

# Tissue-specific activation tagging in *Arabidopsis thaliana* for identification and isolation of genes of agronomic importance

Kushwah Neetu Singh\*

ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi 110 012, India and  
ICAR-Indian Institute of Pulses Research, Kanpur 208 024, India

Activation tagging is used to recover and clone dominant gain-of-function alleles and usually employs a T-DNA vector containing four tandem copies of the CaMV35S enhancer sequence outward usually at the right border. This strategy, however, is not efficient as it could result in overexpression of multiple genes or genes far away from the insertion site. Therefore, we tested constitutive (CaMV35S) and tissue-specific (*TFL1* and *rbcS2B*) promoters for activation tagging. From an initial screening of 400 T<sub>1</sub> plants, we identified several morphological variants which include seedling mutants (single cotyledon, slow seedling growth, long root, short root), leaf mutants (fused leaves, excess rosettes, altered shape, hyponastic leaves, bushy rosette, altered leaf polarity), inflorescence mutants (fasciated inflorescence, inflorescence fused with the petiole of rosette leaf) and flowering time and growth habit mutants (dwarf, tall, miniature, late flowering, sturdy stem). Comparison of different activation tagging populations for different categories of mutants revealed the pre-dominance of seedling mutants in the *TFL1* promoter activation tagged population; leaf and inflorescence mutant frequencies were higher in *rbcS2B* promoter activation tagged population, while the flowering time and growth habit mutant frequency was higher in CaMV35S promoter activation tagged population. Flanking sequence analysis of three of the mutants showed that all the mutants carried an insertion in the intergenic region. Segregation analysis of seedlings for kanamycin resistance showed that on average each mutant carried about 1.3 copies of T-DNA.

**Keywords:** Activation tagging, *Arabidopsis thaliana*, morphological variants, promoters.

DISCOVERY of new genes controlling various traits and metabolic pathways is important both for a basic understanding of various biological processes and for application in genetic engineering of plants. A variety of approaches such as conventional chemical or physical mutagenesis<sup>1</sup> and modern insertional mutagenesis, using

either transposable elements<sup>2</sup> or T-DNA of *Agrobacterium*<sup>3</sup> are currently available for the discovery of genes in different organisms. However, conventional and modern insertional mutagenesis methods yield mostly loss-of-function mutations or mutations with altered efficiency of encoded proteins. As most plants are polyploids, majority of such gene mutations fail to show phenotypes and thus go undetected. Even *Arabidopsis thaliana* with the smallest genome size among angiosperms shows a high degree of structural and functional redundancy; nearly 65% of genes have at least one paralogue, and duplicate segments make up about 60% of the genome<sup>4</sup>. Consequently, less than 2% of gene knockouts show clear mutant phenotype<sup>5</sup>. Therefore, mutational approaches are not efficient for gene discovery. Also, mutagenesis approach fails to recover gene mutations that play an essential role in cell metabolism, or genes that are essential during gametophyte or early embryo stage. Thus, assigning correct functions to a large number of genes remains a constraint. Gene overexpression is a novel strategy of gene discovery that overcomes the limitations posed by mutagenesis. Depending on the way it is employed, it could be a forward or reverse genetic approach. For instance, activation tagging utilizes transformation technology to insert at random enhancer or promoter sequences into the genome. If the promoter or enhancer lands next to a gene, it leads to overexpression of adjacent genes. In many instances, gene overexpression yields a mutant phenotype. Thus, unlike traditional mutagenesis, activation tagging gives dominant mutations that manifest in the T<sub>1</sub> generation itself. On the other hand, in the reverse genetic approach, a putative gene/orf is linked to a strong promoter and is used to transform the host organism or model species. If the organism shows a phenotype due to overexpression, the function could be assigned to the gene.

Activation tagging vectors generally employ enhancer elements for overexpression of neighbouring genes independent of their orientation relative to the genes and hence the chance of gene activation is high. CaMV35S enhancer is the most commonly employed element for activation tagging in plants. It has been reported that 4X35S enhancer strongly enhances endogenous gene expression pattern, rather than constitutive/ectopic overexpression<sup>6-8</sup>.

\*e-mail: neeturajawat@gmail.com

Hence, use of enhancer only reflects the native role of the gene. Further, the enhancer might activate one or more genes located in its vicinity, and thereby complicate analysis and interpretation<sup>9</sup>. Also, since enhancers can influence genes located far away and on either side of the T-DNA/transposon insertion, identification of activated gene/s poses additional challenges. A large number of activation tagged lines have been generated in *Arabidopsis* and rice using 35S enhancer elements<sup>8–13</sup>. However, due to the problems stated above, only a few genes have been characterized using this approach so far.

Activation tagging using tissue-specific or inducible promoters can overcome the above-mentioned problems. For example, genes whose expression causes cell/tissue death could be identified by the inducible expression system. Further, one could target specific types of mutants by choosing appropriate promoters for gene activation. Tissue-specific/inducible activation tagging can generate novel, dominant gain-of-function mutations through neofunctionalization and thereby uncover novel utility to known genes. For example, using  $\beta$ -estradiol-inducible promoter for activation tagging, Zuo *et al.*<sup>14</sup> isolated a mutant with increased frequency of *in vitro* somatic embryogenesis. Further analysis of the mutant revealed overexpression of *WUSCHEL* gene known to be involved in conferring stem cell identity in the shoot apical meristem.

A variety of stress- or chemical-inducible promoters like 17- $\beta$ -estradiol hormone-inducible promoter<sup>14</sup>, heat shock promoter *HSP18.2* (ref. 15), pathogen-inducible promoter *PRI* (ref. 16), sugar-inducible promoter Spomin<sup>17</sup> and drought-inducible promoter *RD29a*<sup>18</sup> have been used in the past for activation tagging. However, there are no reports on the use of developmentally regulated/tissue-specific promoters like *TERMINAL FLOWER1 (TFL1)* and Rubisco small subunit 2B (*rbcS2B*) gene promoter in activation tagging. We selected the *TFL1* promoter as *TFL1* gene is predominantly expressed in the shoot apical meristem (SAM) and *TFL1* expression is found in SAM within 3–4 days after sowing<sup>19</sup>. Hence use of this promoter in activation tagging is expected to give SAM mutants. As SAM is involved in plant development, many developmentally defective mutants are expected in this screen. *TFL1* promoter of *Arabidopsis* has been reported to contain distinct *cis*-regulatory elements required for its correct spatio-temporal expression at the *Arabidopsis* shoot apex. Unlike other protein-coding gene promoters, the *cis*-regulatory elements of *TFL1* promoter are located at the both 5'- and 3'-intergenic regions. The 5'-intergenic region has been reported to contain important *cis*-regulatory elements required for its high expression level in SAM, whereas 3'-intergenic region contains distinct *cis*-regulatory elements required for its expression in the vegetative SAM, main and lateral inflorescence meristem, upregulation and maintenance of its expression in the inflorescence meristem, and for repression of its

expression in the lateral organs like leaves and flowers. Deletion of the 3'-intergenic region of *TFL1* promoter has been reported to retain its consistent and stable expression, but its expression becomes more diverse such as in leaves and flowers where *TFL1* is not normally expressed<sup>20</sup>. The present study was started before the *TFL1* promoter was characterized in detail. Following the general notion that the promoter of most of the protein coding genes lie in the upstream intergenic region of the genes, we used only 5'-intergenic region of the *TFL1* gene for activation tagging. Nevertheless, the chosen sequence was capable of driving gene expression in SAM<sup>21</sup>. According to the eFP browser, *TFL1* appears to be a weak promoter as maximum absolute expression value of the *TFL1* gene is 93.97 ([Supplementary Figure 1](#)), which is very low. Hence, it is less likely that it could give mutants in the organs like leaves and flowers. However, shoot apical meristem contains stem cells from which other organs like leaves and flowers arise. Hence, even a very small change in gene expression caused by activation tagging in SAM could bring about major changes in the plant organ development. The choice of using *rbcS2B* promoter in this study was due to the fact that it is a strong green tissue-specific promoter and hence is expected to give a wide spectrum of mutations, similar to CaMV35S promoter. Also, chances of getting leaf mutations will be high as it drives high-level gene expression in the leaves. Promoters of *rbcS* gene family have been well characterized in many plant species<sup>22</sup>, and are found to contain important *cis*-regulatory elements required for light and tissue-specific expression. I and G box sequences are the two important *cis*-regulatory elements of *rbcS* promoters, and are required for their light-regulated expression<sup>23</sup>. Promoters of *rbcS* gene family contain conserved modular array 5 (CMA5), composed of I and G boxes, that have been shown to act as an enhancer elements capable of activating heterologous minimal promoter in response to light and chloroplast-dependent signals<sup>24</sup>. CMA5 enhancer in *rbcS2B* promoter of *Arabidopsis* is located at –150 to –190 bp upstream of the transcription start site of the gene<sup>25</sup>. Upstream intergenic region of *rbcS2B* gene for promoter activity has been studied both by *GUS* and *GFP* reporter systems<sup>25,26</sup>. In both studies, expression of reporter gene was mainly found in the green tissues like leaves, flowers, stems, siliques, etc. indicating that *rbcS2B* promoter is green tissue-specific. Kim *et al.*<sup>27</sup> examined the tissue specificity of *rbcS2B* promoter using *GFP* reporter gene and found its expression only in leaf mesophyll and guard cells of the epidermis in transgenic *Arabidopsis* leaves. eFP browser data show that the maximum absolute expression value of *rbcS2B* gene is 6639.56 ([Supplementary Figure 2a](#)), indicating the strength of the *rbcS2B* promoter. eFP browser graph of expression values shows its highest expression in leaves ([Supplementary Figure 2b](#)), suggesting that the *rbcS2B* promoter in activation tagging is likely to give

more of leaf mutation. A CaMV35S promoter was also included as control to compare results with the novel strategies.

## Materials and methods

### *Construction of TFL1, rbcS2B and CaMV35S promoter-based plasmids*

*TFL1* (ref. 21), *rbcS2B* promoter<sup>25</sup> of *Arabidopsis* and CaMV35S promoter<sup>28</sup> were used for the construction of activation tagging vector. The forward and reverse primers of *TFL1* promoter were designed and sequences for *SalI* and *BamHI* restriction sites were included at the 5' end of the forward and reverse primers respectively (Table 1). PCR was performed using genomic DNA of *Arabidopsis* with the following reaction mixture: 10 µl 5× buffer, 1 µl 10 mM dNTP, 0.5 µl of each primer (10 µM), 0.5 µl Phusion polymerase, 1 µl DNA (100 ng/µl) and 36.5 µl H<sub>2</sub>O. PCR conditions were: initial denaturation at 98°C for 1 min followed by 36 cycles of denaturation at 98°C for 30 s, annealing at 52°C for 30 s and extension at 72°C for 90 s, and a final extension at 72°C for 10 min. PCR product was cloned into pGEMT-easy vector (Promega, Wisconsin, USA). Clone identity was confirmed by restriction digestion and sequencing. The *TFL1* promoter fragment (2536 bp) was released by *BamHI* and *SalI* restriction enzymes from pGEMT-easy vector and cloned into the binary vector pBinGFP trap replacing GFP to make the pBinTFL1 activation tag vector. For construction of pBinrbcS2B binary vector, blunt end cloning was performed. In a previous study, we have cloned the *rbcS2B* promoter (1817 bp) in pGEMT-easy vector<sup>25</sup>. Here, we released the promoter from pGEMT-easy vector using *NotI* restriction enzyme and GFP from pBinGFP trap vector using *BamHI* and *SalI* restriction enzymes. pBinGFP trap vector as well as *rbcS2B* promoter fragments were gel-eluted and blunt-ended using Klenow fragment (Epicentre®, Madison, Wisconsin, USA). The blunt-ended vector and *rbcS2B* promoter fragments were ligated with T4 DNA ligase (NEB, Massachusetts, USA), and the ligated product was used for transformation of *Escherichia coli*. Clone identity was confirmed by restriction digestion and PCR. For construction of pBin35S plasmid, CaMV35S promoter (835 bp) was released from the pBI121 vector using *HindIII* and *BamHI* restriction enzymes and subcloned into the pBinGFP trap vector predigested with the same enzymes which replaced GFP with the 35S promoter. All the plasmids were mobilized into *Agrobacterium* strain GV3101 separately by freeze–thaw method and confirmed by PCR.

*In planta* transformation of *Arabidopsis*: *A. thaliana* ecotype Columbia plants were grown under controlled condition at 23° ± 2°C day temperature/18°C night temperature under long day conditions (16 h) and used for *in*

*planta* transformation according to the floral dip protocol<sup>29</sup>. T<sub>1</sub> seedlings were selected on semi-solid MS medium containing kanamycin (50 mg/l). Kanamycin resistant putative transformants were transferred to pots for further growth and molecular analysis.

### *Screening of constitutive or tissue-specific promoter activation tagged population for isolation of mutants*

To identify visible mutants, about 400 kanamycin-resistant T<sub>1</sub> plants of pBin35S, pBinTFL1 and pBinrbcS2B tagged population (~150/construct) were observed throughout development for any sign of deviation in morphology from the wild-type *Arabidopsis*, especially for seedling development, leaf shape, plant height, flowering time and overall growth.

### *Isolation of T-DNA flanking sequences by TAIL PCR/adaptor ligation-mediated PCR*

Thermal asymmetric interlaced (TAIL) PCR<sup>30</sup> was performed to amplify the T-DNA flanking plant sequence in 35S-79 and 35S-84 mutants. For this, a set of nested CaMV35S promoter-specific primers along with six short arbitrary degenerate (AD) primers (ADn1–ADn6) of 15- or 16-mers with low melting temperature were used. The forward primers (TR1, TR2 and TR3) were designed from the CaMV35S promoter region in such a manner that they were as close as possible to the right border repeat sequence, and the TR3 primer was designed very close to the right border (27 nucleotide upstream from the right border). Distance between TR2 and TR3 was kept 57 bp apart to facilitate confirmation of product specificity by size comparison in gel electrophoresis. Genomic DNA isolated from the activation tagged mutant lines was used in the primary reaction (25 µl) of TAIL PCR. The PCR reaction mixture contained 2.5 µl of 10× PCR buffer, 200 µM each of dNTPs, about 100 ng of genomic DNA, 1.5 units of *Taq* DNA polymerase (Bangalore Genei), 0.2–0.25 µM TR1 and 4 µM of a given ADn primer. The reaction was performed according to the thermal programme described by Liu *et al.*<sup>30</sup> (Biometra T-Gradient Thermocycler). PCR products of primary TAIL-PCR reaction were diluted to 1:49 and 1 µl of 50-fold diluted PCR product was chosen as template for the secondary TAIL-PCR reaction, which contained all the components of primary PCR except for substitution of TR1 primer with TR2. The secondary TAIL-PCR product was diluted 50-fold and 1 µl aliquot was taken as template for the tertiary TAIL-PCR reaction. This reaction was set-up in 50 µl using the third nested TR3 primer. All three nested PCR products were subjected to electrophoresis in 1% agarose gel. A consistent size difference between secondary and tertiary PCR products according to the primer positions indicated the product to be

insertion-specific. Putative insertion-specific bands were eluted from the gel and cloned into the pGEMT-easy vector (Promega, USA) for sequencing. The sequencing results were analysed using TAIR-BLASTN program to identify the location of T-DNA insertion sites in the *A. thaliana* genome. For isolation of T-DNA flanking plant sequence in 35S-5 mutant, adapter-ligation mediated PCR was followed according to the protocol described by O'Malley *et al.*<sup>31</sup>. T-DNA-specific primers TR1 and TR2 were the same as that used in TAIL PCR.

#### Confirmation of predicted T-DNA insertion site in the mutants

PCR was performed to confirm the insertion sites identified in different mutants by TAIL PCR and adapter ligation-mediated PCR. For this, primers were designed from the predicted T-DNA flanking plant sequences. PCR was performed using one primer from the T-DNA 35S promoter region (TR2) and another primer from the predicted T-DNA insertion site (At5g-R for 35S-5 mutant, At2g-F for 35S-84 mutant, and At3g-F for 35S-79 mutant) (Figure 3) with the following reaction mixture: 2.5  $\mu$ l of 10 $\times$  PCR buffer, 0.5  $\mu$ l of 10 mM dNTP, 1  $\mu$ l of each primer (10  $\mu$ M), 0.5  $\mu$ l *Taq* DNA polymerase (3U/ $\mu$ l) 1  $\mu$ l of DNA (100 ng/ $\mu$ l), 19.5  $\mu$ l of H<sub>2</sub>O, with final volume of 25  $\mu$ l. PCR program was initial denaturation at 94°C for 5 min followed by 36 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1 min, and final extension at 72°C for 10 min. PCR product was analysed through gel electrophoresis and UV images of the gel were captured using Alpha Innotech (Fluorchem® FC2) gel documentation system.

#### Estimation of number of independent T-DNA insertions in individual T<sub>1</sub> plants

To estimate the number of independent T-DNA insertions in the transformants, T<sub>2</sub> seeds of individual transformants were surface-sterilized, stratified and plated on MS medium containing kanamycin (50 mg/l). The number of resistant and sensitive seedlings was counted 15 days after germination. Data were tested for one, two or multiple T-DNA copies using  $\chi^2$  test.

#### Total RNA isolation and real-time PCR analysis

Total RNA was isolated from leaves of five-week-old 35S-84 mutant and wild type *Arabidopsis* plants using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Saint Louis, USA). On column DNase I treatment (ON-COLUMN DNASE I DIGESTION SET, Sigma-Aldrich, Saint Louis, USA) was done to remove genomic DNA

from the RNA preparation. First-strand cDNA synthesis was performed using SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen, Life Technologies, USA). qRT-PCR was performed with KAPA SYBR® FAST Universal 2 $\times$  qPCR Master Mix (KAPA Biosystems) in a Stratagene MX3000P qPCR instrument. PCR reaction was performed with the following programme: 95°C for 10 s, 45 cycles of 95°C for 30 s, 60°C for 45 s and 72°C for 45 s, followed by dissociation curve programme: 95°C for 1 min, 55°C for 30 s, 95°C for 30 s to check for multiple amplicons. Transcript levels in mutant and wild type were normalized using *actin8* (*At1g49240*) as the internal control. Table 1 lists the primers used in qRT-PCR. The qRT-PCR data were analysed using 2<sup>- $\Delta\Delta$ Ct</sup> method<sup>32</sup>. Considering wild type as a calibrator, transcript level of *At2g37440* gene in 35S-84 mutant was estimated. qRT-PCR experiment was conducted with three technical replicates and one biological replicate.

## Results and discussion

#### Construction of activation tagging vectors

Previously, we have characterized *TFL1* and *rbcS2B* promoters of *Arabidopsis* using *GUS* reporter gene<sup>21,25</sup>. Here, we have chosen these promoters along with CaMV35S promoter for the construction of activation tagging vectors<sup>28</sup>. All these promoters were separately placed adjacent to the right border (RB) of T-DNA in the pBinGFP promoter trap binary vector replacing GFP with these promoters (Figure 1). It has been reported that T-DNA flanking plant sequences near the right border of T-DNA are more precise whereas flanking plant sequences adjacent to the left border of T-DNA undergo extensive rearrangement during T-DNA integration<sup>33</sup>. Hence cloning of promoter near the right border of T-DNA could activate the nearby genes more precisely. Even the extensively used activation tagging vectors pSKI015 and pSKI0074 contain the outward facing tetramer of 35S-enhancer adjacent to the right border of T-DNA<sup>8</sup>.

#### Isolation of visible mutants in the T<sub>1</sub> generation

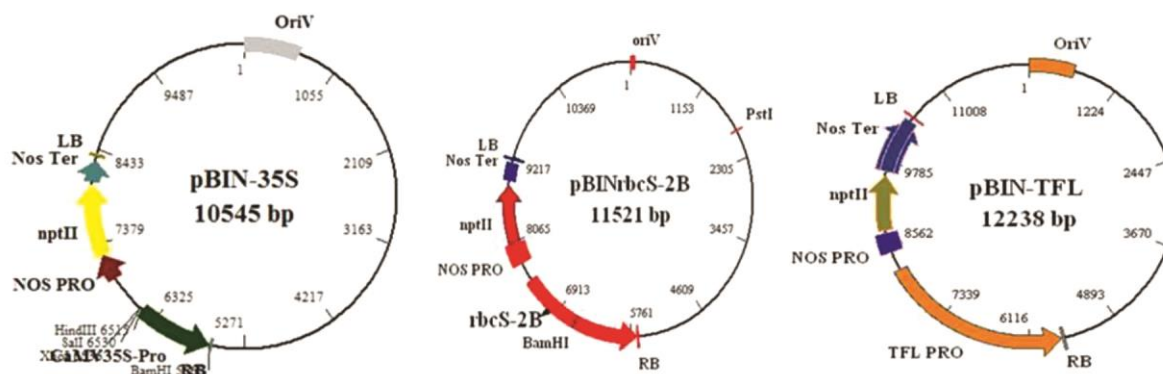
*A. thaliana* seeds were obtained following floral dip transformation using CaMV35S, *rbcS2B* and *TFL1* promoter-carrying constructs. The overall *in planta* transformation frequency was 0.66%, which is comparable to the values (0.5%–3%) reported by Clough and Bent<sup>29</sup>. Kanamycin-resistant plants were observed throughout development for any signs of deviation in morphology. About 400 T<sub>1</sub> plants (100–140 plants of each population) were examined in batches of 20–30 mutants. Since activation tagging generates mutant phenotypes due to

**Table 1.** Sequences of primers used in the experiments

Primer	Sequence (5'-3')
<i>TFL-F</i>	<u>GTCGACTATTCTCCATCTGTTCTACCTGAATC</u> <sup>#</sup>
<i>TFL-R</i>	<u>GGATCCTGTAACTTAGAGGAAAAGAGAGATTG</u> <sup>#</sup>
<i>rbcS2B-F</i>	ACTTCTTCTGTTGTTTCTCTTCTTTT
<i>rbcS2B-R</i>	CTTTACCCTAACTACTCTTCTCAGTTGGC
CaMV35SF	TTCAGAAAAGAATGCTAACCCACAGATGG
CaMV35SR	AGAGGAAGGGTCTTGC GAAGGATAG
TR1	TGGAAAAAGAAGACGTTCCAACCAC
TR2	CACGTCTCAAAGCAAGTGGATTG
TR3	CCCTTCTCTATATAAGGAAGTTCATTTC
ADn1	NTCGA(G/C)T(A/T)T(G/C)G(A/T)GTT
ADn2	NGTCGA(G/C)(A/T)GANA(A/T)GAA
ADn3	(A/T)GTGNAG(A/T)ANCANAGA
ADn4	TG(A/T)GNAG(A/T)ANCA(G/C)AGA
ADn5	AG(A/T)GNAG(A/T)ANCA(A/T)AGG
ADn6	TCTTICGNATCTTNGGA
Long strand of adapter	<u>GTAATACGACTCACTTAGGGCACGCGTGGTCGACGGCCCGGGCTGC</u> <sup>*</sup>
Short strand of adapter <i>EcoRI</i>	5'-Phosphate-AATTGCAGCCCG-amino C7-3' <sup>*</sup>
Adapter primer1 (AP1)	GTAATACGACTCACTATAGGGC <sup>*</sup>
At5g-R	CACCTTCACTGTGTAAAAGATAAAAACCC
At3g-F	GGTCATGATTGTGATTGGTGTGTTG
At2g-F	TTCCCGAAAGTTAACATGTTTGAAGAT
440 RealF	TGATCGGATATTGTGGAAAGGTGACG
440 RealR	TACCGGGTTGGACGATGATTTTG
<i>Act8F</i>	TGGCACCACCCGAGAGGAAAGT
<i>Act8R</i>	AATGTGATCCCGTCATGGAAACG

<sup>\*</sup>Adapter sequences designed according to O'Malley *et al.*<sup>31</sup>.

<sup>#</sup>Restriction sites added at the 5' end of primers is underlined (solid line).

**Figure 1.** Activation tagging vectors used in the experiments.

overexpression/ectopic expression of trapped genes, mutant phenotypes are expected to manifest in the T<sub>1</sub> generation itself. Therefore, we screened the activation tagged population in the T<sub>1</sub> generation. A wide range of mutant phenotypes were observed in the T<sub>1</sub> generation, and we categorized them into four groups, Group I – seedling mutants: this group consists of mutants observed on the selection plate (kanamycin). Group II – rosette and cauline leaf mutants. Group III – inflorescence mutants. Group IV – flowering time and growth habit mutants. Nine lines showed multiple phenotypes which were counted in more than one category (Table 2). A total of

33 lines were identified as abnormal lines. Among them, five belonged to group I, 17 to group II, three to group III and 16 to group IV. Frequency of seedling mutants was higher in pBinTFL1-tagged population (80% of total seedling mutants), leaf and inflorescence mutant frequencies were higher in the pBinrbcS2B-tagged population (47% of total leaf mutants, 66% of total inflorescence mutants), and flowering time and growth habit mutants frequency was higher in pBin35S activation tagged lines (62% of total flowering and growth habit mutants). Seeds from 15 of these mutant lines were germinated to observe mutant phenotypes in the T<sub>2</sub> generation. Only one line

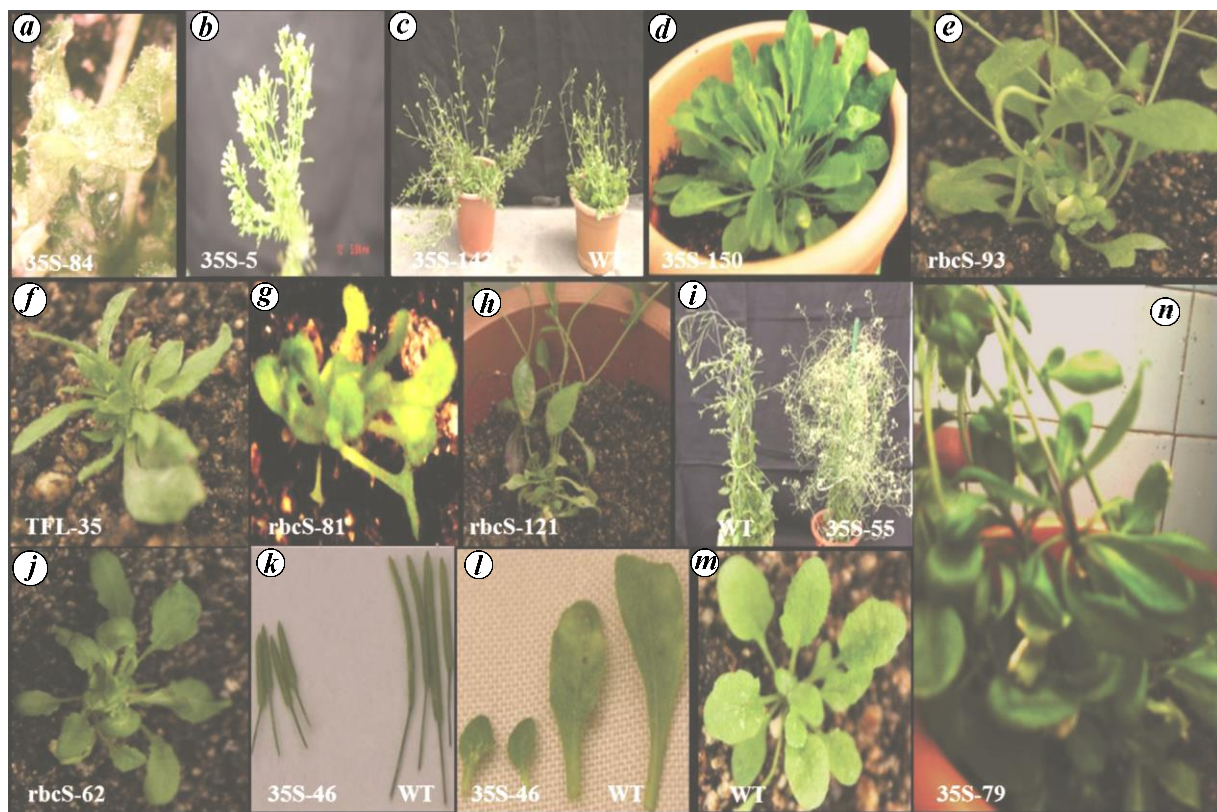
**Table 2.** Mutants observed in T<sub>1</sub> generation of activation-tagged *Arabidopsis*

Mutant	Group I (seedling mutation)	Group II (rosette and cauline leaves)	Group III (inflorescence)	Group IV (flowering and growth habit)
Vector : pBin35S				
# 5			Fasciated inflorescence (Figure 2)	
# 39				Dense vegetative growth, late flowering
# 43		Wide leaves		Dense vegetative growth, late flowering
# 46		Small leaves		Short siliques (Figure 2)
# 55				Bushy, late seed set, short siliques, partial sterile (Figure 2)
# 57				Tall
# 63		Multiple rosettes		
# 65		Curled leaves, multiple rosettes		
# 68		Small leaves		More number of siliques
# 79		Hyponastic cauline leaves (Figure 2)		
# 84		Altered leaf dorsi- ventral polarity (Figure 2)		Late flowering
# 138	Single cotyledon, long root			
# 139				More number of auxillary inflorescences, plants looks like creepers
# 142				Increased number of leaves and rosette inflorescences (Figure 2)
# 150				Increased number of leaves, late flowering (Figure 2)
Vector : pBinrbcS2B				
# 25		Second juvenile leaf with bifurcated blade		
# 62		Leaves just before emergence of primary inflorescence grow vertically from the main axis and spirally rolled downward (Figure 2)		
# 64		Small rosette leaves		Dwarf (7 cm), short siliques
# 81		Leaf lamina of two adjacent rosette leaves fused at their tip (Figure 2)		
# 83			Primary inflorescence converted into leaf- like structure; auxillary buds proliferate and give rise to normal inflorescence	
# 92		Rosette leaves with variable shape		
# 93		Rosette leaves fused together making disc-shaped structure, cauline leaf of primary inflorescence also fuses with the rosette leaves		
# 105				Tall, longer siliques, more number of seeds
# 121		Abnormal rosette growth; some leaves arranged in pairs on the stem (Figure 2)	Sturdy stem	
# 143		Long petiole		Dense vegetative growth

(Contd)

**Table 2.** (Contd)

Mutant	Group I (seedling mutation)	Group II (rosette and cauline leaves)	Group III (inflorescence)	Group IV (flowering and growth habit)
Vector : pBinTFL				
# 35		Narrow and elongated leaf blade (Figure 2)		Late flowering
# 44		Dark green leaves		
# 47	Long root, slow growth			
# 52	Very short root, slow growth			
# 56	Slow seedling growth			
# 58				Late flowering
# 64	Long root, very slow growing with only first two true leaves after 21 days			
# 76				More number of leaves, late flowering



**Figure 2.** Mutant phenotypes observed in T<sub>1</sub> generation of activation tagged population. *a*, 35S-84 showing altered leaf dorsi-ventral polarity. *b*, 35S-5 showing fasciated inflorescence. *c*, 35S-142 showing increased number of leaves and inflorescences compared to wild type. *d*, 35S-150: showing more number of leaves. *e*, rbcS-93 showing fusion of rosette and some cauline leaves making disc-like structure. *f*, TFL-35 showing narrow and elongated leaves and late flowering. *g*, rbcS-81 showing leaf lamina of two adjacent rosette leaves fused at their tip. *h*, rbcS-121 showing disorganized rosette and sturdy stem. *i*, 35S-55 showing bushy appearance, short and aborted siliques compared to wild type. *j*, rbcS-62 showing lamina rolled downward. *k*, *l*, 35S-46 showing short siliques and leaves compared to wild type. *m*, WT showing normal wild type plant. *n*, 35S-79 showing hyponastic cauline leaves.

(#35S-84) showed mutant phenotype in a small percentage of progeny. Table 2 provides a detailed description of mutants and Figure 2 shows images of selected phenotypes.

Lack of repeatability of activation tagged mutants has already been reported by several researchers in the past. Weigel *et al.*<sup>8</sup> reported that only 50% of the T<sub>1</sub> mutants reappeared in the T<sub>2</sub> generation. Ichikawa *et al.*<sup>34</sup> reported

**Table 3.** Estimation of T-DNA copy number by genetic segregation for kanamycin resistance in T<sub>2</sub> generation of activation tagged lines

Line no.	Kanamycin resistance in T <sub>2</sub> generation		Segregation ratio tested	Calculated $\chi^2$ value	Significance at $P = 0.05$ and 1 d.f.	Number of insertions
	Resistant	Sensitive				
pBin35S-5	72	7	15:1	0.95	No	2
pBin35S-37	34	9	3:1	0.37	No	1
pBin35S-46	11	11	1:1	0	No	1
pBin35S-55	75	5	15:1	0	No	2
pBin35S-68	107	41	3:1	0.58	No	1
pBin35S-79	94	39	3:1	1.39	No	1
pBin35S-84	177	49	3:1	1.36	No	1
pBin35S-90	286	38	15:1	17.15	Yes	?
pBin35S-105	94	39	3:1	1.32	No	1
pB in35S-142	134	41	3:1	0.16	No	1
pBin35S-150	23	2	3:1	3.66	No	1
pBinrbcS-65	107	30	3:1	0.70	No	1
pBinrbcS-62	40	11	3:1	0.16	No	1
pBinrbcS-71	115	43	3:1	0.41	No	1
pBinrbcS-74	88	15	3:1	4	Yes	?
pBinrbcS-81	132	9	15:1	0.03	No	2
pBinrbcS-93	103	3	63:1	1.11	No	3
pBinrbcS-101	203	76	3:1	0.73	No	1
pBinrbcS-105	125	52	3:1	1.90	No	1
pBinTFL-35	171	7	15:1	1.60	No	2
pBinTFL-66	152	52	3:1	0.02	No	1
pBinTFL-81	13	3	3:1	0.33	No	1
pER16-38	131	20	15:1	13.4	Yes	?
pER16-118	216	7	63:1	3.64	No	3
Average						1.26

that seven out of 40 morphological mutants reappeared in the T<sub>2</sub> generation, and found both dominant/semi-dominant phenotypes as well as low penetrance phenotype, i.e. showing mutant phenotype in a small percentage of the progeny. They reasoned that this lack of consistency of appearance of mutant phenotype in the T<sub>1</sub> and T<sub>2</sub> generations may be due to transgene suppression or mischaracterization of the mutant phenotype caused by environmental condition, or stresses after transfer of seedlings from the selection plate to soil. Since activation-tagged mutants are likely to be unstable, one has to screen a large population to get a consistent phenotype which is stably inherited over generations.

#### Determination of T-DNA copy number in mutants through genetic segregation analysis

The number of T-DNA insertions in independent T<sub>1</sub> plants was assessed by recording kanamycin-susceptible/resistant progeny in the T<sub>2</sub> generation. Twenty-four primary transformants were randomly selected for analysis. Based on  $\chi^2$  test for one, two or three unlinked T-DNA insertions, 15 lines were found to carry T-DNA at one locus, five lines carried T-DNA at two loci, while one line had three independent copies of T-DNA. The remaining three lines deviated from the expected Mendelian

segregation ratio. Thus transgenic plants contained on an average 1.3 copies of T-DNA (Table 3). Our results are in agreement with those of McElver *et al.*<sup>35</sup>, who reported a mean of 1.5 unlinked T-DNA insertions per line. Similar frequency of T-DNA insertion per event (1.35 insertions/line) has been reported by Robinson *et al.*<sup>11</sup> based on segregation data of 100 T<sub>1</sub> plants in a 50,000 activation tagged population. Using a more robust Southern hybridization approach, Ichikawa *et al.*<sup>34</sup> found an average of 2.6 T-DNA copies per *Arabidopsis* FOX lines which is expected as genetic segregation data fail to identify closely linked T-DNA insertions.

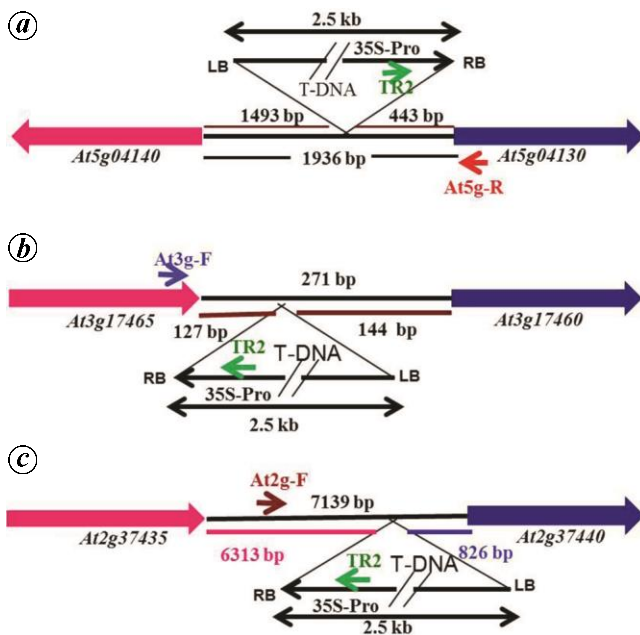
#### Isolation of T-DNA flanking plant sequence

We selected three mutants for the isolation of T-DNA flanking sequences. TAIL PCR<sup>30</sup> and adapter ligation-mediated PCR<sup>31</sup> were employed for the identification of T-DNA flanking sequences. All three mutants were found to carry T-DNA insertion in the intergenic region (Table 4 and Figure 3) and in each mutant, at least one copy of T-DNA insertion was precise, except for a minor rearrangement of sequences at the junction of the right border (RB) repeat sequences with the host genomic sequence (Table 3). Further, PCR was performed using T-DNA-specific (TR2) and flanking genomic region-specific



**Table 4.** Mutants and their corresponding genes

Mutant	Number of insertions identified through PCR	Structural features of T-DNA insertion sites	Insertion location	Adjacent genome region	Annotations in database
pBin35S-5	One	RB repeat sequences highly rearranged; incorporation of 19 nucleotide sequences of unknown origin	Intergenic region	Between <i>At5g04140</i> and <i>At5g04130</i>	Between DNA GYRASE B2 and <i>GLU1</i> gene
pBin35S-79	One	RB repeat sequences deleted completely	Intergenic region	Between <i>At3g17460</i> and <i>At3g17465</i>	Between PHD finger family protein and putative L3 ribosomal protein (RPL3P)
pBin35S-84	Two	RB repeat sequences deleted, inserted 20 bp filler sequence shows 8 bp duplication of adjacent genomic region Imprecise insertion, carries vector backbone	Intergenic region  Unknown	First insertion between <i>At2g37440</i> and <i>At2g37435</i>	Between DNase I-like superfamily protein and cystatin/monellin superfamily protein



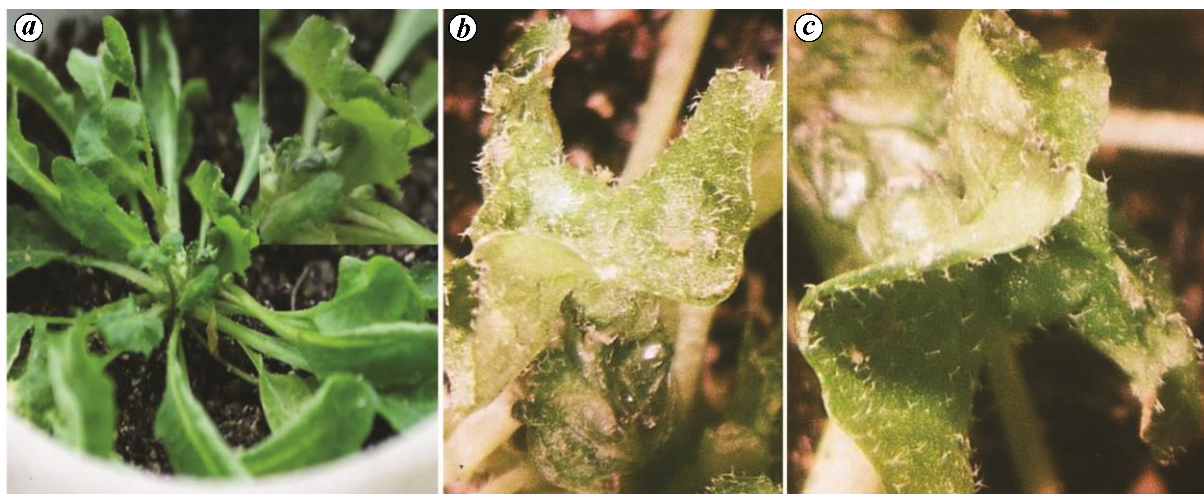
**Figure 3.** Schematic diagram showing location of T-DNA insertion in different mutants. *a*, 35S-5 mutant showing insertion in the intergenic region and 35S promoter orientation is towards the *At5g04130* gene. *b*, 35S-79 mutant showing T-DNA insertion in the intergenic region, but 35S promoter orientation is opposite to the transcription direction of both the adjacent genes. *c*, 35S-84 mutant showing insertion in the long intergenic region and 35S promoter is oriented opposite to both the adjacent genes. Positions of primers used to confirm the insertion are shown using arrows. p35S, CaMV35S promoter.

primers (Figure 3), which gave the expected size of band, confirming the T-DNA insertion sites. Ichikawa *et al.*<sup>9</sup> performed large-scale sequence characterization of 1172 activation lines that showed the phenotypes in the T<sub>1</sub> generation, and reported predominance of integration of T-DNA in the intergenic region (73%) than in the open reading frame (27%). Similarly, Li *et al.*<sup>36</sup> reported that T-DNA insertion in the GABI-Kat population (originally designed for activation tagging carrying CaMV35S pro-

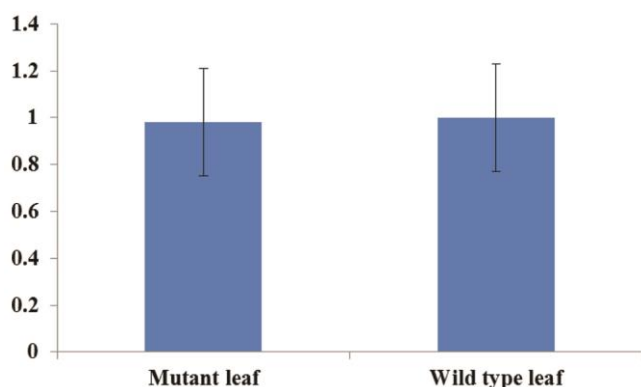
motor at right border of the T-DNA) showed preferential insertion around the transcription start site. The intergenic region or the region around the transcription start site appears most amenable for insertion of activation tagging vectors.

#### *Molecular characterization and expression analysis of a leaf mutant*

This mutant showed altered leaf dorsio-ventral polarity; in the basal half, the abaxial surface of leaf is transformed into adaxial surface (characterized by dark green, glossy and trichome rich-surface), while the apical half portion of the leaf shows abaxial character. Petiole is highly reduced in the affected leaf (Figure 4 *a-c*). In other words, the leaf appears to have twisted half way during development (Figure 4 *a*). This phenotype manifested in the rosette leaves just before the emergence of primary inflorescence. The mutant phenotype was stable. TAIL PCR identified T-DNA insertion in the intergenic region between *At2g37435* and *At2g37440* genes which are arranged in a tail-to-head fashion on chromosome 2 of *Arabidopsis* (Figure 3 *c*). The intergenic region is 7139 bp. T-DNA is located at 826 bp upstream of the transcription start site of *At2g37440* gene and 6313 bp downstream of the 3' UTR of *At2g37435* gene. The orientation of T-DNA in the intergenic region is such a way that its left border is towards the *At2g37440* gene and the right border towards the *At2g37435* gene. The T-DNA is 2562 bp, of which 35S promoter is located at 38 bp inside the right border. Therefore the distance of 35S promoter from the *At2g37440* transcription start site is 3350 bp (2524 bp + 826 bp gene sequence) and the distance of 35S promoter from the transcription start site of the *At2g37435* gene is 7418 bp (6313 bp + 1105 bp gene sequence). CaMV35S promoter was oriented opposite to the transcription direction of both genes (Figure 3 *c*).



**Figure 4.** pBin35S-84 mutant showing altered leaf dorsi-ventral polarity. *a*, Whole plant showing abnormal leaf adjacent to primary inflorescence. (Inset) Abnormal leaf. *b*, Close-up view of abnormal leaf showing adaxial character in the basal half. *c*, Close-up view of a leaf showing normal abaxial character in the apical half.



**Figure 5.** qRT-PCR of *At2g37440* gene in leaves of mutant and wild type *Arabidopsis*. mRNA expression level was normalized with *actin8* gene. Considering wild type as a calibrator, fold change in mutant line was calculated using  $2^{-\Delta\Delta Ct}$  method. Bar indicates standard deviation calculated from three technical replicates.

Hence, overexpression was not expected in any of the genes. However, it is known that CaMV35S promoter contains a strong enhancer at  $-343$  to  $-90$  bp positions, which can activate the transcription of adjacent genes, independent of orientation and distance<sup>37</sup>. Therefore, we tested the overexpression of *At2g37440* gene as it is closer and downstream of CaMV35S promoter than *At2g37435* gene. Mutant leaves as well as normal wild-type leaves were used for RNA isolation and qRT-PCR was performed using gene-specific primers. RT-PCR results showed no significant difference in the transcript level of *At2g37440* gene in leaf tissue of mutant compared to wild type (Figure 5), indicating that proper orientation of the promoter is required in order to activate the gene. The leaf phenotype in this mutant may be due to another copy of T-DNA. The two copies of T-DNA in this mutant were confirmed by Southern blot (Supple-

mentary Figure 3). Insertion of T-DNA in the intergenic region does not always warrant overexpression of adjacent genes. For example, Ichikawa *et al.*<sup>9</sup> performed expression analysis of 20 genes in which 35S enhancer activation tag occurred in the intergenic region, of which only seven genes were found to be overexpressed. To study the activation tag mutant, we need to carefully analyse the promoter distance, its orientation with respect to the adjacent gene, T-DNA copy number and mutant stability before performing overexpression studies.

1. Kazama, Y., Hirano T., Saito, H., Liu, Y., Ohbu, S., Hayashi, Y. and Abe T., Characterization of highly efficient heavy-ion mutagenesis in *Arabidopsis thaliana*. *BMC Plant Biol.*, 2011, **11**, 161.
2. Sundaresan, V., Springer, P., Volpe, T., Harward, S., Jones, J. D. G., Dean, C. and Martienssen, R., Patterns of gene actions in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.*, 1995, **9**, 1797–1810.
3. Krysan, P. J., Young, J. C. and Sussman, M. R., T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, 1999, **11**, 2283–2290.
4. Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 2000, **408**, 796–815.
5. Bouche, N. and Bouchez, D., *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.*, 2001, **4**, 111–117.
6. Neff, M. M. *et al.*, BAS1: a gene regulating brassinosteroid levels and light responsiveness in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 15316–15323.
7. Van der Graaff, E., Dulk-Ras, A. D., Hooykaas, P. J. and Keller, B., Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*. *Development*, 2000, **127**, 4971–4980.
8. Weigel, D. *et al.*, Activation tagging in *Arabidopsis*. *Plant Physiol.*, 2000, **122**, 1003–1013.
9. Ichikawa, T. *et al.*, Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation. *Plant J.*, 2003, **36**, 421–429.
10. Nakazawa, M. *et al.*, Activation tagging, a novel tool to dissect the functions of a gene family. *Plant J.*, 2003, **34**, 741–750.

11. Robinson, S. J. *et al.*, An archived activation tagged population of *Arabidopsis thaliana* to facilitate forward genetics approaches. *BMC Plant Biol.*, 2009, **9**, 101.
12. Jeong, D. H. *et al.*, Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. *Plant J.*, 2006, **45**, 123–132.
13. Hsing, Y. I. *et al.*, A rice activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.*, 2007, **63**, 351–364.
14. Zuo, J., Niu, Q. W., Frugis, G. and Chua, N. H., The *WUSCHEL* gene promotes vegetative-to-embryonic transition in *Arabidopsis*. *Plant J.*, 2002, **30**, 349–359.
15. Matsuhara, S., Jingu, F., Takahashi, T. and Komeda, Y., Heat-shock tagging: a simple method for expression and isolation of plant genome DNA flanked by T-DNA insertions. *Plant J.*, 2000, **22**, 79–86.
16. Grant, J. J., Chini, A., Basu, D. and Loake, G. J., Targeted activation tagging of the *Arabidopsis* NBS-LRR gene, *ADR1*, conveys resistance to virulent pathogens. *Mol. Plant–Microbe Interact.*, 2003, **16**, 669–680.
17. Masaki, T., Tsukagoshi, H., Mitsui, N., Nishii, T., Hattori, T., Morikami, A. and Nakamura, K., Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar inducible genes in *Arabidopsis thaliana*. *Plant J.*, 2005, **43**, 142–152.
18. Koiwa, H., Bressan, R. A. and Hasegawa, P. M., Identification of plant stress-responsive determinants in *Arabidopsis* by large-scale forward genetic screens. *J. Exp. Bot.*, 2006, **57**(5), 1119–1128.
19. Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R. and Coen, E., Inflorescence commitment and architecture in *Arabidopsis*. *Science*, 1997, **275**(5296), 80–83.
20. Serrano-Mislata, A., Fernandez-Nohales, P., Domenech, M. J., Hanzawa, Y., Bradley, D. and Madueno, F., Separate elements of the TERMINAL FLOWER 1 *cis*-regulatory region integrate pathways to control flowering time and shoot meristem identity. *Development*, 2016, **143**, 3315–3327.
21. Kushwah, N. S., Ahmad, I. and Ali, S., Characterization of promoter of Terminal Flower1 (*TFL1*) gene of *Arabidopsis*. *Res. J. Biotechnol.*, 2014, **9**(3), 35–40.
22. Bakhsh, A. and Husnain, T., Endeavours of Rubisco small subunit promoter as a tool of green tissue specific expression. *Czech. J. Genet. Plant Breed.*, 2012, **48**, 1–9.
23. Donald, R. G. and Cashmore, A. R., Mutation of either G box or I box sequences profoundly affects expression from the *Arabidopsis* *rbcs-1A* promoter. *EMBO J.*, 1990, **9**, 1717–1726.
24. Martinez-Hernandez, A., Lopez-Ochoa, L., Arguello-Astorga, G. and Herrera-Estrella, L., Functional properties and regulatory complexity of a minimal RBCS light-responsive unit activated by phytochrome, cryptochrome, and plastid signals. *Plant Physiol.*, 2002, **128**, 1223–1233.
25. Kushwah, N. S., Isolation, cloning and characterization of promoter of Rubisco small subunit 2B (*rbcs2B*) gene of *Arabidopsis thaliana*. *Innov. Farm.*, 2016, **1**(4), 119–128.
26. Sawchuk, M. G., Donner, T. J., Head, P. and Scarpella, E., Unique and overlapping expression patterns among members of photosynthesis-associated nuclear gene families in *Arabidopsis*. *Plant Physiol.*, 2008, **148**, 1908–1924.
27. Kim, J. Y., Yuan, Z. and Jackson, D., Developmental regulation and significance of KNOX protein trafficking in *Arabidopsis*. *Development*, 2003, **130**, 4351–4362.
28. Jefferson, R. A., Kavanagh, T. and Bevan, M. W., GUS fusion:  $\beta$ -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.*, 1987, **6**, 3901–3907.
29. Clough, S. J. and Bent, A. F., Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.*, 1998, **16**, 735–743.
30. Liu, Y. G., Mitsukawa, N., Oosumi, T. and Whittier, R. F., Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.*, 1995, **8**, 457–463.
31. O'Malley, R. C., Alonso, J. M., Kim, C. J., Leisse, T. J. and Ecker, J. R., An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the *Arabidopsis* genome. *Nature*, 2007, **2**(11), 2910–2917.
32. Livak, K. J. and Schmittgen, T. D., Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods*, 2001, **25**, 402–408.
33. Gheysen, G., Van Montagu, M. and Zambryski, P., Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences. *Proc. Natl. Acad. Sci. USA*, 1987, **84**(17), 6169–6173.
34. Ichikawa, T. *et al.*, The FOX hunting system: an alternative gain-of-function gene hunting technique. *Plant J.*, 2006, **48**, 974–985.
35. McElver, J. *et al.*, Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics*, 2001, **159**(4), 1751–1763.
36. Li, Y., Rosso, M. G., Ulker, B. and Weisshaar, B., Analysis of T-DNA insertion site distribution patterns in *Arabidopsis thaliana* reveals special features of genes without insertions. *Genomics*, 2006, **87**, 645–652.
37. Kay, R., Chan, A., Daly, M. and McPherson, J., Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science*, 1987, **236**(4806), 1299–302.

ACKNOWLEDGEMENTS. I thank the Council of Scientific and Industrial Research, New Delhi for funds (JRF). I also thank the National Phytotron Facility, IARI, New Delhi, for providing space to raise *Arabidopsis* plants, and Dr S. R. Bhat for supervising the study and critically reviewing the manuscript.

Received 23 January 2017; revised accepted 28 May 2019

doi: 10.18520/cs/v117/i4/627-637