The Plant Journal (2003) **36**, 421–429 doi: 10.1046/j.1365-313X.2003.01876.x

TECHNICAL ADVANCE

Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation

Takanari Ichikawa[1,†], Miki Nakazawa[1,†], Mika Kawashima[1], Shu Muto[2], Kazushi Gohda[1], Kumiko Suzuki[1], Akie Ishikawa[1], Hiroko Kobayashi[1], Takeshi Yoshizumi[1], Yuko Tsumoto[1], Yumi Tsuhara[1], Haruko Iizumi[1], Yukiko Goto[1] and Minami Matsui[1,*]

[1]*Plant Function Exploration Team, Plant Functional Genomics Research Group, Genomic Sciences Center, RIKEN, 1-7-22 Suehiro-cho, Tsurumiku, Yokohama, Kanagawa 230-0045, Japan, and*

[2]*VALWAY Technology Center NEC Soft Ltd., 1-18-6 Shinkiba, Koto-ku, Tokyo 136-8608, Japan*

Received 8 April 2003; revised 28 July 2003; accepted 4 August 2003.

*For correspondence (fax +81 45 503 9586; e-mail minami@postman.riken.go.jp).

†These authors contributed equally to this study.

Summary

Plant genomic resources harbouring gain-of-function mutations remain rare, even though this type of mutation is believed to be one of the most useful for elucidating the function of unknown genes that have redundant partners in the genome. An activation-tagging T-DNA was introduced into the genome of *Arabidopsis* creating 55 431 independent transformed lines. Of these T1 lines, 1262 showed phenotypes different from those of wild-type plants. We called these lines 'AT1Ps' (activation T1 putants). The phenotypes observed include abnormalities in morphology, growth rate, plant colour, flowering time and fertility. Similar phenotypes re-appeared either in dominant or semi-dominant fashion in 17% of 177 AT2P plants tested. Plasmid rescue or an adaptor-PCR method was used to identify 1172 independent genomic loci of T-DNA integration sites in the AT1P plants. Mapping of the integration sites revealed that the chromosomal distribution of these sites is similar to that observed in conventional T-DNA knock-out lines, except that the intragenic type of integration is slightly lower (27%) in the AT1P plants compared to that observed in other random knock-out populations (30–35%). Ten AT2P lines that showed dominant phenotypes were chosen to monitor expression levels of genes adjacent to the T-DNA integration sites by RT-PCR. Activation was observed in 7 out of 17 of the adjacent genes detected. Genes located up to 8.2 kb away from the enhancer sequence were activated. One of the seven activated genes was located close to the left-border sequence of the T-DNA, having an estimated distance of 5.7 kb from the enhancer. Surprisingly, one gene, the first ATG of which is located 12 kb away from the enhancer, showed reduced mRNA accumulation in the tagged line. Application of the database generated to *Arabidopsis* functional genomics research is discussed. The sequence database of the 1172 loci from the AT1P plants is available (http://pfgweb.gsc.riken.go.jp/index.html).

Keywords: activation tagging, T-DNA, *Arabidopsis*, T1 generation, functional genomics, database.

Introduction

In the *Arabidopsis* post-sequencing era, much effort is being devoted to elucidating the functions of unknown genes (Somerville and Dangl, 2000). Approaches most commonly used in functional genomics are designed to knockout or alter the function of genes by use of a chemical mutagen, transposon or T-DNA insertions, which create mainly recessive mutations (Bouche and Bouchez, 2001). However, the limitation of these approaches became obvious when the *Arabidopsis* and rice genome projects revealed sequence redundancy of more than half of the

© 2003 Blackwell Publishing Ltd

genes in these genomes (Arabidopsis Genome Initiative, 2000; Feng *et al*., 2002; Sasaki *et al*., 2002; Vision *et al*., 2000; Yu *et al*., 2002). It is also not difficult to imagine that a gene unique in terms of its sequence may be functionally redundant. In order to circumvent these limitations, activation tagging can be adopted for functional genomic research in plants as an alternative tool to analyse the functions of redundant genes (Jeong *et al*., 2002; Marsch-Martinez *et al*., 2002; Walden *et al*., 1994; Weigel *et al*., 2000; Wilson *et al*., 1996). As activation tagging creates dominant or semi-dominant phenotypes, the effects of wild-type loci and functionally redundant genes can be ignored in the mutant lines. Another problem that occurs frequently during the screening of recessive mutants in conventionally tagged resources is that the mutant phenotype often does not segregate with the tag. In some populations, the rate of co-segregation of the phenotype and the tag is as low as 20–30% (Budziszewski *et al*., 2001). The reason for this low tagging rate is not known. However, it is well-documented that the integration of T-DNAs or transposons into the genome is often accompanied with an irritation of the genome, such as deletion or re-arrangement of genomic fragments unrelated to the integration site (Nacry *et al*., 1998; Rinehart *et al*., 1997). Such 'genomic scratches' could be good candidate loci, explaining non-cosegregating phenotypes, as homozygous lines are always needed to screen recessive mutants. Activation tagging populations might be better alternative resources for gene tagging approaches because, by screening for only dominant phenotypes, mutants created by the tag can be observed, and hence, the tagging rate may increase. In order to increase the number of gain-of-function mutants in *Arabidopsis* and to assess the utility of generating such lines, we have created 55 431 independent *Arabidopsis* lines transformed with a tetrameric enhancer T-DNA (Walden *et al*., 1994). A total of 1262 lines that showed a phenotype in the T1 generation were isolated with the aim of collecting only dominant mutants. DNA from these lines was analysed to identify the T-DNA positions in the genome. A database of 1172 tagged positions and adjacent candidate genes was created. Fifty-six per cent of candidate genes were transcriptionally upregulated in a defined population. Application of the database to *Arabidopsis* functional genomics research is discussed.

Results and discussion

Transformation and growth of the Arabidopsis *activation-tagging lines*

Arabidopsis thaliana ecotype Columbia (Col-0) and activation-tagging T-DNA vector pPCVICEn4HPT (Walden *et al*., 1994) were used for transformation by floral dipping (Clough and Bent, 1998). A total of 61 591 hygromycin-resistant T1 seedlings on culture plates were transplanted into soil resulting in 55 431 fertile plants. Twenty plants were randomly chosen for Southern blot analysis using the hygromycin gene as a probe. All these lines contained the hygromycin-resistant gene. On average, 2.1 T-DNA inserts per line were detected in the genome of these transgenic lines (data not shown).

Phenotypic screening for gain-of-function mutants in the T1 generation

During the growth of the 61 591 T1 plants, any visible phenotypes were scored. These phenotypes included abnormalities in morphology, growth rate, plant colour, flowering time and fertility (Table 1). A total of 1262 lines were isolated as abnormal lines. When seeds from 177 of these lines were germinated, 29 lines showed the same phenotypes as the parent plant in either a dominant or a semi-dominant fashion. Figure 1 shows one of the representative lines, Z007706, that maintained the dominant phenotype for five generations (Nakazawa *et al*., 2003). However, it is already known that in activation-tagged populations, phenotypes observed in T1 do not necessarily appear in the next generation, probably because of silencing of the target genes (Weigel *et al*., 2000). Therefore, we called these T1 lines that showed activation T1 putant (AT1P) phenotypes.

Hygromycin resistance segregation test

In order to check segregation of hygromycin resistance of AT2P plants, seeds from AT1P plants were germinated on hygromycin-containing basic agar medium (BAM; Nakazawa and Matsui, 2003). Figure 2 shows rate distribution of hygromycin-resistant seedlings of 61 AT2P lines. As can be expected from the average number of T-DNA inserts (2.1 per line), a peak of the resistance rate ranges between 70 and 100%. Surprisingly, there was another peak between 0 and 10%. When genomic PCR was performed on six of these apparently hygromycin-sensitive AT2P lines, using hygromycin gene-specific primers, only one line showed a band of the expected size (data not shown). As the plasmid rescue was successfully performed in the corresponding AT1P lines, these results indicate that the appearance of these hygromycin-sensitive lines might be derived from loss of the T-DNA rather than increased sensitivity towards hygromycin in the AT2P plants caused by their phenotypes.

Identification of the T-DNA insertion sites in the genome of Arabidopsis

Genomic DNA was extracted from all of the AT1P plants for either plasmid rescue or adaptor-PCR. T-DNA

Table 1 Category of phenotypes and number of AT1P lines that showed each phenotype

Rosette leaves		Plant height		Flowering time		Branching	
Pale green	61	Dwarf	190	Early	23	Bushy	133
Dark green	70	Tall	11	Late	32	HAD	16
Anthocyanin	18			No	9	Fasciation	14
Large	134	Cauline leaves				Short internode	45
Small	207	Small	8	Fertility		Long internode	13
Long petiole	19	Large	46	High	6	Others	25
Short petiole	28	Many	36	Low	217		
Epinastic	203	Margin shape	8	Sterile	164	Flowers	
Hyponastic	12					Petal	16
Margin shape	98					Pedicel	5
Round	156					Filament	1
Narrow	144					Sepal	3
Many	188						
Few	18						
Others	42						

Some lines showed several different phenotypes and were counted in each different category. In the column 'Rosette leaves', 'pale green' and 'dark green' indicate the phenotypes that showed significant difference in leaf colour as observed by eye compared to other plants. In the columns of 'Rosette leaves' and 'Cauline leaves', 'many' and 'few' indicate the phenotype of lines that developed more than double and less than half the amount of rosette leaf number compared to the wild type, respectively. In the columns 'Rosette leaves' and 'Branching', 'other' indicates phenotypes that did not fit to any other categories listed in each column. In the column 'Plant height', 'dwarf' indicates phenotype of the lines in which the height of mature plants is less than one-third of the wild-type height. In the column 'Flowering time', 'no' indicates phenotype that did not develop inflorescence. In the column 'Fertility', 'high' indicates the phenotype of lines that set seed more than double the amount of seed compared to the wild type, and 'low' indicates the phenotype of lines that set seed less than 10% of the wild-type seed yield. In the column of 'Branching', 'HAD' represents hyper-apical dominance phenotype.

primers close to the T-DNA left border were used to sequence the adjacent genomic sequences, as the T-DNA sequence close to the right border was prone to deletion between the right-border sequence and enhancer sequences (data not shown). BLASTN was used to localize the insertion positions using the Munich Information Center for Protein Sequences (MIPS) *A. thaliana* genome database. A total of 1172 independent genomic positions had complete matches with the *Arabidopsis* genome sequence. Twenty-three sites matched to the chloroplastic genome and four sites matched to the mitochondrion genome.

Distribution of the insertion sites in the genome

The 1172 positions were mapped on the *Arabidopsis* chromosomes (Figure 3). Distribution of the insertion sites is not random as has been pointed out by other researchers (Sessions *et al.*, 2002; Szabados *et al.*, 2002). Regions close to each centromere tend to lack integration events. There are some regions where insertion sites accumulate, such as 1.0–1.5 Mb of chromosome I, 16.0–16.5, 18.5–19.0 and 19.0–19.5 Mb of chromosome II, 20.0–20.5 Mb of chromosome III and 12.5–13.0 Mb of chromosome IV. Clustering within 0–0.5 Mb of chromosome II may be a consequence of the repeated sequence of ribosomal DNA. However, the pattern of the distribution was quite similar to that observed in other populations (Sessions *et al.*, 2002; Szabados *et al.*, 2002). Also the number of insertion sites on each chromosome is nearly the same. The insertion frequency was 9.5 on chromosome I, 10.0 on chromosome II, 10.0 on chromosome III, 8.7 on chromosome IV and 9.6 on chromosome V sites per Mb. In the line Z007706 (*iso-1D*; Nakazawa *et al.*, 2003), two T-DNA inserts were found in its genome: one in chromosome I and the other in chromosome II. These results indicate that there is no unique nature concerning the chromosomal distribution of the activation-tagging T-DNA, despite the fact that the isolation of the lines was performed according to the visible phenotypes. Types of insertion were categorized into two groups: group 1, where the T-DNA insertion occurred within an open-reading frame (317 cases); and group 2, where the insertion occurred in an intergenic region of the genome (855 cases). In the case of group 1 insertion, gene activation must have occurred within neighbouring genes and not in the inserted gene. The rate of incident of group 1 insertions is slightly lower (27%) in the AT1P plants compared with that observed in other random knockout populations (30–35%; Krysan *et al.*, 2002; Sessions *et al.*, 2002; Szabados *et al.*, 2002). This might be a consequence of the phenotypic screening because, if the target gene is to be activated successfully, the T-DNA should be inserted in the intergenic region of the gene, namely, a group 2 insertion. Therefore, screening for visible phenotypes might bias the isolation of group 2 mutations. Considering that 48% of the *Arabidopsis* genome is occupied with transcribed regions (Sessions *et al.*, 2002), there is a clear and general bias for T-DNA insertion into the intergenic region of the genome. This

Figure 1. Phenotype of Z007706 line.
(a) Phenotype observed in a T5 Z007060 plant. A wild-type (WT) and the Z007706 plants were grown for 48 days in soil. The phenotype co-segregated with hygromycin resistance and appeared in a dominant fashion.
(b) Enlargement of the T5 Z007706 plant.

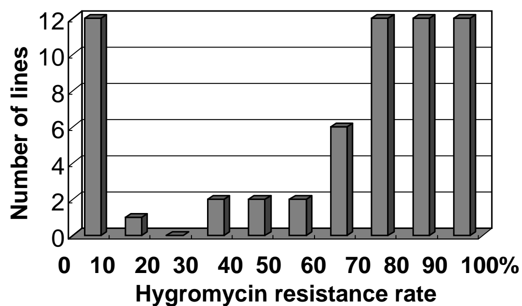

Figure 2. Distribution of hygromycin resistance rate of 61 AT2P seedlings. Sixty-one AT2P lines were selected, and approximately 40 seeds per line were sown on hygromycin-containing BAM. Percentage of seedlings surviving were scored and plotted on the histogram.

favours activation-tagging-type approaches for the same reasons as mentioned before.

Expression analysis of the genes adjacent to the activation tag

In order to monitor the perturbation of gene expression in the AT1P plants caused by the activation tag, 10 AT2P plants that maintained the same phenotype as the AT1P lines were chosen randomly. Insertion sites of the T-DNAs from these AT1P lines were used to deduce the two genes closest to the ends of the T-DNAs: gene R, the start codon of which is closest to the T-DNA right-border sequence; and gene L, the

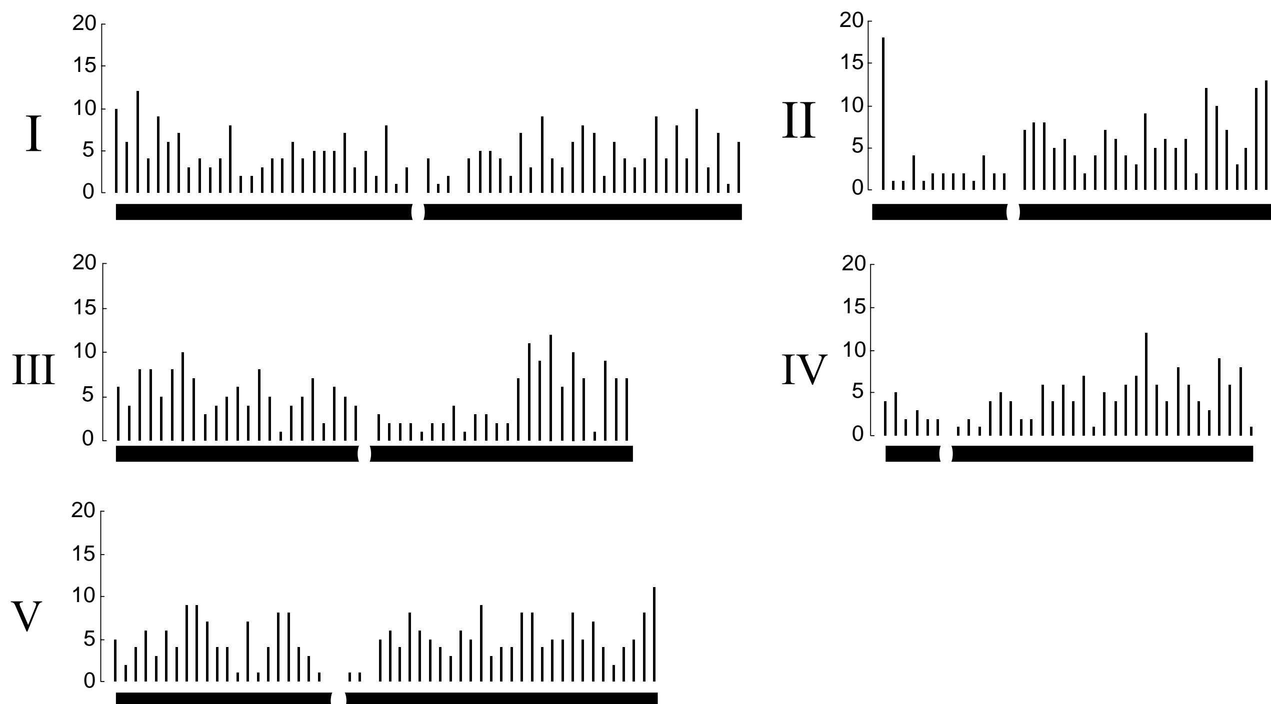

Figure 3. Mapping of 1172 T-DNA insertion positions on *Arabidopsis* chromosomes.
Genomic sequences connected to the T-DNA left border were subjected to BLAST analysis on the MIPS sequence database to obtain the absolute position on each chromosome. The number of insertions on every 0.5 Mb of chromosome was counted. The circles indicate the positions of centromeres.

start codon of which is closest to the T-DNA left-border sequence (Figure 4). None of these genes has the T-DNA integration site within the transcriptional unit. Gene-specific primers corresponding to these 20 genes were designed to allow semi-quantitative PCR to be performed. The distance of the ATG position calculated from the T-DNA enhancer sequence varies from gene to gene but is between 0.7 and 13 kb. RT-PCR products of 17 genes were detected at the predicted molecular weight. No PCR

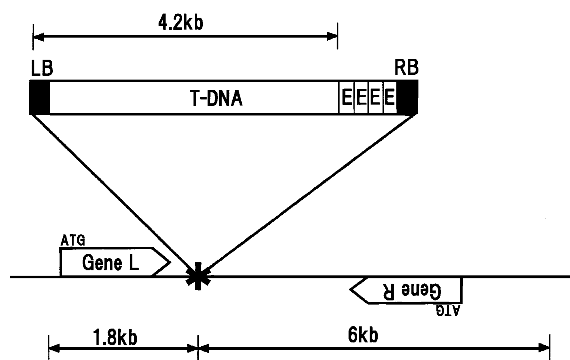

Figure 4. Schematic diagram of the T-DNA insertions in relation to adjacent genes.
Asterisk indicates the T-DNA insertion point in the genome. LB and RB represent the right- and left-border sequences of the T-DNA, respectively. E represents a single repeat of the enhancer sequence. Typical positions of Gene L and Gene R in relation to the T-DNA are indicated.

products were detected for three genes. Figure 5 shows the result of the RT-PCR, and the results are summarized in Table 2. Typical activation of the candidate genes was observed in 7 of the 17 genes for which a PCR product was detected (41%). The distances between the enhancer sequence and the ATGs of these activated genes were between 0.7 and 8.2 kb. One of these genes was located close to the left border of the T-DNA, resulting in an estimated distance of 5.7 kb from the enhancer. As a gene 8.2 kb away from the right-border sequence was successfully activated, it is not surprising that a gene located approximately 1 kb away from the left-border sequence corresponding to less than 6 kb away from the enhancer was activated. The average distance from the enhancer to successfully activated genes was 4.4 kb compared to 6.4 kb in the cases of lines that showed no difference or no band. Although activation by the tag can be observed at a distance of as much as 8 kb, these results indicate that the efficiency of the activation reversely correlates to the distance from the enhancer sequence, and that genes close to the left border can also be activated. Surprisingly, one gene (Z26013L in Figure 5(b)) that is located at a distance of about 12 kb from the enhancer showed reduced mRNA accumulation in the tagged line. This cannot be explained by the positional effect of the genomic region, as another candidate gene (Z26013R in Figure 5(c)) from the same line showed no difference.

One way to explain the gene suppression is the activation of antisense RNA enhanced by the enhancer sequence but driven by a promoter-like DNA sequence located on the corresponding DNA strand (antisense DNA). This observation suggests that the dominant or semi-dominant phenotype of the 'activation'-tagged line could be caused by not only activation but also the suppression of adjacent genes.

Table 2 Summary of the expression analysis

Name of lines	Right or left gene	Distance from the enhancer	Transcription
Z18548	R	0.7	Increased
Z49623	R	2.4	Increased
Z29218	R	2.7	Increased
Z88201	R	3.8	Increased
Z29732	L	5.7	Increased
Z23920	R	7.4	Increased
Z35240	R	8.2	Increased
Z26013	L	12.5	Decreased
Z26013	R	1.2	No difference
Z32731	R	1.8	No difference
Z21637	L	4.9	No difference
Z18548	L	5.3	No difference
Z88201	L	6.6	No difference
Z32731	L	6.7	No difference
Z29732	R	6.9	No difference
Z23920	L	10.3	No difference
Z29218	L	13.6	No difference

Names of lines, relative position of the target gene to the border sequences (whether it is close to the right-border sequence: R, or the left-border sequence: L), estimated distance between the enhancer sequence and the first ATG in kb, 'increased'; 'decreased' and 'no difference' indicate change in transcript level of the corresponding genes in each line compared to those in the wild type.

Integrated sequence and annotation database from AT1P plants

Although activation of adjacent genes located more than 8 kb away from the enhancer were achieved, the average successful gene activation was observed at a distance of 4.4 kb from the enhancer sequence. Therefore, all the annotated genes that have their initiation codons within 6 kb from the enhancer sequence were registered as activation candidate genes in our activation-tagging database (http://pfgweb.gsc.riken.go.jp/index.html). This was performed by searching all genes in the following regions of the MIPS database: the genomic region from the integration site and and up to 6 kb away from the T-DNA right-border sequence, and the genomic region form the insertion site up to 1.8 kb from the T-DNA left-border sequence. Such unsymmetrical genomic scannings from both ends of the T-DNA were needed because the distance between the tandem enhancers and left border end is approximately 4.2 kb (Figure 4). As our expression analysis suggests a reverse correlation between the successful activation of a gene and the distance from the enhancer sequence, candidate genes were categorized according to their relative distance from the enhancer and are shown in different colours. The database also contains the sequence primers and raw sequence data used to localize the insertion sites.

Application of the database

As we have registered all annotated genes located within 6 kb from the enhancer sequence, this database can be used in a reverse genetics approach to find a transgenic line in which the gene of interest could be over-expressed because of the activation-tagging T-DNA. From the expression analysis, if the candidate genes are narrowed down to those within 6 kb from the enhancer sequence, the rate of gene over-expression goes up to 56% (five out of nine genes). This value is similar to the ratio found in the activation lines from rice (50%: four out of eight genes; Jeong *et al.*, 2002). In the database, there are 1837 genes annotated as activation candidates and another 317 genes are knockout candidates. As far as we know, this is the first attempt to introduce a database concept, namely the reverse genetics approach in order to obtain plant materials where the target genes could be over-expressed showing any functions. Another important aspect of the database is that it provides a good base from which to investigate the actual tagging ratio in such population. It is well known that epigenetical variation can cause mutant-like phenotypes in the T1 generation. A conventional approach to eliminate such variegates is to observe the phenotype in the T2 generation. Applying such an approach is not optimal in this activation-tagging system, as the activated genes are known to be suppressed in the next generation (Weigel *et al.*, 2000). This also explains why many activation mutants show a semi-dominant phenotype in the T2 generation and thereafter. The expression level of the target gene might differ between T2 plants. Therefore, the only way to confirm that the candidate gene caused the T1 phenotype is to re-transform the wild-type plant with a construct over-expressing the gene in question and recapitulate the mutant phenotype in the T1 transformants. We are currently introducing these constructs to determine the genes causing the phenotypes and estimate the frequency

Figure 5. RT-PCR of genes adjacent to the T-DNA tag.
RT-PCR was performed on wild-type and AT2P plants. Gene-specific primers were used to monitor the expression levels. Beta-tubulin fragment was first amplified to adjust the amount of template cDNAs.
(a) Results of RT-PCR that showed significant increases in the corresponding gene transcripts in each line (Z18548, Z23920, Z29218, Z29732, Z35240, Z49623 and Z88201) compared to those in the wild-type plant.
(b) Results of RT-PCR that showed significant decrease in the corresponding gene transcript in the line Z26013.
(c) Results of RT-PCR that did not show significant differences in the corresponding gene transcripts (lines Z18548, Z21637, Z23920, Z26013, Z29218, Z29732, Z32731 and Z88201). L and R represent genes closest to the T-DNA left- and right-border sequences, respectively (see Figure 4). Upper bands correspond to gene-specific PCR fragments. Lower bands correspond to beta-tubulin fragments used for template adjustment. Left bands are results from corresponding lines. Right bands are results from the wild-type plants.

of tagging in the dominant phenotype population generated by the activation-tagging T-DNA. Finally, this system can also be used to define a gene function directly from the database, if several lines showing the same or similar phenotype contained different insertion loci, but next to an identical gene or to gene family members (Nakazawa *et al.*, 2003). Such analysis is reminiscent of allelic test in the conventional forward screening of the knockout population.

Experimental procedures

Plant material and harvesting

Wild-type *A. thaliana* (Col-0) and the transformed lines were grown at 22°C in a cultivation container system (ARACON) in long-day conditions (16-h light and 8-h dark) under white fluorescent tubes (FL40SW, Sanyo, Japan). Wild-type plants were transformed by the floral dipping method using *Agrobacterium* GV3101 strains harbouring the activation-tagging T-DNA vector, pPCVICEn4HPT (Clough and Bent, 1998; Walden *et al.*, 1994). Hygromycin-resistant T1 seedlings were selected on BAM containing 1 mM KNO_3, 0.8% agar and 50 mg l^{-1} hygromycin for 7 days and transferred to soil (Nakazawa and Matsui, 2003). Any visible phenotypes were scored and all plants showing phenotypes (AT1P) were transferred to new ARACON trays and observed further. Either rosette leaves or flowers were harvested from all of the AT1P plants for DNA sequence analysis.

Southern blot analysis and hygromycin resistance check

Southern blot analysis was performed as described elsewhere (Meyer *et al.*, 1995). Twenty lines were randomly chosen and T2 plants were grown in soil for 3 weeks as described above. Leaves from seven plants per line were harvested and homogenized in liquid nitrogen. Genomic DNA was isolated by DNeasy plant mini kit (Qiagen, Tokyo, Japan), according to the instruction manual. A 2-kb *Hin*dIII fragment of pPCVICEn4HPT, which contains the hygromycin resistance gene, was used as a template to make a random priming probe by the Rediprime II DNA labelling system (Amersham, Tokyo, Japan). For the hygromycin resistance test, 40 seeds on average from the AT1P lines were germinated on hygromycin-containing BAM. The number of resistant seedlings per germinated seeds was counted 10 days after transfer of the culture plates into the light.

Plasmid rescue, adaptor PCR and identification of T-DNA insertion position in the genome of A. thaliana

Approximately 200 mg FW of rosette leaves were harvested. Five ceramic particles (CERAMICS YTZ ball, D: 2.3 mm; Nikkato, Japan) and 300 μl of the lysis buffer were added to the plant material to homogenize it using Shake Master (Shake Master ver. 1.0, Bio Medical Science Inc., Japan). Genomic DNA was extracted using the Wizard Magnetic 96 DNA Plant System (Promega, Tokyo, Japan). A workstation system (Tecan genesis workstation 150) was adopted to run the extraction protocol described in the kit. For plasmid rescue, the genomic DNA was digested with *Bam*HI restriction enzyme in a volume of 100 μl in an overnight reaction. The digested DNA was ethanol precipitated and dissolved in 8 μl of water. The DNA was self-ligated by T4 ligase (400 U μl^{-1}, New England Biolabs, Beverly, MA, USA) in a volume of 10 μl at room temperature overnight. Twenty microlitres of *Escherichia coli* DH10B (Invitrogen, Carlsbad, CA, USA) was transformed by electroporation with 2 μl of the ligation mixture. Transformed *E. coli* was cultured on LB-Amp agar medium. After overnight culture at 38°C, eight colonies per line were chosen for subculture into LB-Amp liquid medium. Plasmids isolated from the *E. coli* cultures were digested with *Hin*dIII restriction enzyme to select out the plasmids that contained only vector fragments. Plasmids containing the genomic DNA were sequenced using the LB2 primer 5′-TGACCATCATACCCATTGCTGATCC-3′. In the case of the adaptor-PCR method, digestion of the genomic DNA and PCR were carried out as described by Yamamoto *et al.* (2003), except that the genomic DNA was digested with *Bam*HI and *Hin*dIII rather than with *Bgl*II and *Xho*I, respectively, and that the *Xho*I adaptor primer, the first PCR primer C1, and the nested PCR primer C4 were replaced with a *Hin*dIII adaptor primer: 5′-AGCTACCAGCCC-(NH_2)-3′, the LB5 primer: 5′-GATATCGATCGTGAAGTTTCTCA-3′ and the LB6 primer: 5′-GCCCCCATTTGGACGTGAA-3′, respectively. After nested PCR, the reaction mixtures were purified by QIAquick PCR Purification Kit (Qiagen) and then sequenced using the LB2 primer. When the sequence reaction did not work successfully, each PCR fragment was isolated on an agarose gel. All the isolated fragments were separately sequenced using the LB2 primer. The sequence results were analysed using the BLASTN program against the MIPS non-redundant database to identify the position of the T-DNA in the genome of *A. thaliana*.

RT-PCR

Seed from 10 AT1P plants that showed phenotypes in the T2 generation and seed from wild-type plants were sown in soil. Plants were grown for approximately 6 weeks. Rosette leaves from at least three different plants in each line were harvested and mRNA was extracted using the Dynabeads mRNA DIRECT Kit (DYNAL, Oslo, Norway), according to the manufacturer's instructions. The mRNA was treated with RQ1DNase (Promega) for 1 h at 37°C. cDNA was synthesized using the Superscript first-strand synthesis system for RT-PCR (Invitrogen) according to the manufacturer's instructions. RT-PCR was performed using the following beta-tubulin-specific primers (Takahashi *et al.*, 2001) to adjust the ratio of cDNA between wild-type plants and individual lines: TU1: 5′-TTCATATCCAAGGCGGTCAATGTG-3′; and TU2: 5′-CCATGCCT-TCTCCTGTGTACCAA-3′. The amount of cDNA template and the number of PCR cycles were adjusted individually, but the PCR was always performed under the following conditions: 94°C for 30 sec for denaturation; 60°C for 30 sec for annealing; and 72°C for 120 sec for elongation. Experiments were performed at least two times to confirm the results.

Acknowledgements

This study was performed during the project of saturation mutagenesis carried out by the Plant Functional Genomics Group of the Genomic Science Center. We thank Ms K. Bishop for correcting English.

References

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Bouche, N. and Bouchez, D. (2001) *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* **4**, 111–117.

Budziszewski, G.J., Lewis, S.P., Glover, L.W. *et al.* (2001) *Arabidopsis* genes essential for seedling viability: isolation of insertional mutants and molecular cloning. *Genetics*, **159**, 1765–1778.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Feng, Q., Zhang, Y., Hao, P. *et al.* (2002) Sequence and analysis of rice chromosome 4. *Nature*, **420**, 316–320.

Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K. and An, G. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol.* **130**, 1636–1644.

Krysan, P.J., Young, J.C., Jester, P.J., Monson, S., Copenhaver, G., Preuss, D. and Sussman, M.R. (2002) Characterization of T-DNA insertion sites in *Arabidopsis thaliana* and the implications for saturation mutagenesis. *Omics*, **6**, 163–174.

Marsch-Martinez, N., Greco, R., Van Arkel, G., Herrera-Estrella, L. and Pereira, A. (2002) Activation tagging using the En-I maize transposon system in *Arabidopsis*. *Plant Physiol.* **129**, 1544–1556.

Meyer, A.D., Ichikawa, T. and Meins, F., Jr (1995) Horizontal gene transfer: regulated expression of a tobacco homologue of the *Agrobacterium* rhizogenes rolC gene. *Mol. Gen. Genet.* **249**, 265–273.

Nacry, P., Camilleri, C., Courtial, B., Caboche, M. and Bouchez, D. (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics*, **149**, 641–650.

Nakazawa, M. and Matsui, M. (2003) Selection of hygromycin-resistant *Arabidopsis* seedlings. *Biotechniques*, **34**, 28–30.

Nakazawa, M., Ichikawa, T., Ishikawa, A., Kobayashi, H., Tsuhara, Y., Kawashima, M., Suzuki, K., Muto, S. and Matsui, M. (2003) Activation tagging, a novel tool to dissect the functions of a gene family. *Plant J.* **34**, 741–750.

Rinehart, T.A., Dean, C. and Weil, C.F. (1997) Comparative analysis of non-random DNA repair following Ac transposon excision in maize and *Arabidopsis*. *Plant J.* **12**, 1419–1427.

Sasaki, T., Matsumoto, T., Yamamoto, K. *et al.* (2002) The genome sequence and structure of rice chromosome 1. *Nature*, **420**, 312–316.

Sessions, A., Burke, E., Presting, G. *et al.* (2002) A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell*, **14**, 2985–2994.

Somerville, C. and Dangl, J. (2000) Genomics. Plant biology in 2010. *Science*, **290**, 2077–2078.

Szabados, L., Kovacs, I., Oberschall, A. *et al.* (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. *Plant J.* **32**, 233–242.

Takahashi, M., Sasaki, Y., Ida, S. and Morikawa, H. (2001) Nitrite reductase gene enrichment improves assimilation of NO (2) in *Arabidopsis*. *Plant Physiol.* **126**, 731–741.

Vision, T.J., Brown, D.G. and Tanksley, S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290**, 2114–2117.

Walden, R., Fritze, K., Hayashi, H., Miklashevichs, E., Harling, H. and Schell, J. (1994) Activation tagging: a means of isolating genes implicated as playing a role in plant growth and development. *Plant Mol. Biol.* **26**, 1521–1528.

Weigel, D., Ahn, J.H., Blazquez, M.A. *et al.* (2000) Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013.

Wilson, K., Long, D., Swinburne, J. and Coupland, G. (1996) A dissociation insertion causes a semi-dominant mutation that increases expression of TINY, an *Arabidopsis* gene related to APETALA2. *Plant Cell*, **8**, 659–671.

Yamamoto, Y.Y., Tsuhara, Y., Gohda, K., Suzuki, K. and Matsui, M. (2003) Gene trapping of the *Arabidopsis* genome with a firefly luciferase reporter. *Plant J.* **35**, 273–283.

Yu, J., Hu, S., Wang, J. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, **296**, 79–92.