homology of the flanking sequence is 90–95%.

in windows of sizes between one and 200 centered on the insertion position. No striking pattern was observed, except perhaps that the nucleotide immediately after the insertion position was infrequently a T (15%), while the next two nucleotides were T's more frequently than expected by chance alone for the overall base composition of the examined sequences (W. Miller and N. Fedoroff, unpublished).

Regulatory elements that affect transcription have been identified within about a kilobase upstream from the translational start site in a number of studies (Bate *et al.*, 1996; Cardoso *et al.*, 1997; Kim *et al.*, 1997; de Boer *et al.*, 1999; Kim and Guiltinan, 1999; Cunnillera *et al.*, 2000). Maleck *et al.* (2000) considered sequences 1.1 kb upstream of ATG as the promoter region in searching for the *cis* regulatory elements of gene clusters identified from microarray expression

profiles. About 20% of the lines in our collection had *Ds* insertions within 1 kb upstream from the translational start site (Figure 3) and these have the potential of disrupting expression of the adjacent gene (Trifiro *et al.*, 1997; Zhang and Somerville, 1997; Gu *et al.*, 1998; Wisman *et al.*, 1998; Halliday *et al.*, 1999; Jean *et al.*, 1999; Johnson *et al.*, 1999; Papi *et al.*, 2000). The *Ds* insertion was within an intron in 20 lines. There are several reports that insertions in introns can disrupt gene expression (Negruk *et al.*, 1996; Henikoff and Comai, 1998; Amin *et al.*, 1999; Zou *et al.*, 1999). Transcribed, but untranslated sequences (3'-UTR) downstream from the translational stop codon also have been reported to influence translation efficiency and mRNA stability (Kern *et al.*, 1997; Chan and Yu, 1998; Liu and Redmond, 1998). There are 6 lines in the present group in which the *Ds* inserted

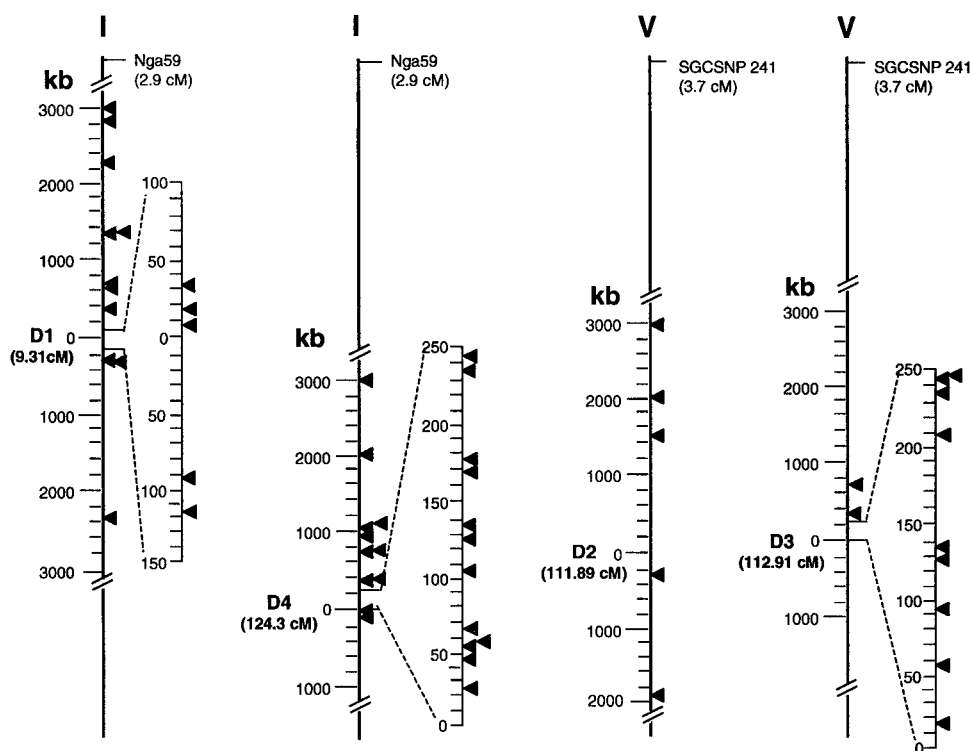


Figure 5. The distribution of reinsertion sites near donor sites. Transposons are represented by triangles. All of the reinsertion sites are for transposons that originated from the donor site shown on the diagram at 0 and the distance is given in kb from the donor site. The location of each donor site on the chromosome is shown in cM in parenthesis.

within 0.2 kb downstream from the stop codon, a reasonable 3'-UTR length (Figure 3). We classified as 'intergenic' all lines in which the *Ds* element inserted more than 1 kb upstream and 0.2 kb downstream from the ORF or in a sequence whose identity was not certain, such as an 'orphan' exon identified by the GRAIL program (<ftp://arthur.epm.ornl.gov/pub/xgrail>). About 20% of the insertions in our collection were in such intergenic regions (Figure 3). Thus 80% of the insertions in the present collection have the potential for affecting expression of genes.

#### Genomic distribution of reinsertion sites

Members of the *Ac* transposon family are known to reinsert preferentially near the donor site in *Arabidopsis*, as they do in maize (Greenblatt and Brink, 1962, 1963; Greenblatt, 1984; Bancroft and Dean, 1993b; Smith *et al.*, 1996; Long *et al.*, 1997; Machida *et al.*, 1997; Dubois *et al.*, 1998). In the present study, the fraction of same-chromosome reinsertions varied between 32% and 48% for different donor sites. There was a typical cluster of reinsertion sites near donor

sites D1, D3, and D4, but not near donor site D2 (Figure 2).

Two-thirds to more than three-fourths of the same-chromosome insertion sites were within 300 kb of the donor sites D1, D3 and D4. A more detailed display of the insertion sites closest to the donor site is shown in Figure 5. At two sites, there is a marked asymmetry in the distribution of reinsertion sites on the two sides of the donor locus. All 9 insertion sites within 300 kb of donor site D3, as well as 12 of 14 reinsertions within 300 kb of donor site D4, are on the centromere-proximal side. While donor site D4 is close to the end of chromosome 1 in a region not yet completely mapped, it appears likely that it is at least several hundred kilobases from the telomere. Thus we find an asymmetric distribution of reinsertion sites at two of the four donor sites used in the present study. A similar asymmetric distribution of reinsertion sites has been described at the maize *P* locus, although the distances involved are likely to be greater in maize (Greenblatt, 1984). By contrast, the reinsertion sites within 300 kb of the donor site D1 are distributed uniformly on both sides. Similar reinsertion site pat-

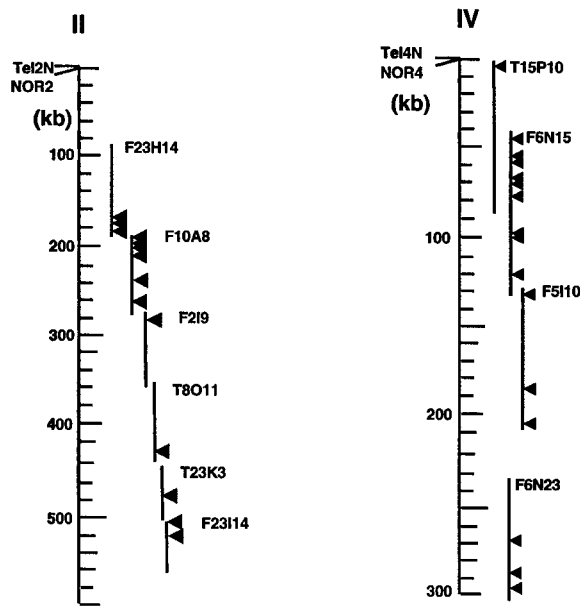


Figure 6. The distribution of the *Ds* insertion sites in hot regions near NORs. Insertions are represented by triangles and have been shown on the individual BACs covering the regions near the NORs on chromosomes 2 and 4.

terns have been reported at the *bronze* locus of maize (Dooner and Belachew, 1989) and for the single donor site previously characterized at the molecular level in *Arabidopsis* (Machida *et al.*, 1997). Interestingly, only one of the reinsertions from the *D2* locus was within 300 kb of the donor site.

In addition to the clusters of reinsertions around three of the four donor sites, there are two regions of strikingly preferential reinsertion near each of the two telomeric nucleolus organizers (NOR) on chromosomes 2 and 4 (Figure 2). Similar clustering of reinsertion sites near the NORs has been reported previously in a collection of *Arabidopsis Ds* transposon lines derived from parental lines in which the selection markers were designed for preferential recovery of reinsertion sites unlinked to the donor site (Parinov *et al.*, 1999). The reinsertion site clusters on chromosomes 2 and 4 are located between 150–300 kb and 40–130 kb from the telomere, respectively (Figure 6).

## Discussion

### Global transposition patterns

The present collection of lines comprises the largest existing collection of *Arabidopsis* lines with newly

transposed *Ds* elements whose composition has not been biased by selection against transposition to closely linked sites (Parinov *et al.*, 1999). Because the precise insertion sites have been identified at the nucleotide sequence level, this collection permits an evaluation of transposition patterns at both chromosomal and local scales. Overall, roughly 40% of transpositions were intrachromosomal and a majority (64%) were to sites within 3 Mb of the donor site, in agreement with previous recombination-based observations (Smith *et al.*, 1996). We did not observe the many very short-range transpositions reported in two recent studies (Machida *et al.*, 1997; Parinov *et al.*, 1999). Lines with reinsertions in the T-DNA comprised less than 2% of the reinsertions. The negative selection scheme used by Parinov *et al.* (1999) is likely to favor the recovery of insertions that inactivate the negative selectable marker adjacent to the donor site. Our selection scheme, by contrast, mitigates against recovery of insertions in the positive selectable marker used to detect transposition, possibly resulting in some underestimation of the short-range transposition frequency.

One of the four donor sites showed no marked bias toward either intrachromosomal or short-range transpositions. Two of the sites with nearby clusters showed asymmetric distributions of reinsertion sites, while the third did not. This variety in the patterns of reinsertion site distribution relative to the donor site suggests a major influence of local chromosome structure. We previously reported significant variation in the frequency of *Ds* transposition from different single-copy T-DNA donor sites, also suggesting that local chromatin structure influences transposition frequency (Smith *et al.*, 1996). But the consistent observation of reinsertion site clusters for *Ac* and *Ds* transposons near donor sites in several different plants (Greenblatt, 1984; Dooner and Belachew, 1989; Jones *et al.*, 1990; Peterson, 1990; Athma *et al.*, 1992; Moreno *et al.*, 1992; Dubois *et al.*, 1998; Ito *et al.*, 1999) suggests that nearby sequences are preferentially incorporated into transposition complexes as a fundamental feature of the transposition mechanism (Feldmar and Kunze, 1991; Becker and Kunze, 1996; Essers *et al.*, 2000).

Striking insertional hotspots near the NORs on both chromosomes 2 and 4 were observed in the present collection, as well as in a population of transposon lines obtained by selecting for unlinked transpositions (Parinov *et al.*, 1999). The clusters are near, but not in, the ribosomal RNA repeats of the subtelomeric NOR regions. At the nucleotide sequence

level, the clusters are spread over a 50–100 kb region. Possible explanations for these hot regions are: (1) that they are in close juxtaposition with donor sites in the 3-dimensional architecture of the nucleus, (2) that they have a particularly open and accessible chromatin structure, and (3) that they contain target sequences preferred by the transposons. Since *Ds* elements have transposed into these regions from many different donor sites, it seems unlikely that the explanation lies in the relative positions of the donor sites with respect to the NORs. The likeliest explanation is that the regions near the NORs are in a chromatin configuration that makes the DNA available for association with the transposition complex.

#### *Local transposition preferences*

The probability that an insertion is within an ORF and between ORFs is roughly proportional to the fraction of the genome occupied by coding and non-coding sequences in *Arabidopsis*. This contrasts with the preferential reinsertion of transposons into active chromosomal areas in maize (Chen *et al.*, 1987). It may be that the different methylation levels of *Arabidopsis* and maize transposon-rich intergenic regions account for this difference (Pruitt and Meyerowitz, 1986; Bennetzen *et al.*, 1994; Jeddeloh and Richards, 1996; SanMiguel *et al.*, 1998). Insertions within ORFs cluster around the translational initiation codon, showing a slight bias toward sequences downstream from the initiation codon. This bias is difficult to explain, since transcription and translation are not coupled and transcription start sites are generally at a distance from the translational start site.

About 70% of the lines in our collection have *Ds* insertions in the ORF or the putative promoter region of a gene, and half of these show GUS expression patterns (unpublished data). Only two of 49 homozygous insertion lines tested have exhibited an obvious mutant phenotype under standard growth conditions (Tsugeki *et al.*, 1996; Lu and Fedoroff, 2000). This frequency of mutations in our transposant lines is in agreement with the previous reports using the *Ac/Ds* system in *Arabidopsis* (Bancroft and Dean, 1993a; Long *et al.*, 1993; Sundaresan *et al.*, 1995; Long *et al.*, 1997). In a systematic search for knockouts in R2R3 MYB genes, Meissner *et al.* (1999) reported that none of the 32 homozygous lines carrying insertions in 26 different MYB genes showed any obvious mutant phenotype under normal growth conditions. In a search for T-DNA tagged lines in the cytochrome P450 genes

Winkler *et al.* (1998) identified 12 different mutants, only one of which showed a mutant phenotype under normal growing conditions.

One explanation for the paucity of mutant phenotypes is that they are manifested only under a restricted set of environmental conditions (Swaminathan *et al.*, 2000). Structural and functional redundancy are also likely to contribute. Only 35% of *Arabidopsis* genes are unique or singletons, while the remaining 65% are members of families comprising two or more genes (Initiative, 2000). Of the 66 lines in our collections with ORF insertions in genes of known function, only one has an insertion in a singleton gene. Plants having insertions in genes with closely related family members may fail to show mutant phenotypes because of functional redundancy.

#### *Ds insertions for functional genomics studies*

Functional genomics is an integrative approach, combining structural genomics information, genome-wide expression monitoring of mRNA or protein, and large-scale screening for isolating mutants to study gene function. To understand the cellular function of every gene, information and biological materials from these diverse sources must be easily accessible. Towards this end, we have created a searchable insertion site database accessible through the Internet (<http://sgio2.biotec.psu.edu/sr>) and made the insertion lines available through the *Arabidopsis* Biological Resource Center, Columbus, OH.

#### **Acknowledgements**

We thank Maya Olsen and Gayle VanTol for technical support. We thank Nigam Shah, Hong Gao and John Szot for assistance in developing the web page.

#### **References**

- Altschul, S.F., Gish, W., Miller, W. and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25: 3389–3402.
- Amin, P., Sy, D.A., Pilgrim, M.L., Parry, D.H., Nussaume, L. and Hoffman, N.E. 1999. *Arabidopsis* mutants lacking the 43- and 54-kilodalton subunits of the chloroplast signal recognition particle have distinct phenotypes. *Plant Physiol.* 121: 61–70.
- Athma, P., Grotewold, E. and Peterson, T. 1992. Insertional mutagenesis of the maize P gene by intragenic transposition of *Ac*. *Genetics* 131: 199–209.

- Bancroft, I., Bhatt, A.M., Sjodin, C., Scofield, S., Jones, J.D. and Dean, C. 1992. Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 233: 449–461.
- Bancroft, I. and Dean, C. 1993a. Factors affecting the excision frequency of the maize transposable element *Ds* in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 240: 65–72.
- Bancroft, I. and Dean, C. 1993b. Transposition pattern of the maize element *Ds* in *Arabidopsis thaliana*. *Genetics* 134: 1221–1229.
- Bate, N., Spurr, C., Foster, G.D. and Twell, D. 1996. Maturation-specific translational enhancement mediated by the 5'-UTR of a late pollen transcript. *Plant J.* 10: 613–623.
- Bechtold, N. and Pelletier, G. 1998. *In planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Meth. Mol. Biol.* 82: 259–266.
- Becker, H.A. and Kunze, R. 1996. Binding sites for maize nuclear proteins in the subterminal regions of the transposable element *Activator*. *Mol. Gen. Genet.* 251: 428–435.
- Bennetzen, J.L., Schrick, K., Springer, P.S., Brown, W.E. and SanMiguel, P. 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* 37: 565–576.
- Bensen, R.J., Johal, G.S., Crane, V.C., Tossberg, J.T., Schnable, P.S., Meeley, R.B. and Briggs, S.P. 1995. Cloning and characterization of the maize *An1* gene. *Plant Cell* 7: 75–84.
- Cardoso, M.I., Meijer, A.H., Rueb, S., Machado, J.A., Memelink, J. and Hoge, J.H. 1997. A promoter region that controls basal and elicitor-inducible expression levels of the NADPH:cytochrome P450 reductase gene (*Cpr*) from *Catharanthus roseus* binds nuclear factor GT-1. *Mol. Gen. Genet.* 256: 674–681.
- Castle, L.A., Errampalli, D., Atherton, T.L., Franzmann, L.H., Yoon, E.S. and Meinke, D.W. 1993. Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. *Mol. Gen. Genet.* 241: 504–514.
- Chan, M.T., Yu, S.M. 1998. The 3'-untranslated region of a rice  $\alpha$ -amylase gene mediates sugar-dependent abundance of mRNA. *Plant J.* 15: 685–695.
- Chen, J., Greenblatt, I.M. and Dellaporta, S.L. 1987. Transposition of *Ac* from the P locus of maize into unreplicated chromosomal sites. *Genetics* 117: 109–116.
- Chuck, G., Meeley, R.B. and Hake, S. 1998. The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1*. *Genes Dev.* 12: 1145–1154.
- Cunillera, N., Boronat, A. and Ferrer, A. 2000. Spatial and temporal patterns of GUS expression directed by 5' regions of the *Arabidopsis thaliana* farnesyl diphosphate synthase genes *FPS1* and *FPS2*. *Plant Mol. Biol.* 44: 747–758.
- Das, L. and Martienssen, R. 1995. Site-selected transposon mutagenesis at the *hcf106* locus in maize. *Plant Cell* 7: 287–294.
- de Boer, G.J., Testerink, C., Pielage, G., Nijkamp, H.J. and Stuitje, A.R. 1999. Sequences surrounding the transcription initiation site of the *Arabidopsis* enoyl-acyl carrier protein reductase gene control seed expression in transgenic tobacco. *Plant Mol. Biol.* 39: 1197–1207.
- Dooner, H.K. and Belachew, A. 1989. Transposition pattern of the maize element *Ac* from the *bz-m2(Ac)* allele. *Genetics* 122: 447–457.
- Dubois, P., Cutler, S. and Belzile, F.J. 1998. Regional insertional mutagenesis on chromosome III of *Arabidopsis thaliana* using the maize *Ac* element. *Plant J.* 13: 141–151.
- Essers, L., Adolphs, R.H. and Kunze, R. 2000. A highly conserved domain of the maize activator transposase is involved in dimerization. *Plant Cell* 12: 211–224.
- Fedoroff, N.V. and Smith, D.L. 1993. A versatile system for detecting transposition in *Arabidopsis*. *Plant J.* 3: 273–289.
- Feldmar, S. and Kunze, R. 1991. The ORF $\alpha$  protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain. *EMBO J.* 10: 4003–4010.
- Fladung, M. 1999. Gene stability in transgenic aspen (*Populus*). I. Flanking DNA sequences and T-DNA structure. *Mol. Gen. Genet.* 260: 574–581.
- Gheysen, G., Villarroel, R. and Van Montagu, M. 1991. Illegitimate recombination in plants: a model for T-DNA integration. *Genes Dev.* 5: 287–297.
- Greenblatt, I.M. 1984. A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element, *Modulator*, in maize. *Genetics* 108: 471–485.
- Greenblatt, I.M. and Brink, R.A. 1962. Twin mutations in medium variegated pericarp maize. *Genetics* 47: 489–501.
- Greenblatt, I.M. and Brink, R.A. 1963. Transpositions of *Modulator* in maize into divided and undivided chromosome segments. *Nature* 197: 412–413.
- Gu, Q., Ferrandiz, C., Yanofsky, M.F. and Martienssen, R. 1998. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125: 1509–1517.
- Halliday, K.J., Hudson, M., Ni, M., Qin, M. and Quail, P.H. 1999. *poc1*: an *Arabidopsis* mutant perturbed in phytochrome signaling because of a T DNA insertion in the promoter of *PIF3*, a gene encoding a phytochrome-interacting bHLH protein. *Proc. Natl. Acad. Sci. USA* 96: 5832–5837.
- Henikoff, S. and Comai, L. 1998. A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics* 149: 307–318.
- Hu, G., Yalpani, N., Briggs, S.P. and Johal, G.S. 1998. A porphyrin pathway impairment is responsible for the phenotype of a dominant disease lesion mimic mutant of maize. *Plant Cell* 10: 1095–1105.
- Hui, E.K.-W., Wang, P.-C. and Lo, S.J. 1998. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. *Cell Mol. Life Sci.* 54: 1403–1411.
- Initiative, A.G. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Ito, T., Seki, M., Hayashida, N., Shibata, D. and Shinozaki, K. 1999. Regional insertion mutagenesis of genes on *Arabidopsis thaliana* chromosome V using the *Ac/Ds* transposon in combination with a cDNA scanning method. *Plant J.* 17: 433–444.
- Jean, M., Pelletier, J., Hilpert, M., Belzile, F., Kunze, R. 1999. Isolation and characterization of *AtMLH1*, a *MutL* homologue from *Arabidopsis thaliana*. *Mol. Gen. Genet.* 262: 633–642.
- Jeddeloh, J.A. and Richards, E.J. 1996. mCCG methylation in angiosperms. *Plant J.* 9: 579–586.
- Johnson, K.R., Cook, S.A., Erway, L.C., Matthews, A.N., Sanford, L.P., Paradies, N.E. and Friedman, R.A. 1999. Inner ear and kidney anomalies caused by IAP insertion in an intron of the *Eyal* gene in a mouse model of BOR syndrome. *Hum. Mol. Genet.* 8: 645–653.
- Jones, J.D., Carland, F., Lim, E., Ralston, E. and Dooner, H.K. 1990. Preferential transposition of the maize element *Activator* to linked chromosomal locations in tobacco. *Plant Cell* 2: 701–707.
- Kern, J.A., Warnock, L.J. and McCafferty, J.D. 1997. The 3'-untranslated region of IL-1 $\beta$  regulates protein production. *J. Immunol.* 158: 1187–1193.
- Kim, H.U., Park, B.S., Jin, Y.M. and Chung, T.Y. 1997. Promoter sequences of two homologous pectin esterase genes from Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) and

- pollen-specific expression of the GUS gene driven by a promoter in tobacco plants. *Mol. Cells* 7: 21–27.
- Kim, K.N. and Guitinan, M.J. 1999. Identification of *cis*-acting elements important for expression of the starch-branching enzyme I gene in maize endosperm. *Plant Physiol.* 121: 225–236.
- Krysan, P.J., Young, J.C. and Sussman, M.R. 1999. T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11: 2283–2290.
- Kumar, S. and Fladung, M. 2000. Transgene repeats in aspen: molecular characterization suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome. *Mol. Gen. Genet.* 264: 20–28.
- Lin, X., Kaul, S., Rounsley, S. *et al.* 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761–768.
- Lister, C. and Dean, C. 1993. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* 4: 457–463.
- Liu, S.Y. and Redmond, T.M. 1998. Role of the 3'-untranslated region of RPE65 mRNA in the translational regulation of the RPE65 gene: identification of a specific translation inhibitory element. *Arch. Biochem. Biophys.* 357: 37–44.
- Liu, Y.-G., Mitsukawa, N., Oosumi, T. and Whittier, R.F. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* 8: 457–463.
- Liu, Y.G., Mitsukawa, N. and Whittier, R.F. 1993. Rapid sequencing of unpurified PCR products by thermal asymmetric PCR cycle sequencing using unlabeled sequencing primers. *Nucl. Acids Res.* 21: 3333–3334.
- Liu, Y.G. and Whittier, R.F. 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* 25: 674–681.
- Long, D., Swinburne, J., Martin, M., Wilson, K., Sundberg, E., Lee, K. and Coupland, G. 1993. Analysis of the frequency of inheritance of transposed *Ds* elements in *Arabidopsis* after activation by a CaMV 35S promoter fusion to the *Ac* transposase gene. *Mol. Gen. Genet.* 241: 627–636.
- Long, D., Goodrich, J., Wilson, K., Sundberg, E., Martin, M., Puangsomlee, P. and Coupland, G. 1997. *Ds* elements on all five *Arabidopsis* chromosomes and assessment of their utility for transposon tagging. *Plant J.* 11: 145–148.
- Lu, C. and Fedoroff, N. 2000. HYL1, a dsRNA-binding nuclear regulatory protein in plant hormone signaling. *Plant Cell* 12: 1–15.
- Machida, D., Onouchi, H., Koizumi, J., Hamada, S., Semiarti, E., Torikai, S. and Machida, Y. 1997. Characterization of the transposition pattern of the *Ac* element in *Arabidopsis thaliana* using endonuclease I-SceI. *Proc. Natl. Acad. Sci. USA* 94: 8675–8680.
- Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K.A., Dangl, J.L. and Dietrich, R.A. 2000. The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nat. Genet.* 26: 403–410.
- Mathur, J., Szabados, L., Schaefer, S., Grunenberg, B., Lossow, A., Jonas-Straube, E., Schell, J.A.Z. and Koncz-Kalman, Z. 1998. Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension. *Plant J.* 13: 707–716.
- Mayer, K., Schuller, C., Wambutt, R. *et al.* 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402: 769–777.
- Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Cramer, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B. and Koncz, C. 1991. T-DNA integration: a mode of illegitimate recombination in plants. *EMBO J.* 10: 697–704.
- Meissner, R.C., Jin, H., Cominelli, E. *et al.* 1999. Function search in a large transcription factor gene family in *Arabidopsis*: assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. *Plant Cell* 11: 1827–1840.
- Moreno, M.A., Chen, J., Greenblatt, I. and Dellaporta, S.L. 1992. Reconstitutive mutagenesis of the maize *P* gene by short-range *Ac* transpositions. *Genetics* 131: 939–956.
- Nacry, P., Camilleri, C., Courtial, B., Caboche, M. and Bouchez, D. 1998. Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* 149: 641–659.
- Negrak, V., Eisner, G. and Lemieux, B. 1996. Addition-deletion mutations in transgenic *Arabidopsis thaliana* generated by the seed co-cultivation method. *Genome* 39: 1117–1122.
- Ohba, T., Yoshioka, Y., Machida, C. and Machida, Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target. *Plant J.* 7: 157–164.
- Osborne, B.I., Wirtz, U. and Baker, B. 1995. A system for insertional mutagenesis and chromosomal rearrangement using the *Ds* transposon and Cre-lox. *Plant J.* 7: 687–701.
- Papi, M., Sabatini, S., Bouchez, D., Camilleri, C., Costantino, P. and Vittorioso, P. 2000. Identification and disruption of an *Arabidopsis* zinc finger gene controlling seed germination. *Genes Dev.* 14: 28–33.
- Parinov, S., Barsky, V., Yershov, G., Kirillov, E., Timofeev, E., Belgovskiy, A. and Mirzabekov, A. 1996. DNA sequencing by hybridization to microchip octa- and decanucleotides extended by stacked pentanucleotides. *Nucl. Acids Res* 24: 2998–3004.
- Parinov, S., Sevugan, M., De, Y., Yang, W.C., Kumaran, M. and Sundaresan, V. 1999. Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell* 11: 2263–2270.
- Peterson, T. 1990. Intragenic transposition of *Ac* generates a new allele of the maize *P* gene. *Genetics* 126: 469–476.
- Pruitt, R.E. and Meyerowitz, E.M. 1986. Characterization of the genome of *Arabidopsis thaliana*. *J. Mol. Biol.* 187: 169–183.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y. and Bennetzen, J.L. 1998. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20: 43–45.
- Smith, D., Yanai, Y., Liu, Y.G., Ishiguro, S., Okada, K., Shibata, D., Whittier, R.F. and Fedoroff, N.V. 1996. Characterization and mapping of *Ds*-GUS T-DNA lines for targeted insertional mutagenesis. *Plant J.* 10: 721–732.
- Speelman, E., Metz, P.L., van Arkel, G., te Lintel Hekkert, B., Stiekema, W.J. and Pereira, A. 1999. A two-component *Enhancer-Inhibitor* transposon mutagenesis system for functional analysis of the *Arabidopsis* genome. *Plant Cell* 11: 1853–1866.
- Stewart, C.N. Jr. and Via, L.E. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* 14: 748–750.
- Sundaresan, V., Springer, P., Volpe, T., Howard, S., Jones, J.D., Dean, C., Ma, H. and Martienssen, R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* 9: 1797–1810.
- Swaminathan, K., Yang, Y., Grotz, N., Campisi, L. and Jack, T. 2000. An enhancer trap line associated with a D-class cyclin gene in *Arabidopsis*. *Plant Physiol.* 124: 1658–1667.
- Takano, M., Egawa, H., Ikeda, J.E. and Wakasa, K. 1997. The structures of integration sites in transgenic rice. *Plant J.* 11: 353–361.

- Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G. and Jones, J.D. 1999. Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* 11: 1841–1852.
- Trifiro, M.A., Lumbroso, R., Beitel, L.K., Vasilou, D.M., Bouchard, J., Deal, C., van Vliet, G. and Pinsky, L. 1997. Altered mRNA expression due to insertion or substitution of thymine at position +3 of two splice-donor sites in the androgen receptor gene. *Eur. J. Hum. Genet.* 5: 50–58.
- Tsugeki, R. and Fedoroff, N.V. 1999. Genetic ablation of root cap cells in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 96: 12941–12946.
- Tsugeki, R., Kochieva, E.Z. and Fedoroff, N.V. 1996. A transposon insertion in the *Arabidopsis SSR16* gene causes an embryo-defective lethal mutation. *Plant J.* 10: 479–489.
- Van Lijsebettens, M., Inzé, D., Schell, J., Van Montagu, M. 1986. Transformed cell clones as a tool to study T-DNA integration mediated by *Agrobacterium tumefaciens*. *J. Mol. Biol.* 188: 129–145.
- Wilton, S.D., Lim, L., Dye, D. and Laing, N. 1997. Bandstab: a PCR-based alternative to cloning PCR products. *Biotechniques* 22: 642–645.
- Winkler, R.G., Frank, M.R., Galbraith, D.W., Feyereisen, R. and Feldmann, K.A. 1998. Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. *Plant Physiol.* 118: 743–750.
- Wisman, E., Hartmann, U., Sagasser, M., Baumann, E., Palme, K., Hahlbrock, K., Saedler, H. and Weisshaar, B. 1998. Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc. Natl. Acad. Sci. USA* 95: 12432–12437.
- Zhang, J.Z. and Somerville, C.R. 1997. Suspensor-derived polyembryony caused by altered expression of valyl-tRNA synthetase in the *twn2* mutant of *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 94: 7349–7355.
- Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G. and Taylor, D.C. 1999. The *Arabidopsis thaliana TAG1* mutant has a mutation in a diacylglycerol acyltransferase gene. *Plant J.* 19: 645–653.