

Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database

Choong-Hwan Ryu^{1,+}, Jung-Hwa You^{1,+}, Hong-Gyu Kang¹, Junghe Hur¹, Young-Hea Kim¹, Min-Jung Han¹, Kyungsook An¹, Byoung-Chull Chung², Choon-Hwan Lee² and Gynheung An^{1,*}

¹National Research Laboratory of Plant Functional Genomics, Department of Life Science, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Republic of Korea (*author for correspondence; e-mail genean@postech.ac.kr); ²Department of Molecular Biology, Pusan National University, Busan 609-735, Republic of Korea; +these authors contributed equally to this work

Received 3 November 2003; accepted in revised form 20 January 2004

Key words: flanking sequence database, gene trap, GFP, GUS, insertional mutagenesis, T-DNA

Abstract

We have developed a binary T-DNA vector, pGA2717, that contains the promoter-less β -glucuronidase (*gus*) gene adjacent to the right border and the promoter-less green fluorescence protein (*gfp*) gene next to the left border of the T-DNA. Therefore, inserting T-DNA into a gene can result in the activation of either *gus* or *gfp*. A total of 12 169 T-DNA insertional lines of japonica rice were generated using this binary vector. Out of 3140 lines examined, 0.5% of their mature seeds and 2.0% of the 3-day-old etiolated seedlings were GFP-positive. However, GUS assays of the same materials resulted in the identification of 151 (4.8%) GUS-positive lines. Using DNA gel blot and reverse transcription (RT)-PCR analyses, we confirmed that the GFP-positive lines were a true indication of gene trapping. A fusion transcript was also obtained between *gfp* and the trapped gene. We isolated 990 genomic sequences flanking T-DNA from our analysis of 2099 transgenic plants. Among the insertions, 625 T-DNAs were integrated into genic regions; 361 were located in intergenic regions. These tagging lines will be valuable in trapping and studying various genes for their expression patterns, as well as providing a useful tool for genetic approaches.

Abbreviations: GFP, green fluorescence protein; GUS, β -glucuronidase; iPCR, inverse PCR; LB, left border; RB, right border; RT-PCR, reverse transcription-PCR; T-DNA, transferred DNA

Introduction

Compared with other techniques for gene identification, insertional mutagenesis has an advantage in that the inserted element acts as a tag. Another benefit is that elements can be modified to trap the tagged gene by using a reporter gene (McElroy and Brettell, 1994; Sundaresan *et al.*, 1995; Springer, 2000). This system is convenient for observing mutant phenotypes because reporter activation indicates the location, condition, and

time of expression for the disrupted gene. The *gus* gene has been most frequently used as a reporter gene in plants because of the accurate detection of its gene products and the tolerance of the N-terminal translational fusion in its enzyme activity (Jefferson *et al.*, 1987). Gene-trap systems contain an intron with multiple splicing acceptor and donor sites in each of three reading frames in front of the promoter-less *gus*-coding region, allowing GUS expression when the insertion occurs in either an exon or intron. Consequently, the

efficiency of reporter activation is higher than that from constructs lacking the introns. In *Arabidopsis*, activation of the *gus* reporter gene in an insertional mutant population can be as high as 30% (Sundaresan *et al.*, 1995). In rice, the reporter gene has been activated in at least 10% of the *Ds*-tagged lines and T-DNA-tagged lines (Chin *et al.*, 1999; Jeon *et al.*, 2000, 2002; Kolesnik *et al.*, 2004).

More recently, the *gfp* gene has become a popular reporter gene because of several desirable features (McElroy and Brettell, 1994; Heim *et al.*, 1995; Gerdes and Kaether, 1996; Pang *et al.*, 1996; Sallaud *et al.*, 2003). First, the GFP (green fluorescent protein) assay does not require exogenous cofactors or substrates (Cody *et al.*, 1993; Inouye and Tsuji, 1994; Prasher, 1995), as opposed to the GUS (Jefferson *et al.*, 1987) and LUC (Williams *et al.*, 1989) methods. Second, because the fluorescence of GFP is cell autonomous, it can be detected non-invasively (Chalfie *et al.*, 1994), thereby allowing induction assays under a variety of environmental stresses such as temperature, metals, or drought (Allen *et al.*, 1999; Li *et al.*, 2001; Zimmer, 2002). Therefore, the *gfp* reporter gene has been used in a variety of organisms, including prokaryotes (Errington, 1993; Baulcombe *et al.*, 1995), yeast (Hampton *et al.*, 1996), *Drosophila* (Yeh *et al.*, 1995), *Xenopus* (Bronchain *et al.*, 1999), mammals (Yang *et al.*, 1998), and higher plants (Sheen *et al.*, 1995; Chiu *et al.*, 1996; Stewart, 1996; Ahlandsberg *et al.*, 1999; Jang *et al.*, 1999; Molinier *et al.*, 2000; Yamamoto *et al.*, 2003).

Sequencing of the rice genome has led to the identification of more than 40 000 putative genes (Feng *et al.*, 2002; Goff *et al.*, 2002; Sasaki *et al.*, 2002; Yu *et al.*, 2002; Rice Chromosome 10 Sequencing Consortium, 2003). The most immediate goal has been to determine biological functions for these genes rather quickly. The most popular method for these functional analyses is to obtain knockout mutants within a given gene and analyze the consequences (Koornneef *et al.*, 1982; Sundaresan, 1996; Krysan *et al.*, 1999). However, to facilitate sharing the knockout mutants, it is critical to generate databases for the knockout sites. This can be achieved by random sequencing of the flanking sequences at the insertion sites. Such databases have been established in rice knockout mutants generated by *Ds*- (Ito *et al.*, 2002), *Tos17*- (Yamazaki *et al.*, 2001), and *gus*-trapped T-DNA lines (Szabados *et al.*, 2002; An

et al., 2003; Chen *et al.*, 2003; Sallaud *et al.*, 2003; Wu *et al.*, 2003).

The vector backbone sequence is co-transferred with T-DNA more than 30% efficiency in rice (De Buck *et al.*, 2000; Yin and Wang, 2000; Kim *et al.*, 2003a; Sallaud *et al.*, 2003). Three types of junctions have been found: (1) some nucleotide overlap, (2) no homology, and (3) having filler DNA between T-DNA end and plant genomic DNA (Gheysen *et al.*, 1991; Mayerhofer *et al.*, 1991; Kumar and Fladung, 2002; Kim *et al.*, 2003a; Vain *et al.*, 2003). Here, we report the generation of T-DNA tagged lines trapped by the bidirectional reporter genes (*gfp* and *gus*) and the establishment of a database for 990 insertion sites.

Materials and methods

Construction of vectors

Plasmid pGA2705 was constructed by inserting the 150 bp *SalI*-*ClaI* fragment of pGA2703, which contains the *OsTubA1* intron3 (I3) with the putative triple splice donor and acceptor sequences, into the *SalI*-*ClaI* site in front of the *sgfpS65T* (*gfp*) gene, encoding a modified form of GFP (Jang *et al.*, 1999). Plasmid pGA2713 was constructed by adding the *SacI*-*XhoI* linker (5'-CTCGA GAGCT-3') into the *SacI* site of pGA2705. Plasmid pGA2717 was constructed by introducing the 1.2 kb *XhoI*-*SalI* fragment, containing the chimeric *OsTubA1:gfp:nos* gene of pGA2713, into the *XhoI* site near LB of the T-DNA in pGA2707 (Jeon *et al.*, 2000; Kim *et al.*, 2003b).

Plant materials

Scutellum-driven embryogenic calluses from *Oryza sativa* var. *japonica* (cv. Dongjin and Hwayoung) were transformed with our binary vector via the *Agrobacterium*-mediated cocultivation method (Lee *et al.*, 1999; Jeong *et al.*, 2002). We used completely dried seeds and 3-day-old etiolated seedling for GFP and GUS assays. Seeds of the transgenic plants were husked and surface sterilized with 70% ethanol for 30 s, followed by treatment with 0.025% Spotack ES (Aventis CropScience, Yongin, Korea) for 12 h. After being washed three times with sterile water, they were germinated for 3 days in tap water at 27 °C in the

dark for the GFP assay. For isolation of sequences flanking T-DNA, seeds were sown in soil and grown for three weeks under greenhouse conditions with a 14+ h photoperiod and a minimum night temperature of 20 °C. For RT-PCR, seedlings were grown in MSO medium containing 0.44% Murashige and Skoog basal salt, 3% sucrose, 0.2% phytigel, and 0.55 mM myo-inositol (Sigma, St. Louis, MO) for 1–2 weeks at 27 °C.

GFP assay

The GFP assay was performed according to the method described by Chung *et al.* (2000), with the following modifications. Excitation light (at 488 nm) was obtained from a 50 mW argon ion laser (NT54-166; Edmund Industrial Optics, NJ). A high-resolution CCD camera (Quantix Air-Cooled and Liquid-Cooled Camera Systems, Roper Scientific, USA) was used for producing images of the samples from a vertical position. The Metric Sized Narrow Band-pass Interference Filter (FILTER INT, 515 nm, 11.80 mm diameter, NT43-120; Edmund Industrial Optics) was positioned in front of the lens to pass a green fluorescence beam. Parameters for the image analysis software (RS image) were as follows: exposure time, 1 s; speed/gain value, 1; frame transfer of CCD parameter clocking, MPP; binning, 1.

Detection of gfp and vector backbone sequences

The *gfp* gene and vector backbone sequences were PCR amplified with the following primers. For detection of the *gfp* gene, the primers were 5'-CCTGAAGTTCATCTGCACCA-3' and 5'-GA-CTCCAGCAGGACCATGT-3'. For amplification of vector backbone sequences near the left border, the primers were 5'-CGATCTTGAGAACTATGCCGA-3' and 5'-GGTGGCCTACTT-CACCTAT-3'. For amplification of vector backbone sequences near the right border, the primers were 5'-TCAGTGAGGGCCAAGTTTTC-3' and 5'-CGATCGCAAATCCGACGCTGT-3'. For identification of the read-through of the T-DNA left border, the primers were 5'-GGTGGCCTACTTACCTAT-3' and 5'-GTTGGTTAGAGAACAGCACAA-3'. For detection of the read-through of the right border, the primers were 5'-TCAGTGAGGGCCAAGTTTTC-3' and 5'-TTGGGGTTTCTACAGGACGTAAC-3'. The

20 μ l PCR-reaction solution contained 1 \times ExTaq buffer, 200 μ M each of dNTPs, 20–100 ng template DNA, 1 unit ExTaq (TaKaRa), and 0.5 μ M of each primer. We used 36 cycles, each consisting of denaturation at 94 °C for 30 s, primer annealing at 55 °C for 30 s, and extension at 72 °C for 45 s.

RT-PCR analysis

Tri-reagent (MRC, Cincinnati, OH) was used to isolate total RNA from immature leaves and roots. For first-strand cDNA synthesis, 2 μ g of total RNA was reverse transcribed in a total volume of 25 μ l that contained 10 ng of oligo(dT)_{12–18} primer, 2.5 mM dNTPs, and 200 units of Moloney murine leukemia virus reverse transcriptase (New England Biolabs, Beverly, MA) in a reaction buffer. PCR was performed in a 50 μ l solution containing a 1 μ l aliquot of the cDNA reaction, 0.2 μ M gene-specific primers, 10 mM dNTPs, 1 unit of ExTaq DNA polymerase (Takara, Otsu, Japan), and 10 \times reaction buffer. The reaction included an initial 5 min denaturation at 94 °C, followed by 24–35 cycles of PCR (94 °C 1 min, 62 °C 50 s, and 72 °C 1 min), and a final 10 min at 72 °C. Primers for RT-PCR were 5'-TTCG-TCACTGCTGATTTTGC-3' (F1, 335 bp upstream T-DNA inserted site in the *OsZFP33* gene) and 5'-ACGCTGAACTTGTGGCCGTT-3' (G1, in the *GFP* gene). PCR products were separated on a 1.2% w/v agarose gel and transferred onto a Hybond N⁺ membrane (Amersham, Buckinghamshire, UK). The blot was hybridized to a ³²P-labeled probe at 60 °C for 12 h in Church buffer (Church and Gilbert, 1984). Membranes were washed once for 10 min with 0.2 \times SSC and 0.1% w/v SDS at 25 °C, and twice, for 10 min each, with 0.2 \times SSC and 0.1% w/v SDS at 58 °C. These washed membranes were then exposed to room temperature for 30 min on X-ray film with intensifying screens.

Isolation of the sequence-flanking T-DNA

Genomic DNAs were extracted from the leaves of seedlings 2–3 weeks old according to the procedure of Chen and Ronald (1999), except that the samples were ground in an MM300 Mixer Mill (Retsch, Haan, Germany). The sequences flanking T-DNA were isolated by the inverse PCR (iPCR) method, as described previously (Hui *et al.*, 1998;

Spertini *et al.*, 1999; Jung *et al.*, 2003). Genomic DNA (1 μ g) was digested with 10 units of *Pst*I or *Xho*I in 50 μ l for 10 h. The PCR reaction was performed in 25 μ l of a mixture containing 20 ng of plant DNA, 0.1 units ExTaq polymerase, 1 \times ExTaq buffer, 0.2 mM dNTPs, and 0.4 μ M of the primers. PCR was carried out with an initial 5 min denaturation at 94 °C, followed by 35 cycles (each cycle: 94 °C, 1 min; 56 °C, 1 min; and 72 °C, 2 min), then a final 10 min at 72 °C. To amplify the flanking region, we performed nested PCR with the primer sequences shown in Table 1. Sequences flanking RB of T-DNA were isolated with the primers R1 and R2 for the *Pst*I-cut genomic DNA or R1 and R3 for the *Xho*I-cut genomic DNA. The nested primers R4 and R5 or R4 and R6 were used for the *Pst*I- or *Xho*I-cut genomic DNA, respectively. Sequences flanking LB of T-DNA were obtained with the primers L1 and L2 for the *Pst*I-cut DNA or L1 and L3 for the *Xho*I-cut DNA. The nested primers were L4 and L5 for *Pst*I and L4 and L6 for *Xho*I. Nested PCR was performed with 0.1 μ l aliquot of the first PCR product as a template.

Sequence data analysis

Sequences flanking the T-DNA were analyzed with the NCBI database (<http://www.ncbi.nlm.nih.gov/>), the TIGR database (<http://tigrblast.tigr.org/euk-blast>), and the RiceGD database (<http://btn.genomics.org.cn/rice/>), as previously described (An *et al.*, 2003). The nucleotide sequence of the amplified DNA fragments was also assessed by RiceGAAS (<http://ricegaas.dna.affrc.go.jp>). If a particular sequence had not yet been

annotated in the public databases, we undertook annotation with the Softberry program (<http://www.softberry.com/berry.phtml>). Functional classification of the T-DNA inserted genes was performed with gene ontology programs (<http://www.geneontology.org>; <http://rgp.dna.affrc.go.jp>; <http://www.edi.ac.uk/interpro/>).

Results

Vector construction and production of T-DNA-tagged transgenic rice plants

We previously reported the construction of a binary plasmid, pGA2707, that contains the promoter-less *gus* reporter gene with an intron and multiple splicing donors and acceptors immediately next to RB (Jeong *et al.*, 2002). In the current study, we constructed another binary vector, pGA2717 (Figure 1), by adding the promoter-less *gfp* reporter gene with an intron and multiple splicing donors and acceptors adjacent to LB. The hygromycin-resistant selectable marker gene, under the rice α -tubulin (*OsTubA1*) promoter, was located between the *gus* and the *gfp* reporter genes. Thus, we demonstrated that T-DNA insertions within a gene could result in activation of either *gus* or *gfp*. We have now produced 12 169 fertile transgenic plants with pGA2717 via the *Agrobacterium*-mediated transformation procedure (Lee *et al.*, 1999).

Assay of gfp expression

We previously presented the development of a detection system that can rapidly screen transgenic

Table 1. Primers used for iPCR.

Primer	Enzyme	Direction	PCR	Sequence
R1	<i>Pst</i> I, <i>Xho</i> I	Right	First, first	TTGGGGTTTCTACAGGACGTAAC
R2	<i>Pst</i> I	Right	First	CCATGTAGTGTATTGACCGATTG
R3	<i>Xho</i> I	Right	First	AGGTTTGATATAAAGTGGCACG
R4	<i>Pst</i> I, <i>Xho</i> I	Right	Second, second	CAAGTTAGTCATGTAATTAGCCAC
R5	<i>Pst</i> I	Right	Second	TCGTCTGGCTAAGATCGGCCGCA
R6	<i>Xho</i> I	Right	Second	AGATCGAGCGGGCTCACA
L1	<i>Pst</i> I, <i>Xho</i> I	Left	First, first	ACGCTGAACTTGTGGCCGTT
L2	<i>Pst</i> I	Left	First	CGACAACCACTACCTGAGCA
L3	<i>Xho</i> I	Left	First	CCGATGATAAGCTGTCAAACATGAG
L4	<i>Pst</i> I, <i>Xho</i> I	Left	Second, second	GTTGGTTAGAGAACAGCACAA
L5	<i>Pst</i> I	Left	Second	GATCACATGGTCCTGCTGGAGTT
L6	<i>Xho</i> I	Left	Second	GGGATCCACTAGTTCTAGA

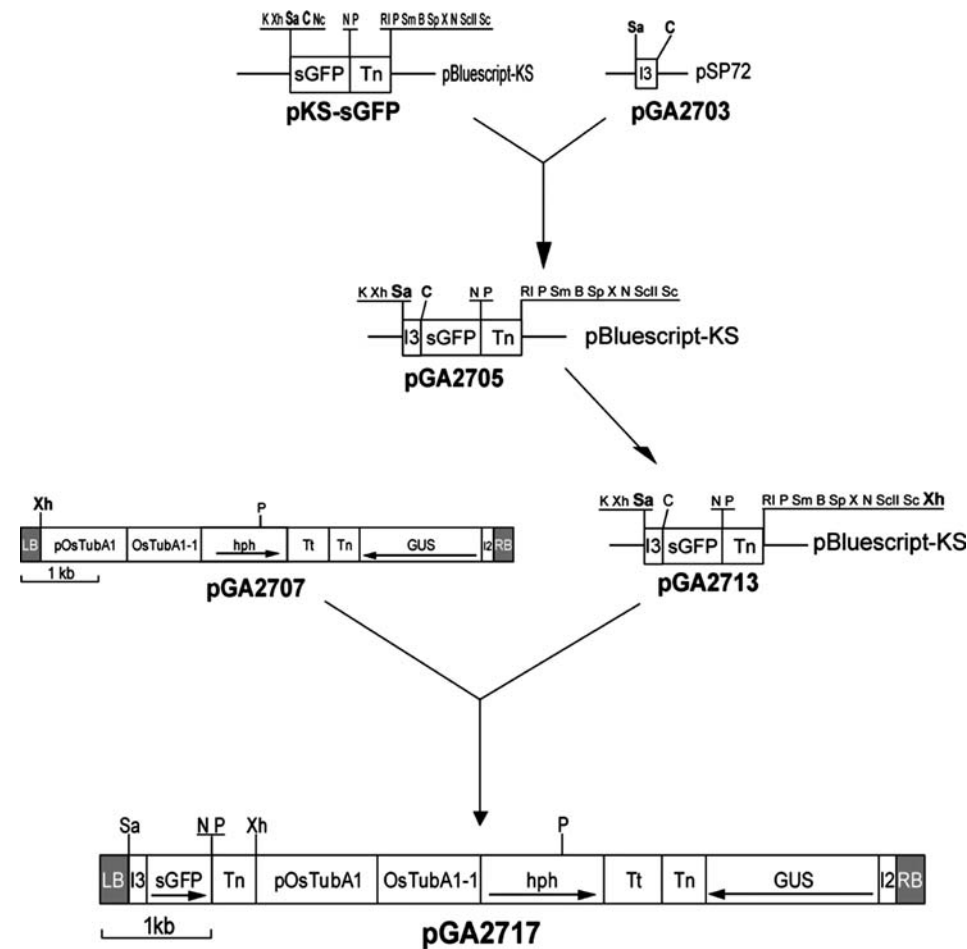


Figure 1. Schematic diagram of the trapping vector pGA2717. RB and LB in the gray bars represent the right and left borders of T-DNA, respectively. I2 and I3 are the intron 2 and intron3 of *OsTubA1*, respectively, carrying three putative splicing acceptor and donor sites. *hph*, hygromycin phosphotransferase gene; pOsTubA1, the promoter of *OsTubA1*; OsTubA1-1, the first intron of *OsTubA1*; Tn, *nos* terminator; Tt, *OsTubA1* terminator; B, *Bam*HI; C, *Cl*aI; K, *Kpn*I; N, *Not*I; Nc, *Nco*I; P, *Pst*I; RI, *Eco*RI; Sa, *Sal*I; Sc, *Sac*I; ScII, *Sac*II; Sm, *Sma*I; Sp, *Spe*I; X, *Xba*I; Xh, *Xho*I site.

plants expressing the *gfp* gene (Chung *et al.*, 2000). This system consists of an argon laser light source (radiating at a 488 nm peak) and a 515 nm interference filter passing GFP fluorescence emitted from transgenic plants. The emitted fluorescence is captured by a charge-coupled device (CCD) camera connected to a personal computer for recording the photo images. The program used in this system then converts the photo-image intensity to quantitative data.

With this device, we screened mature seeds and 3-day-old seedlings that expressed the *gfp* gene. Of the 3140 lines examined, GFP activity was detected in 73 (2.3%). Among them, the GFP signal was preferentially detected in the dry seeds of 11 lines

and in the seedlings of 57 lines. For the remaining five lines, activity was measured in both tissue types. In general, however, GFP was preferentially detected in a certain organ (Figure 2A–C) or ubiquitously in seedlings (Figure 2D). Frequencies of preferential detection were 0.25% (8 lines), 0.19% (6 lines), 0.80% (25 lines), 0.25% (8 lines), and 0.45% (14 lines) in the embryo, endosperm/coat, scutellum, shoots, and roots, respectively. In 12 lines, GFP was found in more than one organ. We also conducted a GUS assay to compare its tagging efficiency with that of GFP. Out of those same 3140 lines, 151 (4.8%) were identified as GUS-positive. This result indicates that tagging efficiency is lower with *gfp* than with *gus*.

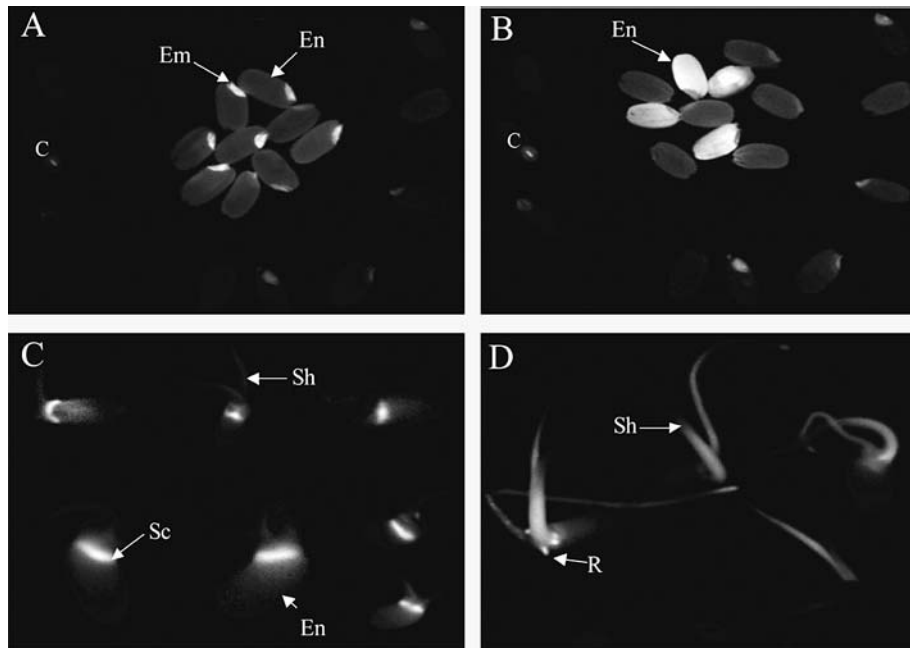


Figure 2. GFP assay of dry seeds or etiolated seedlings. (A) 1B-104-23 seeds; (B) 1B-117-21 seeds; (C) 1B-082-18 seedlings; (D) 1B-086-03 seedlings. Symbols: C, control; Em, embryo; En, endosperm; R, roots; Sc, scutellum; Sh, shoot.

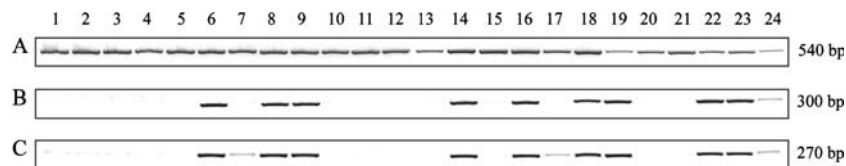


Figure 3. PCR analyses of transgenic plants. Presence of the *gfp* gene (A) and the vector backbone sequences near RB (B) and LB (C) was examined by PCR. Numbers indicate the independent lines transformed by pGA2717.

Detection of the *gfp* gene by PCR and DNA gel-blot analyses

We selected 73 lines that were displayed as GFP-positive, and analyzed them for the presence of the intact *gfp* gene. DNA from T₂ seedlings was extracted and the *gfp* fragment was amplified by PCR with the primers located within the *gfp* gene. Amplification was complete in all the lines, indicating that the plants were truly transgenic and that the GFP emission was due to the expression of that reporter gene (Figure 3A). Previous studies have revealed that the vector backbone located outside of the T-DNA is occasionally co-transferred (Cluster *et al.*, 1996; Kononov *et al.*, 1997; Wenck *et al.*, 1997; Wolters *et al.*, 1998; De Buck *et al.*, 2000; Yin and Wang, 2000; Kim *et al.*, 2003a; Sallaud *et al.*, 2003; Vain *et al.*,

2003). Therefore, we estimated the frequency of co-transfer by determining the presence of the vector backbone sequence. Among 73 samples examined, 29 lines carried the entire vector backbone and two lines carried truncated vectors, lacking sequences near the T-DNA RB (Figure 3B and C).

To further confirm that the GFP positives were due to expression of the *gfp* gene, we performed DNA gel-blot analysis. Genomic DNA prepared from the T₂ seedlings of GFP-positive line 1A-071-28 was digested with *Pst*I and hybridized with the *gfp* probe (Figure 4A). Two bands arose, thereby indicating the presence of two T-DNA copies in the line. In fact, of the ten plants examined, only one (plant 3) carried no T-DNA; all the remaining plants contained both T-DNAs. Furthermore, GFP analysis of these plants showed that all but

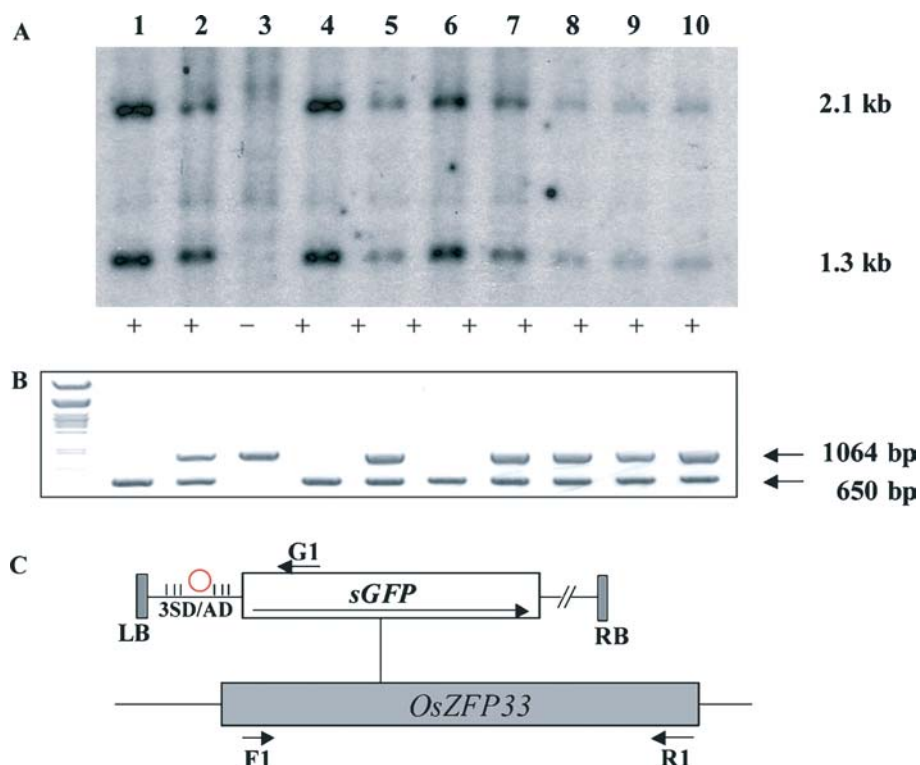


Figure 4. Analyses of transgenic lines. **A.** Correlation between DNA gel-blot analysis (upper part) and GFP assay data (lower part). Genomic DNAs were prepared from mature green leaves and the DNA gel blot was hybridized with the *sGFP* probe. Results of GFP assay are indicated by + or -. Numbers indicate independent T₂ plants of line 1A-071-28. **B.** Genotyping of 10 T₂ plants by PCR. The upper band (1064 bp) is the PCR product of two primers (F1 and R1) located within the tagged gene; the lower band (650 bp) is the product of the F1 primer in the tagged gene and the G1 primer within the *gfp* gene. **C.** Schematic diagram of the T-DNA insertion into the *OsZFP33* gene. The F1, R1, and G1 primers used for genotyping are shown. Symbols: LB, left border of T-DNA; RB, right border of T-DNA; 3SD/AD, third intron of *TubA1* with triple splice donors and acceptors.

plant 3 were positive to the assay, confirming that the GFP analysis is reliable.

Isolation of *gfp*-trapped genes

We employed inverse PCR to isolate the DNA sequences flanking the *gfp* gene. Among the 25 lines that showed preferential expression pattern of GFP, 19 tags were obtained. These lines are marked with symbol # in the supplementary data 1. Analysis of these sequences indicated that the tagged gene in line 1A-071-28 was *OsZFP33*, which encodes a putative zinc-finger protein. Moreover, the tagged gene in Line 1A-068-13 was homologous to a gene encoding putative 3-ketoacyl-CoA thiolase protein; in line 1B-086-33, the *gfp* gene was inserted into a gene containing the Maf domain.

The T₂ seedlings of line 1A-071-28 were genotyped (Figure 4B). Two primers (F1, 5'-TTCGT-

CACTGCTGATTTTGC-3' and R1, 5'-ACAGG-ATTCTGCTTGGCAAT-3') located on the genome and a primer (G1, 5'-ACGCTGAACCTTGTGGCCGTT-3') on the T-DNA were used to PCR-amplify the genomic DNA of the T₂ progeny (Figure 4C). A 1064 bp fragment was amplified by the F1 and R1 primers when no insert was made, whereas a 650 bp fragment was produced by F1 and G1 when a T-DNA insertion was available (Figure 4B). Therefore, when both fragments were present, the plant was heterozygous (Figure 4B, plants 2, 5, 7, 8, 9, and 10). In contrast, homozygotic plants showed a single band of 650 bp (plants 1, 4, and 6), while the wild-type segregants had a single 1064 bp band (plant 3). GFP analysis of these plants revealed that all the progeny, except plant 3, were positive for the assay, demonstrating that the GFP detection system could monitor the stable inheritability of the reporter gene.

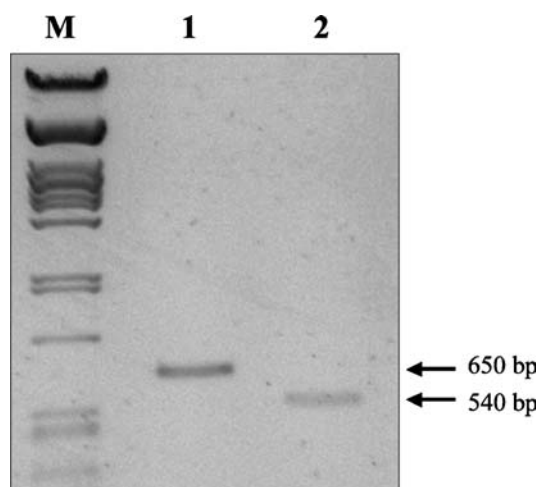


Figure 5. RT-PCR analysis of the *OsZFP33-sGFP* fusion transcript. PCR was performed with the gene-specific forward primer (F1) in the 5'-UTR of *OsZFP33* and the *sGFP* reverse primer (G1), with genomic (sample 1) or cDNA of the *OsZFP33* heterozygous plants (sample 2) as templates. M, lambda DNA size marker cut with *Pst*I.

Detection of the tagged gene-gfp fusion transcript

We performed RT-PCR analysis to isolate a fusion transcript between *OsZFP33* and *gfp*. First-strand cDNA, prepared from the GFP-positive leaves, was PCR-amplified with the G1 primer located in the *GFP* gene and the F1 primer in the tagged gene. This resulted in the amplification of a 540 bp fragment (Figure 5, sample 2), which was shorter than the PCR band from genomic DNA (Figure 5, sample 1). Sequencing this fragment revealed that the fusion transcript was formed by splicing between the third donor and first acceptor (red circle in Figure 4C) of the *OsTubA1* intron3 located between *gfp* and *OsZFP33*. This result indicates that the fusion transcript is under the control of the *OsZFP33* promoter.

Isolation of sequences flanking T-DNA

The sequences flanking T-DNA were isolated via iPCR. Analysis of 2099 lines tagged by pGA2717 resulted in the identification of 990 genomic sequences (Table 2). To learn the distribution of these T-DNA insertions, we analyzed their locations with respect to the putative genes. Among the 990 insertions, 625 (63.1%) were integrated into the genic regions, and 361 (36.5%), into the intergenic regions (Table 2). The remaining four

Table 2. Distribution of T-DNA insertions in genic and intergenic regions.

	Number	Percentage
Genic	625	63.1
Exon	217	22.0
Intron	255	25.7
5'-UTR (within 300 bp)	96	9.7
3'-UTR (within 300 bp)	57	5.76
Intergenic	361	36.5
NA ^a	4	0.4
Total	990	100

^aT-DNA was inserted into a region that cannot be annotated.

loci were not determined because the surrounding sequence of the insertion site was located at the end of the contigs and, therefore, the regions could not be annotated. T-DNAs were inserted into introns and exons at almost equal frequencies. A detailed description of the information is shown in Supplemental Data 1.

The T-DNA-tagged genes were classified into ten functional groups (Table 3) with the software packages INTERPRO (<http://www.edi.ac.uk/interpro>) and Gene Ontology (<http://www.geneontology.org>). When it was difficult to assign functions to some predicted coding proteins, such as for a hypothetical or unknown protein, they were classified as 'other'. The order of frequencies for our tagged genes was similar to that previously reported (Szabados *et al.*, 2002; An *et al.*, 2003).

Investigation of T-DNA insertion events

We analyzed the junction between T-DNA and rice genomic DNA with 329 fragments flanking

Table 3. Functional classification of the tagged genes.

	Number (%)
Metabolism	64 (10.2)
Signaling	46 (7.4)
Transcription	39 (6.2)
Defense	20 (3.2)
Protein fate	17 (2.7)
Cell structure	17 (2.7)
Transporter	9 (1.4)
Retroelement	18 (2.9)
Unclassified	9 (1.4)
Other	386 (62.0)
Total	625

Table 4. Deletion of T-DNA at the junctions.

Nucleotide length	RB (%)	LB (%)
0	136 (41.3)	3 (0.5)
1–5	104 (31.6)	35 (6.2)
6–10	26 (7.9)	63 (11.2)
11–20	39 (11.9)	91 (16.1)
21–30	11 (3.3)	64 (11.3)
31–40	3 (0.9)	54 (9.6)
41–50	1 (0.3)	61 (10.8)
51–60	0 (0.0)	54 (9.6)
61–70	2 (0.6)	28 (5.0)
71–80	3 (0.9)	21 (3.7)
81–90	0 (0.0)	32 (5.7)
91–100	1 (0.3)	11 (1.9)
> 100	3 (0.9)	48 (8.5)
Total	329	565

RB and 565 fragments flanking LB. In 136 (41.3%) events at T-DNA RB, break points were observed between the third and fourth nucleotide of the border sequence (Table 4). The site between nucleotides 3 and 4 of the bottom strand of the T-DNA borders have been known as a cleavage site for generating single-stranded T-DNA fragments (Yanofsky *et al.*, 1986; Stachel *et al.*, 1987; Wang *et al.*, 1987). This fact has been supported by data from *Arabidopsis* (Tinland, 1996), tobacco (Tinland, 1996), aspen (Kumar and Fladung, 2002), barley (Stahl *et al.*, 2002), and rice (Kim *et al.*, 2003a). In the remaining events, small deletions ranging from 1 to 30 nucleotides were predominant (Figure 6A and Table 4). Unlike RB, the cleavage site at LB was remained only in three events (Table 4). The size of deletions was generally longer than at the BR (Figure 6B and Table 4).

We also analyzed the sequence organization at the junctions. It was previously reported that there are three types of junctions (Gheysen *et al.*, 1991; Mayerhofer *et al.*, 1991; Kumar and Fladung, 2002; Kim *et al.*, 2003a). In type I, one to several nucleotide overlaps are found between T-DNA and genomic DNA. In our analysis, 136 out of 329 RB and 328 out of 565 LB belong to this type (Table 5 and Figure 6). In type II, a filler DNA is present at the junctions. We found that ca. 35% of the junctions belong to this category (Table 5). Among 315 filler DNA observed at the junctions, a majority (80%) is less than 30 nucleotides long and 161 (51%) fillers are less than 10 nucleotides long (data not shown). They are from T-DNA, vector

backbone, or plant genomic DNA. In type III junctions, the link between T-DNA and genomic DNA is direct, without any overlaps or filler DNA. This type is less frequent at both RB and LB (Table 5).

Discussion

We have developed a binary T-DNA vector, pGA2717, that contains the promoter-less *gfp* adjacent to LB and the promoter-less *gus* gene next to RB. Using this dual trapping vector, we have produced 12 169 fertile rice lines. Considering that transgenic plants contain an average of 1.4 of T-DNA inserts (Joen *et al.*, 2000), we estimate that about 17 000 T-DNA tags have been generated. Together with the gene-tagging lines previously created with other vectors, including the activation-tagging vector (Jeon *et al.*, 2000, 2002), we have now produced a total of 100 000 T-DNA tagging lines. If T-DNA inserts randomly into rice chromosomes, then ca. 67% of those genes have been tagged by the dual trapping vector (Jeon and An, 2001).

We have also constructed a database for the sequences flanking T-DNA. As observed previously in our investigations, T-DNA appears to prefer genic regions as previously reported (An *et al.*, 2003; Chen *et al.*, 2003), so gene-tagging efficiency is higher than the estimated value. Our database can be searched at website postech.ac.kr/life/pfg. The T₂ seeds will be available upon request.

Fluorescence microscopy and confocal systems have been used for monitoring subcellular localization of the GFP reporter protein (Köhler, 1998; Hanson and Köhler, 2001; Stewart, 2001; Goodin *et al.*, 2002). However, they are not suitable for rapid assessment of a large number of samples. Therefore, to efficiently screen the GFP-positive plants from the pGA2717 tagging lines, we employed a high-throughput device for detecting GFP-positive samples in a non-destructive manner. Molecular analysis of the GFP-positive lines, as identified by screening, showed that they carried the *gfp* gene and expressed the fusion transcript between the tagged gene and the *gfp* reporter, thereby indicating that this method is reliable.

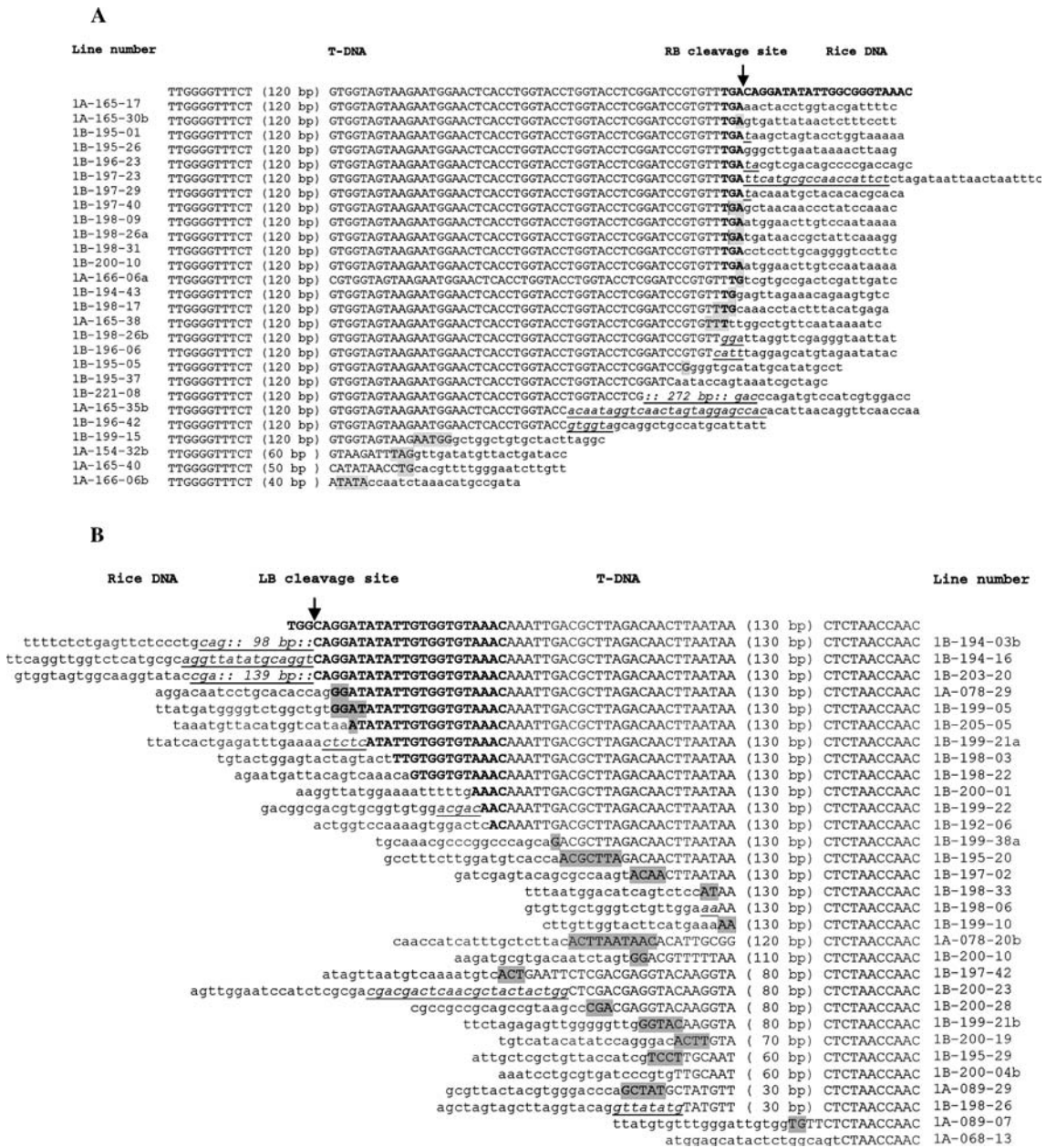


Figure 6. Analysis of junction sequences between the T-DNA right end and rice genomic DNA (A), and between the T-DNA left end and rice DNA (B). T-DNA sequences are marked in capitals with RB and LB sequences in bold. The 20 bp plant DNA sequences are given in lower cases. Shaded characters are homologous sequences between the T-DNA end and the rice genomic DNA. Filler DNAs are represented in underlined italics. Sequence of T-DNA borders and nearby regions in the binary vector is shown on top for comparison. The numbers in blanks indicate the distance between the nucleotides.

Gene-trapping efficiency by the *gfp* reporter was 2.3% for seeds and 3-day-old seedlings. In contrast, efficiency with the *gus* reporter was 4.8% at the same stages. Therefore, the total gene-trapping efficiency when using the dual trap-

ping vector was 7.1%. If we had examined other organs as well, trapping efficiency would have been greatly increased. Because the GFP assay can be conducted *in vivo* and does not require any substrate, it will be a useful method for

Table 5. Three types of junction between T-DNA and genomic DNA.

	RB (%)	LB (%)
Type I	136 (41.3)	328 (58.1)
Type II	116 (35.3)	199 (35.2)
Type III	77 (23.4)	38 (6.7)
Total	329	565

identifying genes that are induced by environmental stimuli.

The efficiency of *gfp* trapping was lower than for *gus*, probably because of the location of the *gfp* reporter within the T-DNA. The left end of T-DNA frequently is deleted during integration into chromosomes and, therefore, a portion of the intron or *gfp* gene may have been lost in the tagging lines (Kim *et al.*, 2003a). In contrast, the right end of T-DNA is more conserved. Tinland (1996) has proposed that this left end, being naked, is not well protected and that the right end is more conserved because it is attached to the VirD2 protein. In addition, it is possible that the GFP assay system used in these experiments was not sensitive enough to detect low levels of expression. The *gfp*-trapping efficiency in the endosperm was quite low compared with that obtained from the GUS assay, suggesting that light intensity was reduced during penetration into the tissue. If this intensity had been increased, we might have detected the *gfp*-trapped lines at a higher frequency.

We studied T-DNA integration patterns in rice chromosomes. The cleavage site was conserved at the majority of RB, but less conserved at LB. This observation is similar to previous reports (Tinland, 1996; Kumar and Fladung, 2002; Stahl *et al.*, 2002; Kim *et al.*, 2003a). The T-DNA deletions were more significant in LB as found previously (Hiei *et al.*, 1994; Jeon *et al.*, 2000; Kim *et al.*, 2003a). All three types of junctions were observed at both borders and the frequency was similar to each other. Windels *et al.* (2003) and Kim *et al.* (2003a) reported that in *Arabidopsis* 27 of 67 T-DNA junctions, and in rice 35 of 114, had filler insertions, respectively. We observed the similar frequency (35%) with much large number of samples. The fillers were mostly from plant genomic sequences, T-DNA or the vector backbone sequences. However, some of them were from unknown origins.

Rice, a major agronomic crop, is recognized as a model monocot species. Recent completion of the draft sequence for its genome has resulted in an explosion of information on rice genes (Goff *et al.*, 2002; Yu *et al.*, 2002). Our gene-trap lines are an important tool for identifying their biological functions. We are now establishing pools of our pGA2717 vector-tagging lines, which can be used for screening mutations in a given gene. Once such a mutant is found, the gene trapped by *gfp* or *gus* can be examined as a means for elucidating expression patterns and functions.

Acknowledgements

We thank J. Kim for providing the *gfp* gene, Sung-Ryul Kim for valuable discussion, and Priscilla Licht for critical reading of the manuscript. This work was supported, in part, by a grant from POSCO and from the Crop Functional Genomic Center, the 21st Century Frontier Program (CG-1111 to GA and CG-1112 to C-H L).

References

- Ahlandsberg, S., Sathish, P., Sun, C. and Jansson, C. 1999. Green fluorescent protein as a reporter system in the transformation of barley cultivars. *Physiol. Plant.* 107: 194–200.
- Allen, G.J., Kwak, J.M., Chu, S.P., Llopis, J., Tsien, R.Y., Harper, J.F. and Schroeder, J.I. 1999. Cameleon calcium indicator reports cytoplasmic calcium dynamics in *Arabidopsis* guard cells. *Plant J.* 19: 735–747.
- An, S., Park, S., Jeong, D.-H., Lee, D.-Y., Kang, H.-G., Yu, J.-H., Hur, J., Kim, S.-R., Kim, Y.-H., Lee, M., Han, S., Kim, S.-J., Yang, J., Kim, E., Wi, S.J., Chung, H.S., Hong, J.-P., Choe, V., Lee, H.-K., Choi, J.-H., Nam, J.-M., Kim, S.-R., Park, P.-B., Park, K.Y., Kim, W.T., Choe, S., Lee, C.-B. and An, G. 2003. Generation and analysis of end-sequence database for T-DNA tagging lines in rice. *Plant Physiol.* 133: 2040–2047.
- 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.
- Bronchain, O.J., Hartley, K.O. and Amaya, E. 1999. A gene trap approach in *Xenopus*. *Curr. Biol.* 9: 1195–1198.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263: 802–805.
- Chen, D.H. and Ronald, P.C. 1999. A rapid DNA miniprep-paration method suitable for AFLP and other PCR applications. *Plant Mol. Biol. Rep.* 17: 53–57.
- Chen, S., Jin, W., Wang, M., Zhang, F., Zhou, J., Jia, Q., Wu, Y., Liu, F. and Wu, P. 2003. Distribution and

- characterization of over 1000 T-DNA tags in rice genome. *Plant J.* 36: 105–113.
- Chin, H.G., Choe, M.S., Lee, S.H., Koo, J.C., Kim, N.Y., Lee, J.J., Oh, B.G., Yi, G.H., Kim, S.C., Choi, H.C., Cho, M.J. and Han, C.D. 1999. Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. *Plant J.* 19: 615–623.
- Chiu, W.-L., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J. 1996. Engineered GFP as a vital reporter in plants. *Curr. Biol.* 6: 325–330.
- Chung, B.-C., Kim, J.-K., Nahm, B.-H. and Lee, C.-H. 2000. *In planta* visual monitoring of green fluorescent protein in transgenic rice plants. *Mol. Cells* 10: 411–414.
- Church, G.M. and Gilbert, W. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81: 1191–1195.
- Cluster, P.D., O'Dell, M., Metzloff, M. and Flavell, R.B. 1996. Details of T-DNA structural organization from a transgenic *Petunia* population exhibiting co-suppression. *Plant Mol. Biol.* 32: 1197–1203.
- Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G. and Ward, W.W. 1993. Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry* 32: 1212–1218.
- De Buck, S., De Wilde, C., Van Montagu, M. and Depicker, A. 2000. T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation. *Mol. Breed.* 6: 459–468.
- Errington, J. 1993. *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. *Microbiol. Rev.* 57: 1–33.
- Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J., Liu, Y., Hu, X., Jia, P., Zhang, Y., Zhao, Q., Ying, K., Yu, S., Tang, Y., Weng, Q., Zhang, L., Lu, Y., Mu, J., Lu, Y., Zhang, L.S., Yu, Z., Fan, D., Liu, X., Lu, T., Li, C., Wu, Y., Sun, T., Lei, H., Li, T., Hu, H., Guan, J., Wu, M., Zhang, R., Zhou, B., Chen, Z., Chen, L., Jin, Z., Wang, R., Yin, H., Cai, Z., Ren, S., Lv, G., Gu, W., Zhu, G., Tu, Y., Jia, J., Zhang, Y., Chen, J., Kang, H., Chen, X., Shao, C., Sun, Y., Hu, Q., Zhang, X., Zhang, W., Wang, L., Ding, C., Sheng, H., Gu, J., Chen, S., Ni, L., Zhu, F., Chen, W., Lan, L., Lai, Y., Cheng, Z., Gu, M., Jiang, J., Li, J., Hong, G., Xue, Y. and Han, B. 2002. Sequence and analysis of rice chromosome 4. *Nature* 420: 316–320.
- Gerdes, H.-H. and Kaether, C. 1996. Green fluorescent protein: applications in cell biology. *FEBS Lett.* 389: 44–47.
- Gheysen, G., Villarroel, R. and Van Montagu, M. 1991. Illegitimate recombination in plants: a model for T-DNA integration. *Genes Dev.* 5: 287–297.
- Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.
- Goodin, M.M., Dietzgen, R.G., Schichnes, D., Ruzin, S. and Jackson, A.O. 2002. pGD vectors: versatile tools for the expression of green and red fluorescent protein fusions in agroinfiltrated plant leaves. *Plant J.* 31: 375–383.
- Hampton, R.Y., Koning, A., Wright, R. and Rine, J. 1996. *In vivo* examination of membrane protein localization and degradation with green fluorescent protein. *Proc. Natl. Acad. Sci. USA* 93: 828–833.
- Hanson, M.R. and Köhler, R.H. 2001. GFP imaging: methodology and application to investigate cellular compartmentation in plants. *J. Exp. Bot.* 52: 529–539.
- Heim, R., Cubitt, A.B. and Tsien, R.Y. 1995. Improved green fluorescence. *Nature* 373: 663–664.
- 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.
- Hui, E.K., Wang, P.C. and Lo, S.J. 1998. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. *Cell. Mol. Life Sci.* 54: 1403–1411.
- Inouye, S. and Tsuji, F.I. 1994. *Aequorea* green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett.* 341: 277–280.
- Ito, T., Motohashi, R., Kuromori, T., Mizukado, S., Sakurai, T., Kanahara, H., Seki, M. and Shinozaki, K. 2002. A new resource of locally transposed *Dissociation* elements for screening gene-knockout lines *in silico* on the *Arabidopsis* genome. *Plant Physiol.* 129: 1695–1699.
- Jang, I.-C., Nahm, B.H. and Kim, J.-K. 1999. Subcellular targeting of green fluorescent protein to plastids in transgenic rice plants provides a high-level expression system. *Mol. Breed.* 5: 453–461.
- 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.
- Jeon, J.-S. and An, G. 2001. Gene tagging in rice: a high throughput system for functional genomics. *Plant Sci.* 161: 211–219.
- Jeon, J.-S., Lee, S., Jung, K.-H., Jun, S.-H., Jeong, D.-H., Lee, J., Kim, C., Jang, S., Yang, K., Nam, J., An, K., Han, M.-J., Sung, R.-J., Choi, H.-S., Yu, J.-H., Choi, J.-H., Cho, S.-Y., Cha, S.-S., Kim, S.-I. and An, G. 2000. T-DNA insertional mutagenesis for functional genomics in rice. *Plant J.* 22: 561–570.
- Jeong, D.-H., An, S., Kang, H.-G., Moon, S., Han, J.-J., Park, S., Lee, H.S., An, K. and An, G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol.* 130: 1636–1644.
- Jung, K.-H., Hur, J.-H., Ryu, C.-H., Choi, Y.-J., Chung, Y.-Y., Miyao, A., Hirochika, H. and An, G. 2003. Characterization of a rice chlorophyll-deficient mutant using the T-DNA gene-trap system. *Plant Cell Physiol.* 44: 463–472.
- Kim, S.-R., Lee, J., Jun, S.-H., Park, S., Kang, H.-G., Kwon, S. and An, G. 2003a. Transgene structures in T-DNA-inserted rice plants. *Plant Mol. Biol.* 52: 761–773.
- Kim, S.-R., Lee, S., Kang, H.-G., Jeon J.-S., Kim K.-M. and An, G. 2003b. A complete sequence of the pGA1611 binary vector. *J. Plant Biol.* 46: 211–214.

- Köhler, R.H. 1998. GFP for *in vivo* imaging of subcellular structures in plant cells. *Trends Plant Sci.* 3: 317–320.
- Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V. and Ramachandran, S. 2004. Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J.* 37: 301–314.
- Kononov, M.E., Bassuner, B. and Gelvin, S.B. 1997. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. *Plant J.* 11: 945–957.
- Koornneef, M., Dellaert, L.W. and van der Veen, J.H. 1982. EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat. Res.* 93: 109–123.
- Krysan, P.J., Young, J.C. and Sussman, M.R. 1999. T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11: 2283–2290.
- Kumar, S. and Fladung, M. 2002. Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration. *Plant J.* 31: 543–551.
- 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.
- Li, Y., Agrawal, A., Sakon, J. and Beitle, R.R. 2001. Characterization of metal affinity of green fluorescent protein and its purification through salt promoted, immobilized metal affinity chromatography. *J. Chromatogr. A.* 909: 183–190.
- Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B. and Koncz, C. 1991. T-DNA integration: a model of illegitimate recombination in plants. *EMBO J.* 10: 697–704.
- McElroy, D. and Brettell, R.I. 1994. Foreign gene expression in transgenic cereals. *Trends Biotechnol.* 12: 62–68.
- Molinier, J., Himber, C. and Hahne, G. 2000. Use of green fluorescent protein for detection of transformed shoots and homozygous offspring. *Plant Cell Rep.* 19: 219–223.
- Pang, S.-Z., DeBoer, D.L., Wan, Y., Ye, G., Layton, J.G., Neher, M.K., Armstrong, C.L., Fry, J.E., Hinshee, M.A. and Fromm, M.E. 1996. An improved green fluorescent protein gene as a vital marker in plants. *Plant Physiol.* 112: 893–900.
- Prasher, D.C. 1995. Using GFP to see the light. *Trends Genet.* 11: 320–323.
- Rice Chromosome 10 Sequencing Consortium. 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300: 1566–1569.
- Sallaud, C., Meynard, D., van Bostel, J., Gay, C., Bes, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., Portefaix, M., Ouwerkerk, P.B.F., Rueb, S., Delseny, M. and Guiderdoni, E. 2003. Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theor. Appl. Genet.* 106: 1396–1408.
- Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.-I., Eun, M.-Y., Yano, M., Jiang, J. and Gojobori, T. 2002. The genome sequence and structure of rice chromosome 1. *Nature* 420: 312–316.
- Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H. and Galbraith, D.W. 1995. Green-fluorescent protein as a new vital marker in plant cells. *Plant J.* 8: 777–784.
- Sperini, D., Beliveau, C. and Bellemare, G. 1999. Screening of transgenic plants by amplification of unknown genomic DNA flanking T-DNA. *Biotechniques* 27: 308–314.
- Springer, P.S. 2000. Gene traps: tool for plant development and genomics. *Plant Cell* 12: 1007–1020.
- Stachel, S.E., Timmerman, B. and Zambryski, P. 1987. Activation of *Agrobacterium tumefaciens vir* gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region: requirement for 5_ *virD* gene products. *EMBO J.* 6: 857–863.
- Stahl, R., Horvath, H., van Fleet, J., Voetz, M., von Wettstein, D. and Wolf, N. 2002. T-DNA integration into the barley genome from single and double cassette vectors. *Proc. Natl. Acad. Sci. USA* 99: 2146–2151.
- Stewart, C.N. Jr. 1996. Monitoring transgenic plants using *in vivo* markers. *Nature Biotechnol.* 14: 682.
- Stewart, C.N. Jr. 2001. The utility of green fluorescent protein in transgenic plants. *Plant Cell Rep.* 20: 376–382.
- Sundaresan, V. 1996. Horizontal spread of transposon mutagenesis: new uses of old elements. *Trends Plant Sci.* 1: 184–191.
- Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H. and Martienssen, R. 1995. Patterns of gene action in plant development revealed by enhancer trap transposable element. *Genes Dev.* 9: 1797–1810.
- Szabados, L., Kovacs, I., Oberschall, A., Abraham, E., Kerekes, I., Zsigmond, L., Nagy, R., Alvarado, M., Krasovskaja, I., Gal, M., Berente, A., Redei, G.P., Haim, A.B. and Koncz, C. 2002. Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. *Plant J.* 32: 233–242.
- Tinland, B. 1996. The integration of T-DNA into plant genomes. *Trends Plant Sci.* 1: 178–184.
- Vain, P., Afolabi, A.S., Worland, B. and Snape, J.W. 2003. Transgene behaviour in populations of rice plants transformed using a new dual binary vector system; pGreen/pSoup. *Theor. Appl. Genet.* 107: 210–217.
- Wang, K., Stachel, S.E., Timmerman, B., van Montagu, M. and Zambryski, P. 1987. Site-specific nick occurs within the 25-bp transfer promoting border sequence following induction of *vir* gene expression in *Agrobacterium tumefaciens*. *Science* 235: 587–591.
- Wenck, A., Czako, M., Kanevski, I. and Marton, L. 1997. Frequent collinear long transfer of DNA inclusive of the whole binary vector during *Agrobacterium*-mediated transformation. *Plant Mol. Biol.* 34: 913–922.
- Williams, T.M., Burlein, J.E., Ogden, S., Kricka, L.J. and Kant, J.A. 1989. Advantages of firefly luciferase as a reporter gene: application to the interleukin-2 gene promoter. *Anal. Biochem.* 176: 28–32.

- Windels, P., De Buck, S., Van Bockstaele, E., De Loose, M. and Depicker, A. 2003. T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. *Plant Physiol.* 133: 2061–2068.
- Wolters, A.A., Trindade, L.M., Jacobsen, E. and Visser, R.G.F. 1998. Fluorescence *in situ* hybridization on extended DNA fibres as a tool to analyse complex T-DNA loci in potato. *Plant J.* 13: 837–847.
- Wu, C., Li, X., Yuan, W., Chen, G., Kilian, A., Li, J., Xu, C., Li, X., Zhou, D.X., Wang, S. and Zhang, Q. 2003. Development of enhancer trap lines for functional analysis of the rice genome. *Plant J.* 35: 418–427.
- Yamamoto, Y.Y., Tsuchihara, Y., Gohda, K., Suzuki, K. and Matsui, M. 2003. Gene trapping of the *Arabidopsis* genome with a firefly luciferase reporter. *Plant J.* 35: 273–283.
- Yamazaki, M., Tsugawa, H., Miyao, A., Yano, M., Wu, J., Yamamoto, Y., Matsumoto, T., Sasaki, T. and Hirochika, H. 2001. The rice retrotransposon *Tos17* prefers low copy sequences as integration targets. *Mol. Gen. Genet.* 265: 336–344.
- Yang, T.T., Sinai, P., Green, G., Kitts, P.A., Chen, Y.T., Lybarger, L., Chervenak, R., Patterson, G.H., Piston, D.W. and Kain, S.R. 1998. Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein. *J. Biol. Chem.* 273: 8212–8216.
- Yanofsky, M.F., Porter, S.G., Young, C., Albright, L.M., Gordon, M.P. and Nester, E.W. 1986. The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease. *Cell.* 47: 471–477.
- Yeh, E., Gustafson, K. and Boulianne, G.L. 1995. Green fluorescent protein as a vital marker and reporter of gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 92: 7036–7040.
- Yin, Z. and Wang, G.-L. 2000. Evidence of multiple complex patterns of T-DNA integration into the rice genome. *Theor. Appl. Genet.* 100: 461–470.
- Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L. and Yang, H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79–92.
- Zimmer, M. 2002. Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem. Rev.* 102: 759–781.