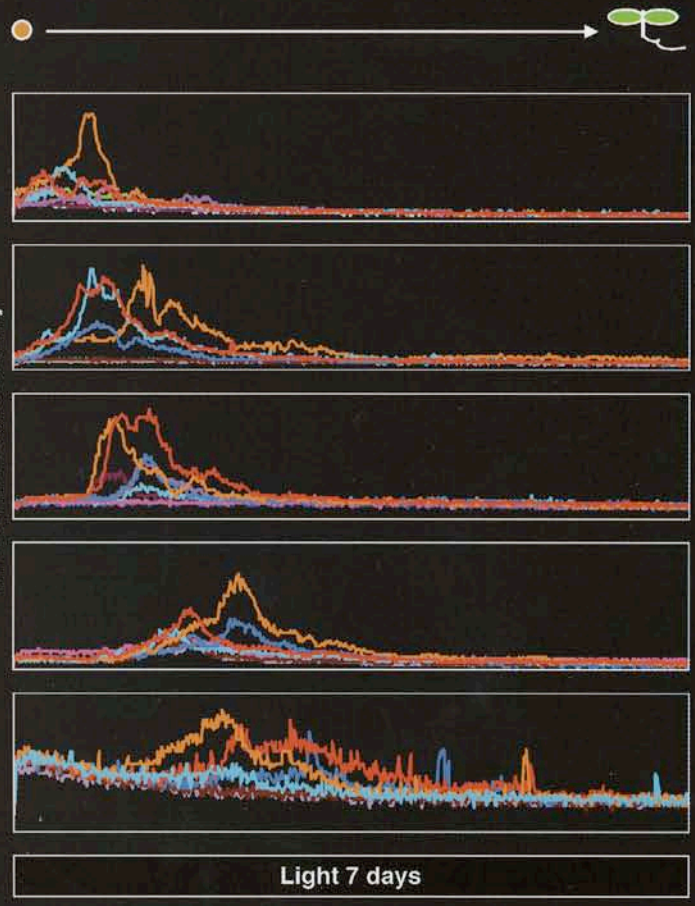
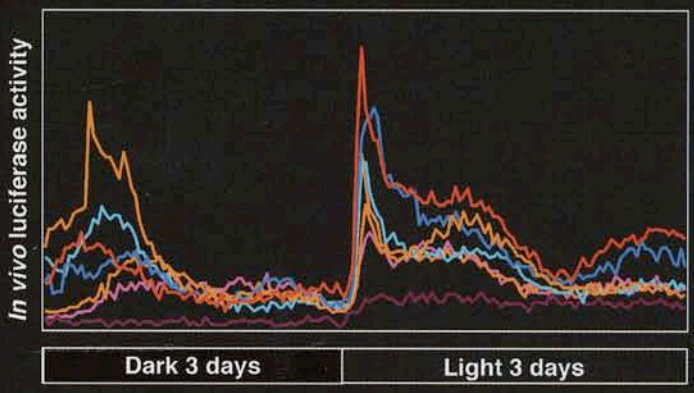
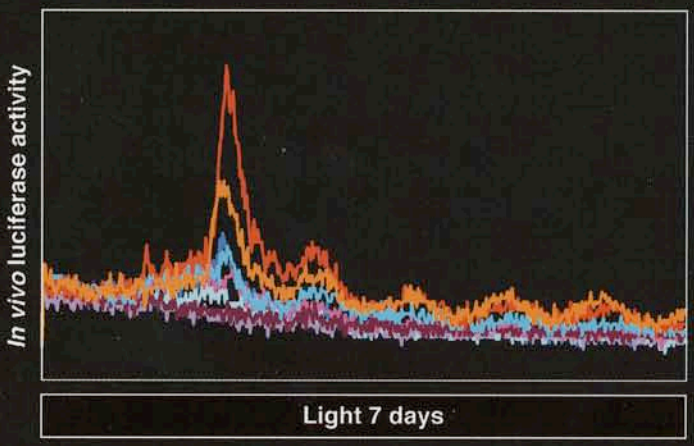
the plant journal







VOLUME 35 | NUMBER 2 | JULY 2003
http:// www.blackwellpublishing.com/journals/tpj | ISSN 0960-7412
Blackwell Publishing

The Plant Journal (2003) **35**, 273–283 doi: 10.1046/j.1365-313X.2003.01797.x

TECHNICAL ADVANCE

Gene trapping of the *Arabidopsis* genome with a firefly luciferase reporter

Yoshiharu Y. Yamamoto[†], Yumi Tsuhara, Kazuhito Gohda, Kumiko Suzuki and Minami Matsui[*]
Plant Function Exploration Team, Plant Functional Genomics Research Group, Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-22 Suyehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Received 22 January 2003; revised 16 April 2003; accepted 1 May 2003.
[*]For correspondence (fax +81 45 503 9586; e-mail minami@postman.riken.go.jp).
[†]Present address: Plant Functions Laboratory, RIKEN, Hirosawa 2-1, Wako, Saitama 351-0198, Japan.

Summary

Experiments with gene-trap vectors containing the firefly luciferase (*LUC*) reporter genes were carried out with the aim of analyzing functions of the *Arabidopsis* genome. Studies with protein fusion-type trap vectors as well as an internal ribosome entry site (IRES)-assisted non-fusion-type vector revealed that both types of vectors were suitable for gene trapping in *Arabidopsis,* although there were some differences in trapping efficiencies. The established trap lines were subjected to analyses for light responses, demonstrating the powerful and unique applications of a *LUC*-trapping system. A systematic survey of the insertion sites of the T-DNAs in *LUC*-expressing lines revealed 12–41% gene-trapping efficiencies depending on the vector. We demonstrate that the *LUC*-trapping system provides a unique system with which to monitor temporal expression of plant genes.

Keywords: gene trap, firefly luciferase, IRES, temporal expression.

Introduction

Gene trapping has been developed as a method to monitor gene expression profiles by random insertion of reporter genes into the genome. At the same time, it also provides knockout mutations of the genes for functional analyses (Springer, 2000). Gene-trapping methods have been used for the following purposes: (i) classification of genes based on expression analysis (Campisi *et al.*, 1999; He *et al.*, 2001); (ii) development of molecular markers for specific cell types or developmental stages (Sabatini *et al.*, 1999; Topping and Lindsey, 1997); (iii) promoter hunting (Plesch *et al.*, 2000); and (iv) preparation of knockout mutants for functional analysis (Rajani and Sundaresan, 2001; Springer *et al.*, 1995). The first of these can be used in conjunction with the microarray technique. The latter technique is efficient in revealing expression patterns of genes with small-scale experiments, but when applied to temporal gene expression analyses, data points tend to be discontinuous, and it is difficult to trace rapid changes in gene expressions. On the other hand, to cover the whole genome with gene-trap lines requires considerable effort. However, more accurate analysis of specific genes (trap lines) can be performed as dense data points can be obtained.

The uniqueness of gene trapping becomes clear when the above purposes are combined. For example, expression profiles of trap lines provide suggestive and useful information about the phenotype of the mutants (Rajani and Sundaresan, 2001; Springer *et al.*, 1995). Gene-knockout phenotypes with specific expression profiles are easily examined with the aid of trapping methods. Therefore, gene-trapping techniques provide important and unique methods for studying the relationship between gene expression and function.

Until now, almost all the gene-trap lines of *Arabidopsis* were generated with the β-glucuronidase reporter gene (*GUS*; Springer, 2000). *GUS* allows fine resolution in histochemical analysis and thus is good for developmental studies. However, it is not suitable for the observation of responses to environmental conditions as this protein has a low turnover rate (Jefferson *et al.*, 1987) and slow induction (Gatz *et al.*, 1992). This is also true for the recently developed green fluorescent protein reporter gene (*GFP*;

Baulcombe *et al.*, 1995). Both these reporter proteins are ideal for the analysis of spatial expression patterns of genes such as tissue specificities and intracellular localization of proteins.

To date, no gene-trap reporter system has been applied for temporal expression analysis. The firefly luciferase (LUC) protein has a high turnover rate and thus subtly responds to changes in transcriptional activities (Millar *et al.*, 1992a). Furthermore, availability of a non-destructive *in vivo* assay of the *LUC* reporter opens up the possibility of novel strategies for the analysis of gene expression (Millar *et al.*, 1992b). In addition, the LUC assay has low background bioluminescence that allows an order of magnitude higher sensitivity for gene expression compared with the GUS assay (Yamamoto and Deng, 1998). This character makes it possible for the LUC reporter to observe temporal expression profiles such as responses to environmental stimuli. To analyze environmental responses, we constructed *LUC* reporter gene-trap lines.

In this study, we have developed several types of gene-trap vectors using *LUC* as the reporter gene, and applied them in *Arabidopsis* gene trapping. The generated gene-trap lines displayed a variety of expression profiles, including responses to light, transient expression during seedling development, and circadian oscillation.

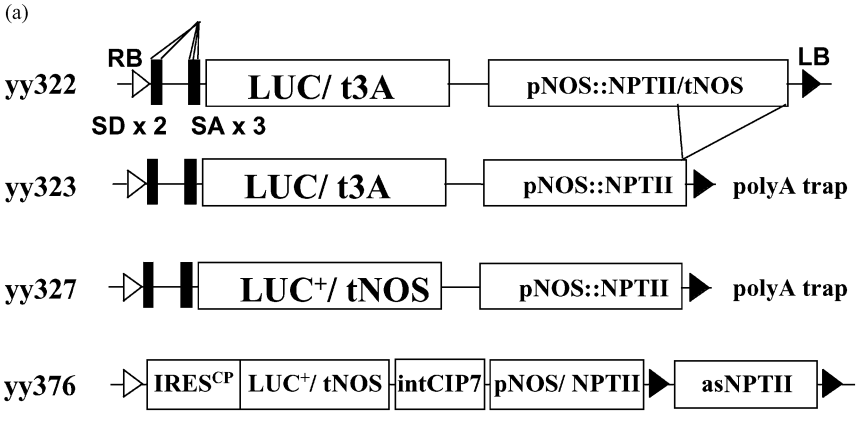
Results

Design of gene-trap vectors

We applied the degenerated splicing strategy that has been developed by Sundaresan *et al.* (1995). Insertion of a pair of splicing donor and acceptor sites between the right border (RB) and the reporter gene ensures fused transcripts not only when the T-DNA is inserted in exons of the recipient genes but also in introns. Furthermore, the multiple splicing unit produces heterologous mRNA species, and some of these will be expected to be fused in frame so that some portion of the various transcripts are always translated as fusion proteins. This strategy is expected to enhance the probability of reporter (Nussaume *et al.*, 1995; Sundaresan *et al.*, 1995).

Secondly, the polyA-trap strategy (Yoshida *et al.*, 1995) was incorporated to enhance the orientation and integration of the T-DNA insertion into the coding region. Because this strategy utilizes the polyA signal of the target gene for the introduced selection marker gene, the marker is expected to be active only when it is integrated within a transcribed region in the sense orientation (Figure 1a). Therefore, precise positioning of the reporter gene relative to the marker gene confers a high trapping efficiency among populations positive for the marker.

(a)



(b)

```
taataaacgctcttttctcttaggtttacccgccaatatatcctgtcaaagcttCAAT:GT
                          F  T  R  Q  Y  I  L  S  K  L  Q  G
                           L  P  A  N  I  S  C  Q  S  F  K   V
                         V  Y  P  P  I  Y  P  V  K  A  S  R   S

CTCTCTTCAAG:GTGAGTTTTTTTCTGTTCACTCTCTTAGATGCCAAAACTTGAGTTATTG
L  S  S  R
 S  L  Q  G
  L  F  K

CTTAATGTTTCAATTGTTGTGGACTCTGTGTATGTGTAG:GTTATATGCAG:GTTATATGCA
                                        V  I  C  R   L  Y  A
G:GTTATATGGGAGGTGGAGGGATCCAAACAatggctatggctgaagacgccaaaaacata
G   Y  M  G  G  G  G  I  Q  T  M  A  M  A  E  D  A  K  N  I
```

Figure 1. Structure of luciferase (LUC)-based trap vectors.
(a) Illustration of T-DNA region. RB: Right border; LB: left border; SD: splicing donor; SA: splicing acceptor; t3A: RbcS3A terminator; tNOS: NOS terminator; pNOS: NOS promoter; NPTII: kanamycin (Km)-resistance marker, internal ribosome entry site ($IRES^{CP}$): IRES element of the tobamo virus coat protein; intCIP7: a *CIP7* intron with a 12-base deletion; asNPTII: an antisense fragment of *NPTII*. The terminator for NPTII is omitted in polyA-trap constructs (yy323, yy327, and yy376).
(b) Nucleotide sequence around the 5′ end of the *LUC* reporter of yy322, yy323, and yy327. Dotted line indicates a *CIP7* intron. Colon indicates splicing junction. The glycine stretch (underlined) is a spacer between the LUC protein and the trapped protein. If *LUC* is expressed without fusion, amino acid residues from 'MAMA' are to be translated.

Luciferase fusion has been reported rarely with a few exceptions (Worley *et al.*, 2000). This lack of successful reports might suggest that LUC activity is easily lost after fusion. One strategy to avoid loss of enzymatic activity in fusion proteins is to insert a spacer region at the junction of two proteins to reduce conformational interference (Cutler *et al.*, 2000; Iwakura and Nakamura, 1998).

Incorporating the strategies described above, we prepared two gene-trap vectors, yy322 and yy323 (with polyA trap; Figure 1). Both constructs contained a multisplicing unit consisting of duplicated splicing donors and triplicated acceptors, and a glycine stretch in front of the *LUC*-coding region to allow free rotation (Figure 1b, underlined).

yy322 and yy323 were introduced into *Arabidopsis* by the floral dipping method (Clough and Bent, 1998). The transformation efficiency was 1.2% (410/32 500) for yy322 and 2.3% (750/32 500) for yy323 (polyA trap). Because polyA trapping produces kanamycin (Km)-resistant plants only when the T-DNA is integrated in a transcribed region in the sense orientation, transformation efficiency is expected to reduce. We observed no reduction in the transformation efficiency of yy323.

This result means that the polyA-trap strategy might not work as expected. Nevertheless, yy323 showed one advantage over yy322 in that its transformants showed stronger Km-resistance than those of yy322. This might be because of the deletion of a 0.4-kbp fragment that was found between *NPTII* gene and the *NOS* terminator in pBIN19 (data not shown). Taking this advantage into consideration, we used yy323 for further analyses.

High-throughput identification of T-DNA copy numbers of transgenic plants using PCR

For gene trapping, a single-copy insertion of the reporter gene is vital for clear identification of the recipient gene that is responsible for the reporter activity of each trap line. In order to facilitate high-throughput determination of T-DNA copy number, we have developed a simple PCR-based method.

Based on the structure of yy323, competitive PCR was designed for amplification of a genomic fragment (*CIP7* intron) and a T-DNA fragment (*CIP7* intron–LUC fusion) in a single reaction of multiplex- (triple primer-) PCR (Figure 2a). Comparison of strength of bands from the T-DNA and from the genomic DNA was expected to give the ratio of the copy number of the T-DNA relative to the genome. Test PCR was performed with the template of an *in vitro* mixture of *Arabidopsis* genomic DNA and yy323 plasmid DNA (Figure 2b). As shown in Figure 2(c), increases in the T-DNA concentration to represent one, two, and three copies of the T-DNA relative to the diploid genome resulted in a clear elevation of the band intensity reflecting the T-DNA/genome ratio.

Next, we applied the same method to transgenic lines containing yy322 and yy323. Leaves from Km-resistant seedlings of the T_1 generation were subjected to PCR analysis. Figure 2(d) shows the histogram of the T-DNA/genome band intensity ratio. As shown in the Figure, the population has several peaks separated by saddle points. From the results of the *in vitro* mix experiments (Figure 2c), the lines that had a ratio of less than 1.4, i.e. the population with 0.95 as a peak, were suggested to have single-copy T-DNAs.

In order to examine the accuracy of this method to estimate the T-DNA copy number, genomic DNA gel blot analysis was performed. As shown in Figure 3(a), the part of the population that showed a PCR ratio of less than 1.4 was rich in single-copy lines, while the majority of the population with a value of more than 1.4 were multicopy lines. However, the analysis revealed that several plants with a PCR value of less than 1.4 did contain multiple T-DNA copies. Taking into account the results of the hybridization, a window between 0.2 and 1.0 for the PCR ratio was set to identify lines with a single copy of the T-DNA with 80% reliability (Figure 3b).

Generation of LUC*-expressing trap lines with single T-DNA insertions*

One of the advantages of the *LUC* reporter is the existence of a non-destructive *in vivo* assay using a high-performance CCD camera (Argus system, Hamamatsu Photonics Co. Ltd, Hamamatsu, Japan; Millar *et al.*, 1992b). This allows screening of T_1 seedlings for LUC activity without killing them. Using this assay, we screened Km-resistant T_1 seedlings to identify LUC-positive ones by visualizing LUC activity on plates, and only positive seedlings were transferred to soil to establish trap lines.

After LUC selection, leaves were harvested for competitive PCR, and seedlings with a PCR value of between 0.2 and 1.0 were allowed to continue growing in soil. As a pilot experiment, about 100 trap lines were established through LUC and PCR screening from 6000 Km-resistant seedlings using the yy322 or yy323 vectors.

Expression profiles identified from the LUC *trap lines*

The established trap lines were subjected to expression analysis. First, we observed *LUC* expression at the seedling stage using a high-performance CCD camera (Argus system, Hamamatsu Photonics Co. Ltd, Hamamatsu, Japan). As shown in Figure 4, this analysis shows the spatial expression profile depending on the tissue or organ. As expected, several types of expression profiles were observed depending on the trap line. There were few lines with root-specific expression (Table 1), as screening was for expression in the aerial parts of the seedlings. However, a small fraction of lines showed root-specific expression or

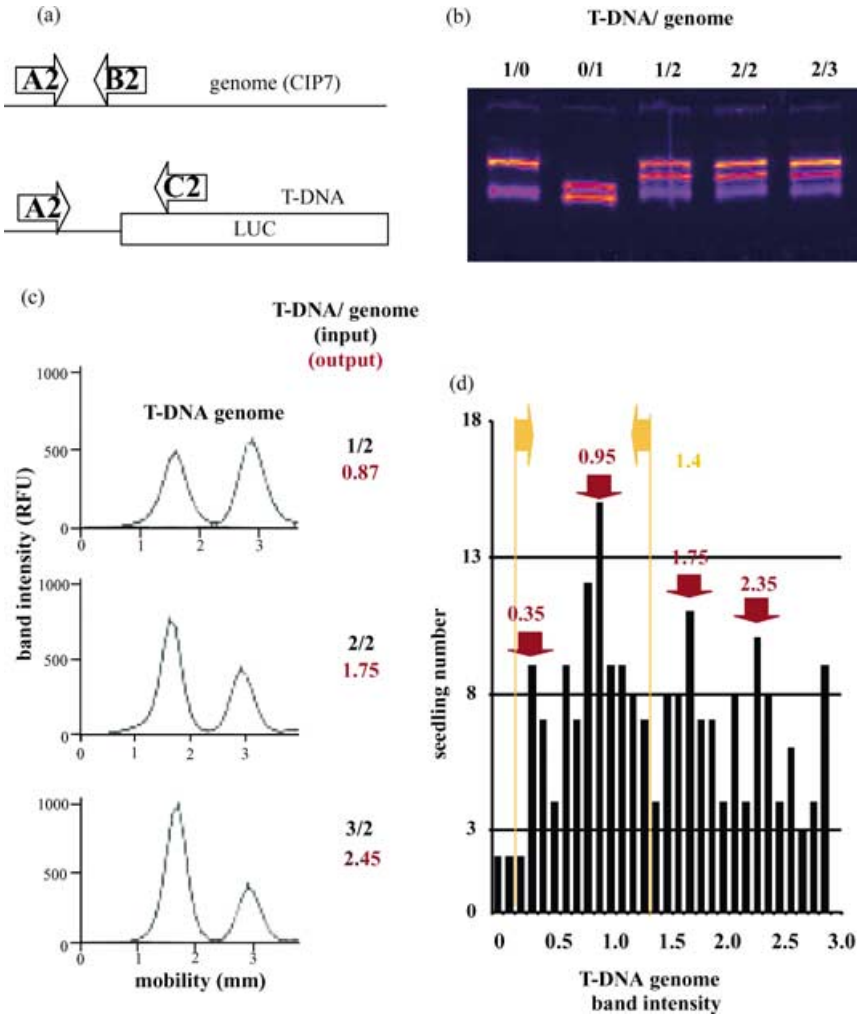
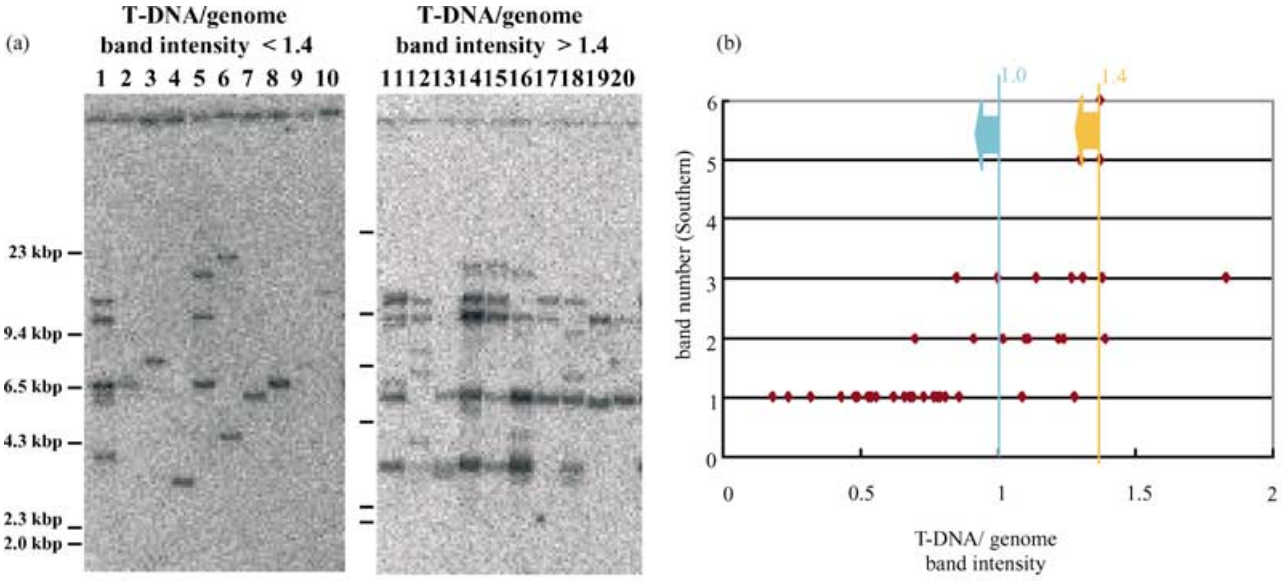


Figure 2. PCR-based detection of T-DNA copy number.
(a) Relationship of the PCR primers used for the multiplex PCR. A2 is a primer for the *CIP7* intron. It anneals with single-copy genomic DNA (CIP7) as well as with the multisplicing unit of the T-DNA. B2 and C2 are specific to genomic DNA and the T-DNA, respectively. These three primers are used in a single reaction to give 99-bp genomic and 150-bp T-DNA fragments.
(b) Electrophoretogram of PCR products of *in vitro* mix experiments. Wild-type genomic DNA and the binary plasmid (yy323) were mixed with the indicated molar ratio and subjected to the multiplex PCR. After separation with gel electrophoresis, PCR products were stained with Vistragreen and detected by a fluorescence scanner. The first and second bands from the top are the T-DNA and genomic products, respectively.
(c) Quantitative data of the electrophoretogram. The band intensity is expressed in relative fluorescence units (RFU). Input is the ratio of the mixed amplicons and the output is the ratio of the two PCR products calculated by the area ratio.
(d) Histogram of T_1 seedlings. One hundred and six independent T_1 seedlings were subjected to multiplex PCR analysis. Yellow vertical bars show the ratio from 0.2 to 1.4.



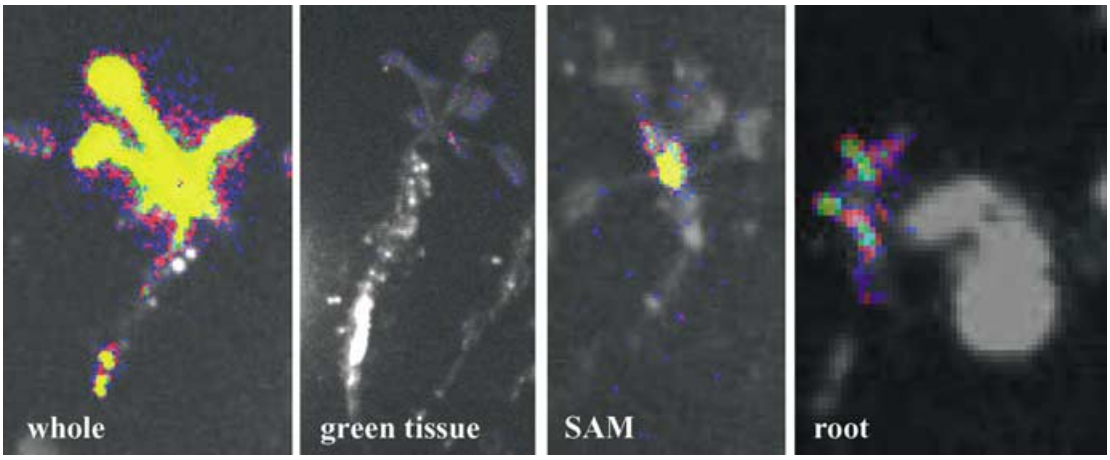


Figure 4. Organ-specific expression of trap lines.
Luciferase (LUC) activity (indicated in color) of trap lines at seedling stage was detected using a high-performance CCD camera. The black and white image is an overlay of the seedling image in the light. Activity was detected in whole seedlings, green tissue, an area around the shoot apical meristem (SAM), and roots.

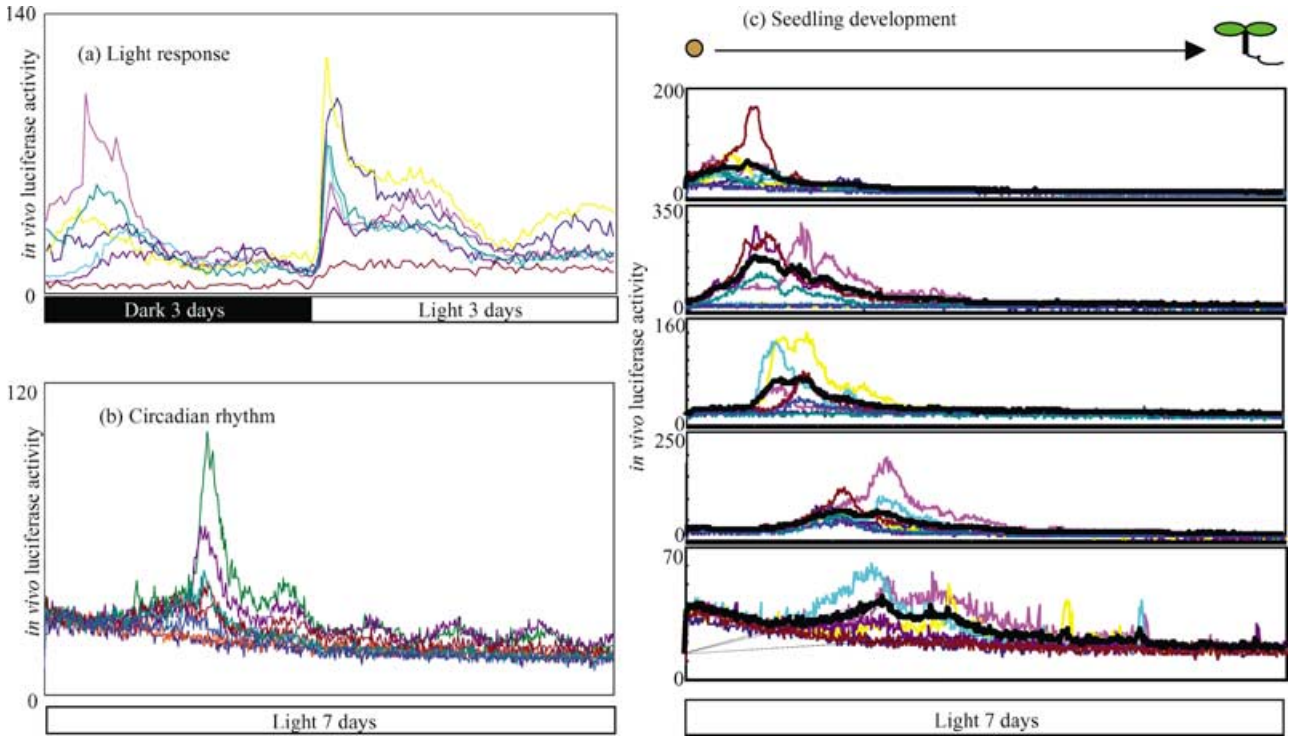


Figure 5. Analysis of light response of trap lines.
Seeds of trap lines were germinated and allowed to develop for a week to the cotyledonal stage and following that the luciferase (LUC) activity of individual seedlings was monitored at 15–25-min intervals. Light conditions in the assay are indicated at the bottom of the panel. Each panel represents one trap line, and each colored line in the panel show the LUC activity of individual seedlings. The black line in (c) indicates the average LUC activity of the seedlings.

Figure 3. DNA gel blot analysis of the trap lines.
(a) Autoradiograph of DNA gel blots. Lanes 1 and 11: size marker; lanes 2–10: plants with T-DNA/genome band intensity between 0.2 and 1.4; lanes 12–20: plants with T-DNA/genome band intensity over 1.4. Total DNA was isolated from independent transformants, digested with *Pst*I, and hybridized with a radiolabeled probe of a 0.4-kbp *luciferase* (*LUC*) fragment.
(b) Relationship of T-DNA/genome band intensity determined by multiplex PCR and copy number of T-DNA established by DNA gel blot hybridization as shown in (a). The population between 0.2 and 1.0 of the T-DNA/genome band intensity gives plants containing a single copy T-DNA with 80% accuracy.

Table 1 Classification of trapped lines with expressed tissues

Expressed tissue	Appearance rate (%)
Whole	45
Green tissue	41
Shoot apical meristem	8.4
Root	6.3

Luciferase (LUC) activity of 2-week-old-seedlings were visualized using a high-performance CCD camera, and the expression patterns were classified. Results of the established 95 lines with LUC activity are shown.

Table 2 Classification of trapped lines with responses to light

Response	Appearance rate (%)
Light	14.7
High light	4.2
Circadian rhythm	3.1

Luciferase (LUC) activity of individual seedlings of the established 95 lines were monitored for a week under continuous light (circadian rhythm), or in the dark for 3 days and then illuminated for 3 days (light), as shown in Figure 5. For observation of the high light response, seedlings grown under weak light (10–50 μE m^{-2} sec^{-1}) were transferred to stronger light (150–300 μE m^{-2} sec^{-1}) for 1 day, and the response of the *LUC* gene was analyzed using a high-performance CCD camera.

no expression. These lines could be escapes of the expressional screening. Surprisingly, as much as 8.4% of the population showed specific expression in the shoot apical meristem (SAM, Figure 4; Table 1).

Subsequently, we analyzed responses of the inserted reporter activities to changes in light conditions by continuous *in vivo* monitoring. In our assays, seedlings were grown for 1 week in 96-well plates in agar medium containing luciferin, and the LUC activities of plants were then measured repeatedly by photomultipliers at 15–25-min intervals for the week. During the week, light conditions were changed and responses were observed.

In Figure 5, each panel shows the results of siblings from the same line in the T_2 generation, and the lines in the Figure indicate the expression of each seedling. Some seedlings did not contain any LUC activity. These are suggested to be T-DNA-negative segregants. Excluding LUC-negative seedlings, overall expression profiles were reasonably reproduced among the siblings. One line shown in Figure 5(a) shows strong light activation of *LUC*. About 15% of the LUC-positive lines showed light activation (Table 2). Figure 5(b) shows circadian oscillation under constant-light conditions. Figure 5(c) shows five lines with transient expression during seedling development under constant light. Depending on the line, several peaks of reporter expression were found. These analyses demonstrate powerful and unique applications of the LUC-trapping system for studying environmental responses as well as developmental regulations.

Determination of the insertion sites

In order to examine whether these LUC-positive lines had trapped responsive genes as expected, we determined the insertion sites by sequencing the T-DNA flanking sequences by an adapter ligation-mediated PCR method. Table 3 summarizes the pattern of T-DNA insertion sites relative to the genes annotated on *Arabidopsis* chromosomes, and further information together with some expressional information can be found in Table S1. The expected frequency of a T-DNA insertion relative to the gene-coding region by random insertion is 18%. First of all, we noticed a considerable amount of reporter expression with intergenic insertions (67%). Because most of the examined lines started the transcription of the reporter gene beyond the integrated T-DNA regions (Hachisu *et al.*, unpublished results), the observed reporter expression by intergenic insertions should not be the result of 'enhancer trap', but regions upstream of the integrated T-DNA contain the promoter activity, which was revealed by insertion of the reporter. The observed reporter expression with intergenic insertions would be explained by cryptic promoter activity

Table 3 Insertion pattern of luciferase (LUC)-positive lines

	Calc.[a]	yy323(LUC)[b]	yy327(LUC$^+$)[b]	yy376 (IRES trap)[b]
Intragenic sense	18%	12% (4/33)	20% (9/46)	41% (16/39)
Intragenic antisense	18%	6.1% (2/33)	11% (5/46)	23% (9/39)
Promoter (<-500 bp)	9.1%	15% (5/33)	15% (7/46)	7.7% (3/39)
Sense + antisense				
Rest	55%	67% (22/33)	54% (25/46)	28% (11/39)

[a]Appearance rate among a population regardless the LUC activity, the calculation is based on the assumption that the average size of a gene (genome size/gene number) and the coding region (exon plus intron) of a gene are 5.5 and 2 kbp, respectively.
[b]Observed appearance rate among populations of LUC positive lines. Insertion site means 5′ end of the reporter construct. In most cases RB was excluded from the integrated T-DNA, and the T-DNA started from the *Hin*dIII site at the 5′ end of the multisplicing unit (Figure 1).

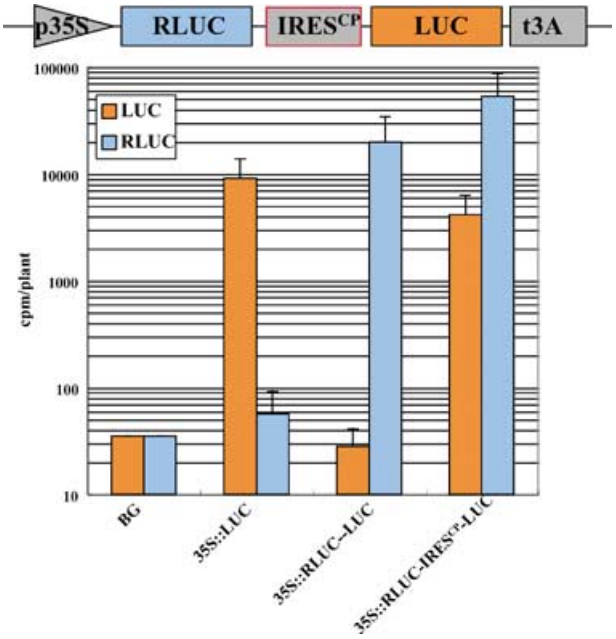


Figure 6. Dual luciferase (LUC) assay to detect internal ribosome entry site (IRES) activity *in vivo*.
Flower meristems of transgenic plants containing *35S::LUC*, *35S::RLUC-LUC*, and *35S::RLUC*-IRES^{CP}-*LUC* were subjected to an *in vitro* dual LUC assay. BG is the background level of the scintillation counter with a blank vial. Activity of non-transgenic plants was at a similar level to BG (data not shown). Average and standard deviation of 4 (*35S::LUC*), 3 (*35S::RLUC-LUC*), and 8 (*35S::RLUC*-IRES^{CP}-*LUC*) independent transformants are shown.

in such regions (Plesch *et al.*, 2000), or from the promoter activity of unannotated genes, which would also explain the *LUC* expression seen with insertions in the antisense orientation in coding regions (6.1%). Secondly, the trapping efficiency of the LUC-positive population was 12%, which is less than that calculated for a random situation (18%). This reduction suggests that some of the translational fusions with the recipient gene may cause suppression or reduction of LUC activity. On the other hand, promoter-trap-type insertions were enriched in LUC-positive lines (15% with yy323 against a calculated figure of 9.1%). This group would be translated as non-fusion proteins so that LUC activity would be conserved.

Improvement of trapping efficiency by use of a more active LUC *gene*

In order to improve trapping efficiency, we switched the reporter from LUC to LUC^{+}. LUC^{+} is an improved LUC protein that has codon usage suitable for mammalian cells and a C-terminal region that is deleted for cytoplasmic targeting (Schenborn and Groskreutz, 1997). The *LUC* gene of yy323 was replaced with *LUC*$^{+}$ to make yy327 (Figure 1a). Examination of the insertion sites of yy327 revealed a significant improvement in the trapping efficiency or integration frequency into intragenic regions in the sense orientation; it reached 20%, that is a twofold increase over that of yy323 (Table 3).

Application of an internal ribosome entry site for the LUC *trap vector*

Internal ribosome entry sites (IRES; Pestova *et al.*, 2001) enable dicistronic expression from a single mRNA species. They were recently used in mammalian expression vectors to monitor the gene expression without reporter gene fusion (Pestova *et al.*, 2001). As an alternative approach to monitor gene expression in plants, we decided to develop IRES-assisted vectors.

Before construction of IRES-type trapping vectors, we compared several established IRES elements (e.g. Excephatomyocarditis virus (EMCV); Urwin *et al.*, 2000) as well as the so-called translational enhancers (e.g. Yamamoto *et al.*, 1995; data not shown). One of them, an IRES from the tobamo virus coat protein (IRES^{CP}; Skulachev *et al.*, 1999) showed the clearest IRES activity in *Arabidopsis*.

Figure 6 showed the results of dicistronic assays of IRES^{CP}. In this assay, transcript with the first (*Renilla* luciferase (RLUC)) and the second (firefly luciferase (LUC)) cistrons was expressed by a single CaMV 35S promoter, and translation of the second cistron was examined. The *LUC* reporter activity of a monocistronic transcript (*35S::LUC*) was high, but activity was not observed when expressed as a second cistron (*35S::RLUC-LUC*) without an IRES, showing that polycistronic transcripts are never expressed in *Arabidopsis*. However, insertion of IRES^{CP} between the first and the second cistrons (*35S::RLUC*-IRES^{CP}-*LUC*) resulted in recovery of LUC activity to almost that of monocistronic ones, demonstrating IRES activity in plants. Therefore, we constructed an IRES-type *LUC*-trapping vector using IRES^{CP} (yy376; Figure 1).

As shown in Table 3, the trapping efficiency of yy376 was as high as 41%, which is considerably higher than yy327 (20%). We found that this trapping vector was also inserted in intergenic regions (28%), as well as in promoter regions, both producing LUC activity. Further information can be found in Table S1. These data demonstrate that yy376 is highly effective in gene trapping with a *LUC* reporter gene for *Arabidopsis*.

Discussion

LUC *reporter gene trap system is suitable for tracing temporal gene expression*

Several systems for gene trapping or promoter trapping have been reported, but all of them are useful for analysis of the spatial expression of genes.

To date, *LUC* is the only reporter gene that is suitable for the analysis of temporal gene expressions. There are several reports that have used the *LUC* reporter for the analysis

of gene expression. In most cases, it has been used for the analyses of stress responses or circadian rhythm.

We constructed two kinds of gene-trap vectors using *LUC* as the reporter gene. First, we made protein-fusion-type trap vectors. This kind of vector has been reported using *GUS* as the reporter gene. We introduced multiple splicing donor and acceptor sites in front of the *LUC* reporter gene in a T-DNA vector. When this T-DNA is inserted into a gene in the sense orientation, some of the transcripts make protein fusion with LUC. To enhance the possibility of T-DNA inserts in the sense orientation and in a gene, we employed a polyA-trap strategy. In this vector, the polyA signal sequence from the target gene will be required for proper expression of the Km-resistance gene, and only when this gene is inserted in an intragenic region (exons and introns) in the sense orientation will the transformants become Km-resistant.

The second type of vector is the IRES-type polycistronic one. IRES has not been used in a gene-trap method in plants, although a few reports, especially from the studies of plant viruses, have indicated that there are functional IRESs in plants. We tried several IRES sequences to establish whether they were functional when located polycistronically. We proved that the IRES from tobamo virus coat protein ($IRES^{CP}$) functions as a real IRES sequence in *Arabidopsis*, and we constructed an IRES-type vector using this $IRES^{CP}$ sequence. We introduced these gene-trap vectors and examined their integration profiles by making transgenic lines that had LUC activity.

Advantage of T_1 screening

For high-throughput analysis of trap lines, single insertions of the reporter gene are vital. This is achieved elegantly in the Ac/Ds trapping system (Bancroft *et al.*, 1992; Fedoroff and Smith, 1993; Sundaresan *et al.*, 1995). However, it is laborious to generate independent lines with this system, so it is not feasible to use it to produce the large number of lines required to saturate the *Arabidopsis* genome. Furthermore, the tendency for local transposition of the Ds element also makes it unsuitable for such a task (Ito *et al.*, 1999; Parinov *et al.*, 1999). In contrast, the *LUC*-trapping system described here makes large-scale preparation of T-DNA-containing lines by *Agrobacterium* transfection much easier than relying on Ds transposition to generate the lines. In addition, as the T-DNA integration is random, it is suitable for saturation mutagenesis. However, frequent multicopy integrations of the T-DNA complicate later analysis of the trap lines. In this sense, the PCR-based method for screening for single T-DNA-insertion lines described in this report has increased the value of a T-DNA-tagging system. Screening in the T_1 rather than in the T_2 generation means that no effort is wasted in growing and harvesting multicopy lines.

One advantage of the utilization of the *LUC* reporter for gene trapping is the ability to screen in the T_1 generation without killing the plants. Using this feature, it is possible to establish only LUC-positive lines as demonstrated in this study. While this strategy restricts the population to ones reporter-positive at the screened stage and in the screened tissues (i.e. aerial parts at the seedling stage), it concentrates on trapping active genes. Therefore, if focussing on a specific developmental stage, this strategy will reduce the effort for saturated mutagenesis of the active genes. Furthermore, if gene trapping is to be used for specific research, e.g. studying responses to light, screening in the T_1 generation allows effort to be concentrated on specific responses.

LUC *reporter for analysis of environmental responses*

The advantage of the *LUC* reporter is the possibility of real-time monitoring of expression within a single plant. This kind of monitoring cannot be done with other reporter genes. This assay is achieved by a non-destructive *in vivo* assay using automated monitoring of bioluminescence with the aid of photomultipliers (TopCount, Packard, Tokyo, Japan) or a high-performance CCD camera (Argus system, Hamamatsu Photonics Co. Ltd, Hamamatsu, Japan). Taking advantage of this feature, the dynamics of the reporter gene expression were monitored in response to the light conditions, seedling growth, as well as autonomous circadian oscillation (Figure 5). The small size of *Arabidopsis* seedlings enables the assay to be performed in 96-well plates for 1 week with high-throughput processing. The material is not necessarily restricted to seedlings, but detached flowers, roots, leaf disks, and also cultured cells can be assayed as well. Using these materials, environmental responses can be monitored including those to light, drought, salt, pathogens, wounding, as well as responses to plant hormones. In summary, our analyses using vectors of the protein-fusion-type as well as an IRES-assisted non-fusion-type have demonstrated the applicability of the *LUC* reporter for gene trapping in plants, and we also used them for unique applications in monitoring environmental and developmental responses. In combination with a PCR-based method for establishing the copy number of T-DNAs in transgenic plants it is possible to establish a large population of single T-DNA-inserted lines.

Experimental procedures

Construction of LUC-trapping vectors

The multiple splicing unit utilized in the vectors yy322 and yy323 is a chimeric sequence of donor sites of *CIP4* (Yamamoto *et al.*, 2001), *CIP7* (Yamamoto *et al.*, 1998), the fourth intron of *CIP7*, an acceptor site of *CIP7*, and the tandem acceptor from pSLJ5002

(Nussaume *et al*., 1995). The multiple splicing unit is followed by a glycine stretch as shown in Figure 1. The LUC/t3A in the vectors comes from pG6LUC (Aoyama and Chua, 1997), and pNOS::NPTII/tNOS from pBIN19 (Bevan, 1984). All the fragments were synthesized *in vitro* or amplified using PCR with appropriate primers, subcloned into pPZP200 (Hajdukiewicz *et al*., 1994), and confirmation of the inserts was obtained by sequencing. The final structure of the T-DNA region is as shown in Figure 1, and the whole sequence data are available at DDBJ (accession numbers AB086433 for yy322 and AB086434 for yy323).

yy327: LUC/t3A region of yy323 was replaced with LUC^{+}/tNOS of 221-LUC^{+} (K. Hiratsuka, Nara Institute of Science and Technology). The DDBJ accession number is AB086435.

yy376: As shown in Figure 3, multicopy T-DNA lines appear to contain the whole binary vector (the common 10-kbp band), an observation that has been reported previously by Galbiati *et al*. (2000). Therefore, a counterselection marker was inserted outside the left border in order to reduce integration of the whole vector. Antisense *NPTII* sequence and a left border were inserted downstream of the original left border of yy327 to make yy331. The $IRES^{CP}$ fragment ($IRES^{CP,148}$; Skulachev *et al*., 1999) and a modified *CIP7* intron with a 12 base-deletion were synthesized *in vitro* and sequentially inserted into yy331. The final construct, yy376, is shown in Figure 1, and the sequence data are available at DDBJ (accession number AB086436).

35S::LUC (yy300), 35S::RLUC-LUC (yy289), and 35S::RLUC-$IRES^{CP}$-LUC (yy366) have the *NPTII* marker from pBIN19 (Bevan, 1984) with the pPZP200 backbone (Hajdukiewicz *et al*., 1994). The CaMV 35S promoter, LUC/t3A, and RLUC are derivatives of pBI221 (Jefferson *et al*., 1987), yy211 (Kimura *et al*., 2001), and pRL-TK (Promega, Tokyo, Japan), respectively. $IRES^{CP}$ (Skulachev *et al*., 1999) in yy366 was synthesized *in vitro*, and the sequence is the same as in yy376. During subcloning, the t3A fragment dropped out from yy289 and yy300, leaving them without a terminator for *LUC*. However, both of them gave similar LUC activity to ones with the *NOS* terminator in transgenic *Arabidopsis* (data not shown). Further information of these plasmids will be supplied upon request.

Plant transformation

Transformation of *Arabidopsis* (Col) was achieved with the aid of *Agrobacterium* (GV3101, pMP90; Clough and Bent, 1998). T_1 transformants were screened for the presence of 50 μg ml^{-1} Km and 100 μg ml^{-1} carbenicillin. Carbenicillin was required because the presence of *Agrobacterium* on plants disrupted the LUC and PCR screenings (data not shown). Typically, 0.3 g seeds were plated on a 10 cm × 13 cm rectangular plate, and hundreds of seedlings showed Km-resistance.

PCR and gel blot analysis

For determining the copy numbers of the T-DNA, multiplex PCR was performed. Genomic DNA of T_1 seedlings was prepared according to Klimyuk *et al*. for use as PCR templates (Klimyuk *et al*., 1993) and subjected to PCR (Sambrook *et al*., 1989) with primers A2 (5′-GCC AAA ACT TGA GTT ATT GCT-3′), B2 (5′-GAA TTT TCT TCC ACA GTG TCT CCA TCA GT-3′), and C2 (5′-GGG CCT TTC TTT ATG TTT TTG GCG TCT TCA-3′). Conditions of PCR were: (94°C for 1 min, 80°C for 4 min) × 1 cycle, (94°C for 15 sec, 50°C for 15 sec, 72°C for 30 sec) × 40 cycles, and (72°C for 5 min) × 1 cycle. Primer length was designed, so annealing of the common primer (Figure 2, Primer A2) for the genomic fragment and the T-DNA was the rate-limiting step, and both genomic and T-DNA fragments were amplified at the same rate. The PCR products were mixed with an equal amount of 1 μg ml^{-1} solution of Vistra Green (Amersham Biosciences, Tokyo, Japan), 1/10 volume of 10× sample buffer (Sambrook *et al*., 1989), separated by gel electrophoresis and analyzed by a fluorescence scanner (FluorImager, Amersham Biosciences, Tokyo, Japan). The scanned gel image was analyzed with a prepared template set for the position of the gel slots, and the peak area and the position of each lane were incorporated into Excel files (Microsoft, Tokyo, Japan). Band identification based on the peak position and calculation of the T-DNA/genome ratio was both achieved using a VBA program (EXCEL MACRO). The *in vitro* mix experiments shown in Figure 2 were performed with a DNA template of total *Arabidopsis* DNA and plasmid DNA (yy323), both of which had been digested with *Xho*I to equalize their template activity.

Genomic DNA was isolated from mature leaves (50 mg) using the Nucleon PhytoPure Kit (Amersham Pharmacia Biotech, Tokyo, Japan). 1.3–3 μg of purified genomic DNA was digested with *Pst*I, separated by gel electrophoresis, blotted onto a nylon membrane, and hybridized with a ^{32}P-labeled 0.4-kbp *LUC* fragment, which had been amplified by PCR from yy323 using primers (forward 5′-ATG GGA GGT GGA GGG ATC CAA-3′ and reverse 5′-GGC TGC GAA ATG TTC ATA CTG-3′). DNA gel blot analysis was performed essentially according to Church and Gilbert (1984). Hybridization was performed at 65°C overnight, and washed three times with buffer (40 mM phosphate buffer pH 7.2, 1% SDS) for 30 min at 65°C. Radioactivity of the membranes was visualized using BioImage Analyzer, BAS2000 (Fuji Photo Inc., Tokyo, Japan).

LUC assay

Methods for visualization of LUC activity with a high-performance CCD camera (Argus system, Hamamatsu Photonics Co. Ltd, Hamamatsu, Japan) and sequential and repeated *in vivo* assays using an automated scintillation counter (TopCount, Packard, Tokyo, Japan) are described elsewhere (Kimura *et al*., 2001, 2003). A variety of expression profiles regarding tissue or organ specificity were observed not only with yy323 (Figure 4) but also with yy327 and yy376 (data not shown). The LUC activity of the trap lines was confirmed in the T_2 generation. *In vitro* dual assays as shown in Figure 6 were performed with coelenterazine and luciferin as substrates (Dual-Luciferase Reporter Assay System, Promega, Tokyo, Japan) according to the manufacturer's protocol.

Determination of T-DNA insertion sites

For sequencing of the T-DNA border, a modified version of the adaptor ligation-mediated PCR method was performed (Siebert *et al*., 1995; J. Alonso and J. Ecker, in preparation). Genomic DNA was prepared from T_2 seedlings based on a magnetic beads method (Wizard Magnetic 96 DNA Plant, Promega, Tokyo, Japan) according to the manufacturer's manual. The prepared DNA samples were digested with *Bgl*II, *Xho*I, and *Eco*RI, ligated with mixed adaptors (5′-GTA ATA CGA CTC ACT ATA GGG CAC GCG TGG TCG ACG GCC CGG GCT GGT-3′, 5′-AAT TAC CAG CCC-(NH_2)-3′, 5′-GAT CAC CAG CCC-(NH_2)-3′, and 5′- TCG AAC CAG CCC-(NH_2)-3′), and subsequently subjected to PCR and nested PCR. The first PCR was performed with a primer set of AP2 (5′-ACT ATA GGG CAC GCG TGG T-3′) and C1 (5′-TGG CGT CTT CAG CCA TAG CCA TTG TTT GGA-3′), and the cycle conditions were (94°C for 25 sec, 72°C for 3 min) × 7 cycles and (94°C for 25 sec, 67°C for 3 min) × 32 cycles. The resultant PCR products were diluted approximately 30-fold in the second PCR mixture with the aid of disposable 96 pin-replicators. The nested PCR was performed with primers AP1 (5′-GTA ATA CGA CTC ACT ATA GGG C-3′) and C4 (5′-TCT CCA GCG

GTT CCA TCC TCT AGA GGA TAG-3′), and the cycle conditions were (94°C for 30 sec, 63°C for 30 sec with a touch-down of 0.5°C per cycle, 72°C for 3 min) × 14 cycles, (94°C for 30 sec, 56°C for 30 sec, 72°C for 3 min) × 25 cycles, and (72°C for 10 min) × 1 cycle. The product of the nested PCR was precipitated in the presence of 11.9% PEG (8000), 0.85 M NaCl, and 3 mM $MgCl_2$ at room temperature for 30 min. The pellet was washed with 70% ethanol, dried, and subjected to sequencing analysis with the C4 primer, using automated DNA sequencers (ABI Prism 377XL, Applied Biosystems Tokyo; Megabase 1000, Amersham Biotech, Tokyo). The sequence data was mapped to the *Arabidopsis* genome at the TAIR blast site (http://www.arabidopsis.org/Blast/). The interpretation of the insertion sites was done using the BAC and P1 annotations at GenBank around September 2000–September 2001. Mapping of the insertion sites and the interpretation as described above were achieved with the aid of PERL and VBA programs.

Acknowledgements

We would like to acknowledge Drs J. Alonso and J. Ecker for the protocol of the adaptor ligation-mediated PCR, Dr P. Maliga for pPZP200 vector, Dr K. Hiratsuka for 221-LUC$^+$, Mr M. Kimura for technical assistance with *in vivo* LUC assays, and Ms. Yasue Ichikawa and Rie Nakazawa for DNA sequencing (Bioarchitect Research Group, RIKEN). We also thank Drs S. Kay and T. Oyama for LUC$^+$ information and technical advice about TopCount utilization, Dr J. Obokata for discussion about cryptic promoters, Dr T. Hirayama for critical reading of the manuscript, Drs Y. Yoshioka and T. Kato for discussion about gene trapping efficiency. Finally, we acknowledge the TAIR (http://www.arabidopsis.org/) for the web service.

Supplementary Material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/TPJ/TPJ1797/TPJ1797sm.htm

Table S1 Positions of T-DNA insertion

References

Aoyama, T. and Chua, N.-H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.* **11**, 605–612.

Bancroft, I., Bhatt, A.M., Sjodin, C., Scofield, S., Jones, J.D. and Dean, C. (1992) Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **233**, 449–461.

Baulcombe, D.C., Chapman, S. and Santa Cruz, S. (1995) Jellyfish green fluorescent protein as a reporter for virus infections. *Plant J.* **7**, 1045–1053.

Bevan, M. (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucl. Acids Res.* **12**, 8711–8721.

Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A.J., Allen, D.W., Xiang, H. and Jack, T. (1999) Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. *Plant J.* **17**, 699–707.

Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.

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.

Cutler, S.R., Ehrhardt, D.W., Griffitts, J.S. and Somerville, C.R. (2000) Random GFP: cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proc. Natl. Acad. Sci. USA*, **97**, 3718–3723.

Fedoroff, N.V. and Smith, D.L. (1993) A versatile system for detecting transposition in *Arabidopsis*. *Plant J.* **3**, 273–289.

Galbiati, M., Moreno, M.A., Nadzan, G., Zourelidou, M. and Dellaporta, S.L. (2000) Large-scale T-DNA mutagenesis in *Arabidopsis* for functional genomic analysis. *Funct. Integr. Genomics*, **1**, 25–34.

Gatz, C., Fronberg, C. and Wendenburg, R. (1992) Stringent repression and homogenous de-repression by tetracyclin of a modified CaMV 35S promoter in intact transgenic tobacco plants. *Plant J.* **2**, 397–404.

Hajdukiewicz, P., Svab, Z. and Maliga, P. (1994) The small, versatile *pPZP* family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989–994.

He, Y., Tang, W., Swain, J.D., Green, A.L., Jack, T.P. and Gan, S. (2001) Networking senescence-regulating pathways by using *Arabidopsis*-enhancer trap lines. *Plant Physiol.* **126**, 707–716.

Ito, T., Seki, M., Hayashida, N., Shibata, D. and Shinozaki, K. (1999) Regional insertional mutagenesis of genes on *Arabidopsis thaliana* chromosome V using the *Ac/Ds* transposon in combination with a cDNA scanning method. *Plant J.* **17**, 433–444.

Iwakura, M. and Nakamura, T. (1998) Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase. *Protein Eng.* **11**, 707–713.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Kimura, M., Yoshizumi, T., Manabe, T., Yamamoto, Y.Y. and Matsui, M. (2001) *Arabidopsis* transcriptional regulation by light stress *via* hydrogen peroxide-dependent and -independent pathways. *Genes Cells*, **6**, 607–617.

Kimura, M., Manabe, K., Abe, T., Yoshida, S., Matsui, M. and Yamamoto, Y.Y. (2003) Analysis of hydrogen peroxidase-independent expression of a high light-inducible gene for *Arabidopsis Early Light-Inducible Protein* with the aid of the *ELIP2* promoter-luciferase fusion. *Photochem. Photobiol.* **77**, 668–674.

Klimyuk, V.I., Carroll, B.J., Thomas, C.M. and Jones, J.D.G. (1993) Alkali treatment for rapid preparation of plant material for reliable PCR analysis. *Plant J.* **3**, 493–494.

Millar, A.J., Short, S.R., Chua, N.-H. and Kay, S.A. (1992a) A novel circadian phenotype based on firefly luciferase expression in transgenic plants. *Plant Cell*, **4**, 1075–1087.

Millar, A.J., Short, S.R., Hiratsuka, K., Chua, N.-H. and Kay, S.A. (1992b) Firefly luciferase as a reporter of regulated gene expression in higher plants. *Plant Mol. Biol. Rep.* **10**, 324–337.

Nussaume, L., Harrison, K., Klimyuk, V., Martienssen, R., Sundaresan, V. and Jones, J.D.G. (1995) Analysis of splice donor and acceptor site function in a transposable gene trap derived from the maize element *Activator*. *Mol. Gen. Genet.* **249**, 91–101.

Parinov, S., Sevugan, M., De, Y., Yang, W.C., Kumaran, M. and Sundaresan, V. (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell*, **11**, 2263–2270.

Pestova, T.V., Kolupaeva, V.G., Lomakin, I.B., Pilipenko, E.V., Shatsky, I.N., Agol, V.I. and Hellen, C.U. (2001) Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. USA*, **98**, 7029–7036.

Plesch, G., Kamann, E. and Mueller-Roeber, B. (2000) Cloning of regulatory sequences mediating guard-cell-specific gene expression. *Gene*, **249**, 83–89.

Rajani, S. and Sundaresan, V. (2001) The *Arabidopsis myc/bHLH* gene *ALCATRAZ* enables cell separation in fruit dehiscence. *Curr. Biol.* **11**, 1914–1922.

Sabatini, S., Beis, D., Wolkenfelt, H. *et al.* (1999) An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell*, **99**, 463–472.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schenborn, E. and Groskreutz, D. (1997) Reporter gene vectors and assays. *Mol. Biotechnol.* **13**, 29–44.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucl. Acids Res.* **23**, 1087–1088.

Skulachev, M.V., Ivanov, P.A., Karpova, O.V., Korpela, T., Rodionova, N.P., Dorokhov, Y.L. and Atabekov, J.G. (1999) Internal initiation of translation detected by the 5′-untranslated region of the tobamo virus subgenomic RNA I_2. *Virology*, **263**, 139–154.

Springer, P.S. (2000) Gene traps: tools for plant development and genomics. *Plant Cell*, **12**, 1007–1020.

Springer, P.S., McCombie, W.R., Sundaresan, V. and Martienssen, R.A. (1995) Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in *Arabidopsis*. *Science*, **268**, 877–880.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D.G., Dean, C., Ma, H. and Martienssen, R. (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* **9**, 1797–1810.

Topping, J.F. and Lindsey, K. (1997) Promoter trap markers differentiate structural and positional components of polar development in *Arabidopsis*. *Plant Cell*, **9**, 1713–1725.

Urwin, P., Yi, L., Martin, H., Atkinson, H. and Gilmartin, P.M. (2000) Functional characterization of the EMCV IRES in plants. *Plant J.* **24**, 583–589.

Worley, C.K., Zenser, N., Ramos, J., Rouse, D., Leyser, O., Theologis, A. and Callis, J. (2000) Degradation of Aux/IAA proteins is essential for normal auxin signalling. *Plant J.* **21**, 553–562.

Yamamoto, Y.Y. and Deng, X.-W. (1998) A new vector set for GAL4-dependent transactivation assay in plants. *Plant Biotechnol.* **15**, 217–220.

Yamamoto, Y.Y., Tsuji, H. and Obokata, J. (1995) 5′ leader of a photosystem I gene in *Nicotiana sylvstris*, *psaDb*, contains a translational enhancer. *J. Biol. Chem.* **270**, 12466–12470.

Yamamoto, Y.Y., Matsui, M., Ang, L.-H. and Deng, X.-W. (1998) Role of COP1-interactive protein in mediating light-regulated gene expression in *Arabidopsis*. *Plant Cell*, **10**, 1083–1094.

Yamamoto, Y.Y., Deng, X.-W. and Matsui, M. (2001) CIP4, a new COP1 target, is a nuclear-localized positive regulator of *Arabidopsis* photomorphogenesis. *Plant Cell*, **13**, 399–411.

Yoshida, M., Yagi, T., Furuta, Y., Takayanagi, K., Kominami, R., Takeda, N., Tokunaga, T., Chiba, J., Ikawa, Y. and Aizawa, S. (1995) A new strategy of gene trapping in ES cells using 3′ RACE. *Transgenic Res.* **4**, 277–287.

Table S1 Positions of T-DNA insertion

Line[1]	Insertion pattern[2]	Chr	Insertion site (bp)[3]	BAC/P1/ Contig	BAC GI	Insertion site relative to the neighboring gene[4]	MIPS protein code	gene annotation	LUC expression[5]
YB007	R	4	130488	Chr 4 contig No. 82	7270418	-526	At4g35030	leucine rich repeat receptor-like	SAM
YB009	R	1	66532	T6A9	7958959	+648	At1g02360	putative chitinase	transient
YB011	R	3	23775	F8J2	7629988	inetrgenic			light, circadian
YB023	IA	3	23037	MFE16	5041964	coding	At5g44880	similar to unknown protein	
YB025	P	4	147246	Chr.4 contig No.67	7269590	-295	At4g27730	putative protein	transient
YB026	R	2	30036	Chr II section 3 of 255	6598472	inetrgenic			light, transient
YB030	R	1	58840	F22K20	2477521	-999	At1g77000	F-box protein family, AtFBL5	
YB031	R	5	32530	F15L12	4757388	inetrgenic			
YB036	R	4	55834	Chr.4 contig No.75	7270043	+175	At4g31550	putaive DNA-binding protein	SAM
YB041	R		8118	F9H16	4757678	+666	At1g20990	Unknown protein	
YB042	R	5	45428	MRG21	3985954	inetrgenic			
YB043	R	3	24625	F16B3	6957700	inetrgenic			light
YB046	P	1	95405	F9L1	5051726	-681 (381)	At1g15350	expressed protein	root
YB050	IS	1	74027	F12B7	10998864	coding	At1g67620	unknown protein	
YB051	R	3	91746	F21F14	6899881	+119	At3g61910	strong similarity to no apical meristem (NAM)	
YB052	R	3	27940	MXC7	4757411	+787	At3g23040	unkonwn protein	
YB056	IS	3	15181	T20N10	7630060	coding	At3g58710	DNA-binding WRKY-like protein	transient
YB057	R	2	31797	Chr II section	6598400	inetrgenic			SAM

				183 of 255					
YB065	IS	2	75126	T9I4	6598462	coding	At2g29090	putative cytochrome P450	
YB071	R	1	56382	F6F9	7839909	inetrgenic			light
YB074	P	4	148232	Chr.4 contig No.33	7267889	-235	At4g12250	nucleotide sugar epimerase-like protein	
YB075	R	2	30479	Chr II section 223 of 255	6598478	inetrgenic			
YB083	R	2	19073	T1E2	6598639	+642	At2g27920	putative carboxypeptidase	
YB084	R	5	49937	MRN17	2264315	inetrgenic			transient
YB087	R	5	35747	T9L3	9755738	-505	At5g14760	L-aspartate oxidase-like protein	
YB097	R	2	70144	Chr II section 243 of 255	6598396	inetrgenic	At2g45660/ At2g45670		
YB107	R	3	29625	K16N12	5541654	-504	At3g27883	Ta11-like non-LTR retroelement protein-like, pseudogene	
YB108	P	2	2596	T20D16	6598369	-470 (62)	At2g23130	putative proline-rich protein	
YB110	IA	4	45162	Chr. 4 contig No.67	7269590	coding	At4g27460	hypothetical protein	
YB114	R	3	45559	T29H11	7630033	intergenic			
YB115	R	1	19820	F5A18	12324739	intergenic			root
YB116	IS	1	60295	T25K16	6587720	coding	At1g01130	hypothetical protein	
YB118	P	1	51314	F14J16	7798719	-441 (95)	At1g55900	unknown protein	
YC006	IS	3	32021	MGH6	7025846	coding	At3g12970	unknown protein	
YC007	R	3	57303	MJK13	7021719	-837	At3g15510	putative jasmonic acid regulatory protein	
YC008	IA	4	148794	FCA contig No.0	5280985	coding	At4g14210	phytoene dehydrogenase precursor (phytoene desaturase)	SAM
YC012	R	3	76409	T8B10	7287982	+270	At3g60540	transport protein subunit-like	
YC016	P	5	36698	MDC12	2618600	-174	At5g63140	similarity to unknown protein	
YC017	IS	2	4553	T26J13	6598427	-55 (118)	At2g41430	ERD15 (dehydration-induced protein)	

YC018	IS	5	47838	MTH12	9758826	coding	At5g59710	unknown protein	
YC021	P	4	167535	Cr. 4 contig No.2	4006885	-85 (43)	At4g36410	ubiquitin-conjugating enzyme 17 (UBC17)	
YC022	R	2	105228	Chr II section 166 of 255	6598426	+57	At2g29500	putative small heat shock protein	
YC024	R	5	27283	MHJ24	2618601	+698	At5g64110	peroxidase ATP3a homolog	
YC025	R	5	23059	MHF15	9758404	inetrgenic			
YC026	R	3	16078	MPE11	4220640	-903	At3g25890	AP2 domain transcription factor-like protein	
YC032	R	3	27519	MOJ10	4757405	+835	At3g27020	similarity to unknown protein	
YC038	R	1	16620	F17M19	12324531	inetrgenic			
YC041	IS	4	6864	AP2 contig.No. 1	4376087	coding	At4g37320	cytochrome P450-like protein	
YC043	R	1	35379	T18K17	12324309	inetrgenic			
YC045	IA	1	41683	F24J13	12325034	coding	At1g70530	putative protein kinase	
YC046	R	2	42278	T13H18	6598517	+219	At2g11180	putative retroelement pol polyprotein	
YC047	IA	1	65227	T23E18	6425618	coding	At1g76250	hypothetical protein	SAM
YC048	IA	2	26176	Chr II section 190 of 255	6598403	coding	At2g34490	putative cytochrome P450	SAM
YC051	R	3	63761	F24K9	12322889	+400	At3g11500	putative small nuclear ribonucleoprotein	
YC054	R	2	25989	F13A10	6598561	inetrgenic	At2g46520/ At2g46530		
YC055	R	2	11579	Chr II section 129 of 255	6598474	+872	At2g22760	putative bHLH transcription factor	
YC057	IS	5	16134	MJM18	4589429	coding	At5g51920	strong similarity to unknown protein	SAM
YC058	IA	5	30558	MUL3	4220641	-110 (114)	At5g57120	similar to unknown protein	
YC059	R	5	22401	MIF21	4220638	inetrgenic			

YC061	R	2	15581	Chr II section 213 of 255	6598394	+237	At2g39670	unknown protein	
YC067	R	4	100246	T22A6	5051759	inetrgenic			
YC070	P	3	52998	F18O21	7572902	-29	At3g56200	similarity to neuronal glutamine transporter -like	
YC071	R	4	23314	T19F6	2262097	+119	At4g24140		
YC074	R	1	47960	F7G19	2342673	+32	At1g09010	expressed protein	
YC076	R	5	41645	K17N15	3702724	inetrgenic			
YC077	R	5	29631	F21J6	4063730	inetrgenic			
YC088	IS	1	41753	F19K6	12323115	coding	At1g52340	putative short chain alcohol dehydrogenase	
YC089	P	2	18620	Chr II section 145 of 255	6598463	-400	At2g25770	hypothetical protein	
YC090	R	5	25921	MSJ1	2618602	inetrgenic			
YC091	R	3	4658	T28A8	7362771	inetrgenic			
YC092	R	1	47065	T25N20	4757411	inetrgenic			SAM
YC093	R	3	32770	MXC7	4757411	inetrgenic			
YC096	P	3	40928	T21L8	5541692	-464 (197)	At3g47340	glutamine-dependent asparagine synthetase	
YC104	P	2	33833	Chr II section 131 of 255	6598413	-317	At2g23010	putative serine carboxypeptidase I	
YC106	IS	2	40485	F14M13	6598614	-16 (213)	At2g22430	homeodomain transcription factor (ATHB-6)	
YC109	R	1	10930	F7H2	8099275	-758 (164)	At1g15690	inorganic pyrophosphatase, putative	
YC112	IS	5	12260	K22F20	3449314	coding	At5g37790	protein kinase - like protein	
YC117	P	5	15632	T21H19	9755818	-92	At5g16140	CRS2-like protein	
YC120	IS	4	78917	M4I22	3269280	coding	At4g27380	hypothetical protein	
YF009	IS	3	50862	MDC11	4519193	-128 (142)	At3g13310	128-, DnaJ-like protein	
YF017	IS	5	54702	MAH20	2351062	coding	At5g08600	strong similarity to unknown protein	
YF024	IA	2	50167	Chr II	6598430	-125 (399)	At2g35190	unknown protein	

				section 193 of 255				
YF026	IA	1	50298	F24D7	12324933	-8 (113)	At1g63690	expressed protein
YF028	R	1	115561	F25P12	9954738	+466	At1g56590	Putative clathrin-associated adaptor protein
YF036	IS	3	41734	T1B9	12408738	coding	At3g07210	hypothetical protein
YF043	IA	3	9278	T29H11	7630033	coding	At3g48270	similar to cytochrome P450-like protein
YF050	IS	5	58597	MUD21	2828185	coding	At5g66900	disease resistance protein (CC-NBS-LRR class), putative
YF080	IA	3	31330	F2K15	6723384	coding	At3g49170	putative protein
YF095	R	3	42184	MAA21	7573320	-530	At3g63430	putative protein
YF097	IS	3	87709	T6K12	12408745	coding	At3g04090	hypothetical protein
YF108	R	1	63970	F1M20	12324786	+230	At1g74510	hypothetical protein
YF109	P	1	39336	F12K8	6056182	-147	At1g22550	peptide transporter, putative
YF110	R	4	32430	T6K21	5738375	+391	At4g17910	putative protein
YF129	IS	5	84212	MNJ8	3510345	coding	At5g37370	putative protein
YF131	IA	3	31331	F2K15	6723384	coding	At3g49170	putative protein
YF135	R	4	7970	F26P21	3688169	intergenic		
YF138	IS	5	84212	MNJ8	3510345	coding	At5g37370	putative protein
YF141	R	5	4306	F24B18	4757390	+271	At5g54390	putative protein
YF144	IS	1	5173	T25K16	6587720	coding	At1g01010	NAC domain protein, putative
YF150	IA	1	56883	T23E23	9369387	coding	At1g23970	unknown protein
YF157	IS	3	87709	T6K12	12408745	coding	At3g04090	hypothetical protein
YF164	IS	3	41629	T1B9	12408738	coding	At3g07210	expressed protein
YF170	IA	4	184570	ATFCA8	2245073	coding	At4g17620	hypothetical protein
YF173	R	1	17002	F7H2	8099275	+483	At1g15700	ATP synthase gamma-subunit, putative
YF175	IA	5	15834	MGN6	3510342	coding	At5g53740	unknown protein
YF178	IS	2	61455	Chr II section 77 of 255	20197792	-46 (703)	At2g13800	putative receptor-like protein kinase
YF183	R	4	71246	T9A21	2832689	-333	At4g18340	glycosyl hydrolase family 17
YF187	IS	2	8110	F23H14	20197903	transcribed		Chromosome 26S ribosomal RNA

YF189	IS	5	40040	MUL8	2656030	coding	At5g39440	AKin11
YF191	P	5	12532	K18G13	3128134	-224	At5g54150	putative protein
YF192	P	1	34267	F25C20	4760411	-303	At1g11730	Avr9 elicitor response-like protein
YF193	R	2	16602	Chr II section 113 of 255	6598497	+386	At2g19440	putative beta-1,3-glucanase
YF196	IS	4	61999	T12H17	2827538	coding	At4g22780	similarity to translation factor EF-1 alpha genfamily
YF198	IA	3	3846	F15B8	4678266	coding	At3g57790	similarity to polygalacturonase
YF200	R	5	48640	MHK7	2924654	-827	At5g40870	uridine kinase-like protein
YF209	R	2	582	Chr II section 11 of 255	6598377	intergenic		
YF213	IS	4	118252	IG002N01	2191126	coding	At4g01040	glycosyl hydrolase family 18
YF216	IS	3	27096	MAG2	6045155	coding	At3g14130	glycolate oxidase

[1]Prefix indicates type of the vector; YB: yy323, YC yy327, YF: yy376

[2]Insetion types are classified into the following categories. IS: intragenic sense insertion, IA: intragenic antisense insertion, P: promoter insertion, R: rest.

[3]Insertion site of the BAC/P1/Contig sequence.

[4]Upstream insertion is indicated as minus prefix with the distance (bp) from the translation start site, and length of the 5' UTR (bp) is shown within parenthesis when mRNA information is available. Downstream insertion is indicated as plus prefix with the distance (bp) from the end of CDS.

[5]Noticed expression profiles. SAM: expression around shoot apical meristem, root: expression in roots, transient: transient expression during seedling development under continuous light, light: light activation, circadian: circadian oscillation under continuous light.

It should be mentioned that the expression analysis has not been achieved in all the lines in the list. The lack of expressional information in the table does not mean no specific expression patterns.