Multiple Independent Defective Suppressormutator Transposon Insertions in Arabidopsis: A Tool for Functional Genomics

Alain F. Tissier,a1 Sylvestre Marillonnet,a Victor Klimyuk,a Kanu Patel,a Miguel Angel Torres,a George Murphy,b and Jonathan D. G. Jonesb2

a Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom
b John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom

A new system for insertional mutagenesis based on the maize Enhancer/Suppressor-mutator (En/Spm) element was introduced into Arabidopsis. A single T-DNA construct carried a nonautonomous defective Spm (dSpm) element with a phosphinothricin herbicide resistance (BAR) gene, a transposase expression cassette, and a counterselectable gene. This construct was used to select for stable dSpm transpositions. Treatments for both positive (BAR) and negative selection markers were applicable to soil-grown plants, allowing the recovery of new transpositions on a large scale. To date, a total of 48,000 lines in pools of 50 have been recovered, of which ~80% result from independent insertion events. DNA extracted from these pools was used in reverse genetic screens, either by polymerase chain reaction (PCR) using primers from the transposon and the targeted gene or by the display of insertions whereby inverse PCR products of insertions from the DNA pools are spotted on a membrane that is then hybridized with the probe of interest. By sequencing PCR-amplified fragments adjacent to insertion sites, we established a sequenced insertion-site database of 1200 sequences. This database permitted a comparison of the chromosomal distribution of transpositions from various T-DNA locations.

INTRODUCTION

The availability of newly compiled databases that store complete genome sequences and expressed sequence tags confronts molecular biologists with the problem of understanding the function of thousands of previously unknown genes. Computer studies can assign putative functions by relatedness to known genes, but only the biological characterization of mutants can provide concrete answers (Miklos and Rubin, 1996). The process of identifying a gene’s function subsequent to establishing its DNA sequence is known as reverse genetics, for which a number of methods have been explored. In plants, these methods include homologous recombination, gene silencing, and insertional mutagenesis. A recent study (Kempin et al., 1997) shows further that gene replacement in Arabidopsis is feasible. However, the requirement for hundreds or even thousands of transformants makes the approach unsuitable for reverse genetics on a large scale. Gene silencing, although widely used to gain insight into the function of plant genes (Baulcombe, 1996), similarly imposes the need for significant numbers of transgenic lines. In addition, the resulting genotypes are dominant, thereby preventing the study of essential genes (Martienssen, 1998). Also, gene silencing may affect entire gene families, making it difficult to pinpoint the role of individual members of the family. Finally, gene function may not be obliterated totally but rather only partially silenced.

Insertion mutagens, primarily transposons or T-DNA, are well suited for large-scale reverse genetics because the insertions can be easily detected by polymerase chain reaction (PCR). Recent progress in the efficiency of Agrobacterium transformation in Arabidopsis has allowed the production of thousands of transgenic lines (Feldmann, 1991; Bechtold et al., 1993). These lines have in turn been assembled into libraries, and reverse genetic screens have been performed successfully (McKinney et al., 1995; Krysan et al., 1996; Winkler et al., 1998). T-DNA insertions, however, are often complex, characterized by multiple inverted or tandem copies or truncated T-DNA, making molecular analysis difficult (Gheysen et al., 1990; Nacry et al., 1998).

Transposons offer several advantages in insertional mutagenesis as compared with T-DNA. For example, they can
be remobilized germinally, thereby producing revertants that can confirm the phenotypic consequences of the insertion, or somatically, thereby permitting mosaic analysis. Plant transposons additionally show a marked preference for insertion into genetically linked sites (Greenblatt and Brink, 1962; Jones et al., 1990; Bancroft and Dean, 1993; James et al., 1995). This property can be used to isolate new mutant alleles or to perform local mutagenesis in a particular region of interest (Das and Martienssen, 1995). Moreover, multiple insertions can be readily generated by highly active endogenous elements, such as the Mutator system in maize (Chandler and Hardeman, 1992; Das and Martienssen, 1995), the Tam elements in Antirrhinum (Saedler et al., 1984; Hehl et al., 1991; Nacken et al., 1991), or dTph in petunia (Koes et al., 1995). Indeed, each plant that undergoes mutagenesis by means of Mutator may contain up to 200 insertions, so that a few thousand plants will ensure complete coverage of the genome.

Of particular interest are libraries generated by introducing the maize Enhancer/Suppressor-mutator (En/Spm) system into Arabidopsis (Aarts et al., 1995; Wisman et al., 1998; Speulman et al., 1999, in this issue). These multicopy libraries are compact, compiled from a relatively small number of lines, but the somatic insertional activity of the element may lead to the detection of insertions that are not transmitted to the germ line. Moreover, the presence of multiple inserts may render the genetic and molecular analysis of the mutant difficult. In contrast, a system for the recovery of stable single-copy insertions circumvents these problems but again requires larger numbers of plants to ensure genome-wide coverage. In such a system, based on the maize Activator/Dissociation (Ac/Ds) element, Arabidopsis lines carrying a gene trap or enhancer trap Ds element on a T-DNA are crossed to lines expressing the transposase on another T-DNA (Sundaresan et al., 1995). Stable unlinked insertion events are recovered by a combination of positive selection for kanamycin resistance on the Ds element and negative selection against both T-DNAs.

Here, we describe the use of a similar system of positive/negative selection for transposition in Arabidopsis based on the En/Spm element. Unlike the system of Sundaresan et al. (1995), ours is contained within a single T-DNA, and the positive and negative selections can be applied to soil-grown plants. These characteristics allowed us to rapidly assemble a large collection of inserts to be used for reverse and forward genetic screens.

RESULTS

Constructs and Selection of Unlinked Transposition Events

The constructs introduced into Arabidopsis ecotype Columbia via Agrobacterium vacuum infiltration are shown in Figure 1. The defective Spm (dSpm) element carries a phosphinothricin (PPR) resistance gene (BAR) used to select for T-DNA integration and for transposon reinsertion. A terminator region of the nopaline synthase (nos) gene at the 3’ end of the BAR gene occurs in the opposite orientation to the termination region of the Spm element. Thus, insertions in an intron of a mutagenized gene should abolish gene function. dSpm lies between the cauliflower mosaic virus 35S promoter and the ATG codon of a β-glucuronidase (GUS) gene, which acts as an excision marker to monitor the activity of the element. The transposase gene fragment, coding for both the TnpA and TnpD proteins, was cloned under the control of a 35S promoter, the Spm promoter, or a meiosis-specific promoter from the AtDMC1 gene (constructs 8313, 8337, and 8353, respectively; see Methods; see also Klimyuk and Jones, 1997). The SU1 gene from Streptomyces griseolus encodes a cytochrome P450 that confers sensitivity to R7402, which is a sulfonylurea proherbicide from Du Pont (O’Keefe et al., 1994). The counterselection cassette with the SU1 gene was cloned between the left border and the transposase gene in the same orientation. The selectable genes were chosen for their ability to be screened on soil, thus avoiding the need to prepare sterile media and seeds.

The principle for recovery of unlinked transposition events is represented in Figure 2. Seeds from a plant heterozygous for the T-DNA are collected and allowed to germinate on soil. The plantlets are then treated with PPT (100 mg/L) and R7402 (100 μg/L) several days after germination. If before or at meiosis there is transposition of the dSpm element to an unlinked site, on either the same or a different chromosome, then the T-DNA and the newly transposed element may segregate in the progeny. In this case, the plants that carry the dSpm and lack the T-DNA will be resistant to both PPT and R7402. On the other hand, if the dSpm transposes to a linked site, double-resistant (DR) plants can be recovered only if there is recombination between the T-DNA and dSpm. DR individuals could also result from inactivation of the SU1 gene, either by insertion of the dSpm or by silencing. We observed that the majority of plants did not survive...
Selection of Active Lines for Large-Scale Production of Insertions

Multiple transformants arising from the three different constructs were recovered, and their progeny were tested in vitro for segregation of the BAR and SU1 markers. Because the presence of multiple unlinked T-DNA inserts would greatly reduce the probability of recovering DR plants, we screened for transformants with T-DNA inserts at a single locus; two out of three transformants were in this class (data not shown). The progeny of these primary transformants were then analyzed by different means to evaluate the activity of the constructs in these lines. In a first assay, germinated seedlings were stained for GUS activity of the constructs in these lines. In a first assay, germinal excision events were identified, and the complete T-DNA should remain. This assay led us to eliminate lines such as 8337 plant 4 in which 28 of 30 DR plants stained positive for GUS (Table 2). On the other hand, no GUS-staining individuals were found in DR plants from other lines, such as 8313 plant 1. Finally, the independence of the recovered insertions was checked by DNA gel blot analysis. DNA from DR plants was digested by two restriction enzymes separately and hybridized with a dSpm probe to reveal fragments flanking the insertions (data not shown). This treatment confirmed the GUS staining analysis (see Table 2).

These preliminary results allowed us to select the lines subsequently retained for large-scale production of transposition events. The general trend is that lines with relatively low frequencies of double resistance, such as 8337 plant 2, 8337 plant 9, 8313 plant 4, and 8353 plant 4, give rise to

<table>
<thead>
<tr>
<th>Transformant</th>
<th>Total Seed</th>
<th>GUS+</th>
<th>%GE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8337 No. 1</td>
<td>1160</td>
<td>55</td>
<td>4.7</td>
</tr>
<tr>
<td>8337 No. 3</td>
<td>1120</td>
<td>0</td>
<td>&lt;0.0009</td>
</tr>
<tr>
<td>8337 No. 4</td>
<td>1200</td>
<td>18</td>
<td>1.5</td>
</tr>
<tr>
<td>8337 No. 5</td>
<td>1160</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>8337 No. 6</td>
<td>1080</td>
<td>16</td>
<td>1.5</td>
</tr>
<tr>
<td>8337 No. 7</td>
<td>1200</td>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>8337 No. 8</td>
<td>1080</td>
<td>0</td>
<td>&lt;0.0006</td>
</tr>
<tr>
<td>8337 No. 9</td>
<td>1080</td>
<td>7</td>
<td>0.6</td>
</tr>
<tr>
<td>8313 No. 1</td>
<td>1320</td>
<td>24</td>
<td>1.8</td>
</tr>
<tr>
<td>8313 No. 2</td>
<td>1240</td>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>8313 No. 4</td>
<td>1240</td>
<td>10</td>
<td>0.8</td>
</tr>
<tr>
<td>8353 No. 2</td>
<td>1280</td>
<td>0</td>
<td>&lt;0.0008</td>
</tr>
<tr>
<td>8353 No. 4</td>
<td>1440</td>
<td>5</td>
<td>0.3</td>
</tr>
<tr>
<td>8353 No. 5</td>
<td>1560</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8353 No. 7</td>
<td>1200</td>
<td>31</td>
<td>2.6</td>
</tr>
</tbody>
</table>

*Number of completely blue seedlings after GUS staining.

a Percentage of germinal excision events as estimated from the GUS staining data.
more independent events, whereas lines such as 8337 plant 1 or 8337 plant 4 gave a relatively high frequency of DR plants, few of which were independent. Line 8313 plant 1 appeared to be a good compromise, with 66% independent insertions (eight of 12).

**Strategy for Large-Scale Mutagenesis**

Unlinked transposition events can be recovered only in the progeny (called F1) arising from the self-pollination of plants that are heterozygous for the T-DNA locus. Such parental heterozygotes can be generated by crossing the transgenic line to a wild-type plant and testing the progeny of individual plants for segregation of the T-DNA. The parental heterozygotes can also be selected if the T-DNA insertion in the homozygous state produces a phenotype that can be easily scored and thus eliminated. This was the case for transformant 8337 plant 6, which has a flowering defect linked to the T-DNA insertion. However, for most lines there is no phenotype, and production of pure heterozygous lines necessitates crossing or progeny testing. We first used outcrossing; however, this is labor intensive, and so we devised the strategy shown in Figure 3.

The strategy presented in Figure 3 relies on the fact that the progeny of homozygous plants should not contain individuals resistant to the double selection of R7402 and PPT. The seeds that will be subjected to this double selection can then be harvested not from heterozygous plants (F1 plants) but from their PPT-resistant progeny (F2). The ratio of heterozygotes to homozygotes from this F2 population will thus be one to three. Although one out of three F3 seeds therefore should not contribute any transposition events, this strategy allows rapid production of massive numbers of selectable seeds. One heterozygous plant can produce up to 20,000 seeds, of which 10,000 are heterozygous for the T-DNA. When these are grown with PPT-resistant homozygous individuals, they can produce 1000 to 10,000 seeds each (depending on the density of the plants). Three hundred plants per tray (21 × 35 cm2) will produce a total of 300,000 seeds, that is, 1000 per plant, which are harvested from sectors of the tray. Under those conditions, within two generations, a total of 10^7 selectable seeds can be generated from a single heterozygous plant. Transposition events that occurred in the F1 generation will be transmitted and may give rise to large numbers of DR plants in the F3 progeny. Therefore, only one or two plants are collected from each half tray (20,000 seeds) in which >50 DR F3 plants are present. In addition, plants expressing the SU1 gene in the

---

**Table 2. Frequency of DR Plants and Independent Transposition Events in the Progeny of Primary Transformants**

<table>
<thead>
<tr>
<th>Transformant</th>
<th>Total Seed</th>
<th>DR Frequency</th>
<th>GUS +</th>
<th>DNA Gel Blot Analysis</th>
<th>% Independent/DR</th>
<th>% Independent/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>8337 No. 1</td>
<td>16,000</td>
<td>28</td>
<td>1.75 × 10^{-3}</td>
<td>2</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>8337 No. 2</td>
<td>4,000</td>
<td>1</td>
<td>2.5 × 10^{-4}</td>
<td>0</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>8337 No. 4</td>
<td>16,000</td>
<td>30</td>
<td>1.88 × 10^{-3}</td>
<td>28</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8337 No. 6</td>
<td>17,600</td>
<td>6</td>
<td>3.4 × 10^{-4}</td>
<td>4</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>8337 No. 8</td>
<td>17,600</td>
<td>1</td>
<td>6.10^{-5}</td>
<td>0</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>8313 No. 1</td>
<td>21,600</td>
<td>12</td>
<td>5.5 × 10^{-4}</td>
<td>0</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>8313 No. 4</td>
<td>19,600</td>
<td>10</td>
<td>5 × 10^{-5}</td>
<td>1</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>8353 No. 2</td>
<td>19,200</td>
<td>5</td>
<td>2.6 × 10^{-4}</td>
<td>2</td>
<td>40.0</td>
<td></td>
</tr>
<tr>
<td>8353 No. 3</td>
<td>8,000</td>
<td>1</td>
<td>1.2 × 10^{-4}</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8353 No. 4</td>
<td>12,400</td>
<td>2</td>
<td>1.6 × 10^{-4}</td>
<td>NA</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>8353 No. 5</td>
<td>24,000</td>
<td>1</td>
<td>4 × 10^{-5}</td>
<td>NA</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

*Number of GUS-staining plants.

b Number of independent insertions as estimated by DNA gel blot analysis.

*NA, not applicable.

---

**Figure 3.** Strategy for Large-Scale Production of DR Individuals.

The numbers at right represent the number of plants at the corresponding stage of the selection procedure. From the F2 generation, the numbers of heterozygous and homozygous plants are given separately, although the plants were actually grown together.
absence of selection exhibit a slightly different growth habit than do untransformed plants. Specifically, they are shorter, bushier, and darker green (data not shown). In a PPT-resistant F2 population, the wild-type individuals, which may carry unlinked transposition events, were removed to prevent contamination of the F3 seeds.

Pooling Strategy and Organization of the Transposant Library

Seeds and DNA from plants that survived the double selection were harvested in pools of ~50 individuals. These pools of 50 were then assigned to a superpool of 48 pools. Each pool can thus be defined according to a two-coordinate address (x,y; e.g., 01.01), where x is the superpool number and y the pool number. To date, 20 superpools have been collected, which represent a total number of ~48,000 DR individuals in 960 pools.

Estimation of the Frequency of Independent Events by Adapter PCR

To estimate the number of independent transposition events in the collection, we performed adapter PCR on a representative sample of the pools (see Methods). The principle of adapter PCR was described previously (Lagerstrom et al., 1991). The amplified fragments are displayed on a denaturing polyacrylamide gel, as shown in Figure 4. The number of fragments displayed is inevitably less than the number of independent events (Cavrois et al., 1995) because very long or short fragments will be less easily visualized. To alleviate this problem, we used two enzymes to obtain a better representation of the real number of inserts. The results of this survey are summarized in Table 3. The percentage of independent events per pool determined by this assay varied from 30 to 40%, depending on the primary transformant.

Sequencing of Plant DNA Flanking dSpm Insertions

Fragments displayed on the adapter PCR gels can be extracted, reamplified, and directly sequenced with a Spm primer, which allowed us to obtain the sequence of 1200 insertion sites. The ensemble of these sequences constitutes the Sequenced Insertion Sites (SINS) database. The complete database is available at http://www.jic.bbsrc.ac.uk/sainsbury-lab/jonathan-jones/SINS-database/sins.htm and can be browsed in annotated form, subjected to keyword search, and also obtained for local searches. Analysis of BlastN searches (Altschul et al., 1997) indicates that 50% of the sequences are identical to known Arabidopsis genomic sequences, which is to be expected in view of the current status of Arabidopsis genomic sequencing. Of these, 70% are in coding regions, the remainder being in either intergenic regions or repeated sequences. Less than 1% of the sequences correspond to the counterselectable marker or the 35S:GUS cassette. One of the insertion sequences is identical to the APETALA 3 (AP3) gene, and plants from the corresponding pool were screened for the phenotype of the ap3 mutant. ap3 plants were indeed found, confirming the existence of the insertion and its transmission to the progeny (data not shown).

Insertion sequences matching published Arabidopsis genomic sequences were placed on the physical map, as summarized in Figure 5. Two T-DNA inserts (8337 plant 6 and 8313 plant 1) that contributed significantly to the selection of transpositions were physically mapped. Such physical mapping was performed by amplifying flanking DNA by thermal asymmetric interlaced (TAIL) PCR (Liu et al., 1995) and hybridizing the Texas A & M University Bacterial Artificial Chromosome Center library (Choi et al., 1995) filters with the products. Both T-DNA inserts map to chromosome 1. The apparently uneven distribution of transpositions in Figure 5 is a reflection of the current status of the Arabidopsis genome

Figure 4. An Example of Adapter PCR Performed with Various DNA Pools.

Each band represents a different insertion and allows for a minimal estimate of the number of independent insertions in the pool. The intensity of the bands may reflect the relative abundance of the insertion within the pool (i.e., some events are clonal) and/or the relative efficiency of the amplification. These bands can then be excised and the DNA reamplified for direct sequencing. Each lane is the result of adapter PCR with a different pool. The pools used are (03.01) to (03.28). Molecular length markers are given at right in base pairs.
sequencing program. Thus, the largest number of insertions is found on chromosome 2, which is almost completely sequenced, whereas the smallest number corresponds to chromosome 3, for which there is comparatively little sequence available. Some sequence homologies are remarkable. For example, some insertions are into mitochondrial and chloroplast DNA sequences. This must reflect insertions into regions of plastid DNA that have become incorporated into the Arabidopsis chromosomal DNA. There are also insertions in various repetitive DNA sequences, including 25S and 18S rDNA, 5S rRNA intergenic regions, retrotransposons, and minisatellites.

Reverse Genetics: PCR Screens

Insertions into genes of known sequence can be detected by PCR with a combination of primers on the transposon and on the gene of interest. This approach was successfully used with DNA from T-DNA-mutagenized populations (McKinney et al., 1995; Krysan et al., 1996; Winkler et al., 1998), and similar methods were applied here. Briefly, PCR is performed on the superpools, with primers at the 5' and 3' ends of the gene and of Spm. Each combination is done separately. The presence of specific amplification products is verified by hybridization to a probe spanning the gene or a part thereof. Nested primers increase the specificity of the amplifications (see Meissner et al., 1999, in this issue). If a superpool gives a positive signal, PCR with the same primers is then repeated on an 8 × 6 matrix of the 48 individual pools from that superpool. The presence of a specific PCR product in one of the columns and one of the rows gives the address of the pool containing the insertion. To isolate single plants carrying the insertion, further screening of the progeny of the corresponding pool is effected. Because 50 plants most likely heterozygous for the dSpm are contained

### Table 3. Percentage of Independent Transpositions as Estimated by Adapter PCR

<table>
<thead>
<tr>
<th>Transformant</th>
<th>DRa</th>
<th>Poolsb</th>
<th>Mslec</th>
<th>Bfalc</th>
<th>Msel/DRd</th>
<th>Bfal/DRd</th>
<th>Max (Mse, Bfa)e</th>
<th>Max/DRf</th>
</tr>
</thead>
<tbody>
<tr>
<td>8313 No. 1</td>
<td>1237</td>
<td>25</td>
<td>323</td>
<td>329</td>
<td>26.1%</td>
<td>26.6%</td>
<td>384</td>
<td>31.0%</td>
</tr>
<tr>
<td>8313 No. 1</td>
<td>1974</td>
<td>40</td>
<td>NDa</td>
<td>611</td>
<td>ND</td>
<td>31.0%</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>8313 No. 4</td>
<td>274</td>
<td>5</td>
<td>ND</td>
<td>60</td>
<td>41.7%</td>
<td>21.7%</td>
<td>115</td>
<td>41.7%</td>
</tr>
<tr>
<td>8313 No. 4</td>
<td>421</td>
<td>8</td>
<td>ND</td>
<td>106</td>
<td>25.2%</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Mixed</td>
<td>1058</td>
<td>20</td>
<td>313</td>
<td>245</td>
<td>29.6%</td>
<td>23.2%</td>
<td>342</td>
<td>32.3%</td>
</tr>
<tr>
<td>Mixed</td>
<td>1720</td>
<td>33</td>
<td>ND</td>
<td>434</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>8337 No. 1</td>
<td>1236</td>
<td>3</td>
<td>48</td>
<td>29</td>
<td>35.3%</td>
<td>21.3%</td>
<td>48</td>
<td>35.3%</td>
</tr>
<tr>
<td>8337 No. 2</td>
<td>191</td>
<td>4</td>
<td>69</td>
<td>52</td>
<td>36.1%</td>
<td>27.2%</td>
<td>74</td>
<td>38.7%</td>
</tr>
<tr>
<td>8337 No. 5</td>
<td>50</td>
<td>1</td>
<td>13</td>
<td>10</td>
<td>26.0%</td>
<td>20.0%</td>
<td>13</td>
<td>26.0%</td>
</tr>
<tr>
<td>8337 No. 9</td>
<td>150</td>
<td>3</td>
<td>43</td>
<td>ND</td>
<td>28.7%</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>8353 No. 5</td>
<td>100</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>4%</td>
<td>5%</td>
<td>5</td>
<td>5%</td>
</tr>
</tbody>
</table>

*a Number of DR plants analyzed.

*b Number of pools of 50 DR plants analyzed.

*c Total number of bands from these pools observed with either Msel or Bfal digests.

*d Percentage of independent transposition events calculated as the ratio (number of bands with either Msel or Bfal)/(number of DR plants).

*e As given for footnote c, except that for each pool of 50 DR plants, the highest value of either the Msel or the Bfal digests was taken.

*f As given for footnote d, except the number of bands used to calculate the ratio is the corresponding number from the column Max (Mse, Bfa).

*g ND, not determined.

---

Figure 5. Distribution of Sequenced Insertion Sites and T-DNAs on the Arabidopsis Chromosomes.

Black and white triangles represent insertions from line 8313 plant 1 and mixed lines, respectively. The two mapped T-DNA inserts (8313 plant 1 and 8337 plant 6) appear as hatched triangles. cM, centimorgans.
in a pool, at least 200 progeny plants should be analyzed. In practice, isolation of single plants carrying the insertion is done on populations of 200 to 500 individuals. Again, DNA is collected in pools to minimize the number of PCR amplifications. One of the first screens was done with the six Arabidopsis Atboh (for A. thaliana respiratory burst oxidase homologs) genes, which are homologs of mammalian NADPH oxidase (Torres et al., 1998) and hypothesized to play a role in the generation of the oxidative burst during plant-pathogen interactions. An example of a successful screen for the AtbohE gene is shown in Figures 6A and 6B.

Reverse Screening by Inverse Display of Insertions

PCR screens for insertions consume significant quantities of time and DNA. Therefore, we devised a new method, named inverse display of insertion (IDI), that relies on inverse PCR (IPCR). IPCR is a widely employed technique to clone unknown DNA fragments adjacent to known sequences (Ochman et al., 1988). Successful IPCR depends on choosing the appropriate enzyme to digest the DNA. If the restriction site is very close to the inserted DNA, the amplification product will not be very informative. On the other hand, if the restriction site is too far from the inserted DNA, amplification will not compete with that of smaller fragments.

IPCR was performed with all of the DNA pools for both ends of the dSpm element and with a combination of three enzymes, namely, HindIII, ApoI, and BstYI, which greatly increases the probability of amplifying the flanking DNA for each insertion present in the pool. The six IPCRs for each pool were then mixed and spotted in a gridded array onto a nylon membrane. Hybridization of this membrane with a labeled genomic DNA fragment in which insertions are sought directly indicates which pools, if any, contain an insertion. PCR can then be performed on the positive pools to confirm the presence and position of the insertion, and isolation of single plants carrying the insertion can be performed as described above.

Validation of the IDI method was provided in that all PCR “hits” were also detected by the IDI method (data not shown). An example of a new screen for insertions in AtbohC is shown in Figures 7A and 7B. Subsequent PCR analysis showed that these signals corresponded to two different insertions (data not shown), both of which were subsequently identified in single plants. This approach is now routinely used for reverse genetic screens of our collection and has the advantage of portability; filters can be mailed to interested laboratories. Insertions in all six Atboh genes have been obtained (M.A. Torres, data not shown) and are now being analyzed through intercrossing and self-pollination for phenotypic evaluation. Multiple probes can be used on each filter, and the identity of any particular insertion can be characterized subsequently from pools of 50. Multigene families could in principle cause problems, because cross-hybridization on the filter could be due to a member of the family that will not be amplified by primers

Figure 6. Example of a PCR Screen for Insertions in the AtbohE Gene.

(A) Primary screen. A putative insertion was found in superpool 02 (lane 1) using DNA gel blots (not shown) and confirmed by nested PCR with internal primers (lane N). The authenticity of the amplified product was checked by hybridization with an AtbohE probe (bottom). Simultaneously, the individual pool containing the insertion was identified by screening an 8 × 6 matrix of superpool 02. In this case, the number of the positive pool is 02.43. Lane M shows 1-kb ladder molecular weight markers.

(B) Screen for individual plants carrying the insertion. Seeds from pool 02.43 were germinated and DNA extracted from the plants. PCR was performed with pools of diminishing size (50, 20, and 5) and finally with individual plants.
designed for the specific member of interest. However, the IPCR can be repeated, fragments cloned, colonies hybridized, and inserts sequenced to establish which multigene family member was tagged. We now have several examples of the validity of this approach (S. Marillonnet and K. Patel, data not shown).

Estimation of the Total Number of Independent Insertions Based on the Reverse Genetics Screens

The results of the reverse genetics screens, which incorporate some data from Meissner et al. (1999, in this issue), are as follows. A total of 25 insertions were found for 41 genes searched. The total length of the DNA searched was ~120 kb. This implies an average of one insertion every 4.8 kb, which is consistent with a total of 25,000 inserts, assuming a 120-Mb genome. These screens were conducted on 12 superpools (~29,000 plants). Thus, one can conclude that ~80% of the plants carry independent insertions. This value is higher than the estimate derived from the adapter PCR method (30 to 40%), indicating that the number of insertions detected by the latter method is underestimated (Cavrois et al., 1995).

DISCUSSION

We have introduced a new transposon-tagging system in Arabidopsis that is designed for the rapid collection of stable independent insertions over the whole genome. Because plant DNA transposons preferentially move to linked loci, genome-wide coverage can be attained by two distinct approaches. Specifically, one can either introduce a sufficient number of transposon-carrying T-DNAs throughout the genome to be used as “launching sites” for local mutagenesis or select for many unlinked transposition events. The first method requires the mapping of the T-DNAs, before using them for mutagenesis, to ensure that the entire genome is covered. In addition, to recover stable insertions, the transposase source has to be located on a separate T-DNA that can segregate away. These demands are placed on the system developed by Sundaresan et al. (1995), for example. This particular system was designed to recover unlinked transposition events with a counterselectable marker, namely, the indole acetic acid hydrolase (iaaH) gene, so as to eliminate both T-DNAs and a positive selection on the disarmed transposon. Because both T-DNAs carry the counterselectable marker, only one of 16 progeny of the crosses between the two lines is resistant to the counterselection. Therefore, we chose to employ a system using a single T-DNA with a positive selection for the disarmed transposon and counterselection for the T-DNA. In this case, one of four progeny of a plant heterozygous for the T-DNA is resistant to the counterselection.
We chose the BAR and SU1 genes as positive and negative selectable markers, respectively, because they allow for the use of soil-grown plants, thereby eliminating the need to sterilize seeds and media. In addition, both selections are very efficient (see Figure 2), and up to 20,000 seeds can be selected per half tray (21 x 16 cm²).

Finally, the En/Spm element was chosen, rather than Ac/Ds, because the frequencies of excision and independent transposition events are reportedly high in Arabidopsis (Aarts et al., 1995). However, compared with the study of Aarts et al. (1995), germlinal excision (10⁻²) and unlinked transposition (2.5 to 10 x 10⁻⁴) frequencies in our system are lower than expected. We can only speculate as to reasons for the low rates of excision/transposition that we observe. One possibility is that the presence of the BAR gene and the bacterial spectinomycin resistance gene on the dSpm interferes with transposition. Indeed, actively transcribed genes on a dSpm element can reduce the frequency of excision (Cardon et al., 1993). Assuming that all excision events give rise to insertions, the ratio between germlinal excision and unlinked transposition frequencies (1:20 to 1:50) suggests that 8 to 20% of transpositions are to unlinked loci.

Two methods were employed to evaluate the number of independent transpositions in our collection. The first method is based on the direct detection of inserts in pools of 50 by adapter PCR followed by polyacrylamide gel electrophoresis. This approach suggested that 30 to 40% of transpositions were independent events. The second method relied on the detection of insertions into genes of known sequence. Although indirect, this method provides a more reliable and significant estimate because it evaluates the probability of finding an insertion in a given gene. Based on a search for 41 genes, spanning a total length of 120 kb, we can thus estimate that 80% of the plants selected harbor independent insertions. Although these reverse screens were performed with 12 superpools, we now have 20 superpools (i.e., 48,000 plants) and ~38,000 independent inserts.

We also initiated systematic sequencing of insertion flanks from the products of adapter PCR products displayed on gels. Thus far, 1,200 sequences have been obtained, assembled into a database, and compared with existing sequences. Currently, nearly 50% match known Arabidopsis genomic sequences and can be positioned on the physical map (Figure 5). The distribution of these sequences is a good indicator of the random nature of the transposition events because insertions, 70% of which fell within putative or confirmed coding regions, were found on all chromosomes. In the future, with the advent of more extensive databases of this type, reverse genetics will be greatly facilitated. Indeed, it will be possible to search for insertions in genes of interest from the complete genome sequence by performing a sequence search in silico, thereby reducing the need to perform PCR or other types of molecular screens. Even with our limited data set of 1200 insertions, the approach has revealed insertions into the genes for protein kinases, transcription factors, cytochrome 450s, and many others for which the real function is unknown. To facilitate global function searches, the expansion of the SINS database is now a high priority.

METHODS

Plasmids

Binary Vector

T-DNA binary vector SLJ 491 (Jones et al., 1992) was digested with Clal and HpaI, and the large gel-purified fragment was ligated to the annealed and phosphorylated oligonucleotides AT3 (5' - AAGCTT-TATGAATTCTATGTTAACTATGGATCCTA-3') and AT4 (5' - CGTAGGATCATGTTAACATGAATCCATGTTAACATGAATCCATGTTACATGTA-3') to create SLJ 7693. The resulting T-DNA vector contains unique sites for BamHI, HpaI, EcoRI, and HindIII, in that order, from the left border to the right border.

Excision Marker

Plasmid SLJ 4K1 (Jones et al., 1992) was cut with BamHI, treated with the Klenow fragment of DNA polymerase I, and recircularized, thereby effectively removing the BamHI site to yield SLJ 8271. The EcoRI-HindIII fragment of SLJ 8271 was then inserted into SLJ 7693, having been cut with the same enzymes, to make SLJ 8281.

dSpm

A full-length Suppressor-mutator (Spm) fragment was cloned into the Clal site of pMUCClα, a low-copy-number derivative of pUC119 with a unique Clal site at the ATG site, to make SLJ 651. From SLJ 651, we then created a new defective Spm (dSpm) element, SLJ 7648, by deleting the internal fragment bordered by the most proximal and most distal PstI sites of the full-length Spm element. The BamHI site between BAR and octopine synthase (ocs) sequences of plasmid SLJ 0512 (that carries a nos:BAR:ocs 3' expression cassette) was removed by digestion followed by filling in with the Klenow fragment to give SLJ 7652. A fragment of pSP118 (Svab et al., 1990) obtained upon digestion by SstI, treatment with T4 polymerase, and subsequent HindIII digestion, thereby carrying the spectinomycin resistance marker, was cloned into SLJ 7652; the resulting plasmid was digested with PstI and then treated with T4 polymerase and HindIII to yield SLJ 7662. The dSpm element carrying the BAR gene and the spectinomycin resistance gene, SLJ 7674, was obtained by ligation of a BclI-HindIII fragment (obtained by partial digestion with BclI and treatment with the Klenow enzyme) of SLJ 7662 into the SLJ 7641 plasmid that had been cut by PstI, treated with BclI and polymerase, and then digested with HindIII. The resulting dSpm element was cloned as a Clal fragment into SLJ 8281 to yield SLJ 828'3.

Counterselectable Marker

A BglII adapter was inserted at the EcoRI site of plasmid AGS465 (O’Keefe et al., 1994) to generate plasmid SLJ 8241. A BglII-BamHI
fragment containing the counterselectable marker could then be excised and ligated into the BamHI site of SLJ 8283. Of the two possible orientations, the one with the BamHI site internal to the T-DNA was retained and designated plasmid SLJ 8291.

Promoters
The minimal cauliflower mosaic virus 35S promoter was recovered from plasmid SLJ 4D4 (Jones et al., 1992) as a BglII-XhoI fragment. The Spm promoter was obtained by amplification of 180 bp from full-length Spm with primers AT8 (5'-AAAGGTATCCGTCAAAGGAGTGTCAG-3') and AT9 (5'-AAACCTGAGCTCTGCGTGGCGAGG-3'), digested with BamHI and XhoI, and cloned in plasmid SLJ 1721. Site-directed mutagenesis was performed with the EcoRV fragment of the AtDMC1 gene (Kimjyk and J ones, 1997) to introduce a BglII site into the fragment. The AtDMC1 promoter was then excised as a BglII-XhoI fragment.

Transposase
The internal BssHII fragment of full-length Spm was treated with the Klenow fragment and ligated into pMUCBS digested with EcoRV. This yielded plasmid SLJ 7713. The transposase fragment could then be excised with SmaI (3' end) and XhoI (5' end).

Final Constructs
To obtain constructs SLJ 8313, SLJ 8337, and SLJ 8353, three-way ligations were performed with SLJ 8291 that had been digested with HpaI and BamHI (T-DNA vector with SU1 and dSpm); the large fragment of SLJ 7713 after treatment with SmaI and XhoI (to yield the fragment that encodes the transposase); and the cauliflower mosaic virus 35S, Spm, and AtDMC1 promoter fragments, respectively.

Plant Transformation
Plant transformation was performed by modification of the method of Bechtold et al. (1993). Arabidopsis thaliana plants were grown over long days in 50-cm² square pots until the first siliques were visible. The plants were then vacuum infiltrated by inverting the pots into the Agrobacterium tumefaciens suspension. Vacuum was applied for ~2 to 3 min until the suspension started boiling. The vacuum was then released. The inflorescences were rinsed with water, and the plants were covered with plastic and brought back to the greenhouse. The plastic bags were removed after 2 to 3 days. Seeds were collected after 3 weeks and germinated in soil. Transformants were selected by spraying phosphinothricin (PPT) at 100 mg/L 4 to 5 days after germination. Spraying was repeated two or three times.

Determination of the Zygosity of Primary Transformants
Approximately 100 seeds from each transformant were plated on Murashige and Skoog salts (Sigma)/0.8% Bacto agar (Difco) containing either 10 mg/L PPT or 10 μg/L R7402. Only the transformants segregating as a single locus for both markers (BAR and SU1) were retained for production of transposition events.

Selection of Transposition Events
Seeds from plants heterozygous for a T-DNA locus or from PPT-resistant progeny of heterozygous plants were sown on soil at a density of ~20,000 per half trays (21 × 16 cm²) and sprayed 4 to 5 days after germination with a solution containing 100 mg/L PPT and 100 μg/L R7402. Spraying was repeated every other day, for a total of three to five times, until the selection was clearly effective. DNA was extracted from pools of 50 plants, using one inflorescence per plant. Seeds were harvested from the same pools of 50 plants. Pools were assigned a number that identified their position in the 48 × 48 matrix (x,y; 1 ≤ x ≤ 48; 1 ≤ y ≤ 48), where x and y are the superpool and the pool numbers, respectively.

DNA Extraction, Adapter Polymerase Chain Reaction, and Sequencing of Plant DNA Flanking dSpm Insertions
DNA from inflorescences was extracted as described previously (Carroll et al., 1995). Genomic DNA (0.5 μg) was digested with either MseI or BfaI and ligated to the adapter made with the biotinylated dSpml primer (5'-TTATTTACGTAAGTGTTGGTTTGG-3') and either the MseI primer (5'-GACGATGAGTCCGAGA-3') or the BfaI primer (5'-GACGATGAGTCCGAGTAA-3'). Cycling conditions were as follows: 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, for 30 cycles. The products were then selected on streptavidin beads resuspended in 200 μL, of which 5 μL was used for nested PCR with the 32P end-labeled Spm3 primer (5'-ACCCGTGACTAGCTTCACTCTTGTAG-3') and the MseI or BfaI primers. The success- sion of cycling conditions was as follows: (1) 95°C for 3 min; (2) three cycles of 94°C for 30 sec, 70°C for 30 sec, and 72°C for 1 min; (3) a single cycle of 94°C for 30 sec, 65°C for 30 sec, and 72°C for 1 min; (4) 12 cycles of 94°C for 30 sec, 65°C (3.0°C per cycle) for 30 sec, and 72°C for 1 min; and (5) 23 cycles of 94°C for 30 sec, 56°C for 30 sec, and 72°C for 1 min. The reactions were then resolved on a 0.5% denaturing acrylamide gel in a Bio-Rad apparatus at constant wattage (100 W) for 1.5 hr. The gel was then vacuum dried and exposed to X-OMAT film (Kodak) overnight. For sequencing, bands were excised from the dried gel and soaked in 100 μL of Tris-EDTA buffer for >5 hr at 37°C. Five microtrols of the eluted DNA was then used to reamplify the fragment with MseI or BfaI and ~20 universal Spm3 primers. The PCR products were purified in batches of 96 by using the QIAquick 96 PCR purification kit (Qiagen, Chatsworth, CA). Five microtrols of the purified DNA was then used for sequencing with the ABI Dye terminator reaction kit (Perkin-Elmer) and 20 universal primer before analysis on an ABI377 automatic sequencer.

Mapping of T-DNA Insertion Sites
Fragments flanking the T-DNA inserts were amplified by thermal asymmetric interlaced (TAIL) PCR (Liu et al., 1995) with three rounds of amplification with the following nested T-DNA primers: L4 (5'-CGTGAAGTCTCTCATAAGCCCC-3'), L1 (5'-TTAGAAATTT-
ACKNOWLEDGMENTS

This work was funded by research grants to J.D.G.J. from the Biotechnology and Biological Sciences Research Council (No. G05848), the GAIT program (No. GAIT09113), the Gatsby Charitable Foundation, and the European Union Framework IV Arabidopsis Insential Mutagenesis (No. CT950183) program. A.F.T. was supported by an EMBO long-term postdoctoral fellowship. M.A.T. and S.M. were supported by a Training and Mobility of Researchers postdoctoral fellowship from the European Union. We thank J. Jonathan Darby, Richard Gould, Tim Glister, Pat Walker, Pat Theobold, and Doris Walker for help in the greenhouse and seed collection. We also thank Ruth Meissner and Hailing J. in help in tissue collection for DNA preparations and for providing unpublished data.

REFERENCES


