Radhamony, Resmi Nath; Mohan Prasad, Anand; Srinivasan, Ramamurthy
T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics
Electronic Journal of Biotechnology, vol. 8, núm. 1, abril, 2005, pp. 82-106
Pontificia Universidad Católica de Valparaíso
Valparaíso, Chile

Available in: http://www.redalyc.org/articulo.oa?id=173314714012
T-DNA insertional mutagenesis in *Arabidopsis*: a tool for functional genomics

Resmi Nath Radhamony*
National Research Centre on Plant Biotechnology
Indian Agricultural Research Institute
New Delhi, India
Tel: 91 11 258 417 87 223
Fax: 91 11 258 439 84
E-mail: mradhamony@ucdavis.edu

Anand Mohan Prasad
National Research Centre on Plant Biotechnology
Indian Agricultural Research Institute
New Delhi, India
Tel: 91 11 258 417 87 223
Fax: 91 11 25843984
E-mail: prasad_anand@yahoo.com

Ramamurthy Srinivasan*
National Research Centre on Plant Biotechnology
Indian Agricultural Research Institute
New Delhi, India
Tel: 91 11 258 417 87 222
Fax: 91 11 25843984
E-mail: sri@iari.res.in

**Financial support:** CGP project of NATP (CGP II/253) to RS and Senior Research Fellowships of CSIR to RNR and AMP.

**Keywords:** *Arabidopsis*, functional genomics, insertional mutagenesis, promoter trapping, T-DNA tagging.

**Present address:** Plant Biology, Division of Biological Sciences, University of California, Davis, CA, USA. Tel: 530 754 9852, Fax: 530 752 5410.

With the availability of complete genome sequences of several organisms, the focus has shifted from structural genomics to functional genomics, specifically in plants where the complete genomic sequences are becoming available *i.e.*, *Arabidopsis* and rice. *Agrobacterium* mediated transformation which is exploited for transgenic technology is also being used as an effective mutagen and as a tool for functional genomics in higher plants. Besides the fact that the insertion of T-DNA element into a gene can lead to loss or gain of function, ingenious use of a variety of vectors have led to the identification of genes and regulatory elements in *Arabidopsis*. In this review, we highlight the progress made in the field of functional genomics of *Arabidopsis* using T-DNA tagging. Since this strategy has been very successfully employed in *Arabidopsis* and is now being extended to other plant species, we discuss the various vectors and experimental approaches employed to tag, identify and clone genes and promoter elements in *Arabidopsis* using T-DNA as a tool.

*Corresponding author*

The completion of sequencing of the genome of *Arabidopsis thaliana* (AGI, 2000) marks a major milestone in plant biology research. The determination and analysis of the genome sequence of *Arabidopsis* have provided the first detailed description of the genetic blue print of a higher plant (Lin et al. 1999) and revealed several novel processes involved in plant growth and development. The availability of genome sequence of *Arabidopsis* is likely to broaden and accelerate further research in plant sciences. With the availability of complete genome sequences, the focus of research world wide is now on functional genomics and to assign functions to newly identified DNA sequences. Having identified a new sequence, the comparison with sequences in the databases is the simplest way to obtain functional information. Although efficient bioinformatics tools are becoming available for the annotation of genome sequences, *in silico* analysis by itselfs only indicative and is generally not sufficient to define function of a gene. Even in those cases where some indications become available from *in silico* analyses, experimental evidences are required. One of the most challenging tasks before plant
scientists is assigning functions to a large number of plant genes. The functions of ~69% of the genes were classified according to sequence similarity to proteins of known function in all organisms (AGI, 2000). Definitive functions for individual genes have been thoroughly established for less than 10% (Ostergaard and Yanofsky, 2004). Out of the ~26000 genes identified in Arabidopsis, the functions of only a few thousand have been defined with great confidence (Bouche and Bouchez, 2001) and more than 30% of the predicted Arabidopsis genes could not be assigned any specific function (AGI, 2000).

Table 1. T-DNA insertion collections by various groups and their availability in stock centers.

<table>
<thead>
<tr>
<th>Submitted by</th>
<th>Background</th>
<th>Selectable marker</th>
<th>Promoter/enhancer/reporter gene</th>
<th>Population size</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALK</td>
<td>Columbia-0</td>
<td>Kanamycin</td>
<td>---</td>
<td>145589</td>
<td>Alonso et al., 2003</td>
</tr>
<tr>
<td>GABI-Kat(^2)</td>
<td>Columbia-0</td>
<td>Sulfadiazine</td>
<td>---</td>
<td>59455</td>
<td>Rosso et al., 2003</td>
</tr>
<tr>
<td>Czaba Koncz</td>
<td></td>
<td>Hygromycin</td>
<td>---</td>
<td>300</td>
<td>Szabados et al., 2002</td>
</tr>
<tr>
<td>Syngenta</td>
<td>C24</td>
<td>Hygromycin</td>
<td>GUS</td>
<td>1250</td>
<td>Sessions et al., 2002</td>
</tr>
<tr>
<td>INRA-Versailles</td>
<td>Ws, (Wassil-evskija)</td>
<td>Basta</td>
<td>GUS</td>
<td>1480</td>
<td>Balzergue et al., 2001</td>
</tr>
<tr>
<td>Le Clere and Bartel</td>
<td>Columbia</td>
<td>Basta</td>
<td>CaMV35S-cDNA</td>
<td>33100</td>
<td>LeClere and Bartel, 2001</td>
</tr>
<tr>
<td>Haseloff</td>
<td>C24</td>
<td>Kanamycin</td>
<td>GAL4-GFP</td>
<td>8000</td>
<td>Kiegle et al., 2000</td>
</tr>
<tr>
<td>Weigel</td>
<td>Columbia</td>
<td>Basta</td>
<td>Multimerised CaMV35S enhancers</td>
<td>&gt;20000</td>
<td>Weigel et al., 2000</td>
</tr>
<tr>
<td>Sussman and Amasino(^3)</td>
<td>Ws-2</td>
<td>Kanamycin</td>
<td>Ap2::GUS</td>
<td>37800</td>
<td>Sussman et al., 2000</td>
</tr>
<tr>
<td>Jack</td>
<td>Columbia</td>
<td>Kanamycin</td>
<td>CaMV-GUS</td>
<td>11370</td>
<td>Campisi et al., 1999</td>
</tr>
<tr>
<td>Ehrhardt</td>
<td>Col-2 (CS907)</td>
<td>Basta</td>
<td>CaMV 35S -GFP</td>
<td>108</td>
<td>Cutler et al., 2000</td>
</tr>
<tr>
<td>Gallois</td>
<td>C24</td>
<td>Kanamycin</td>
<td>GUS</td>
<td>283</td>
<td>Devic et al., 1995</td>
</tr>
<tr>
<td>Keith Lindsey(^4)</td>
<td>C24, Lands-berg erecta</td>
<td>Kanamycin</td>
<td>GUS</td>
<td>100</td>
<td>Lindsey et al., 1993</td>
</tr>
<tr>
<td>Bressan and Yokoi</td>
<td>C24</td>
<td>Basta</td>
<td>Multimerised CaMV35S enhancers</td>
<td>27330</td>
<td>ABRC website</td>
</tr>
<tr>
<td>Scheible and Somerville</td>
<td>Columbia</td>
<td>Basta</td>
<td>Multimerised CaMV35S enhancers</td>
<td>&gt;145600</td>
<td>ABRC website</td>
</tr>
</tbody>
</table>

\(^1\) All these lines are available with ABRC and NASC except \(^2\), \(^3\) & \(^4\).
\(^2\) collections are available with GABI-Kat.
\(^3\) lines are available only from ABRC.
\(^4\) lines are available only from NASC.
T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics

A variety of approaches are used to clone and gather information about the function(s) of gene(s). Among these, insertional mutagenesis has been extensively used for cloning genes, promoters, enhancers and other regulatory sequences from Arabidopsis. In this review, we highlight the significance of insertional mutagenesis in functional genomic studies and compile the developments on cloning, characterization and identification of genes and promoter elements in A. thaliana, using the T-DNA tagging strategy. In addition to describing the various experimental approaches, we also point out some of the common difficulties and problems associated with the methodology.

Strategies used for cloning and characterization depend upon the information available about the gene or its product. The information, which can be exploited, includes the spatial and temporal expression patterns of a gene; for example, the presence of its mRNA and/or protein in different cell types, during development, during pathogen infection, or in different environments will have to be examined. Expressed sequence tags (EST’s) and microarray-based techniques (Cooke et al. 1996; Singh-Gasson et al. 1999) are some of the powerful approaches in this direction. Methods utilizing the tools of proteomics (Kmieciak et al. 2002; Borner et al. 2003) like SDS-PAGE, MS (Kawamura and Uemura, 2003), MALDI-TOF (Chivasa et al. 2002; Egelhofer et al. 2002; Fukao et al. 2002; Bae et al. 2003), Yeast two-hybrid system (Teige et al. 2004) etc. have proved very useful in elucidating gene function (Bouchez and Hoffer, 1998). Gene suppression (deletion / mutation knock out) or over-expression permits the gene sequence to be linked to a phenotype from which the function of the gene can be deduced (Matzke and Matzke, 1995). TILLING (Targeting Induced Local Lesions In Genomes), another strategy involves traditional chemical mutagenesis followed by high throughput screening for point mutations (McCallum et al. 2000a; McCallum et al. 2000b; Perry et al. 2003; Till et al. 2003; Henikoff et al. 2004) and allows identification of allelic series of induced point mutations in genes of interest. Conversely map based cloning is another tool to identify a gene known only by its phenotype (Komori et al. 2004; Ohno et al. 2004; Sun et al. 2004).

Gene disruption is a powerful tool for obtaining knock out mutants that helps in ascertaining biological function of the numerous uncharacterized open reading frames (ORF’s), revealed by the genome sequencing project or represented in expressed sequence tags. Gene knock out systems provide a direct route to determining function. Most other approaches to gene function are correlative and do not necessarily prove a causal relationship between gene sequence and function. For example, DNA chips provide an exciting means to discover conditions under which gene expression is regulated on a genome wide scale (Becker et al. 2003; Birnbaum et al. 2003). However, because factors other than mRNA level also determine the activity of a gene product in situ, expression studies even when done on a genome wide scale cannot prove a causal relationship. By contrast, the availability of a null mutation for the gene of interest allows one to directly monitor the effect of this deficiency on the organism’s ability to function. Whereas gene replacement through homologous recombination is now a routine in yeast and mice, in plants exploiting homologues recombination for targeted insertion has not been easy. Although Terada et al. (2002) describe a strategy to achieve homologous recombination in rice, the procedure needs further refinement for routine application in plants. Antisense or over expression is another strategy, to identify gene function through interference. However, it is a laborious approach and results are often inconclusive or impossible to interpret (Azpiroz-Leehan and Feldmann, 1997). Recent reports of successful use of double stranded RNA mediated interference (RNAi) approach in C. elegans (Ashrafi et al. 2003; Kamath and Ahringer, 2003), Arabidopsis (Wang and Waterhouse, 2001; Masclaux et al. 2003),MALDI-TOF (Kawamura and Uemura, 2003), MALDI-TOF (Chivasa et al. 2002; Egelhofer et al. 2002; Fukao et al. 2002; Bae et al. 2003), Yeast two-hybrid system (Teige et al. 2004) etc. have proved very useful in elucidating gene function (Bouchez and Hoffer, 1998). Gene suppression (deletion / mutation knock out) or over-expression permits the gene sequence to be linked to a phenotype from which the function of the gene can be deduced (Matzke and Matzke, 1995). TILLING (Targeting Induced Local Lesions In Genomes), another strategy involves traditional chemical mutagenesis followed by high throughput screening for point mutations (McCallum et al. 2000a; McCallum et al. 2000b; Perry et al. 2003; Till et al. 2003; Henikoff et al. 2004) and allows identification of allelic series of induced point mutations in genes of interest. Conversely map based cloning is another tool to identify a gene known only by its phenotype (Komori et al. 2004; Ohno et al. 2004; Sun et al. 2004).

Mutational approaches have been successfully used for the study of genetic and molecular bases for many traits in plant biology. Access to the mutation is obtained using positional cloning strategies. This strategy is facilitated in model species such as Arabidopsis, for which dense genetic maps with many visible and molecular genetic markers exist, and for which complete physical map consisting of a collection of overlapping cloned DNA fragments and total genome sequence are already available. The limiting factors for this approach are the time and effort required for creating the mapping population and the fine mapping of the mutant locus.

### Table 2. Gene disruption resources in Arabidopsis.

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>Sequenced insertions</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salk Institute</td>
<td>94 947</td>
<td><a href="http://signal.salk.edu/cgi-bin/tdnaexpress">http://signal.salk.edu/cgi-bin/tdnaexpress</a></td>
</tr>
<tr>
<td>TMRI</td>
<td>100 000</td>
<td><a href="http://www.tmri.org/en/partnership/sail_collection.aspx">http://www.tmri.org/en/partnership/sail_collection.aspx</a></td>
</tr>
<tr>
<td>GABI-KAT</td>
<td>20 764</td>
<td><a href="http://www.mpiz-koeln.mpg.de/GABI-Kat/">http://www.mpiz-koeln.mpg.de/GABI-Kat/</a></td>
</tr>
<tr>
<td>FLAG</td>
<td>11 500</td>
<td><a href="http://flagdb-genoplante-info.infobiogen.fr/">http://flagdb-genoplante-info.infobiogen.fr/</a></td>
</tr>
</tbody>
</table>
The use of insertional mutagenesis in principle provides a more rapid way to clone a mutated gene. DNA elements that are able to insert at random within chromosomes such as transposons (Sundaresan et al. 1995; Martienssen, 1998) or the T-DNA of Agrobacterium tumefaciens (Azpiroz-Leehan and Feldmann, 1997), can be used as mutagens to create loss of function mutations in plants. Since the sequence of the inserted element is known, the gene in which the insertion has occurred can be recovered, using various cloning or PCR-based strategies. An advantage of using T-DNA as the insertional mutagen as compared to transposons is that the T-DNA insertions do not transpose subsequent to insertion and are chemically and physically

### Table 3. Functionally characterized genes from *Arabidopsis* using T-DNA.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Genes</th>
<th>Functions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GL1</td>
<td>Trichome development</td>
<td>Marks and Feldmann, 1989</td>
</tr>
<tr>
<td>2</td>
<td>CPR-20</td>
<td>Disease resistance signalling</td>
<td>Silva et al. 1999</td>
</tr>
<tr>
<td>3</td>
<td>GI</td>
<td>Phytochrome signalling</td>
<td>Huq et al. 2000</td>
</tr>
<tr>
<td>4</td>
<td>MYB75</td>
<td>Anthocyanin production</td>
<td>Borevitz et al. 2000</td>
</tr>
<tr>
<td>5</td>
<td>LEP</td>
<td>Leaf development</td>
<td>Van der Graaff et al. 2000</td>
</tr>
<tr>
<td>6</td>
<td>DPE1</td>
<td>Starch breakdown</td>
<td>Critchley et al. 2001</td>
</tr>
<tr>
<td>7</td>
<td>Dfi1</td>
<td>Auxin control of root development</td>
<td>Nakazawa et al. 2001</td>
</tr>
<tr>
<td>8</td>
<td>KAN</td>
<td>Organ polarity</td>
<td>Kerstetter et al. 2001</td>
</tr>
<tr>
<td>9</td>
<td>BRS1</td>
<td>Brassinosteroid signaling</td>
<td>Li et al. 2001</td>
</tr>
<tr>
<td>10</td>
<td>QUA1</td>
<td>Biosynthesis of pectins</td>
<td>Bouton et al.2002</td>
</tr>
<tr>
<td>11</td>
<td>ATE1</td>
<td>Leaf senescence</td>
<td>Yoshida et al. 2002</td>
</tr>
<tr>
<td>12</td>
<td>PRPS17</td>
<td>Leaf senescence</td>
<td>Woo et al. 2002</td>
</tr>
<tr>
<td>13</td>
<td>PLS</td>
<td>Auxin-cytokinin homeostasis to modulate root growth and leaf vascular patterning</td>
<td>Casson et al. 2002</td>
</tr>
<tr>
<td>14</td>
<td>CPL1 and CPL3</td>
<td>Abiotic stress signalling growth and development</td>
<td>Koiiwa et al. 2002</td>
</tr>
<tr>
<td>15</td>
<td>OSM1 and SYP61</td>
<td>Stress tolerance and ABA regulation of stomatal response</td>
<td>Zhu et al. 2002</td>
</tr>
<tr>
<td>16</td>
<td>HCF152</td>
<td>RNA processing</td>
<td>Meierhoff et al. 2003</td>
</tr>
<tr>
<td>17</td>
<td>EXO</td>
<td>Cell division</td>
<td>Farrar et al. 2003</td>
</tr>
<tr>
<td>18</td>
<td>OMT1</td>
<td>Lignin and synapoyl ester biosynthesis</td>
<td>Goujon et al. 2003b</td>
</tr>
<tr>
<td>19</td>
<td>ULI3</td>
<td>UVB light response</td>
<td>Suesslin and Frohnmeyer, 2003</td>
</tr>
<tr>
<td>20</td>
<td>TPT</td>
<td>Acclimation of photosynthesis to environment</td>
<td>Walters et al. 2003</td>
</tr>
<tr>
<td>21</td>
<td>AMS</td>
<td>Tapetal and Microspore development</td>
<td>Sorensen et al. 2003</td>
</tr>
<tr>
<td>22</td>
<td>BXL1</td>
<td>cell wall metabolism</td>
<td>Goujon et al. 2003a</td>
</tr>
<tr>
<td>23</td>
<td>Cpn60 beta</td>
<td>Molecular chaperonin</td>
<td>Ishikawa et al. 2003</td>
</tr>
</tbody>
</table>
Table 3. Functionally characterized genes from *Arabidopsis* using T-DNA.

(b). Reverse genetic approach.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Genes</th>
<th>Functions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PGP1</td>
<td>Photosynthesis</td>
<td>Hagio et al. 2002</td>
</tr>
<tr>
<td>2</td>
<td>SEC and SPY</td>
<td>Gametogenesis and embryogenesis</td>
<td>Hartweck et al. 2002</td>
</tr>
<tr>
<td>3</td>
<td>PGP1</td>
<td>Photosynthesis</td>
<td>Hagio et al. 2002</td>
</tr>
<tr>
<td>4</td>
<td>AMT1;1</td>
<td>NH4+transporter</td>
<td>Kaiser et al. 2002</td>
</tr>
<tr>
<td>5</td>
<td>LACS9</td>
<td>Chloroplast Acyl-coenzyme A (CoA) synthetase</td>
<td>Schnurr et al. 2002</td>
</tr>
<tr>
<td>6</td>
<td>DIR1</td>
<td>systemic resistance signalling</td>
<td>Maldonado et al. 2002</td>
</tr>
<tr>
<td>7</td>
<td>GEM1-1 and GEM1-2,</td>
<td>cytokinesis</td>
<td>Twell et al. 2002</td>
</tr>
<tr>
<td>8</td>
<td>ECA1</td>
<td>Ca^{2+} pump</td>
<td>Wu et al. 2002</td>
</tr>
<tr>
<td>9</td>
<td>MRE11 and KU70</td>
<td>Telomere maintenance and DNA repair</td>
<td>Bundock and Hooykaas, 2002</td>
</tr>
<tr>
<td>10</td>
<td>IRT1</td>
<td>Metal transporter</td>
<td>Henrique et al. 2002</td>
</tr>
<tr>
<td>11</td>
<td>AtTOP6B</td>
<td>DNA replication</td>
<td>Hartung et al. 2002</td>
</tr>
<tr>
<td>12</td>
<td>MMT</td>
<td>Selenium volatilisation</td>
<td>Tagmount et al. 2002</td>
</tr>
<tr>
<td>13</td>
<td>RCN1</td>
<td>ABA signal transduction</td>
<td>Kwak et al. 2002</td>
</tr>
<tr>
<td>14</td>
<td>CKI1</td>
<td>Megagametophyte development</td>
<td>Pischke et al. 2002</td>
</tr>
<tr>
<td>15</td>
<td>MMP</td>
<td>Flowering and early senescence</td>
<td>Golldack et al. 2002</td>
</tr>
<tr>
<td>16</td>
<td>AtOPT3</td>
<td>Embryo development</td>
<td>Stacey et al. 2002</td>
</tr>
<tr>
<td>17</td>
<td>AtpOMT1, AtpDCT1, AtpDCT2</td>
<td>Oxoglutarate/Malate and Dicarboxylate Transporters</td>
<td>Taniguichi et al. 2002</td>
</tr>
<tr>
<td>18</td>
<td>PEX10</td>
<td>Peroxisome biogenesis</td>
<td>Schumann et al. 2003</td>
</tr>
<tr>
<td>19</td>
<td>UVH6</td>
<td>DNA repair (UV resistance) and growth</td>
<td>Liu et al. 2003</td>
</tr>
<tr>
<td>20</td>
<td>UGT73C6 and UGT78D1</td>
<td>Flavonol glycoside biosynthesis</td>
<td>Jones et al. 2003</td>
</tr>
<tr>
<td>21</td>
<td>CAP-E1</td>
<td>Embryogenesis, meiosis and Meristem organization</td>
<td>Siddiqui et al. 2003</td>
</tr>
<tr>
<td>22</td>
<td>LUT1</td>
<td>Carotenoid hydroxylase</td>
<td>Tian et al. 2003</td>
</tr>
<tr>
<td>23</td>
<td>SCD1 and Ap2/3</td>
<td>Polar cell expansion</td>
<td>Falbel et al. 2003; Li et al. 2003</td>
</tr>
<tr>
<td>24</td>
<td>CLAVATA1</td>
<td>Organ development</td>
<td>Dievart et al. 2003</td>
</tr>
<tr>
<td>25</td>
<td>ATHB5</td>
<td>Abscisic acid signal transduction</td>
<td>Johannesson et al. 2003</td>
</tr>
<tr>
<td>26</td>
<td>TDS4</td>
<td>Anthocyanidin synthesis</td>
<td>Abrahams et al. 2003</td>
</tr>
<tr>
<td>27</td>
<td>AtFUT1</td>
<td>Xyloglcan fucosilation (cell wall biosynthesis)</td>
<td>Perrin et al. 2003</td>
</tr>
<tr>
<td>28</td>
<td>ARR15</td>
<td>Cytokinin mediated signal transduction</td>
<td>Kiba et al. 2003</td>
</tr>
<tr>
<td>29</td>
<td>ATM</td>
<td>Meiosis and somatic response to DNA damage</td>
<td>Garcia et al. 2003</td>
</tr>
</tbody>
</table>
stable through multiple generations. The T-DNA not only disrupts the expression of the gene into which it is inserted, but also acts as a marker for subsequent identification of the mutation. Since *Arabidopsis* introns are small and there is very little intergenic material, the insertion of a piece of T-DNA on the order of 5 to 25 kb in length generally produces a disruption of gene function. If a large enough population of T-DNA transformed lines is generated, there are reasonably good chances of finding a transgenic plant carrying a T-DNA insert within any gene of interest. Mutations that are homozygous lethal can also be obtained and maintained in a population in the form of heterozygous plants. Low copy number and random nature of insertions are considered to be the advantages of T-DNA induced mutagenesis approach.

Extensive work in T-DNA tagging of *Arabidopsis* has become possible because of improvements in techniques for *Agrobacterium* mediated transformation. The original root explant method of Valvekens et al. (1988) allowed one to isolate a few transformed plants, via a laborious tissue culture process. Tens of thousands of transformed plants were beyond reach, until Feldmann and Marks (1987) devised a method for producing independent T-DNA transgenic lines via seed transformation. The development of transformation methods based on dipping whole plants into *Agrobacterium* suspensions has made it possible to generate hundreds of thousands of insertional mutations necessary for saturation of the genome while minimizing the effect of somaclonal variation associated with the process of *in vitro* culture and regeneration (Clough and Bent, 1998).

### Forward and reverse genetics using T-DNA insertion lines

The advantage of T-DNA over classical mutagens is that the *Arabidopsis* sequences flanking the insertion site can be isolated easily. This simplifies the identification of genes corresponding to interesting mutants. When screening populations that have been mutagenised with an insertion element, not all of the new phenotypes will be due to insertion mutations. It is important to establish linkage between the mutant phenotype and the insertion element. Even if some linkage can be established, this does not exclude the possibility that the phenotype is caused by a nearby but independent mutation. Consequently it is necessary to identify the gene disrupted by the insertion and test for complementation of the mutant phenotype by the wild type allele in transgenic plants. Because it is not difficult to identify sequences flanking insertion sites, it is reasonable to dispense with co-segregation testing, and proceed directly to identify the gene interrupted by the insertion, followed by a test for complementation of the mutant phenotype by the wild-type version of the gene. Such complementation data provide definitive proof that the phenotype of interest is in fact caused by the insertion mutation. Another way of confirming the same would be by characterizing an independent mutation in the same gene displaying an identical phenotype. Allelism tests can be used with any recessive mutants exhibiting the same phenotype.

A variety of strategies have been devised to generate and isolate mutants in known genes of *Arabidopsis* by T-DNA or transposon insertional mutagenesis. In these methods, large populations of tagged mutants are generated, which can then be screened for insertions in specific genes. Alternatively, the insertion tags can be individually sequenced and compiled in databases that can be searched for a gene disruption event of interest. This is becoming easier now as genomic sequences of many insertion sites are becoming available (Parinov et al. 1999; Qin et al. 2003).

A highly efficient procedure for obtaining mutants in genes identified in sequencing programs takes advantage of the availability of large collections of plants mutagenised by an insertion element. This procedure makes use of the specificity and sensitivity of the PCR reaction to screen for insertions within regions of interest in a large population of mutagenised plant lines (Young et al. 2001). Using oligonucleotide primers from the insertion element and from the gene of interest, it is possible to detect an insertion event within the gene, even in case of a pooled DNA sample. The sensitivity of the PCR is so high that it is...

<table>
<thead>
<tr>
<th>Insertion Site</th>
<th>Function</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAF2</td>
<td>Vernalisation</td>
<td>Ratcliffe et al. 2003</td>
</tr>
<tr>
<td>Apyrases (Atapy1, Atapy2)</td>
<td>Pollen germination</td>
<td>Steinebrunner et al. 2003</td>
</tr>
<tr>
<td>GORK</td>
<td>Stomatal movements and transpiration</td>
<td>Hosy et al. 2003</td>
</tr>
<tr>
<td>FATB</td>
<td>Fatty acid biosynthesis</td>
<td>Bonaventure et al. 2003</td>
</tr>
<tr>
<td>GGAT</td>
<td>Photorespiration</td>
<td>Igarashi et al. 2003</td>
</tr>
<tr>
<td>ACT7</td>
<td>Germination and root growth</td>
<td>Gilliland et al. 2003</td>
</tr>
<tr>
<td>PIP2;2</td>
<td>Hydraulic conductivity of roots</td>
<td>Javot et al. 2003</td>
</tr>
<tr>
<td>ADA2b and GCN5</td>
<td>Growth, development, and gene expression</td>
<td>Vilachonasios et al. 2003</td>
</tr>
</tbody>
</table>
T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics

It is possible to detect such an event in large pools (up to a few thousand) of mutagenised plants (Bouchez and Hofte, 1998). The pool is then repeatedly subdivided until a single plant carrying the desired insertion is identified. The identified mutant plants are then tested for phenotypes that are predicted to result from loss of function of the gene. This is becoming easier now as genomic sequences of many insertion sites are becoming available (Parinov et al. 1999; Qin et al. 2003). Winkler et al. (1998) have developed an efficient reverse-genetics protocol that uses expedient pooling and hybridization strategies to identify individual transfer-DNA insertion lines. They have screened a collection of 6000 independently transformed lines using this protocol to systematically isolate Arabidopsis lines containing insertional mutations in individual cytochrome P450 genes.

Feldmann (1991), after segregation analysis of a large number of transformants, concluded that the average number of independent inserts is 1.5 per diploid genome with 57% of the transformants containing a single insert and 25% containing two. Also within the context of the target gene, insertion appeared to be random, as T-DNA inserts were found in exons, introns, and in the 5’ and 3’ flanking regions of the more than 30 genes that were characterized. In addition, after screening of the first 14000 transformants no obvious hotspots for insertion of T-DNA were detected (Azpiroz-Leehan and Feldmann, 1997). However, genome wide analysis of the distribution of integration sites conducted by Alonso et al. (2003) on a T-DNA mutagenised population of Arabidopsis revealed the existence of a large integration site bias at both chromosome and gene levels. Alonso et al. (2003) generated over 225000 independent T-DNA insertions, representing near saturation of the entire gene space, out of these, locations of 88,000 insertions were precisely identified. This study conducted at the Salk institute of biological studies, La Jolla CA, USA lead to creation of mutants in ~73% of predicted Arabidopsis genes. The number of integrations was reported to decrease dramatically from the gene rich chromosome arms towards the centromeres. And at the gene level a significant bias was observed against integration events in exons and introns, in favour of 5’UTR, 3’ UTRs and promoters. A similar observation was made by the German plant genomics research programme GABI-KAT. The GABI-KAT

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Gene</th>
<th>Homology</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shooting induction CKI1</td>
<td>Histidine kinase</td>
<td>Kakimoto, 1996</td>
<td></td>
</tr>
<tr>
<td>Early flower induction Flowering Locus T (FT)</td>
<td>Meristem identity gene LEAFY</td>
<td>Kardailsky et al.1999</td>
<td></td>
</tr>
<tr>
<td>Leaves without petiole Leafy Petiole, LEP</td>
<td>AP2/EREBP family</td>
<td>Van der Graaff et al. 2000</td>
<td></td>
</tr>
<tr>
<td>Suppresses late flowering in winter annual ecotypes AGL20</td>
<td>MADS-domain gene</td>
<td>Lee et al. 2000</td>
<td></td>
</tr>
<tr>
<td>Wide heart shaped ovary, larger siliques CYP7889</td>
<td>Cytochrome P450</td>
<td>Ito and Meyerowitz, 2000</td>
<td></td>
</tr>
<tr>
<td>Intense purple pigmentation in vegetative organs pap1-D</td>
<td>MYB transcription factor</td>
<td>Borevitz et al. 2000</td>
<td></td>
</tr>
<tr>
<td>Stiff inflorescence stem, thicker leaves, larger seeds Sturdy</td>
<td>Patatin</td>
<td>Huang et al. 2001</td>
<td></td>
</tr>
<tr>
<td>Constitutive expression of a number of key defense marker genes ADR1</td>
<td>PR-(pathogenesis related)</td>
<td>Grant et al. 2003</td>
<td></td>
</tr>
<tr>
<td>Attenuated red and far-red photoresponses COG1</td>
<td>Dof family of transcription factors</td>
<td>Park et al. 2003</td>
<td></td>
</tr>
<tr>
<td>Shoot formation in the absence of exogenous cytokinins AtIPT8</td>
<td>isopentenyl transferase</td>
<td>Sun et al. 2003</td>
<td></td>
</tr>
<tr>
<td>Dwarf AtGA2ox7 AtGA2ox8</td>
<td>Gibberellin 2 oxidases</td>
<td>Schomburg et al. 2003</td>
<td></td>
</tr>
</tbody>
</table>
population of T-DNA mutagenised *Arabidopsis thaliana* lines with sequence-characterized insertion sites is being used extensively for efficient progress in plant functional genomics. After PCR-based amplification of DNA fragments spanning insertion site borders and sequencing, the data were placed in a flanking sequence tag (FST) database describing which mutant allele was present in which line. Analysis of the distribution of T-DNA insertions revealed a clear bias towards intergenic regions. Insertion sites appeared more frequent in regions in front of the ATG and after stop codons of predicted genes (Rosso et al. 2003).

**Channelling genome resources for public use**

Large collections of the T-DNA insertion lines are being provided to the *Arabidopsis* stock centres at Ohio State University (USA) and Nottingham (UK) by the individuals/groups who constructed them. More than 175,000 T-DNA insertion lines of various types are already available from the *Arabidopsis* Biological Resource Center (ABRC). Table 1 represents the lists of T-DNA insertion lines generated by various groups and are available with different centers. The non-profit organizations can obtain these lines for a fee of £2.25 (£8.50, additional charge) from NASC (Nottingham Arabidopsis Stock Center also known as the European Arabidopsis Stock Center) and $4 ($15, additional charge), from ABRC. The fee structure for a profit organization for obtaining a line is four times than that of non-profit organization. ABRC sends the stock free of charge to any laboratory that cannot afford the associated charge. The stock centres have a distribution agreement. According to which, North and South American users are expected to order seed stocks from ABRC, and European users are expected to order seed stocks through NASC. The laboratories in other locations may order seed stocks from

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Promoter</th>
<th>Specificity of expression</th>
<th>Source species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>At EM</td>
<td>Embryo</td>
<td><em>Arabidopsis</em></td>
<td>Topping et al. 1994</td>
</tr>
<tr>
<td>2</td>
<td>Cryptic</td>
<td>Seed coat specific</td>
<td><em>Nicotiana tabacum</em></td>
<td>Fobert et al. 1994</td>
</tr>
<tr>
<td>3</td>
<td>HVT1</td>
<td>Tapetum and vascular tissue</td>
<td><em>Arabidopsis</em></td>
<td>Wei et al. 1997</td>
</tr>
<tr>
<td>4</td>
<td>Pyk20</td>
<td>Nematode feeding structure</td>
<td><em>Arabidopsis</em></td>
<td>Puzio et al. 1999</td>
</tr>
<tr>
<td>5</td>
<td>tcup</td>
<td>Constitutive</td>
<td><em>Nicotiana tabacum</em></td>
<td>Foster et al. 1999</td>
</tr>
<tr>
<td>6</td>
<td>Cryptic</td>
<td>Guard Cell</td>
<td><em>Arabidopsis</em></td>
<td>Plesch et al. 2000</td>
</tr>
<tr>
<td>7</td>
<td>Cryptic</td>
<td>Roots</td>
<td><em>Arabidopsis</em></td>
<td>Mollier et al. 2000</td>
</tr>
<tr>
<td>8</td>
<td>Lj Cbp 1</td>
<td>Roots</td>
<td><em>Lotus Japonicus</em></td>
<td>Webb et al. 2000</td>
</tr>
<tr>
<td>9</td>
<td>eIF-4A1</td>
<td>Growing tissues, young leaves</td>
<td><em>Arabidopsis</em></td>
<td>De Greve et al. 2001</td>
</tr>
<tr>
<td>10</td>
<td>EXORDIUM (EXO)</td>
<td>Meristematic cells</td>
<td><em>Arabidopsis</em></td>
<td>Farrar et al. 2003</td>
</tr>
</tbody>
</table>

**Figure 1. Activation tag vector.** The vector contains multimerized cauliflower Mosaic Virus (CaMV) 35S enhancer that can function in either orientation and can cause transcriptional activation of nearby genes resulting in dominant gain of function mutations. RB and LB represent right and left borders of T-DNA, respectively. The activation of plant gene by enhancer elements is indicated by broken arrow (black). The flanking plant DNA is shown in green.
either centre. A number of lines require the signing of a material transfer agreement (MTA). For obtaining GABI-Kat T-DNA lines the costs is € 499/- per line for academic institutions and are available only for research purposes. However, for profit institutions a contract needs to be set up between the company and Garching Innovation (the MPG technology transfer organisation).

Furthermore several groups are working to isolate and sequence DNA fragments flanking insertion sites and are depositing sequences in databases. When these databases reach sufficient size, it will be, possible to find an insertion in any gene of interest simply by consulting the databases and ordering seed from the stock centre. Systematic sequencing of insertion sites in various populations have been initiated by several group (Krysan et al. 1999; Speulman et al. 1999; Tissier et al. 1999; Galbiati et al. 2000; Brunaud et al. 2002; Samson et al. 2002; Session et al. 2002; Szabados et al. 2002; Qin et al. 2003; Ostergaard and Yanofsky, 2004). Table 2 provides the list of variety of databases from where users can access the information regarding the T-DNA insertion lines.

Recently 150,000 transformed plants carrying 225,000 (1.5 insertion per line) independent T-DNA integration events was generated and precise locations determined for >88,000 loci by the Salk Institute of Biological Studies (Alonso et al. 2003). Analysis of the insertion site sequences revealed that insertional mutations had been created in ~74% of the annotated Arabidopsis genes.

Surprisingly, although PCR based reverse genetic screens have been available to Arabidopsis researchers, for several years, relatively few informative knockouts obtained in Arabidopsis have given rise to a visible, directly scorable phenotypes. Several reports have shown that this lack of phenotypes is presumably because of partial and complete functional redundancy besides the ability of higher plants to adapt their physiology to various stresses and constraints without undergoing morphological changes, and by our inability to detect slight physiological alterations and/or weak reductions in fitness (Bouche and Bouchez, 2001). In order to identify conditional phenotypes, the mutants have to be tested under a wide range of environmental conditions (Meissner et al. 1999), as plants being sessile organisms have evolved many adaptive traits that allow them to cope with changes in their biotic and abiotic environment. Many loss of function mutants are therefore expected to be conditional and revealed only by a specific combination of environmental parameters.

Using T-DNA tagging several genes have been identified and cloned. Table 3a and 3b provide a list of genes cloned in Arabidopsis using T-DNA tagging in the recent years, using forward and reverse genetic approaches, respectively.

**SPECIFIC USES OF T-DNA LINES USING SPECIALIZED VECTORS**

Simple insertional mutagenesis, like all gene disruption approaches, has some limitations. It is difficult to identify the function of redundant genes or of genes required in early embryogenesis or gametophytic development. To overcome these limitations modified insertional elements have been developed.

Activation trap, enhancer trap and gene / promoter trap vectors are specialized versions of insertional mutants. In addition to screens for loss of function phenotypes, gene function can also be identified based on expression patterns or gain-of-function phenotypes. These screens are particularly useful for genes that have redundant and multiple functions during development.

**Activation tagging**

Enhancer and gene trap insertions identify genes on the basis of their expression pattern. In activation tagging insertion alleles are generated using T-DNA or transposons that carry either a constitutive promoter such as CaMV 35S (Wilson and Somerville, 1995) or multimeric enhancers from the CaMV 35S promoter (Weigel et al. 2000). Because enhancers can function in either orientation and at a considerable distance from the coding regions, they can cause transcriptional activation of nearby genes, resulting in dominant gain of function mutations (Figure 1). Such gene activation may produce novel phenotypes that are either redundant members of a gene family or are essential for survival. Activation tagging strategy, developed by Walden et al. (1994) has been used to generate collections of morphologically diverse dominant mutants resulting in cloning of corresponding genes (Table 4).
Although the paper published by Walden et al. (1994) was later retracted as the major findings of the work could not be substantiated. However, the basic procedure, namely activation tagging described by these workers is a valid technique and has been used by other workers (Balter, 1999). Majority of the on gain of function mutations are on alterations of phenotypes related to plant defence responses (Weigel et al. 2000). More recently 45000 T-DNA insertion lines were generated using an activation tag vector. Out of the 1194 T-DNA insertion sites 1010 have their tags inserted in or close to a predicted gene and nearly half of these genes were those with unknown functions (Qin et al. 2003). The interfered genes were distributed in all five chromosomes and were involved in a variety of processes including metabolism, transcription, plant defence, signal transduction etc.

**Promoter and enhancer trap vectors**

With the ever increasing reports of creation of transgenic in various crop plants, the search for novel genes and variety of regulatory elements of DNA required for controlled expression of the introduced genes is gaining more importance. Regulatory elements that impart a tissue-specific, stage-specific and/or environmental-stimuli-specific expression to the transgene are being identified and cloned using T-DNA based vectors that have been designed to identify and clone such regulatory sequences.

The general principle behind this approach is to integrate a reporter gene that either lacks a promoter (gene/promoter trap) or carries only a minimal promoter (enhancer trap), at random sites in the wild type genome. A reporter gene cassette containing a minimal promoter (enhancer trap) close to the end of the insertion element can be cis activated when inserted close to a transcriptional enhancer that will drive the expression of the reporter gene (Figure 2). The gus (uid A) reporter gene is the most commonly used reporter gene system in plants, because of the absence of endogenous β-glucuronidase (GUS) activity in most plants and the opportunity to visualize the presence of the enzyme by sensitive histochemical techniques (Jefferson et al. 1987).

Mutagenised populations can be screened for lines expressing reporter gene in specific cell types or in specific environmental conditions. Genes with interesting expression patterns and their promoters can be isolated from such lines. Small populations of promoter trap lines have been generated in our laboratory using promoter trap vectors carrying promoterless GUS gene (Resminath et al. 2005). The T-DNA tagged lines generated are being used for identifying novel genes and tissue specific promoters. Apart from the ease of identifying redundant genes this method helps detection of insertions in UTR’s. An insertion in the 3’ UTR’s of genes generally will not lead to suppression of gene expression as the coding region remains intact, but can lead to reporter gene expression enabling detection of the otherwise hidden mutation. Repeated observations of a significant bias towards T-DNA integration in promoters and UTR’s (Alonso et al. 2003; Qin et al.2003) justify a high frequency of GUS expressing lines in the mutant population.

Promoter trap lines can also be exploited to mark certain cell types for developmental studies (Sundaresan et al. 1995). Enhancer traps allow an ingenious way of identifying genomic sequences expressed in precise developmental patterns. In the case of GUS as the reporter gene, the presence of GUS activity in a particular organ or cell at a particular developmental stage will identify sequences expressed at this place and time. Indeed, GUS enhancer traps have proved successful in detecting novel genes in Arabidopsis. A disadvantage, however, is that it may not be easy to identify and locate the enhancer causing the specific expression pattern, since the enhancer element could be in either side of the T-DNA and could also be at a far off distance.

Plants for mutant screens are usually grown from T2 seed because most mutant phenotypes result from homozygous recessive mutations and first generation of mutagenised plants, which will be heterozygous for induced mutations, will not show mutant phenotype. The T1 generation of T-DNA mutants can be screened for dominant mutation, as these plants are hemizygous for the insertions. Hemizygotes arise because the ovule is the target (Ye et al. 1999) for T-DNA transformation of flowers. Thus, the T1 generation seeds obtained after Agrobacterium mediated floral dip transformation method can be used for the screening of...
T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics

Figure 4. Prompter trap. The promoterless reporter gene can be expressed when insertion of a promoter trap vector occurs in an exon so as to result in a transcriptional fusion and disruption of native gene product.

for the creation of marker lines (Kerthundit et al. 1991) and for the isolation of regulatory sequences and eventually the isolation of specifically expressed genes (Topping et al. 1994; Musket and Lindsey, 1995; Ferreira da Rocha et al. 1996; Wei et al. 1997) including the isolation of environmental and hormonal stress responsive regulatory sequences and genes (Mandal et al. 1995). The promoter trap approach is particularly relevant to identification and cloning of genes (and their regulatory sequences) expressed in tissues that are difficult to analyze, by traditional methods relying on RNA extraction, as demonstrated by the cloning of regulatory sequences driving reporter gene expression in nematode feeding structures (Barthels et al. 1997) or the identification of molecular markers for embryogenesis (Topping et al. 1994; Topping and Lindsey, 1997).

A T-DNA tagged mutant showing intense GUS expression in young leaves and rapidly growing stem tissues was used to isolate an eukaryotic translational initiation factor eIF-4Al (De Greveet al. 2001) Promoter-like sequences, known as cryptic promoters which are not associated with a detectable transcript have also been revealed by promoter trapping (Mollier et al. 2000). The popularity of the promoter trap method can be judged from the various collections developed worldwide as well as the variety of reporters and constructs in use. A collection of 20,261 transgenic lines of Arabidopsis generated with the promoter trap vector pTHuc, which carries a promoterless firefly luc (luciferase) reporter gene linked to the right T-DNA border is the latest report in line. By detection of bioluminescence in three week-old seedlings, 753 lines were identified showing constitutive, organ-specific, and stress-responsive luciferase expression patterns. Several lines showed sugar, salt, and abscisic acid (ABA)-inducible luciferase activity (Alvarado et al. 2004). Moreover the technique has now been adopted for the ongoing intensive functional genomics efforts in rice demonstrated by the trapping of two cold inducible genes from screening of T-DNA tagged rice lines for cold responsive GUS expression (Lee et al. 2004). Table 5 provides a list of promoter elements cloned using promoter trap vectors. Jeong et al. (2002) have created a total of 13,450 T-DNA insertion lines in rice that could be used to screen for promoter activity as well as for creating gain of function mutants. In addition to the promoter less GUS reporter gene located next to the right border, the binary vector used in the study carries multimerised transcriptional enhancers from the CaMV35S promoter near the left border. Interestingly histochemical GUS assays revealed the GUS-staining frequency in these lines to be about twice as high as that from lines transformed with vectors, which lack the enhancer elements suggesting that enhancer sequence present in T-DNA improves GUS tagging efficiency.

Strategies for isolating and cloning sequences flanking T-DNA

The promoterless reporter gene can be expressed when insertion of a promoter trap vector occurs in an exon so as to result in a transcriptional fusion and disruption of native gene product. In the gene/promoter system (Figure 3 and Figure 4), insertion of the promoter-less reporter not only disrupts normal gene function but also activates expression of the reporter gene. Because expression can be monitored in heterozygous plants, gene trap system is useful for studying expression patterns of most plant genes, including essential genes that cause lethal mutations. This system is convenient for observing mutant phenotypes because reporter activation indicates the location, condition and the time of expression for the disrupted gene. In Arabidopsis, activation of reporter genes has been observed in ~30% of the transgenic. Some of the recently developed gene trap systems contain an intron with multiple splicing acceptor and donor sites in each of the three reading frames in front of the coding region of the reporter gene (Sundaresan et al. 1995).

These constructs allow reporter gene expression even when the insertion occurs in intron. As a consequence, expression is also observed at high frequencies in Arabidopsis plants transformed with such a gene trap system. Selection or screening for gene specific insertions can also be accomplished independently of promoter activity, as the expression of the marker in the gene trap vector is dependent not only on the activity of an endogenous promoter but also on transcript processing (splicing) and formation of translational fusions (Maes et al. 1999).

Some ten years of promoter trapping experiments show that frequency of reporter gene expression is generally much higher than expected for random integration in the plant genome (Mollier et al. 1995). Promoter trap strategy provides a means of identifying genes, and characterizing in vivo, their expression patterns, throughout the plant life cycle in viable heterozygotes. The insertion of T-DNA into the tagged gene also facilitates its cloning. During the last few years promoter trap lines have proved to be invaluable for the isolation of regulatory sequences and eventually the isolation of specifically expressed genes (Topping et al. 1994; Musket and Lindsey, 1995; Ferreira da Rocha et al. 1996; Wei et al. 1997) including the isolation of environmental and hormonal stress responsive regulatory sequences and genes (Mandal et al. 1995). The promoter trap approach is particularly relevant to identification and cloning of genes (and their regulatory sequences) expressed in tissues that are difficult to analyze, by traditional methods relying on RNA extraction, as demonstrated by the cloning of regulatory sequences driving reporter gene expression in nematode feeding structures (Barthels et al. 1997) or the identification of molecular markers for embryogenesis (Topping et al. 1994; Topping and Lindsey, 1997).
Once co-segregation of T-DNA along with the phenotype has been established for a given mutant, isolation of the disrupted gene can be achieved by a number of strategies. Sequences flanking the insertion can be easily identified from single or low copy number lines using inverse PCR (IPCR) and by thermal asymmetric interlaced (TAIL) PCR (Figure 5). The latter consists of three consecutive rounds of PCR amplification, performed with a set of three nested insertion-specific primers and a small, arbitrary primer, which anneals nearby in the insertion flanking sequence (Liu et al. 1995). Recently, Antal et al. (2004) have reported a simplified version of TAIL PCR called, single oligonucleotide nested (SON)-PCR, which involves only two rounds of PCR with two or three nested sequence primers.

Another strategy employed called, plasmid rescue (Yanofsky et al. 1990; Feldmann, 1992) makes use of insertion within T-DNA of an ‘ori’ sequence a replication of origin required for replication of plasmids in E. coli, along with a bacterial selection marker. Genomic DNA from mutant plant is digested with an appropriate enzyme and religated at high dilution to favour intramolecular reactions. The ligation products are introduced into E. coli and clones are selected on appropriate antibiotic selection. Because of the presence of the proper origin of replication and the presence of bacterial selectable markers, the plasmid is able to replicate in a proper E. coli host. The plasmid isolated from such bacteria thus contains the linked plant DNA as a passenger (Figure 6). A similar rationale applies to inverse polymerase chain reaction. In this procedure, plant DNA is digested with a restriction enzyme that cuts within the T-DNA and subjected to self-ligation. The circularized ligation products are subjected to PCR using appropriate primers from the T-DNA such that the flanking plant sequences, which are now part of the circularized product, are amplified (Figure 7).

Adaptor PCR or anchor PCR ligates adaptor molecules to the ends of the digested DNA fragments carrying part of the insert. A primer specific to the adaptor region and nested primers specific to T-DNA are used to amplify the flanking region (Figure 8). This method has been efficiently used in the German plant genome project GABI and flanking sequence database has been created using this data.

There is another simple and efficient PCR strategy to amplify the T-DNA flanking region known as T-linker specific ligation PCR (T-linker PCR) described by Yuanxin et al. (2003). This strategy amplifies the template molecules in three steps. First genomic DNA is digested with 3’ overhang enzymes. Secondly, primed with by a specific primer, a strand of target molecule is replicated by Taq DNA polymerase and a single A tail is generated on the 3’ unknown end of the target molecule, and then a 3’ overhang-T linker (named T-linker) is specifically ligated on to the target. Thirdly, the target is amplified by two rounds of nested PCR with specific primers and T-linker primers. It differs from other methods in specific TA ligation instead of arbitrary ligation or random annealing.

Finally one can simply construct a genomic library of the mutant and isolate clones containing T-DNA sequences. Some of these clones will contain plant DNA as well. In any case unequivocal proof about the function of gene is obtained by performing molecular complementation, i.e., by demonstrating that the introduction of a wild type allele would eventually restore the phenotype.

PROBLEMS ASSOCIATED WITH THE T-DNA LINES
A variety of problems are encountered while analyzing the T-DNA tagged lines, in particular while attempts are made to clone flanking sequences from a T-DNA tagged mutant. Difficulty can arise because of multiple insertions, complex arrangement of T-DNA, insertion of vector backbone sequences, chromosomal duplication and rearrangements or a combination of these (Jorgensen et al. 1987, Veluthambi et al. 1988, Tax and Vernon, 2001).

**Multiple insertions and complex T-DNA loci**

Complex T-DNA inserts composed of two or more T-DNA repeats are frequently found in transgenic lines (De Buck et al. 1999). Methods like the TAIL PCR allow for the direct identification of insertions sequences co-segregating with the mutant phenotype even in the presence of more than one insertion. The flanking sequences representing the part of the gene of interest can be readily used to identify the corresponding gene from the *Arabidopsis* genome database.

**Transfer of vector backbone**

Although T-DNA tagging and isolation of flanking sequences may appear a rather simple and straightforward method to clone genes and promoters, the inconsistency in T-DNA integration can complicate efforts to clone and identify adjacent plant DNA. Studies have indicated that *Agrobacterium* T-DNA can be present in the genomes of transformed host plants as single units or in multiple tandem arrays (Jorgensen et al. 1987). Some of them have also shown the existence of truncated T-DNA regions. T-DNA regions beyond the border repeat were also found to be stably integrated into plant genomes at high frequencies. Read through of border sequences during T strand formation in the bacterium has been reported (Stachel et al. 1987; Veluthambi et al. 1988). Even DNA sequences of the vector, from far beyond the defined T-DNA region delimited by the border sequences have been detected. Therefore, flanking sequence amplification from T-DNA tagged mutants has the potential of amplifying considerably larger tags of vector backbone sequences than expected.

This problem can be solved by designing primers outside the T-DNA borders and using special Taq polymerases.
capable of performing long PCR amplification. Long PCR techniques has become routine in most of the large scale T-DNA tagging projects owing to the relatively high prevalence of vector backbone sequences in the mutant population. These sequences transfer to the plant either independently of (unlinked to) the T-DNA or linked to the T-DNA across either the left or right T-DNA border. High frequency transfers of sequences beyond the T-DNA border have been reported by Kononov et al. (1997). This work showed that as high as 75% of transgenic tobacco plants generated using Agrobacterium mediated T-DNA transfer contain binary vector backbone sequences integrated into the plant genome. Therefore, integration of backbone sequences besides the T-DNA sequences into the genomes of plant may also be responsible for causing mutations in T-DNA tagged populations. A number of groups have reported that in a large percentage of T-DNA tagged Arabidopsis plants, the T-DNA did not co segregate with the mutant phenotype (Koncz et al. 1989; Errampalli et al. 1991; Feldmann, 1991). In these cases it is possible that the mutation could have been caused by the insertion of backbone sequences independent of the T-DNA in these lines.

**Chromosomal duplication/rearrangement**

Internal chromosomal duplication/rearrangements have been identified in the course of both reverse and forward-genetics approaches. An embryo-defective88 mutation was characterized with simple T-DNA inserts that exhibited normal Mendelian segregation and detected through a combination of detailed molecular analysis and mapping of the kanamycin-resistance phenotype imparted by the T-DNA (Tax and Vernon, 2001). Although emb88is located on chromosome 1, molecular analysis of genomic DNA directly adjacent to the T-DNA left border revealed sequence from chromosome 5. Mapping of the T-DNA confirmed that a >40 kb region of chromosome 5 had inserted with the T-DNA into the emb88 locus on chromosome 1. A similar scenario was observed in a reverse-genetic study of the LR-RPK gene. A prospective insertion mutation was identified by PCR screening of T-DNA lines. Whereas wild-type LR-RPK is encoded on lower chromosome 4, mapping of the T-DNA localized the mutant allele to chromosome 5. In both these cases, sequence of T-DNA flanking regions did not provide an accurate picture of DNA disruption. Therefore, T-DNA insertion lines- even those that exhibit straightforward genetic behaviour with regard to T-DNA and developmental phenotypes may contain an unexpectedly high frequency of duplication/transpositions. Such rearrangements can interfere with reverse-genetic analyses and provide misleading information on the molecular basis of mutant phenotypes. This would necessitate, genetic mapping of T-DNA insertion sites, in addition to standard genetic and molecular characterization of mutant alleles for meaningful interpretation of phenotypes or sequence information from T-DNA mutant.

**Insertional mutagenesis in other angiosperms**

The progress of genomic analysis of plants has made huge leaps in recent years and a large body of information on genomic sequences, sequences of cDNA's and physical maps are now available for several plants including rice, maize, tomato and Brassica. Databases of many if not all, gene sequences of model plants are available and bioinformatics is enabling the prediction of gene function.

On the basis of comparisons among genomes, it is still necessary to validate these predictions. So various methods used for model Arabidopsis genome have been optimized and suitably streamlined for the systematic identification of the biological functions of genes in other more economically important crop species.

Rice genome has been extensively covered because of its small size among cereal genomes, nearly complete genome sequence and economic importance, as a major crop. Maize Ac/Ds transposon system has been proved to be efficient for gene tagging and functional analysis in this crop (Hiroyuki et al. 1999). Kolesnik et al. (2004) has reported a collection of unlinked single copy Ds transmutants. Analysis of 2057 Ds flanking sequences has revealed preferential transposition of Ds into regions of genome rich in expressed sequences. This coupled with a high germinal and independent transposition frequency establishes the efficiency of this system for large-scale mutagenesis in rice.
T-DNA insertional mutagenesis in *Arabidopsis*: a tool for functional genomics

T-DNA tagging has been used successfully for gene discovery in rice. An initial database has also been constructed using T-DNA flanking sequences (Sha et al. 2004). T-DNA tags have also been observed to insert preferentially into gene rich regions (Chen et al. 2003; Sha et al. 2004). A database of insertion sites in rice is publicly available at http://www.genomics.zju.edu.cn/riceetdna. In addition a rice retrotransposon Tos17 that is activated in tissue culture has also been used for this purpose.

In tomato new tools for functional analysis based on insertional mutagenesis, with Ac/Ds system in the background of the miniature cultivar Micro-Tom has been established (Meissner et al. 2000). Micro-Tom is well suited for large scale mutagenesis in tomato owing to its small size, rapid life cycle, easy transformability, and efficient activity of the Ac/Ds elements (Meissner et al. 1997). In addition, promoter trapping using firefly luciferase reporter gene and enhancer trapping using GUS was also developed. A high luciferase activity in the flowers, fruits and seedlings (69%) suggested preferential insertion of Ds elements into genes. Data was also supported by flanking sequence analysis. This well initiated high-throughput functional analysis of genes and promoter isolation in tomato was complemented by a T-DNA insertional mutagenesis programme using activation tagging to identify genes that regulate metabolic pathways, which produced >10,000 independent activation tagged transgenic tomato lines and identified large number of leaf and fruit colour mutants. Analysis of an activation tagged insertion line with intense purple pigmentation led to cloning of ANT1 gene, a transcriptional regulator of anthocyanin biosynthesis, modification and transport (Mathews et al. 2003).

T-DNA tagging strategy for isolation of new promoters has been tried in *Brassica napus* employing hypocotyl transformation with a promoterless gus: nptII tagging construct. This study led to the isolation of new constitutive promoters suitable to drive expression of transgenes in crops as well callus specific promoters that can be used to drive selectable marker genes enabling selection in early phase of transformation procedure and eliminating marker gene products from mature transgenic (Bade et al. 2003). Considering the high efficiency gained by the reverse genetic programs it appears likely that all of the 20,000 to 25,000 genes that make up the basic angiosperm genome will be assigned function on the basis of experimental evidence by the turn of the decade. Understanding the molecular basis of developmental and metabolic diversity among angiosperms is the key to effective and rational improvements in the productivity and utility of crop species (Somerville and Somerville, 1999).

**CONCLUDING REMARKS**

The advent of high through put DNA sequencing has meant that the ability to identify genes has far outstripped the ability to determine gene function. The consequently, 90% of the predicted genes in the *Arabidopsis* genome are of unknown function. Therefore, an important companion to gene identification are functional genomic investigations aimed at determining the pattern of the expression of genes in the whole organism coupled with assigning functions to individual genes. The use of mutagenesis to find and study plant genes is increasingly being used in functional genomics. T-DNA insertion mutations are a valuable resource for studies of gene function in *Arabidopsis*, serving as the basis for both forward and reverse genetic strategies and as a source of sequence tags from large collections of mutant lines. However mutagenesis allows only partial analysis of genes that function at multiple stages of development, genes that are functionally redundant or highly pleiotropic. Additionally it is difficult to identify genes required in the gametophytic generation. Traps and reporter genes allows for the discovery and analysis of these types of genes also.

Therefore use of activation elements, enhancer traps and promoter traps in gene tagging studies is complementary to loss of function studies, because it provides the opportunity to generate new types of gain of function mutants. Thus promoter traps and enhancer traps provides a means of identifying genes, and characterizing in vivo, their

---

**Figure 8. Schematic representation of adaptor PCR.** Specific adaptor molecules are ligated to the ends of the restricted DNA fragments carrying part of the insert. The restricted fragments are end filled and subjected to PCR using primers specific to the adaptor region and nested primers specific to T-DNA. The flanking plant DNA region is represented by line (green). The adaptor molecules are shown in red. The adaptor specific primer is shown in blue arrow. The T-DNA specific nested primers (P1, P2 and P3) are indicated by red, green and purple arrows, respectively.
expression patterns, throughout the plant life. Activation
tags containing enhancer elements provoke tissue specific
up regulation of activation tagged genes which potentially
reveals more about the function of the gene than constitutive
over expression. In this way even redundant
genes might display an over expression phenotype, if their
product is limiting or a change in concentration of gene
products creates an imbalance that is manifested as a
phenotype. These phenotypes can either directly reveal
gene function or provide a clue to the pathway in which the
concerned gene is involved. Moreover it might be possible
to study the function of essential genes that display lethal
knockout phenotypes by using these special vectors. The
promise of this technology as illustrated by the cloning and
characterization of several genes in Arabidopsis and the
world wide efforts in progress should enable coverage of this
model genome in a few years to come. Interestingly, T-
DNA mutational approach is now being successfully
modified to tag genes in a number of economically
important plant species, rice (430 Mbp) (Hiroi yuki et al.
1999; Chen et al. 2003; Kolesnik et al. 2004; Sha et al.
2004), tomato (953 Mb) (Meissner et al. 2000), Brassica
napus (1182 Mbp) (Bade et al. 2003), Medicago truncatula
(~454 to 526 Mbp)(Trieu et al. 2000) and poplar (550 Mbp)
(Groover et al. 2004) (550 Mbp). Thus T-DNA tagging in
conjunction with other mutation based techniques like
transposon insertion, TILLING etc., would not only
continue to provide useful information in Arabidopsis but is
likely to prove an efficient tool for functional genomics in
other plants also.

REFERENCES

ABRAHAMS, S.; LEE, E.; WALKER, A.R.; TANNER,
G.J.; LARKIN, P.J. and ASHTON, A.R. The Arabidopsis
TDS4 gene encodes leucoanthocyanidin dioxygenase
(LDOX) and is essential for proanthocyanidin synthesis and
vacuole development. The Plant Journal, September 2003,
vol. 35, no. 5, p. 624-636.

AGI. The Arabidopsis Genome Initiative. Analysis of

ALONSO, J.M.; STEPANOVA, A.N.; LEISSE, T.J.; KIM,
C.J.; CHEN, H.; SHINN, P.; STEVENSON, D.K.;
ZIMMERMAN, J.; BARAJAS, P.; CHEUK, R.;
GADRINAB, C.; DUBREUCQ, B.; CHAUVIN, S.; LE-
CLAIMINCE, I.; LE BOULAIRE, F.; DEROSE, R.;
SAMSON, F.; BIAUDET, Y.; LECHARNY, A.;
CRUAUD, C.; WISSENBACH, J.; CABOCHE, M. and
LEPINIEC, L. Improved PCR-walking for large scale
isolation of genes and their flanking regions from expressed
4, p. 240-246.

ANTAL, Z.; RASCLE, C.; FEYRE, M. and BRUEL, C.
Single oligonucleotide nested PCR: a rapid method for the
isolation of genes and their flanking regions from expressed
4, p. 240-246.

ASHRAFI, K.; CHANG, F.Y.; WATTS, J.L.; FRASER,
A.G.; KAMATH, R.S.; AHRINGER, J. and Ruvkun, G.
Genome-wide RNAi analysis of Caenorhabditis elegans fat

AZPIROZ-LEEHN, R. and FELDMANN, K.A. T-DNA
insertion mutagenesis in Arabidopsis: going back and forth.

BADE, J.; VAN GRINSVEN, E.; CUSTERS, J.;
HOEKSTRA, S. and PONSTEIN, A. T-DNA tagging in
Brassica napus as an efficient tool for the isolation of new
promoters for selectable marker genes. Plant Molecular
Biology, May 2003, vol. 52, no. 1, p. 53-68.

BAE, M.S.; CHO, E.J. and PARK, O.K. Analysis of the Arabidopsis nuclear proteome and its
response to cold stress. The Plant Journal, December 2003,
vol. 36, no. 5, p. 652-663.

BALTER, M. Plant Science: data in key papers cannot be

BALZERGUE, S.; DUBREUCQ, B.; CHAUVIN, S.; LE-
CLAIMINCE, I.; LE BOULAIRE, F.; DEROSE, R.;
SAMSON, F.; BIAUDET, Y.; LECHARNY, A.;
CRUAUD, C.; WISSENBACH, J.; CABOCHE, M. and
LEPINIEC, L. Improved PCR-walking for large scale
isolation of plant T-DNA borders. Biotechniques, March

BARTHELS, N.; VAN DER LEE, F.M.; KLAPE, J.;
GODDIJN, O.J.; KARIMI, M.; PUZIO, P.; GRUNDLER,
ROBERTSON, W.M., VAN MONTAGU, M.; GHEYSEN,
G. and SUMONS, P.C. Regulatory sequences of
Arabidopsis drive reporter gene expression in nematode
12, p. 2119-2134.

BECKER, J.D.; BOAVIDA, L.C.; CARNEIRO, J.;
HAURY, M. and FEIJO, J.A. Transcriptional profiling of
Arabidopsis tissues reveals the unique characteristics of the
133, no. 2, p. 713-725.

ALVARADO, M.C.; ZSIGMOND, L.M.; KOVACS, I.;
CSEPLO, A.; KONCZ, C. and SZABADOS, L.M. Gene
trapping with firefly luciferase in Arabidopsis-Tagging of
stress-responsive genes. Plant Physiology, January 2004,
vol. 134, no. 1, p. 18-27.

BARTHELS, N.; VAN DER LEE, F.M.; KLAPE, J.;
GODDIJN, O.J.; KARIMI, M.; PUZIO, P.; GRUNDLER,
ROBERTSON, W.M., VAN MONTAGU, M.; GHEYSEN,
G. and SUMONS, P.C. Regulatory sequences of
Arabidopsis drive reporter gene expression in nematode
12, p. 2119-2134.


Radhamony, R. N. et al.


GROOVER, A.; FONTANA, R. J.; DUPPER, G.; MA, C.;


Radhamony, R. N. et al.


MOLLIER, P.; HOFFMAN, B.; ORSEL, M. and PELLETIER, G. Tagging of a cryptic promoter that confers


SCHOMBURG, F.M.; BIZZELL, C.M.; LEE, D.J.; ZEEVAART, J.A. and AMASINO, R.M. Over expression of a novel class of gibberellin 2-oxidases decreases...
T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics


Radhamony, R. N. et al.


T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics

Molecular Plant Microbe Interaction, June 2000, vol. 13, no. 6, p. 606-616.


