Target selected insertional mutagenesis on chromosome IV of Arabidopsis using the En–I transposon system

Elly Speulman, Ronald van Asperen, Jessica van der Laak, Willem J. Stiekema, Andy Pereira *

CPRO, Department of Molecular Biology, PO Box 16, 6700 AA Wageningen, The Netherlands

Received 9 February 1999; received in revised form 29 November 1999; accepted 3 December 1999

Abstract

Reverse genetics using insertional mutagenesis is an efficient experimental strategy for assessing gene functions. The maize Enhancer–Inhibitor (En–I) transposable element system was used to develop an effective reverse genetics strategy in Arabidopsis based on transposons. To generate insertion mutations in a specific chromosomal region we developed a strategy for local transposition mutagenesis. A small population of 960 plants, containing independent I transpositions was used to study local mutagenesis on chromosome IV of Arabidopsis. A total of 15 genes, located on chromosome IV, were tested for I insertions and included genes identified by the European ESSA I sequencing programme. These genes were of particular interest since homologies to other genes and gene families were identified, but their exact functions were unknown. Somatic insertions were identified for all genes tested in a few specific plants. Analysis of these progeny plants over several generations revealed that the ability to generate somatic insertions in the target gene were heritable. These genotypes that show high levels of somatic insertions can be used to identify germinal insertions in the progeny. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Reverse genetics; En-I; Transposon; Pectinesterase; Arabidopsis

1. Introduction

*Corresponding author. Tel.: +31-31-777114; fax: +31-31-7418094.
E-mail address: a.pereira@cpro.wag-ur.nl (A. Pereira)

Arabidopsis thaliana is used as a model in plant science to study many aspects of plant biology. Due to its small genome and its well-studied genetics, Arabidopsis was the first plant species chosen to generate a complete physical map of the genome (Choi et al., 1995; Liu et al., 1995; Schmidt et al., 1996, 1997; Zachgo et al., 1996; Camilleri et al., 1998; Kaneko et al., 1998; Mozo et al., 1998). This effort was followed by the multinational co-ordinated sequencing of the Arabidopsis genome using bacterial artificial chromosomes (BACs) and P1 clones which had been mapped to defined regions of the genome (Goodman et al., 1995; Bevan et al., 1998). However, while the complete nucleotide sequence of the Arabidopsis genome is expected to become available at the end of the year 2000 (Meinke et al., 1998) the function of most of the unravelled gene sequences remains unclear.
Information on the function of a particular gene might be obtained from the analysis of mutants in which the function of the gene is abolished. Insertional mutagenesis is a convenient method to create knockout mutants to study gene function. Transposon or T-DNA insertions can result in complete inactivation of the target gene and have been used for the isolation and identification of genes (Feldmann, 1991; Koncz et al., 1992; Aarts et al., 1993; Bancroft et al., 1993; Jones et al., 1994; Okuley et al., 1994; Azpiroz-Leehan and Feldmann, 1997). Insertional mutagenesis can also be used to identify insertions in specific target genes with known sequence but unknown function. This reverse genetics procedure termed target selected insertional mutagenesis uses T-DNA or transposable elements to mutagenize the entire genome with insertions at random positions in the genome. This is followed by the identification of an insertion in the gene of interest making use of the sensitivity of the polymerase chain reaction (PCR). A primer, specific for the gene of interest is used in combination with primers specific for the insertion sequence and amplification of a gene-specific product occurs when an insert is present in the gene. This technique has been shown to work in Drosophila, Caenorhabditis, maize and petunia populations saturated with transposable elements in which knockout mutations in specific genes were identified (Ballinger and Benzer, 1989; Kaiser and Goodwin, 1990; Rushforth et al., 1993; Zwaal et al., 1995; Koes et al., 1995; Mena et al., 1996).

For reverse genetics purposes in Arabidopsis, populations of T-DNA insertions have been used successfully (McKinney et al., 1995; Krysan et al., 1996; Azpiroz-Leehan and Feldmann, 1997; Bouché and Höfte, 1998; Winkler et al., 1998). Although T-DNA insertions are in general single or low copy and stable, an advantage of using transposon over T-DNA insertions is their ability to transpose. Transposons can excise from the gene of interest and the resulting reversion of a mutation can be used to confirm that it was caused by the transposon. Also, characteristic of the maize Ac/Ds and En/Spm transposable elements is their preference to transpose to linked sites and this property can be advantageously used in a local mutagenesis strategy to tag genes located in the vicinity of the donor site (Peterson, 1970; Bancroft and Dean, 1993; Cardon et al., 1993; Aarts et al., 1995; James et al., 1995).

In order to further establish a system that could be used in reverse genetics screens in Arabidopsis, to complement the already available T-DNA populations, we used a modified two component maize Enhancer–Inhibitor (En–I or Spm–dSpm) transposable element system. A stable transposase locus T-En5, was shown to mediate frequent transposition of I elements which occurred continuously throughout plant development (Aarts et al., 1993, 1995). The high levels of activity could result in rapid amplification and proliferation of insertions to linked and unlinked sites. These characteristics were utilised for the generation of a small population of 960 lines containing multiple I elements per line. This population was in particular established to study local transposition of the En–I system by using an I element (I-201), which had previously been mapped on chromosome IV, as a 'launch pad' for further transposition events. The genes tested for I insertions were therefore located at relatively near distances (0–15 cm) from I-201 and included genes identified by the ESSA I (European Scientists Sequencing Arabidopsis) programme. These genes were of particular interest since they are homologous to other genes and gene families but their function remains to be analysed. Two other well-characterised genes, GA1 and APETALA 2 (AP2) located at much further distances (31–55 cm) from I-201, were used to test the long distance mutagenesis potential of this transposon system. Insertions were identified for all genes tested, and transpositions of I-201 had therefore occurred across the chromosome. However, all insertions were identified as somatic insertions and the tendency to insert somatically was inherited in some plants and their progeny. Insertion rates and insertion
stability are influenced by many factors and include the proximity of the transposon to the target, the timing and level of transposase expression and other element-encoded gene products, and also host-encoded products and regulation mechanisms (Fedoroff and Banks, 1988; Donlin et al., 1995; Eisses et al., 1997). Somatic insertions will be transmitted germinally to the progeny at a lower frequency and can be identified when more progeny of plants are screened and selected for inheritance.

2. Material and methods

2.1. Plant material

In earlier experiments described by Aarts et al. (1995), the ‘in cis’ two-element En–I’ transposon system was introduced into the A. thaliana ecotype Landsberg erecta using a T-DNA construct (cwEnN::I), which contained an unmarked 2.2 kb I element inserted in the open reading frame of an NPTII marker gene. This construct also harboured an En element as stable transposase source, from which the 5’ and 3’ ends were removed and replaced by the cauliflower mosaic virus (CaMV) 35S promoter (5’) and terminator (3’) sequences. A primary transformant displaying active transposition was used as source of the transposon population. This transformant contained two T-DNA loci harbouring the transposase T-En5 on chromosome II, and T-En2 on chromosome I, conferring 10–30% and 2–5% transposition, respectively. Plants containing the T-En5 locus were selfed over several generations resulting in lines in which new transpositions of the original I element(s) had taken place. The positions of I elements were determined for several of such lines by using inverse PCR to isolate flanking sequences of I elements and subsequently mapping these products on the genome using recombinant inbred lines (RILs) (Lister and Dean, 1993; Aarts et al., 1995). One line harboured I-201 on chromosome IV and was chosen for further experiments. A F2 segregating population was obtained after a cross with wildtype Landsberg erecta. One plant homozygous for I-201, and hemizygous for T-En5 was selected, after probing genomic Southern blots with an I-201 and T-En5 flanking probe. A total of 50 hygromycin resistant progeny plants were transferred to the greenhouse and used as pollen donors in crosses with a marker stock (F2 of ecotype Columbia X Landsberg erecta; phenotype erecta, cer2, ms1). Around 100 crosses were made with every pollen donor plant, resulting in a total of 5000 siliques with different genetic backgrounds. Seed (960), from these crosses was chosen randomly and grown in the greenhouse as a F1 population and planted in ten trays/blocks, and young inflorescence material was harvested in a three dimensional pooling strategy (as outlined in Results and Discussion (Section 3)). As the pollen parent was a hygromycin resistant selfed progeny of a T-En5 hemizygous plant, the resultant F1 population consisted of one third T-En5 transposase free stable plants and 2/3 hemizygous for the T-En5 locus. The CIC-library (Creusot et al., 1995) was later used to fine map I-201 on CIC1H1 and CIC11C1 at position 62 cM on chromosome IV to establish a more precise correlation with the genes and sequences which were tested for insertions and which had also been anchored to CICs on chromosome IV (Schmidt et al., 1996).

2.2. Isolation of genomic DNA

Inflorescence or leaf material of a single plant was used for DNA extraction. Extractions were performed in Eppendorf tubes using liquid nitrogen and twice a volume of 250 μl of DNA extraction buffer (0.3 M NaCl, 50 mM Tris pH 7.5, 20 mM EDTA, 2% sarkosyl (v/v), 0.5% SDS (w/v), 5% phenol (v/v), 5% H2O (v/v)) as described by Pereira and Aarts (1998). Pools of inflorescence material were ground with a mortar and pestle in liquid nitrogen and twice a volume of 2.5 ml of DNA extraction buffer. After phenol/chloroform extraction and centrifugation at 4000 rpm for 15 min, the DNA was precipitated after adding 0.8 volumes of isopropanol to the supernatant, incubation at RT for 10 min and centrifugation at 4000 rpm for 10 min. The pellet was washed twice with 70% ethanol and dissolved in 1 ml TE pH
8.0 and 10 µg ml⁻¹ RNase. After a second round of phenol/chloroform extractions, the pellet was dissolved in 200–400 µl TE pH 8.0 to a final concentration of 50–100 ng µl⁻¹ DNA, depending upon the size of the pool.

2.3. Detection and analysis of I insertions

To detect I insertions in the genes tested, two I specific primers were used; itir2 (CTT TGA CGT TTT CTT GTA GTG) and itir3 (CTT GCC TTT CTT GTA GTG) (Life Technologies), complementary to the 5′ and 3′ terminal inverted repeats of I. Both primers were used in one PCR reaction together with one gene-specific primer. Gene-specific primers were designed complementary to 5′ and 3′% regions of the coding sequence, in such a way that putative insertions could be identified within the gene. The gene-specific primers used were: Ag-6 (ACTCCA GGC CAT TTC CTT CAG) for AGAMOUS, primer 1010 (CGT CCC AAA AAT GCT CTG TTC) and primer 1011 (TGC TAC ACG TCA TGT TCG TTG) for pectinesterase (accession number: Z97340-g2244956), primer 1014 (CTA CCA AGA GAC ATC ATG) and primer 1015 (TCC GGC CAC ACG TTT ACA CGA) for HAT-1 (access. no: Z97343-g2245105), primer 1016 (GGG TGA CTC TTT TAG TGT AGC) and primer 1017 (GTC TCC GTA GAT GAG TCC AGC) for membrane channel protein (access. no: Z97343-g2245093), primer 1030 (GCA GTC GGA TGT GAT GAG GTT) and 1031 (CAA CCC TAA GGC CTC CCA GTT) for glycerol-3-phosphate permease (access. no: Z97343-g2245113), primer 1022 (CGGCAG GCC AGG AAC GTT TCA) and primer 1023 (TCC GAT CTG GAC CGT TGG TGG) for GTP-binding protein (access. no: Z97343-g2245111), primer 1040 (CTG ATG ATGAGG AGA GAA TCC) for AP-2, primer 1052 (CGCCGA CGG AAC TCG AAG GGG) and primer 1053 (CAA GAT AAA CTT CGC CGG GTG) for CH42, primer 1050 (TCA TCT GGC CTG ACG ACG AAA) and primer 1051 (GCC CGA GAA GCT CCA TGA TCT) for GAs, primer 1032 (TGT CGC TAG AGA CAA ATC CAG) and primer 1034 (GTG AGT TTG GAG ATG ATC GCC) for GAI and primer 1048 (TGG AGG GAA GCC CAG TGA CCA) and primer 1049 (TCC GAC CCC ATG AAT CGT GTA) for Cer-2. Primers for PRHA, Mek1 and DD-1 were sent by collaborators for insertion screens. PCR reactions were performed in a total volume of 50 µl (0.5 µg of each primer, 250 µM of each nucleotide, 1 × Supertaq buffer), containing 50–100 ng of DNA and 1U Taq DNA polymerase (Supertaq, HT Biotechnology Ltd, UK). The reactions were performed in a PTC-200 Peltier thermal cycler (MJ Research), using 1 cycle at 94°C for 4 min and 31 cycles of 45 s at 94°C, 45 s at 60°C, 3 min at 72°C, followed by a 7 min extension step at 72°C. Amplification products were size-separated in 1% TBE agarose gels containing ethidium-bromide, alkali treated, then vacuum transferred onto Hybond-N⁺ membranes and hybridised to a gene-specific probe. Primers corresponding to both 5′ and 3′% regions of the coding sequence, as mentioned above, were used in combination to amplify genomic DNA products, which were used as probes. For AGAMOUS, its corresponding cDNA clone was used as a probe. After gel-electrophoresis and isolation of the DNA with a DNA gel-purification kit (QIAGEN), the DNA was used as a probe. Membranes were pre-hybridised at 65°C in 1 M NaCl, 1% SDS, 10% dextran sulphate for 1–4 h and subsequently hybridised with a 32P labelled probe at 65°C, o/n in the same solution. The membranes were washed at room temperature in 2× SSC for 5 min and at 65°C in 2× SSC, 1% SDS for 1 h.

After identification of a positive three-dimensional insertion address, 12 seeds of each three-dimensional address were sown as individuals in the greenhouse and one leaf was isolated from all 12 individuals and pooled. After reconfirmation of an insertion in a pool of a putative three-dimensional address by DNA extraction, PCR, Southern blotting and hybridisation: inflorescences of individual plants were subsequently analysed separately.

Amplification products from DNA of individual plants, obtained with one gene-specific primer and both I primers were gel-purified (QIAGEN) and subjected to another round of PCR amplifications. The purified fragments were cloned into the
PGEM-T-vector (Promega), sequenced via an automated sequencing system (Applied Biosystems 373 DNA Sequencer), and analysed with the BLAST algorithm at the National Centre for Biological Information (NCBI).

To determine if individual plants were either stable or homo- or hemizygous for the T-En5 locus, primers were designed which were complementary to the flanking sequences of T-En5 using primers Ten3 (GGA CCA CAA AAC ATG GAA CAT G) and Ten5 (CAC AGC ACA TGT ACA CAA CTG GCG) and a third primer TEn (CAC CAG TTA TCA TTA CGG CC) complementary to sequences within T-En5. PCR conditions were as described above with three primers in one reaction, using 1 cycle at 94°C for 1 min and 35 cycles of 30 s at 94°C, 30 s at 60°C and 3 min at 68°C.

3. Results and discussion

3.1. Three-dimensional pooling and PCR strategies

To identify insertions in genes of interest using PCR-based reverse genetics screens, the population of 960 plants was divided into three-dimensional pools in order to minimise the amount of PCRs necessary to identify an insertion. In a three-dimensional pooling strategy, one or more plants carrying an insertion in the gene of interest can be identified directly in a large population by using just one round of PCRs on all the pools representing the population. The population was divided over ten blocks of 96 plants each (8 rows × 12 columns) shown in Fig. 1A. A total of six flower buds were taken from the main inflorescence of each plant and divided over three pools.

---

**Fig. 1.** Schematic representation of three-dimensional pooling and PCR strategies. A small population of 960 T element containing *Arabidopsis* lines was divided over 10 blocks (trays) and each block contained 96 plants (8 rows and 12 columns). Inflorescence material of every plant in the population was harvested in three dimensions. Two flower buds of one particular plant in the population were taken for the pool of its block; another two were taken for the pool of its row and two more were used for the pool of its column. A. This resulted in 30 DNA pools (12 columns + 8 rows + 10 blocks). For the identification of an T insertion in a gene of interest, one gene-specific and two T specific primers were used in 30 PCR reactions. B. After gel-electrophoresis, blotting and hybridisation with a gene-specific probe, three-dimensional bands can be observed. The autoradiogram shows the result of a PCR reaction with a primer (1010) specific for pectinesterase. Putative three-dimensional insertions were identified, but numbered are only the three-dimensional insertions of 800 bp identified in column 9 and 10, block IV and VI and row B and F (L represents the negative control Landsberg erecta). These positives correspond to plants IV−9-F, IV−9-B, IV−10-F, IV−10-B, VI−9-F, VI−9-B, VI−10-F and VI−10-B in the population. C. PCR analysis of progeny plants (F2) of putative positives IV−9-F, VI−10-F and VI−10-B (plant 3 of IV−9-F, and plant 12 of VI−10-B did not germinate).
Two flower buds were used to represent the pool of its block (96 plants), another two were taken for the pool of its row (10 × 12 = 120 plants) and the last two for the pool of its column (8 × 10 = 80 plants). In this way material from each plant is represented in a unique combination of one block, one row and one column. A total of 30 pools were obtained after DNA extraction of all pooled material. To further minimise the effort needed to identify an insertion in a gene of interest we used two I element and one gene-specific primer in one reaction. The I primers were complementary to the left and right terminal inverted repeat ends of the I element. After gel-electrophoresis and blotting, the membrane was hybridised with a gene-specific probe (Fig. 1B). Hybridising bands are the result of an amplification reaction between one of the I-element primers and the gene specific primer and indicate the number of different insertion positions in the target gene. However, only those bands which appear to be of equal size in a column, a block as well as a row can be considered as a three-dimensional insertion and correspond to one particular plant in the population.

3.2. Identification and analyses of insertion sites

To identify insertions in all genes of interest on chromosome IV, we used two gene-specific primers that were complementary to 5' and 3' regions of the coding sequence. Insertions were found in all genes tested and a relatively high number of independent three-dimensional insertions (hybridising bands of independent sizes) were obtained as displayed in Table 1. Given the nature of the population used, we did not expect to find any more than one three-dimensional insertion per gene tested. Putative three-dimensional insertions were usually identified in a number of blocks, columns and row pools. Consequently, all possible combinations of one block, one row and one column (= one three-dimensional address and one plant in the population) had to be considered and this resulted on average a total of 15 plants with a putative insertion in the target gene (Table 1, ≠ total insertions). When positives were identified in e.g. three different blocks, one row and one column, three different combinations of one block, one row and one column had to be analysed, although only one such address could have been positive, unless multiple insertions were present in these pools. Because of the high number of putative positives found in these screens, 12 progeny plants of each address were grown in the greenhouse the leaves harvested of all plants and analysed as a pool. Pools of addresses that were positive after a new round of PCR reactions and Southern analysis were subsequently chosen for further individual analysis. Insertions were identified in progenies of 90% of all genes tested. These insertions usually differed in size from the ones identified in the three-dimensional screen. Multiple putative three-dimensional insertions were also identified in a screen with a gene-specific primer (see Material and Methods) for the putative pectinesterase enzyme (Fig. 1B). The results obtained with this gene will be used further to describe this population.

One independent insertion was identified in a number of three-dimensional pools and was located at a distance of 800 bp from the gene-specific primer. Pectinesterases and pectin methylesterases exist as multi-gene families in the genome (Richard et al., 1996) and the multiple insertions found for this pectinesterase could therefore also result from insertions in close family members. However, genomic Southern analysis showed that this gene was not a member of a multi-gene family. Homology searches in the databases also did not identify DNA sequence homologies of any significance to other known pectinesterases although amino acid homologies did exist and were as high as e⁻⁵⁴ over 165 amino acids for a pectinesterase on chromosome IV (BAC T24M8). A maximum of two homologous genes is present in the genome according to Southern analysis, and insertions are therefore most likely insertions in the pectinesterase gene that was tested.

A total of 25 possible three-dimensional addresses were tested for insertions in pectinesterase, which corresponds to an initial insertion frequency of 2.6% (25 addresses/960 lines). After PCR and Southern analysis 11 progeny pools were positive and this corresponds to a frequency of 44% (11 positive pools/25 putative positive
Table 1
Genes tested for insertions

<table>
<thead>
<tr>
<th>Gene</th>
<th># Insertions</th>
<th># Total insertions</th>
<th>Insertions in progeny</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGAMOUS</td>
<td>2</td>
<td>19</td>
<td>No</td>
<td>Flower development</td>
<td>Yanofsky et al. (1990)</td>
</tr>
<tr>
<td>Pectinesterase (ESSA I)</td>
<td>2</td>
<td>25</td>
<td>Yes</td>
<td>Unknown</td>
<td>Schena and Davis (1992)</td>
</tr>
<tr>
<td>HAT-1 (ESSA I)</td>
<td>1</td>
<td>12</td>
<td>Yes</td>
<td>Plant development</td>
<td></td>
</tr>
<tr>
<td>Palmitoyl protein thioesterase (ESSA I)</td>
<td>3</td>
<td>3</td>
<td>Yes</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Membrane channel protein (ESSA I)</td>
<td>4</td>
<td>12</td>
<td>Yes</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Glycerol-3-phosphate permease (ESSA I)</td>
<td>2</td>
<td>15</td>
<td>Yes</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>GTP-binding protein (ESSA I)</td>
<td>2</td>
<td>28</td>
<td>Yes</td>
<td>Unknown</td>
<td>Morris et al. (1997)</td>
</tr>
<tr>
<td>DD1</td>
<td>4</td>
<td>13</td>
<td>Yes</td>
<td>Unknown</td>
<td>Korfhage et al. (1994)</td>
</tr>
<tr>
<td>Mek1</td>
<td>5</td>
<td>&gt;30</td>
<td>Yes</td>
<td>MAP-kinase</td>
<td></td>
</tr>
<tr>
<td>PRHA (a)</td>
<td>3</td>
<td>&gt;30</td>
<td>Yes</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>PRHA (b)</td>
<td>3</td>
<td>15</td>
<td>Yes</td>
<td>Flower development</td>
<td>Jofuku et al. (1994)</td>
</tr>
<tr>
<td>Ap-2 (5')</td>
<td>2</td>
<td>9</td>
<td>Yes</td>
<td>Flower development</td>
<td></td>
</tr>
<tr>
<td>Ap-2 (5')</td>
<td>1</td>
<td>1</td>
<td>Yes</td>
<td>Chloroplast protein</td>
<td>Koncz et al. (1990)</td>
</tr>
<tr>
<td>CH42</td>
<td>2</td>
<td>4</td>
<td>Yes</td>
<td>Gibberellin biosynthesis</td>
<td>Xu et al. (1995)</td>
</tr>
<tr>
<td>GA5</td>
<td>1</td>
<td>4</td>
<td>Yes</td>
<td>Gibberellin biosynthesis</td>
<td>Sun and Kamiya (1994)</td>
</tr>
<tr>
<td>GAI</td>
<td>1</td>
<td>1</td>
<td>Yes</td>
<td>Lipid synthesis</td>
<td>Xia et al. (1997)</td>
</tr>
<tr>
<td>Cer-2</td>
<td>2</td>
<td>&gt;30</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The total number of three-dimensional insertions of distinct sizes obtained (# insertions), and the total number of corresponding putative three-dimensional combinations (# total insertions) are indicated. Progenies of three-dimensional insertions were tested and it is indicated when insertions were identified.

...pools). However, the previously obtained 800 bp band was identified in only 10% of the progeny pools, while bands of different sizes and thus different insertion sites, were identified in the other positive pools (results not shown). In this way 14 addresses could be discarded. Of the 11 positive pools, five that showed strong hybridisation signals were further analysed on individual plant level and PCR progeny analyses of addresses IV-9-F, VI-10-F and VI-10-B (F2 generation) are shown in Fig. 1C. Insertions were detected in the progenies of all five positive pools and this corresponds to a frequency of 100% (5/5). Insertions were confirmed in 25–50% of the individuals at these addresses. However, as was found for the pools, most of these insertions were present at other positions in the gene. Bands of around 800 bp were identified in 10–67% of positive individuals of an address. The frequencies with which insertions in pectinesterase recur in the progenies suggest that the initial three-dimensional insertion was somatic and that the property to insert somatically was heritable. These were insertion events that had occurred in at least several cells of the flower and hence insertions could be visualised only after hybridisation with a gene-specific probe. To assess whether the insertions were somatic, we analysed the individual progeny plants by genomic Southern analysis. Insertions that are transmitted through the gametes to one or more of the progeny are considered germinal events and can be identified as...
band shifts of the wild-type allele (the size of an \( I \) element is 2.2 kb) when restriction enzymes are used to digest the genomic DNA that do not cut within the \( I \) element. No band-shifts were identified for any of the progeny plants and this confirms the somatic but heritable nature of these insertions.

Progeny of plant IV-9-F-4 (F3 generation), that was homozygous for the T-En5 locus, were also analysed (results not shown) and also displayed a similar pattern of somatic insertions. The F2 generation segregated as 1:2:1 for the T-En5 locus. It was found that transposase free plants did not display somatic or stable insertions; thus the somatic insertions are a characteristic of transposase containing families.

To identify the position of one such insertion in pectinesterase and also to confirm that this insertion was not present in a homologue, a weak band of 400 bp was isolated from the gel and after purification re-amplified using one gene-specific primer and both \( I \) element primers. These products were cloned and sequenced. Sequence analysis showed that in one clone the \( I \) element insertion was present in an exon of the pectinesterase gene identified by ESSA I and had generated a 3 bp target site duplication, characteristic for the \( En/Spm \) transposon family (Fig. 2A and B) (Aarts et al., 1993).

Other clones identified five different sequences of 400 bp between two \( I \) elements that were in the same orientation, since itir2 and itir3 sequences were found at either end of the fragment. This data suggests that transposition had resulted in clustering of \( I \)-elements. Analysis of the sequences showed that these insertions were present on chromosome II (BAC T3F17) in a putative Ap2 domain containing protein, on chromosome II (BAC T11A7) in an unknown protein, on chromosome V (P1 clone MTH16) in a gene which has similarity to the transcriptional activator Ra, and two insertions were identified in gene sequences which have not yet been identified by the \textit{Arabidopsis} sequencing programmes but which showed homologies to a phenylalanine ammonia lyase and a protein serine/threonine kinase like gene, respectively.

### 3.3. Local mutagenesis on chromosome IV

The other genes that were tested for insertions on chromosome IV are \textit{APETALA2} (\textit{AP2}), \textit{PRHA}, \textit{DD1}, \textit{Mek1}, \textit{GA5}, \textit{Cer2}, \textit{CH42}, \textit{AGAMOUS} (\textit{AG}), \textit{GAI} and genes identified by ESSA I (Fig. 3). The genes described by ESSA I were annotated as having a ‘strong similarity’or ‘similarity’ to known genes. The three-dimensional insertion patterns obtained for all these genes were similar to what was observed for pectinesterase. All insertions identified were also somatic and heritable and patterns were complex. The number of insertions identified varied between the genes, but a correlation between the relative distance of \( I \)-201 and the tested genes and the number of insertions identified was not observed. The total number of putative three-dimensional insertions was low for \textit{GA1} and \textit{AP2}, positioned at 45 cM and 31 cM from \textit{I}-201, respectively. However, equally low numbers were also observed for \textit{GA5}, very near (3cM) \textit{I}-201 and \textit{CH42} at the same genetic position as \textit{I}-201. This might be due to PCR primer discrepancies or insertion-site preferences of \( I \)-elements.

![Fig. 2. Analysis of a pectinesterase::\( I \) tagged site. A. Drawing of the pectinesterase gene showing the position and orientation of an \( I \) insertion and the relative positions of both gene-specific primers. The bar represents 125 bp. B. Nucleotide sequences surrounding the insertion site in the somatic pectinesterase::\( I \) allele. The target site duplication (TSD) is underlined.](image-url)
Fig. 3. Genes tested for \( I \) insertions. Schematic representation of chromosome IV. Genes tested for \( I \) insertions are shown in bold. Sequences representing genes found by ESSA I are shown on the left. Numbers represent the position (cM) of the gene or marker on chromosome IV (Lister and Dean, 1997).

To determine if similar patterns of somatic insertions were obtained with a population which also contained multiple \( I \) insertions per line but with 75,000 \( I \) insertions distributed randomly over the genome (Speulman et al., 1999), the genes \( HAT-I \) and pectinesterase were tested for insertions using identical PCR conditions as described here. Similar multiple patterns of somatic insertions were obtained, suggesting that the obtained patterns are gene-specific and or a consequence of the multiple \( I \) element lines, c.q. the genetic background used.

The most likely explanation for heritability of somatic insertions is that a donor \( I \) element, possibly originating from \( I-201 \) has transposed close to the target gene and is selected for in the process of the population screen. In the next generation, transposition events from this linked site will again produce somatic insertion events at random in the gene tested when the transposase source is present, although certain positions might be preferred. Overall, this data indicates that \( I-201 \) transposed locally as well as over longer distances and that the \( En-I \) system can be used in \( Arabidopsis \) for local as well as long-distance mutagenesis.

A high degree of somatic insertions does not guarantee detection of germinal insertions at a high frequency in the next generation. However, use of DNA from germinal tissue for testing would greatly increase the likelihood of finding progeny with germinal insertions. Although we used inflorescence material for the analysis of the progeny, it might be necessary to isolate DNA from pollen of plants showing strong hybridisation signals as described for the isolation of \( Ds \)-insertions in the gene polygalacturonase, using similar multiple insertion lines of tomato (Cooley and Yoder, 1998). Alternatively, it might be sufficient to analyse very small flower buds for insertions in the gene of interest, followed by selective crosses of several flowers, originating from the same inflorescence meristem and located in the near vicinity of the analysed flower bud, with wild-type pollen to segregate out the transposase. When using very small flower buds, the amount of germline specific tissue and the probability of finding a germinal insertion is considerably higher as opposed to older flower buds.

Systematic selection for germinal events as a system was demonstrated by Cooley and Yoder (1998), who isolated germinal insertions in the polygalacturonase (PG) gene from plants that displayed high somatic transposition. Four thousand progeny were screened in this experiment and five were found containing germinal transmitted \( Ds \) insertions in PG. This procedure can also be applied to the population we produced for chromosome IV transpositions.

The population we generated was established to identify germinal insertions in genes located on chromosome IV using active transposition from the \( I-201 \) launching pad. In addition to the \( I-201 \) element, four other \( I \) elements, including one mapped on chromosome V and another on chromosome III are present in the background. Assuming uniform transposition in the target region on chromosome IV, from \( I-201 \) with 50% transposition frequency (data not shown) and the other elements, we estimate about 500 independent germinal insertions in this region of approximately 21,000 kb (http://genome-www3.stanford.edu/Arabidopsis/chromosome). If insertions are distributed randomly over this chromosome, one
insertion should be present every 40 kb. For a target of an average gene of about 2 kb (using PCR from one side) this gives a chance of about 1/20 or 5% to hit a specific gene. The targeted genes in the ESSA I region are located in a 430 kb interval, with a total of 121 genes. The average length of a gene in this region is 2.3 kb and an average of one insertion every 2 kb is necessary to be able to identify an insertion in any random gene chosen on chromosome IV. As the chance of identifying a germinal insertion within the coding sequences of any of the genes in this region is only about 5%, our failure to identify germinal insertions directly is therefore due to a low genome coverage of the transposons. In a larger population for randomly distributed genome wide coverage of transposons with multiple transposons per plant (Speulman et al., 1999), about 50% insertions in genes were found that were heritable. The evidence of these insertion mutants was also provided with phenotypes of tagged mutants, showing that this transposon system can be useful for reverse genetics methods.

Despite an absence of germinal insertions in any of the genes tested, we show that this population contains a genetic background with a very high insertion activity that can be used for insertional mutagenesis of any gene of interest on chromosome IV. However, more refined and thorough screening procedures and larger progeny populations will be necessary to subsequently identify germinal insertions in the gene of interest.

Acknowledgements

The authors would like to thank Bert Schipper and Bas te Lintel Hekkert for technical assistance.

References


