

TECHNICAL ADVANCE

# T-DNA insertional mutagenesis for functional genomics in rice

Jong-Seong Jeon, Sichul Lee, Ki-Hong Jung, Sung-Hoon Jun, Dong-Hoon Jeong, Jinwon Lee, Chanhong Kim, Seonghoo Jang, Shinyoung Lee, Kiyoung Yang, Jongmin Nam, Kyungsook An, Min-Jung Han, Ryo-Jin Sung, Hyun-Sook Choi, Jung-Hwa Yu, Jung-Hwan Choi, Se-Yu Cho, Sang-Su Cha, Shi-In Kim and Gynheung An\*

National Research Laboratory of Plant Functional Genomics, Division of Molecular and Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Republic of Korea

Received 14 January 2000; revised 21 March 2000; accepted 21 March 2000.

\*For correspondence (fax +82 562 279 2199; e-mail genean@postech.ac.kr).

---

## Summary

We have produced 22 090 primary transgenic rice plants that carry a T-DNA insertion, which has resulted in 18 358 fertile lines. Genomic DNA gel-blot and PCR analyses have shown that approximately 65% of the population contains more than one copy of the inserted T-DNA. Hygromycin resistance tests revealed that transgenic plants contain an average of 1.4 loci of T-DNA inserts. Therefore, it can be estimated that approximately 25 700 taggings have been generated. The binary vector used in the insertion contained the promoterless  $\beta$ -glucuronidase (GUS) reporter gene with an intron and multiple splicing donors and acceptors immediately next to the right border. Therefore, this gene trap vector is able to detect a gene fusion between GUS and an endogenous gene, which is tagged by T-DNA. Histochemical GUS assays were carried out in the leaves and roots from 5353 lines, mature flowers from 7026 lines, and developing seeds from 1948 lines. The data revealed that 1.6–2.1% of tested organs were GUS-positive in the tested organs, and that their GUS expression patterns were organ- or tissue-specific or ubiquitous in all parts of the plant. The large population of T-DNA-tagged lines will be useful for identifying insertional mutants in various genes and for discovering new genes in rice.

---

## Introduction

There has been much progress in the development of strategies to discover the function of plant genes. Development of the strategies has been largely based on genetic approaches such as mutant identification and map-based gene isolation (reviewed in Martin, 1998). Gene inactivation by insertion of a transposon has been employed for functional studies in several plant species. The use of transfer DNA (T-DNA) as a mutagen has also been developed for tagging genes in *Arabidopsis* (Babiychuk *et al.*, 1997; Feldmann, 1991; Krysan *et al.*, 1999). It is believed that T-DNA insertion is a random event, and that the inserted genes are stable through multiple generations (reviewed in Azpiroz-Leehan and Feldmann, 1997).

Insertional mutagenesis has become an attractive method for functional analysis due to the development of several strategies for screening T-DNA or transposon insertions in a known gene and recovering sequences

flanking the insertions (Cooley *et al.*, 1996; Couteau *et al.*, 1999; Frey *et al.*, 1998; Koes *et al.*, 1995; Krysan *et al.*, 1999; Liu and Whittier, 1995). Through sequencing PCR-amplified fragments adjacent to the inserted element, a flanking sequence database has been constructed in *Arabidopsis* (Parinov *et al.*, 1999; Tissier *et al.*, 1999).

Reporter genes as insertional elements have been utilized to aid in the identification of insertions within functional genes (Campisi *et al.*, 1999; Kertbundit *et al.*, 1991; Kertbundit *et al.*, 1998; Sundaresan *et al.*, 1995; Topping *et al.*, 1991). An enhancer trap contains a weak minimal promoter fused to a reporter gene, and a gene trap contains multiple splicing sites fused to a reporter gene. The GUS gene is the most frequently used as a reporter gene because of the accurate detection of its gene products and the tolerance of N-terminal translational fusions in its enzyme activity (Jefferson *et al.*, 1987).

Rice is a model plant of cereal species because of its relatively small genome size, efficient tools for plant transformation, construction of physical maps, large-scale analysis of expressed sequence tags (ESTs) and international genome sequencing projects, as well as economic importance (Hiei *et al.*, 1994; Sasaki, 1998). Therefore, development of insertional mutant lines will be extremely valuable for the functional genomics of rice. To this end, we are producing a large population of rice lines that are tagged with T-DNA. In this study, we report the generation of 18 358 independent fertile transgenic lines, which have been tagged with a gene trap system.

## Results

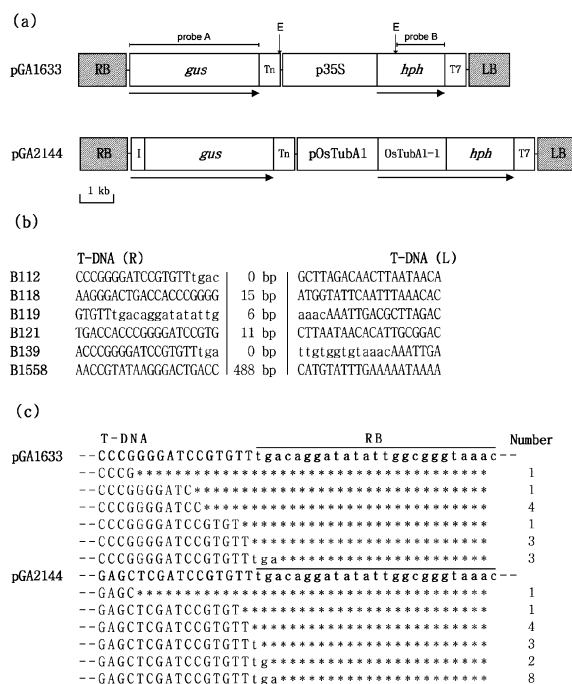
### Vector construction for insertional mutagenesis in rice

Two binary vectors were constructed for T-DNA insertional mutagenesis of rice (Figure 1a). The first plasmid pGA1633 contains the promoterless GUS gene immediately next to the right border and the cauliflower mosaic virus (CaMV) 35S promoter–hygromycin phosphotransferase (*hph*) chimeric gene as a selectable marker. The second plasmid, pGA2144, was constructed to increase the gene trap efficiency. In this plasmid, an intron carrying three putative splicing donors and acceptors was placed in front of GUS (see Experimental procedures). In pGA2144, we replaced the CaMV 35S with the strong promoter of the rice  $\alpha$ -tubulin gene *OsTubA1* along with its first intron (J.-S. Jeon, S. Lee, K.-H. Jung, S.-H. Jun, C. Kim and G. An, unpublished data) for expression of the selectable marker *hph* gene.

### Production of T-DNA-tagged transgenic rice plants

Scutellum-derived embryonic calli were co-cultivated with *Agrobacterium tumefaciens* LBA4404 carrying the binary tagging vector. Approximately 20–40% of the co-cultivated calli produced hygromycin-resistant cells. The frequency of plant regeneration from the calli ranged from 50–85% (data not shown). *Agrobacterium*-mediated rice transformation procedures have been developed using the system based on the super-virulent strain and super-binary vectors carrying the virulence region of pTiBo542 (reviewed in Hiei *et al.*, 1997). Our result showed that the transformation efficiency of our system was as high as the super-binary vector system, indicating that the *Agrobacterium* strain LBA4404 and a common binary vector can be used for efficient transformation of rice. With this system, we have produced 1590 transgenic plants transformed with pGA1633 and 20 500 transgenic plants transformed with pGA2144.

The seed fertility of the primary transgenic plants varied significantly, ranging from complete sterility to full fertility.



**Figure 1.** Maps of the T-DNA tagging vectors and analysis of T-DNA integration patterns.

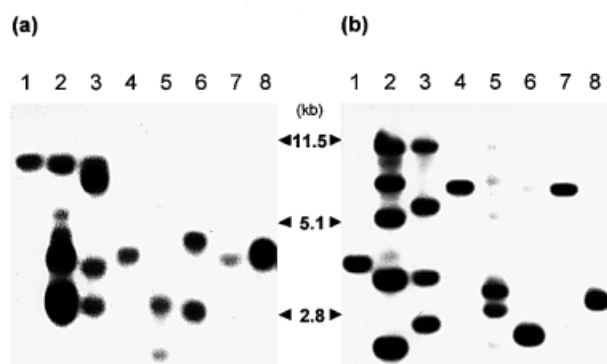
(a) Schematic diagrams of pGA1633 and pGA2144. The RB and LB in grey bars represent the right and left borders of T-DNA, respectively. E, *EcoRI* site; *gus*,  $\beta$ -glucuronidase; Tn, nopaline synthase (*nos*) terminator; p35S, cauliflower mosaic virus 35S promoter; *hph*, hygromycin phosphotransferase; T7, transcription termination region of gene 7 of the pTiA6; I, the *OsTubA1* intron3 carrying three putative splicing acceptor and donor sites; pOsTubA1, the promoter of the rice  $\alpha$ -tubulin gene, *OsTubA1*; OsTubA1-1, the first intron of *OsTubA1*. The GUS probe (probe A) and *hph* probe (probe B) used for DNA blot analyses (Figure 2) are indicated. (b) Sequences at the junctions between two directly integrated T-DNA copies. The capitals indicate T-DNA sequences immediately next to the right (R) and left (L) border sequences (lower case). The length of filler sequences is depicted. (c) Sequences of T-DNA at the junction with plant DNA. The T-DNA sequences of pGA1633 and pGA2144 are shown in capitals and the right border (RB) sequences are shown in lower case. Asterisks signify any nucleotides presumed to originate from the rice genome.

Of the 22 090 primary transgenic plants, 1338 lines (84%) of pGA1633 and 17 020 lines (83%) of pGA2144 produced fertile seeds. Seventeen per cent of the population was sterile, and 13% generated fewer than 10 seeds. About half of the population produced more than 100 seeds and 8% generated 50–100 seeds. The remaining plants generated 10–50 seeds. The pGA1633 lines were amplified, and the majority of the transgenic plants became fully fertile in the next generation. However, approximately one half of the transgenic plants, which showed partial sterility (fewer than 50 seeds) at the primary generation remained partially sterile, suggesting that the low fertility was due to genetic alteration by either T-DNA or other mutations. The pGA2144 lines are being amplified to produce enough seeds to be utilized for further studies.

**Table 1.** Estimation of T-DNA copy numbers

Transformant	Copy numbers in T1 generation <sup>a</sup>		Hygromycin resistance in T2 generation		Number of genetic loci <sup>b</sup>	$\chi^2$
	GUS	<i>hph</i>	Resistant	Sensitive		
B49	2	2	81	4	2	0.35
B71	3	1	58	14	1	1.19
B111	1	1	88	22	1	1.47
B112	2	2	94	7	2	0.08
B113	5	3	79	22	1	0.56
B116	1	1	73	19	1	0.93
B117	1	1	68	25	1	0.18
B118	2	2	77	18	1	1.86
B119	3	2	59	29	1	2.97
B120	1	1	59	14	1	1.32
B121	2	2	82	10	2	3.35
B123	2	3	66	16	1	1.32
B126	1	1	36	18	1	2.00
B128	3	2	53	10	1	2.80
B129	4	2	90	10	2	2.40
B120	3	2	84	8	2	0.94
B131	1	1	64	22	1	0.02
B132	1	1	53	17	1	0.02
B134	3	4	83	20	1	1.71
B835	2	2	44	19	1	0.89
B914	3	3	64	2	3	0.92
B1019	1	1	91	25	1	0.74
B1024	3	5	70	31	1	1.75
B1031	1	1	74	19	1	1.04
B1032	4	4	58	1	3	0.01
B1037	2	2	96	8	2	0.37
B1062	3	4	65	19	1	0.25
B1064	1	1	61	31	1	3.71
B1065	1	1	55	20	1	0.11
B1071	2	1	50	20	1	0.48
B1558	3	2	70	33	1	2.72
B1559	3	2	68	29	1	1.24
B1560	4	6	62	8	2	2.65
B1561	2	4	129	7	2	0.28
Average	2.2	2.1			1.4	

<sup>a</sup>The copy number was determined in the T1 generation by genomic DNA gel-blot analysis with the GUS or *hph* probe. <sup>b</sup>The number of loci was determined by hygromycin-resistant ratios that give the smallest  $\chi^2$  values.



**Figure 2.** DNA blot analyses of the T-DNA tagged lines. Genomic DNAs from the primary transgenic lines were digested with *Eco*RI and hybridized with the GUS probe (a) or *hph* probe (b). Lane 1, B1019; lane 2, B1024; lane 3, B1032; lane 4, B1031; lane 5, B1037; lane 6, B1071; lane 7, B1065; lane 8, B1064.

#### Molecular characterization of T-DNA integration pattern in transgenic rice plants

The number of integrated T-DNA in each plant was estimated from randomly selected primary transformants (Figure 2). Table 1 is a summary of the genomic DNA gel-blot analysis using the GUS or *hph* coding region as a probe. Among the 34 transgenic lines examined, 11 lines carried a single copy of the GUS gene and 13 carried a single copy of the *hph* gene. The remaining lines carried two or more copies of GUS or *hph*. This result indicates that approximately 35% of the transgenic lines carry a single T-DNA insert. In several lines, the numbers of GUS and *hph* genes were different from each other, probably due to T-DNA re-arrangement during the transformation process (Ohba *et al.*, 1995; see below).

**Table 2.** The frequency of GUS expression in leaves, roots, flowers and immature seeds of transgenic rice plants

Organs	Number of lines tested	GUS-positive lines	
		<i>n</i>	%
Leaves	5353	106	2.0
Roots	5353	113	2.1
Flowers	7026	133	1.9
Immature seeds	1948	31	1.6

**Table 3.** GUS assay in leaves of transgenic rice plants<sup>a</sup>

GUS staining pattern	GUS-positive lines	
	<i>n</i>	%
Mesophyll	14	13.2
Spots	14	13.2
Veins	46	43.4
Veins, mesophyll	28	26.4
Veins, spots	2	1.9
Mesophyll, spots	2	1.9
Total	106	100

<sup>a</sup>5353 lines were tested.

The number of T-DNA insertion loci was analysed by scoring hygromycin-resistant progeny (T2) of the primary transgenic plants (T1). Twenty-four of 34 lines appeared to carry T-DNA at one locus, while the remaining 10 lines contained unlinked T-DNA insertion (Table 1). This indicates that transgenic plants contain an average of 1.4 loci of T-DNA inserts. These data are quite similar to the results observed in *Arabidopsis* indicating that T-DNA tagged plants contain an average of 1.4 inserts (Feldmann, 1991). The number of insertion loci that was estimated by hygromycin resistance was smaller than the number of T-DNA copies evaluated by the DNA gel-blot analysis (Table 1). This result was probably due to tandem integration of two or more T-DNA copies into a single chromosome as observed previously in dicot plants (Krizkova and Hroudá, 1998). A PCR approach was undertaken to investigate T-DNA arrangement of the lines that carry multiple T-DNAs at a single chromosome. The result showed that T-DNA copies were arranged in direct or inverted repeats (data not shown). We carried out sequence analysis of the regions between the T-DNA borders from six lines that carry multiple T-DNA copies at a single locus. The results revealed that two lines did not contain any DNA sequences between the T-DNAs (Figure 1b). The remaining four lines carried 6–488 bp of filler DNA. Interestingly, the 488 bp of the longest filler DNA in the B1558 line was found to be a portion of the GUS

gene. A DNA gel-blot analysis confirmed that the B1558 line had one more copy of GUS than *hph* (Table 1). Such a partial T-DNA was previously reported from dicots, such as tobacco (Krizkova and Hroudá, 1998). It may be explained by the suggestion that the formation of repeated T-DNA copies might result from co-integration of several intermediates into one target site.

It has previously been reported that a majority of the T-DNA insertions occur within the right border at a specific locus (reviewed in Tinland, 1996). To examine whether the same was true for our tagging lines, the junction regions between rice genomic DNA and the T-DNA right border were sequenced (Figure 1c). The sequencing results revealed that the boundaries in most of the rice lines did not correspond to the T-DNA nicking position found in *Arabidopsis* and tobacco transgenic plants. In dicot species, most T-DNAs were nicked after the first or second base of the right border. In our tagging lines, five were similar to those of *Arabidopsis* and tobacco. However, the most frequent junction point (11 out of 32 lines) was after the third base of the right border. In seven lines, the junction was at the boundary between T-DNA and the right border. The remaining nine lines showed deletion of 1–12 bases of T-DNA. It was previously reported that two of three right boundaries in transgenic rice plants and four of ten in transgenic maize plants carried three bases originated from the right border (Hiei *et al.*, 1994; Ishida *et al.*, 1996).

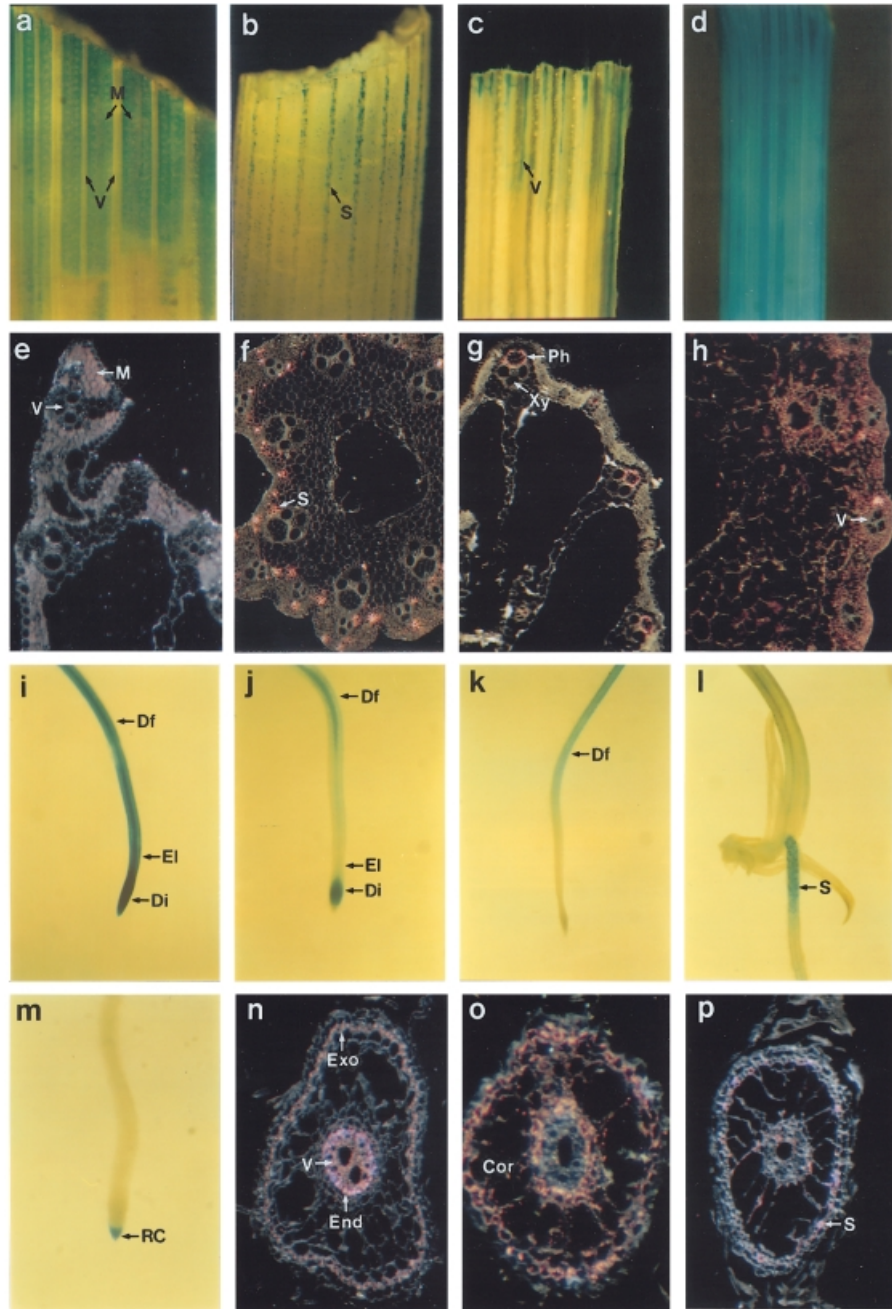
#### Expression of the GUS gene in transgenic rice plants

To evaluate the efficiency of the gene trap system, the GUS expression pattern was examined from various organs of primary transgenic plants transformed with pGA2144. We have analysed GUS activities in the leaves and roots from 5353 lines, mature flowers from 7026 lines, and developing seeds from 1948 lines. The results revealed that the efficiency of GUS staining was 2.0% (106/5353) for leaves, 2.1% (113/5353) for roots, 1.9% (133/7026) for flowers, and 1.6% (31/1948) for immature seeds (Table 2). Among the 106 GUS-positive lines in leaves, 15 (14.2%) were leaf-specific. Likewise, 25 (22.1%) lines were root-specific among the 113 GUS-positive lines in roots. We also obtained data indicating that the efficiency of GUS expression in pGA1633 lines was 1.1% (8/750) for leaves and 0.9% (7/750) for roots. These values are lower than that of pGA2144, indicating that the modified *OsTubA1* intron increased GUS tagging efficiency.

The staining patterns of the 106 lines that showed GUS activity in leaves were observed in detail (Table 3). The vein-preferential GUS staining pattern was the most frequently observed (43.4%), and 14 (13.2%) lines were stained preferentially in mesophyll cells between veins (Figure 3). In most samples, GUS staining was observed

**Figure 3.** Analysis of GUS activity in vegetative tissues of transgenic rice plants.

(a) Line G006034 exhibits mesophyll cell-specific GUS staining in the leaf. (b) Line G006167 exhibits dispersed spots of GUS staining in the leaf. (c) Line B000071 exhibits preferential GUS activity in leaf veins. (d) Line G006006 exhibits GUS activity broadly in the stem. (e) A cross-section of line G006034. (f) A cross-section of the leaf sheath of line G006167 displaying strong GUS activity in a few cells of chlorenchyma tissue between the vascular bundle and surrounding cells. (g) A cross-section of line B000071 displaying phloem-specific GUS staining. (h) A cross-section of line G006006 showing that GUS activity was localized in most tissues except for vascular bundles. (i) Line G006063 exhibits GUS activity strongly in veins in the zone of cell division, elongation and differentiation. (j) Line G006023 exhibits GUS activity strongly in the zone of cell division and cell differentiation. (k) Line G006024 exhibits strong GUS activity in the cortex in the zone of cell differentiation. (l) Line G006060 exhibits GUS activity in spots. (m) Line G006044 exhibits GUS activity in root caps. (n) A cross-section of the zone of cell differentiation of line G006023. (o) A cross-section of line G006024. (p) A cross-section of the spotted region of line G006060. GUS activity under the dark field appeared as pink or red. Cor, cortex; Df, zone of cell differentiation; Di, zone of cell division; El, zone of cell elongation; End, endodermis; Exo, exodermis; M, mesophyll; Ph, phloem; RC, root cap; S, GUS spots; V, veins; Xy, xylem.



strongly in boundary regions exposed by cutting. It is likely that a high concentration of cellulose, lignin, silica cells, and wax in rice leaves could have obstructed penetration of the GUS substrates. The GUS staining patterns in roots are summarized in Table 4. A majority of the lines showed GUS staining in the area of cell differentiation, and more than half of the lines exhibited GUS activity in the area of cell elongation or cell division. We also characterized the GUS staining patterns in transgenic flowers (Table 5). Among the 133 lines that showed GUS activity in flowers, 50 (37.6%) displayed

intense GUS staining primarily in the palea and lemma (Figure 4). One line exhibited GUS activity only in glumes (Figure 4h), eight lines showed GUS activity only in lodicules, and four lines only in a carpel (Figure 4j). Of the 11 lines exhibiting stamen-specific GUS activity, seven showed pollen-specific GUS staining (Figure 4i). The developing seeds were also subjected to GUS staining 5–10 days after flowering. A large portion of these lines showed a tissue-preferential expression pattern. For example, line G930726 exhibited an aleurone layer-preferential GUS staining pattern (Figure 5a,b), indicating that

**Table 4.** GUS assay in roots of transgenic rice plants

GUS staining pattern	GUS-positive lines	
	<i>n</i>	%
GUS activity in various regions <sup>a</sup>		
Cap	1	0.9
Differentiation	36	31.9
Cap, division	4	3.5
Cap, differentiation	3	2.7
Division, differentiation	2	1.8
Elongation, differentiation	7	6.2
Differentiation, spots	3	2.7
Cap, division, elongation	2	1.8
Cap, division, differentiation	2	1.8
Division, elongation, differentiation	16	14.2
Cap, division, elongation, differentiation	37	32.7
Total	113	100
GUS activity in various tissues <sup>b</sup>		
Epidermis	3	4.1
Vein	14	18.9
Epidermis, vein	1	1.4
Exodermis, endodermis, vein	4	5.4
Exodermis, cortex, endodermis	11	14.9
Exodermis, cortex, endodermis, vein	34	45.9
Epidermis, exodermis, cortex, endodermis, vein	7	9.4
Total	74	100

<sup>a</sup>5353 lines were tested.

<sup>b</sup>Seventy-four of the GUS-positive lines were analysed by thin section.

**Table 5.** GUS assay in flowers of transgenic rice plants<sup>a</sup>

GUS staining pattern	GUS-positive lines	
	<i>n</i>	%
Palea/lemma	50	37.6
Lodicules	8	6.0
Stamens	11	8.3
Carpel	4	3.0
Palea/lemma, lodicules	9	6.8
Palea/lemma, stamens	8	6.0
Palea/lemma, carpel	5	3.8
Lodicules, stamens	1	0.8
Lodicules, carpel	3	2.3
Stamen, carpel	9	6.8
Palea/lemma, lodicules, stamens	4	3.0
Palea/lemma, lodicules, carpel	2	1.5
Palea/lemma, stamens, carpel	3	2.3
Lodicules, stamens, carpel	1	0.8
Palea/lemma, lodicules, stamens, carpel	8	6.0
Rachilla	6	4.5
Glumes	1	0.8
Total	133	100

<sup>a</sup>7026 lines were tested.

the trapped gene might be involved in formation of the aleurone layer or in a specific function in the tissue.

## Discussion

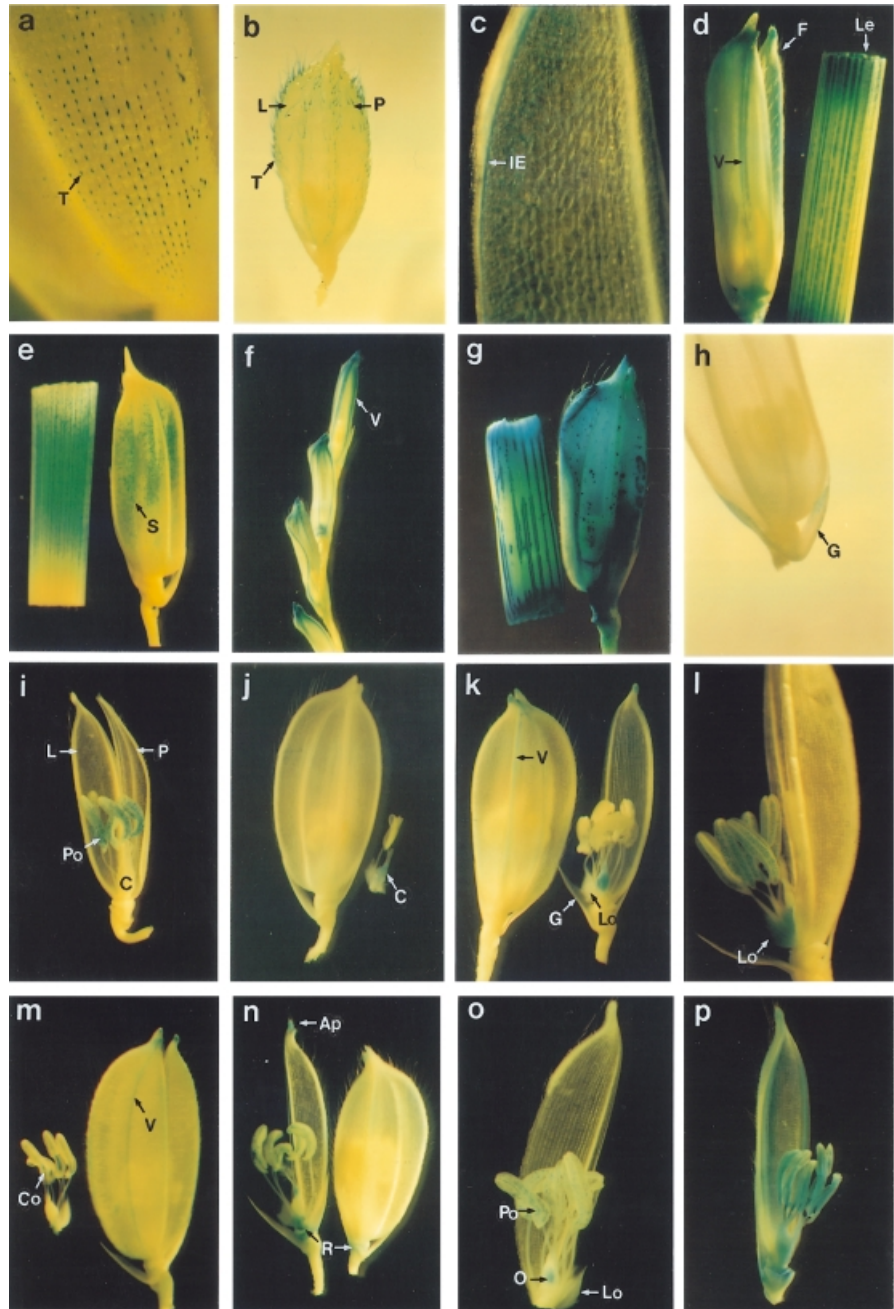
In this study, 18 358 fertile rice lines, which were tagged by T-DNA, were generated. DNA gel-blot analysis of randomly selected transgenic lines has indicated that they carry an average of 2.2 T-DNA copies. This value is lower than that observed from transgenic rice plants resulting from direct DNA transfer procedures (Kohli *et al.*, 1998). The low copy value of T-DNA insertion should facilitate characterization of mutants and isolation of the sequence flanking the T-DNA.

The number of T-DNA-tagged lines that would be required for saturating the rice genome can be estimated using the formula suggested in Krysan *et al.* (1999). The following three facts determine the number. First, the mean size of rice genes can be deduced from the 1 766 754 bp of genomic sequence that has been published in the DDBJ/EMBL/GenBank databases (AB023482, AB026295, AP000391, AP000399, AP000492, AP000559, AP000616, AP157903, AP000815, AP000816, AP000836, AP000837). Within these reported sequences, there are 331 putative genes that have been identified functionally or by exon prediction algorithms. The mean size of the rice genomic DNAs between the start and stop codons including introns is 2.6 kb. Because the upstream and downstream sequences flanked by the start and stop codons were not included, an average length of rice genes should be at least 3.0 kb. Second, the mean number of T-DNA loci distributed among the transgenic rice population was 1.4. Third, the haploid genome size of rice is  $4.3 \times 10^8$  bp (Arumuganathan and Earle, 1991). If we consider a 99% probability that a T-DNA is located within a given gene, we would require approximately 660 000 insertions or 471 000 tagging lines. Therefore, it would be difficult to generate a transgenic rice population in which every gene has been mutated. As the probability is lowered, the number of transgenic plants required becomes exponentially lower (Krysan *et al.*, 1999). It can be estimated that our current tagging lines would provide a 20% probability of finding a T-DNA insertion within a given gene of size 3 kb.

The GUS activation frequency ranged between 1.6 and 2.1% in various organs. Since GUS activity was observed from more than one organ in a number of lines, the GUS activation frequency of the T-DNA-tagged lines would be smaller than sum of the values obtained from each organ. Because we have not examined GUS activity after induction by certain environmental conditions or chemicals such as growth substances, the total GUS tagging efficiency will be higher. We have observed that around 7% of transgenic calli showed GUS staining (data not shown). Analysis of the reported 1 766 754 bp genomic sequence indicated that up to 50% of the genomic DNA is intragenic. Considering that insertion could occur in both

**Figure 4.** Analysis of GUS activity in flowers of transgenic rice plants.

(a) Line E000261 exhibits trichome-specific GUS expression. (b) Line G904527 exhibits GUS activity preferentially in long trichomes on veins. (c) Line G915918 exhibits GUS activity in the inner epidermis of the palea/lemma. (d) A palea/lemma and a leaf in line G006034 exhibit GUS activity in the boundary regions of veins. (e) Line G010031 exhibits GUS activity in a pattern of spots in the leaf and palea/lemma. (f) Line B000071 exhibits GUS activity in veins specifically. (g) Line G008193 exhibits an overall GUS activity in palea/lemma with densely stained spots. A leaf also shows strong GUS activity in the same line. (h) Line G905901 exhibits GUS activity in glumes. (i) Line G906024 exhibits pollen-specific GUS expression. (j) Line G915417 exhibits carpel-specific GUS expression. (k) Line G915412 exhibits GUS activity strongly in the carpel and weakly in veins of the palea/lemma. (l) Line G919302 exhibits strong GUS activity in lodicules, pollens of anthers, and the base of a carpel. (m) Line G922021 exhibits GUS activity in veins of the palea/lemma and the connective tissues of anthers. (n) Line G915827 exhibits GUS activity in the apiculi of the palea/lemma, stamens, a carpel and rachilla. (o) Line G917211 exhibits GUS activity in pollen and an ovary. (p) Line G916433 exhibits GUS activity in all floral organs. Ap, apiculi of palea/lemma; C, carpel; Co, connective tissue; G, glumes; IE, inner epidermis of palea/lemma; F, flower; L, lemma; Le, leaf; Lo, lodicules; O, ovary; P, palea; Po, pollens; R, rachilla; S, GUS spots; T, trichome; V, veins.

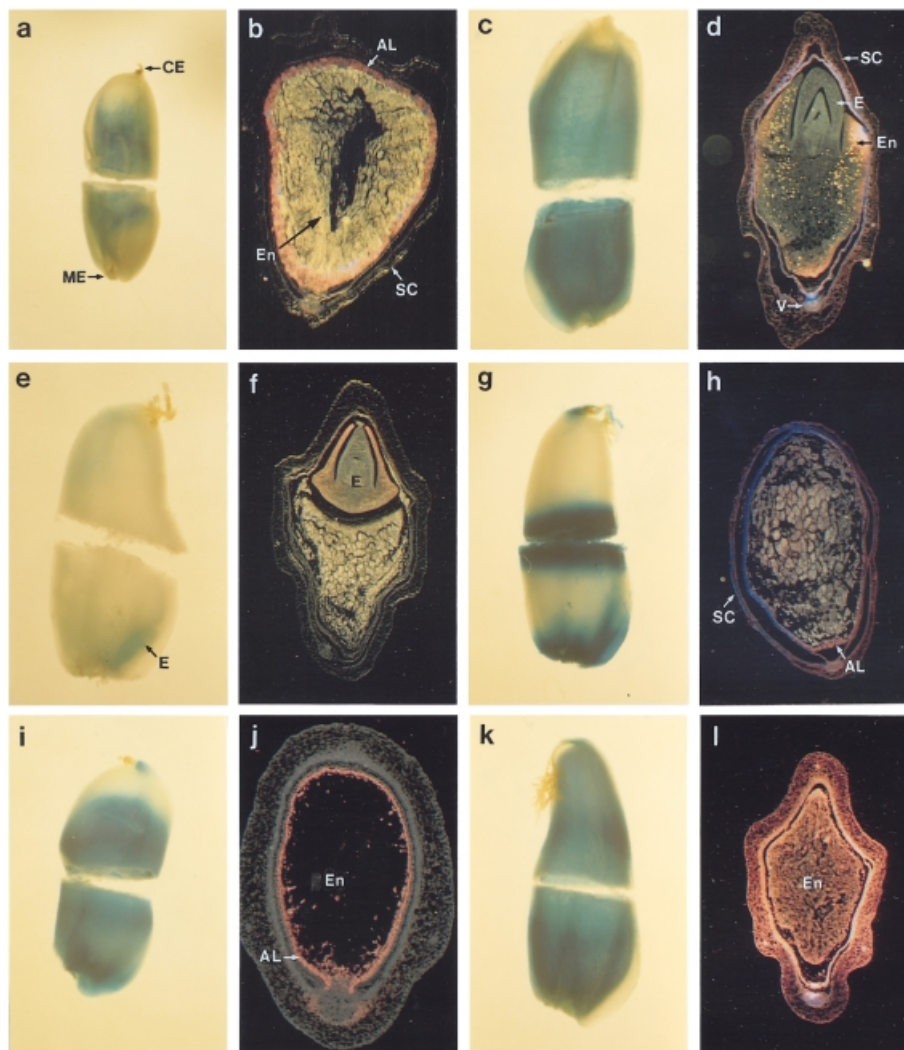


orientations, the maximum GUS tagging efficiency would be 25% of the total population.

A variety of GUS staining patterns was observed from the tagged lines. Some were tissue- or organ-specific, while others were expressed ubiquitously. This observation supports the fact that T-DNA insertion is a random event. We are currently isolating the flanking sequences of the GUS-positive lines. We will study in the next generation whether these lines display any mutant phenotypes in

the organs where the GUS gene was activated, and whether the phenotypes co-segregate with the T-DNA.

Insertional lines that exhibit a particular GUS staining pattern should facilitate identification of genes that are regulated spatially and temporally for plant development. The *Arabidopsis LRP1* (*lateral root primordium 1*) gene, which may play a role in lateral root development, was identified by expression of promoterless GUS expression in tagging plants (Smith and Fedoroff, 1995). The



**Figure 5.** Analysis of GUS activity in developing seeds of transgenic rice plants.

(a) Line G930726 exhibits GUS activity strongly in the aleurone layer. (b) A cross-section of line G930726. (c) Line G931910 exhibits GUS activity strongly in marginal regions of endosperm, aleurone layer, inner seed coats and veins. (d) A cross-section of line G931910. (e) Line G933524 exhibits GUS activity preferentially in the embryo. (f) A cross-section of line G933524. (g) Line G932817 exhibits GUS activity strongly in the cut region, and chalazal and micropylar ends. (h) A cross-section of the cut region in line G932817. (i) Line G932825 exhibits GUS activity in the aleurone layer and endosperm. (j) A cross-section of line G932825. (k) Line G928607 exhibits GUS activity in all tissues. (l) A cross-section of line G928607. All tissues including embryo, seed coat, and endosperm exhibit GUS activity. Samples were cut to allow an efficient penetration of substrates, and then subjected to GUS staining. AL, aleurone layer; CE, chalazal end; E, embryo; En, endosperm; ME, micropylar end; SC, seed coat; V, veins.

*Arabidopsis PROLIFERA* gene, which is related to the *MCM2-3-5* family of yeast genes, was also cloned by gene trap transposon mutagenesis (Springer *et al.*, 1995).

The progeny phenotypes of the pGA1633 transgenic lines were examined. Some of the tagged lines showed mutant phenotypes, including early flowering, tallness, dwarfism, spotted leaves, chlorophyll deficiency, depressed palea, filamentous flowers, extra glumes, long sterile glumes, zebra coloration (transverse green and chlorotic bands in leaves), etc. (data not shown). Whether any of these mutant phenotypes are due to the T-DNA insertion is under investigation.

It is expected that the genome sequence of rice will be completed in the near future. This will produce a large number of genes whose function is unknown. One of the most efficient ways to obtain information on the function of a gene is to create a loss-of-function mutation and to study the phenotype of the resulting mutant. If a large population of mutagenized plants is available, it is possible to detect an insertion within the gene of interest by PCR using oligonucleotide primers from the insertional element and the gene of interest (Coureau *et al.*, 1999; Krysan *et al.*, 1999; Sato *et al.*, 1999). Identification of the desired mutant could be accomplished efficiently using a super-

pooling strategy as suggested by Krysan *et al.* (1999). They estimated that the maximum useful pool size is 2350 lines in *Arabidopsis* based upon the sensitivity for detecting a specific T-DNA insert and the total amount of template DNA. We are performing experiments to determine the upper size limit on DNA pools of the rice tagged lines.

Our T-DNA tagging lines will be useful in analysing the function of a number of valuable genes by various approaches. With the increasing availability of rice genome sequences from public databases, it would be valuable to construct a database for the flanking sequences. The development of effective methods for amplification of sequences adjoining the insertion would enable us to construct the database effectively.

## Experimental procedures

### Construction of vectors

The pGA1633 vector was constructed by insertion of the GUS gene derived from pBI101.1 into the *Bam*HI site of pGA1605 (Lee *et al.*, 1999), which contains multi-cloning sites, *Bam*HI, *Hind*III, *Xba*I, *Sac*I, *Hpa*I, *Asp*718 and *Clal*, and 35S-*hph*. There is no translation initiation or stop codon between the right border of the T-DNA and *Bam*HI site in pGA1633.

Another binary vector pGA2144 was constructed by inserting the modified intron 3 of *OsTubA1* (accession number AF182523) in front of the GUS gene. The *OsTubA1* intron 3 was used as a template. The PCR primers used were 5'GGGTCGACGAGG-TACAAGGTACAAGGTACAGACTTGATCCTT3' and 5'CG-GGTACCCTGCATATAACCTGCATATAACCTGCACATTAGCAATAAA3'. The underlined sequences correspond to *Sal*I and *Asp*718 sites. The primers were designed according to the splicing donor and acceptor sites of Sundaresan *et al.* (1995). The amplified fragment was digested with *Sal*I and *Asp*718, and then cloned between *Xho*I and *Asp*718 in front of GUS in pGA1942, which contains multi-cloning sites (*Sac*I, *Xho*I, *Asp*718 and *Clal*) and 0.5 kb of *OsTubA1* promoter-*OsTubA1* intron 1-*hph*. The resulting plasmid was named pGA2020. Finally, the GA2144 plasmid was constructed from pGA2020 by replacing the 0.5 kb *OsTubA1* promoter fragment with the 1.0 kb *OsTubA1* promoter.

### Production and growth of transgenic rice plants

Rice transformation was performed by *Agrobacterium*-mediated co-cultivation methods as previously described (Jeon *et al.*, 1999; Lee *et al.*, 1999). All transgenic rice plants were generated on a 40 mg l<sup>-1</sup> hygromycin B-containing medium. The regenerated plants were grown in a greenhouse of typically 30°C during the day and 20°C at night. The light/dark cycle in the greenhouse was 14/10 h.

### Test of the progeny for hygromycin resistance

Sterilized seeds were sown on a 70 mg l<sup>-1</sup> hygromycin B-containing MS medium and cultured under continuous illumination. Hygromycin resistance was scored 14 days after germination.

### DNA gel-blot analysis

Genomic DNA was isolated from mature leaves at the heading stage as described previously (Dellaporta *et al.*, 1983). Genomic DNA (5 µg) was digested with *Eco*RI, separated on a 0.7% agarose gel, blotted onto a nylon membrane, and hybridized with a <sup>32</sup>P-labelled probe. The GUS probe was prepared from the 1.8 kb *Bam*HI-*Eco*RI fragment and the *hph* probe was from the 0.7 kb *Eco*RI fragment. All blot analysis procedures were carried out as described previously (Kang *et al.*, 1998).

### Isolation of the sequence flanking T-DNA and the junction sequence between two integrated T-DNAs

The sequence flanking T-DNA was isolated by thermal asymmetric interlaced PCR as previously described (Liu and Whittier, 1995). The specific primer for the first cycle was 5'GCCGTAATGAGTGACCGCATCG3' (Gus1); the second was 5'ATCTGCATCGGCGAACTGATCG3' (Gus2); and the third was 5'CACGGGTTGGGTTTCTACAGG3' (Gus3). The junction between two integrated T-DNAs was amplified by PCR using primers Gus3 and 5'GCTTGACTATAATACCTGAC3' (T7). PCR products were sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA, USA).

### GUS assay

Histochemical GUS staining was performed according to Dai *et al.* (1996), except for addition of 20% methanol to the staining solution. After staining, tissues were fixed in a solution containing 50% ethanol, 5% acetic acid and 3.7% formaldehyde, and embedded in a Paraplast (Sigma). The samples were sectioned to 10 µm thickness and observed under a microscope using dark-field illumination.

### Acknowledgements

We thank Gi-Hwan Yi for sharing us with the Dongjin seeds, Steve Goffs, Bob Dietrich, Lili Chen, Shiping Zhang and Hong Gil Nam for helpful discussions, and Chahm An for critical reading of the manuscript. This work was supported in part by grants from Novartis Crop Protection Inc., and from the NRL Program from Korea Institute of Science and Technology Evaluation and Planning (KISTEP).

### References

- Arumuganathan, K. and Earle, E.D. (1991) Nuclear DNA contents of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218.
- Azpiroz-Leehan, R. and Feldmann, K.A. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. *Trends Genet.* **13**, 152–156.
- Babiychuk, E., Fuangthong, M., Van Montagu, M., Inze, D. and Kushnir, S. (1997) Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach. *Proc. Natl Acad. Sci. USA*, **94**, 12722–12727.
- 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.

- Cooley, M.B., Goldsbrough, A.P., Still, D.W. and Yoder, J.I. (1996) Site-selected insertional mutagenesis of tomato with maize *Ac* and *Ds* elements. *Mol. Genet.* **252**, 184–194.
- Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D. and Doutriaux, M.-P. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of *Arabidopsis*. *Plant Cell*, **11**, 1623–1634.
- Dai, Z., Gao, J., An, K., Lee, J.M., Edwards, G.E. and An, G. (1996) Promoter elements controlling developmental and environmental regulation of a tobacco ribosomal protein gene *L34*. *Plant Mol. Biol.* **32**, 1055–1065.
- Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983) A plant DNA miniprep: version 2. *Plant Mol. Biol. Rep.* **1**, 19–22.
- Feldmann, K.A. (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J.* **1**, 71–82.
- Frey, M., Stettner, C. and Gierl, A. (1998) A general method for gene isolation in tagging approaches: amplification of insertion mutagenized sites (AIMS). *Plant J.* **13**, 717–721.
- Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282.
- Hiei, Y., Komari, T. and Kubo, T. (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **35**, 205–218.
- Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T. and Kumashiro, T. (1996) High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. *Nature Biotech.* **14**, 745–750.
- Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions:  $\beta$ -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.
- Jeon, J.-S., Chung, Y.-Y., Lee, S., Yi, G.-H., Oh, B.-G. and An, G. (1999) Isolation and characterization of an anther-specific gene, *RA8*, from rice (*Oryza sativa* L.). *Plant Mol. Biol.* **39**, 35–44.
- Kang, H.-G., Jeon, J.-S., Lee, S. and An, G. (1998) Identification of class B and class C floral organ identity genes from rice plants. *Plant Mol. Biol.* **38**, 1021–1029.
- Kertbundit, S., De Greve, H., Deboeck, F., Van Montagu, M. and Hernalsteens, J.-P. (1991) *In vivo* random  $\beta$ -glucuronidase gene fusions in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **88**, 5212–5216.
- Kertbundit, S., Linacero, R., Rouze, P., Galis, I., Macas, J., Deboeck, F., Renckens, S., Hernalsteens, J.-P. and De Greve, H. (1998) Analysis of T-DNA-mediated translational  $\beta$ -glucuronidase gene fusions. *Plant Mol. Biol.* **36**, 205–217.
- Koes, R., Souer, E., van Houwelingen, A., et al. (1995) Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. *Proc. Natl Acad. Sci. USA*, **92**, 8149–8153.
- Kohli, A., Leech, M., Vain, P., Laurie, D.A. and Christou, P. (1998) Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hotspots. *Proc. Natl Acad. Sci. USA*, **95**, 7203–7208.
- Krizkova, L. and Hroudka, M. (1998) Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. *Plant J.* **16**, 673–680.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, **11**, 2283–2290.
- Lee, S., Jeon, J.-S., Jung, K.-H. and An, G. (1999) Binary vectors for efficient transformation of rice. *J. Plant Biol.* **42**, 310–316.
- Liu, Y.-G. and Whittier, R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics*, **25**, 674–681.
- Martin, G.B. (1998) Gene discovery for crop improvement. *Curr. Opin. Biotech.* **9**, 220–226.
- Ohba, T., Yoshioka, Y., Machida, C. and Machida, Y. (1995) DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target. *Plant J.* **7**, 157–164.
- Parinov, S., Sevugan, M., Ye, D., 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.
- Sasaki, T. (1998) The rice genome project in Japan. *Proc. Natl Acad. Sci. USA*, **95**, 2027–2028.
- Sato, Y., Sentoku, N., Miura, Y., Hirochika, H., Kitano, H. and Matsuoka, M. (1999) Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants. *EMBO J.* **18**, 992–1002.
- Smith, D.L. and Fedoroff, N.V. (1995) *LRP1*, a gene expressed in lateral and adventitious root primordia of *Arabidopsis*. *Plant Cell*, **7**, 735–745.
- 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.
- Tinland, B. (1996) The integration of T-DNA into plant genomes. *Trends Plant Sci.* **1**, 178–184.
- Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G. and Jones, J.D.G. (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell*, **11**, 1841–1852.
- Topping, J.F., Wei, W. and Lindsey, K. (1991) Functional tagging of regulatory elements in the plant genome. *Development*, **112**, 1009–1019.