Journal of Plant Biology, August 2007, 50(4) : 455-460

Analysis of Junctions between T-DNA and Plant Genome in Transgenic *Arabidopsis thaliana*

Jeong Hoe Kim[1] and Sangman Lee[2]*
[1]Department of Biology, Kyungpook National University, Daegu 702-701, Korea
[2]Department of Agricultural Chemistry, Kyungpook National University, Daegu 702-701, Korea

To understand the mechanism for *Agrobacterium*-mediated transformation of plants, we analyzed the junctions between T-DNA and plant genome, using 12 individual transgenic lines transformed with 7 different plant expression constructs. After performing TAIL-PCR, we sequenced 42 PCR products for analysis. All of the RBs were nicked by VirD1/VirD2 proteins whereas only 62% of the LBs were. Additional deletions of the adjacent T-DNA region were found in 50% of the RBs. For the LBs, only two showed such additional deletions. Filler DNAs were observed in 60% of the RBs (ranging from 1 to 132 nucleotides) versus 54% for the LBs. We also found that only 25% of the RBs were integrated into the plant genome while the rest showed integration into the expression constructs. In comparison, all of the LBs were integrated, except for one that was considered intact. Our results suggest that the origin for a binary vector backbone (BVB) in the plant genome is due not only to a mistake in the VirD1/VirD2 proteins within the T-DNA borders but also because of the linkage of RBs to either the T-DNA or BVB.

Keywords: *Agrobacterium*, binary vector backbone, filler DNA, left border, right border, T-DNA

Agrobacterium tumefaciens-mediated transformation is a frequently used method for generating transgenic plants (Ku et al., 2006), as well as for transforming other bacteria (Kelly and Kado, 2002), yeast (Piers et al., 1996), fungi (de Groot et al., 1998), and human cells (Kunik et al., 2001). The transferred DNA (T-DNA), as part of the Ti-(tumor-inducing) plasmid of *A. tumefaciens*, is delimited by two 25-bp direct repeats called T-DNA borders (right border RB and left border LB). VirD1 and VirD2 proteins recognize the border sequences and produce a single-stranded (ss) T-DNA through nicking between the third and the fourth bases at the bottom strand of the T-DNA borders. The VirD2 protein is covalently linked to the 5′-end of the ss T-DNA, which is transported to the plant cell nucleus in the form of a nucleoprotein complex (T-complex). Finally, the T-DNA is integrated randomly into the plant genome through illegitimate recombination (Gheysen et al., 1991; Gorbunova and Levy, 1997). This integrated T-DNA is single and full-length, but is sometimes truncated or rearranged in multiple copies or may include vector DNA from outside the T-DNA borders (Kononov et al., 1997; Krizkova and Hrouda, 1998). The transfer of a binary vector backbone (BVB) could be the result of read-through at the LB, which would then prevent the normal termination of the T-DNA transfer (Buck et al., 2000; Meza et al., 2002).

Modeling this T-DNA integration mechanism has been based on studies of the junction between T-DNA and plant DNA, and the structures of integrated T-DNA. Two possible models for T-DNA integration are single-strand-gap repair (SSGR) and double-strand-break repair (DSBR) (Mayerhofer et al., 1991). In the SSGR model, the target DNA is nicked and produces a single-stranded gap via exonuclease activity. Then, both the 5′- and 3′-ends of ss T-DNA are annealed to microhomology areas within the target DNA. This is followed by trimming of the T-DNA overhang and ligation to the target DNA. In the DSBR model, the target DNA is broken and produces a double-stranded (ds) gap, and the ss T-DNA is then converted to ds DNA. Both ends of the target DNA and T-DNA are processed by exonuclease to form ss ends, followed by the annealing of each other in areas of microhomology. The ss overhangs are trimmed by exonuclease, and T-DNA is ligated to the target DNA.

Although T-DNA integration is an important step in the processes of *Agrobacterium*-mediated transformation, our understanding of its mechanism is still poor. Therefore, this study analyzed the junctions between T-DNA and target DNA to produce more information for improving the current models of integration.

MATERIALS AND METHODS

Plant Materials and Growing Conditions

Transgenic seeds of *A. thaliana* (L.) Heynh. (ecotype Columbia) were germinated and grown on a half-strength Murashige and Skoog (MS) (Murashige and Skoog, 1962) agar medium (pH 5.8) containing 2% (w/v) sucrose. Plates were maintained in a growth chamber at 23°C under a 12-h photoperiod provided by cool-white fluorescent tubes (photon flux density approx. 80 mE m^{-2} s^{-1}) (Kim and Lee, 2007).

Plant Expression Constructs

APS1, a cDNA encoding *Arabidopsis* ATP sulfurylase (Leustek et al., 1994), was cloned into a plant transformation vector (pBI101) for expression of sense mRNA. A 1490-bp *Xho*I/*Psp*1406I fragment was cloned into pUC18. The *Xho*I and *Bam*HI sites were made compatible with two-base fill-ins of the ends. *Psp*1046I cuts 27 bp 3′ of the translation termination codon, thus eliminating the poly-A tail from the construct. The end created with *Psp*1046I was compatible

*Corresponding author; fax +82-53-953-7233
e-mail sangman@knu.ac.kr

with *Acc*I. *APS1* was then cloned into pFF20, a modified version of the plant expression vector pFF19, in which the *Hin*dIII site is replaced with *Sal*I and the *Sal*I site of the polylinker is eliminated. The *APS*1 expression cassette was cloned into the *Sal*I/*Eco*RI sites of pBI101 so that pAPS(+) replaced the GUS cassette in this vector. *APK* is a 1077-bp cDNA encoding *Arabidopsis* APS kinase (Jain and Leustek, 1994). It was cloned into a plant transformation vector (pBI101) for expression of sense mRNA. An 1100-bp *Xho*I fragment was cloned into the *Bam*HI site of pFF20. The *Xho*I and *Bam*HI sites were made compatible with two-base fill-ins of the ends, and plasmids containing the insert in the sense orientation were screened according to their restriction patterns. The expression cassette was cloned into pBI101 so that pAPK(+) replaced the GUS cassette in this vector. The procedures to construct pPCS(+), pPCS, and pM3 were previously described by Lee et al. (2002, 2003). Briefly, expression of *Arabidopsis* phytochelatin synthase (*AtPCS1*) cDNA was under the control of the cauliflower mosaic virus 35S promoter in pPCS. In pPCS(+), expression of the genomic *AtPCS1* gene was under the control of a 2.0-kb promoter of its own *AtPCS1*. Moreover, in pM3, expression of the *uidA* gene was under the control of a 2.0-kb *AtPCS1* promoter. To prepare p3B, we carried out PCR with *Taq* DNA polymerase, using the forward primer 5′-CACTCTAGAATGGAGTGTGAGTGTGATTGT-3′ and the reverse primer 5′-CACGAGCTCACCACAATCACAATCACACTCACACTC-3′. No templates were needed because the primers had partial sequences complementary to each other. Finally, a 27-bp synthetic gene (3B) of the PCR prod-

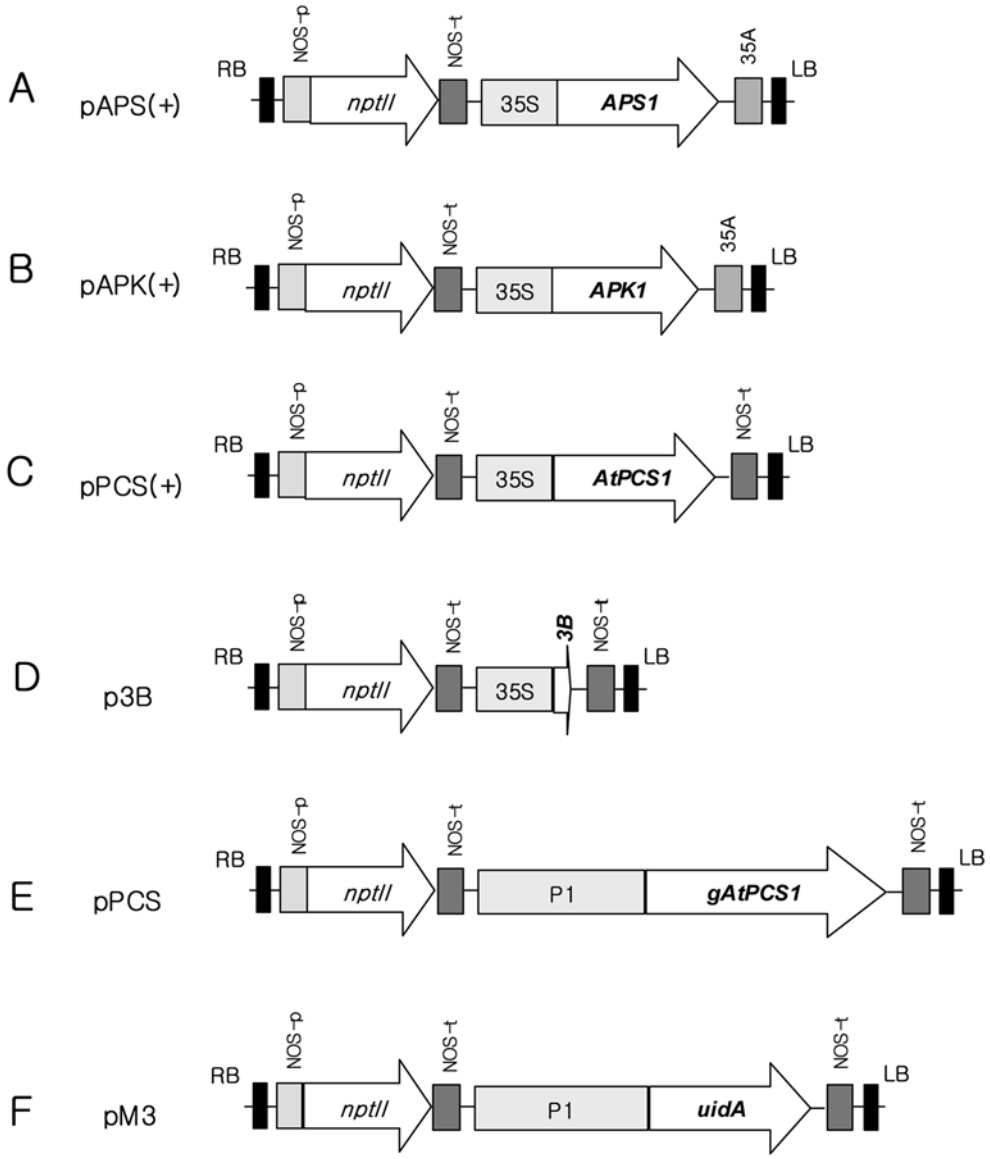

Figure 1. Plant expression constructs. *Arabidopsis* (**A**) ATP-sulfurylase, *APS1*; (**B**) APS kinase, *APK1*; (**C**) phytochelatin synthase, *AtPCS1*; and (**D**) synthetic gene, *3B*, were fused to the 35S promoter of pBI121 in sense orientation. Both (**E**) genomic *AtPCS1* and (**F**) *uidA*; β-glucuronidase (GUS) were fused in sense orientation to 2.0-kb promoter (P1) of *AtPCS1* in pBI121. Arrows indicate direction of transcription. RB, right border; LB, left border; NOS-p, nopaline synthase promoter; *nptII*, neomycin phosphotransferase gene; NOS-t, nopaline synthase terminator; *35S*, cauliflower mosaic virus 35S promoter; *gAtPCS1*, genomic *A. thaliana* PC synthase1.

uct was cloned into the *Xba*I/*Sst*I sites of pBI121, thus replacing the GUS.

Plant Transformation

Wild-type *Arabidopsis* plants were raised in a growth chamber for 7 weeks in 10-cm^2 pots containing Sunshine Mix #1 (Sun Gro Horticulture, USA), under the conditions described above. They were transformed by the floral-dip method (Clough and Bent, 1998) using *A. tumefaciens* GV3101 (pMP90), which harbored the plant gene expression constructs shown in Figure 1. Transgenic lines were selected on an MS agar medium containing 50 mg L^{-1} kanamycin, and their homozygous T_3 seeds were used for analysis.

Thermal Asymmetric Interlaced-PCR Analysis

Genomic DNA was extracted from axenically grown 10-d-old transgenic *A. thaliana* seedlings with a Nucleon Extraction and Purification Kit (Amersham Life Science, UK) according to the manufacturer's instructions. Experimental conditions for TAIL-PCR were the same as those described by Liu and Whittier (1995) except for the use of several different primers. These included TR1 (5′-TTGACAAGAA-ATATTTGCTAGCTG-3′), TR2 (5′-GATAGTGACCTTAGGCGA-CTTTT-3′), TR3 (5′-CTGTCAGTTCCAAACGTAAAACG-3′), TL1 (5′-CCTATCTCGGGCTATTCTTTTGA-3′), TL2 (5′-CTTTTGA-TTTATAAGGGATTTTGC-3′), and TL3 (5′-GTCTCACTGGT-GAAAAGAAAAAC-3′). The PCR products were sub-cloned into a pGEM vector (Promega, USA) and sequenced with an ABI 310 sequencer (Perkin-Elmer, USA).

RESULTS AND DISCUSSION

Seven plant expression constructs (Fig. 1 and pBI121) were used in our *Arabidopsis* transformations, and the homozygous T_3 seeds were screened to examine T-DNA integration into the plant genome. We performed TAIL-PCR with 12 individual transgenic lines and sequenced 42 PCR products. All of the RBs were nicked by the VirD1/VirD2 proteins during the integration process compared with the nicking of only 62% (13/21) of the LBs (Table 1). Moreover, additional deletions of the TR region (i.e., the T-DNA region adjacent to the RB) were observed in 50% (10/20) of the RBs. In Line 121-R2, 100 nucleotide deletions occurred in

Table 1. Analysis of T-DNA integration into plant genome in both right-border and left-border regions. TAIL-PCR was performed on 12 individual transgenic lines transformed with 7 different plant expression constructs, and the PCR products were sequenced. The numbers in parentheses indicate the additional deletion of nucleotides located in either the TL (T-DNA region close to LB) or TR (T-DNA region close to RB) regions. LB, left border; RB, right border; FD, filler DNA; D1, AD1 primer; D2, AD2 primer; D3, AD3 primer; R2, TR2 primer; R3, TR3 primer; L2, TL2 primer; and L3, TL3 primer. No cut means intact LB.

Construct	Line	PCR product	RB deletion	Filler DNA	PCR product	LB deletion	Filler DNA
pM3	M3-1	D3R3-1	20	32	D3L3-1	3	104
		D3R3-2	25(13)	63	D3L3-2	20	26
		D3R3-3	22	60			
pPCS	PCS25	D1R3	23	0	D1L3-1	No cut	
		D2R3	25(24)	0	D1L3-2	22	30
		D3R3	23	0	D1L3-3	21	3
					D3L3-1	25(12)	0
					D3L3-2	16	0
p3B	3B-30	D2R2	22	0	D2L3	No cut	
p3B	3B-49	D2R2	25(5)	24	D2L2	No cut	
pBI121	121-R1	D1R3-1	25	0	D1L3	No cut	
		D1R3-2	25(28)	0			
pBI121	121-R2	D2R3-1	25(100)	47	D2L3	No cut	
		D2R3-2	25(3)	7			
pBI121	121-R3	D3R3	25(67)	0	D3L3	No cut	
pBI121	121-R4	D1R3	24	39	D1L3	3	6
pAPK(+)	K(+)Ab	D1R3	25(38)	21	D3L3	No cut	
pPCS(+)	PC36	D2R2	22	132	D1L3	5	0
					D2L2	8	7
					D3L3-1	No cut	
					D3L3-2	12	6
pPCS(+)	PC40-4	D1R3	23	1	D2L3	10	0
pAPS(+)	S(+)Ab	D1R3	25(13)	30	D1L2	3	0
		D2R3	25(6)	0	D2L3	25(42)	0
		D3R3	24	53			

the D2R3-1 PCR product. For the LBs, only two [D3L3-1 in PCS25 and D2L3 in S(+)Ab] showed additional deletions of the TL region (i.e., the T-DNA region adjacent to the LB). Filler DNA was also found in 60% of the RBs (12/20; ranging from 1 to 132 nucleotides), while the LBs showed either 38% (7/21) or 54% (7/13) of the filler DNA (the latter being true if the intact, or uncut, LBs were excluded).

Deletions may be more severe at the 3′-end of the T-DNA because the VirD2 protein protects the 5′-end (Gheysen et al., 1991; Mayerhofer et al., 1991). However, our findings differed from those reported in which deletions of the 5′-end of T-DNA occurred more frequently. Thus, we suggest that, even though the VirD2 protein is involved in protecting the 5′-end during T-DNA transport into the host cell nucleus, it may not provide any protection in the T-DNA integration step.

We also analyzed the fate of the RB region. Here, only 25% (5/20) of the RBs were integrated into the plant genome while the remaining 75% showed integration into the plant expression constructs (Fig. 2). The LB region also was analyzed (Fig. 3). All of the LBs were integrated into the plant genome, except for one (PCS25, D3L3-1) because we did not consider the intact LBs. Whereas 75% of the RBs had been integrated into either the T-DNA region or the binary vector backbone (BVB) instead of into the plant genome, only one case of LB showed integration into the T-

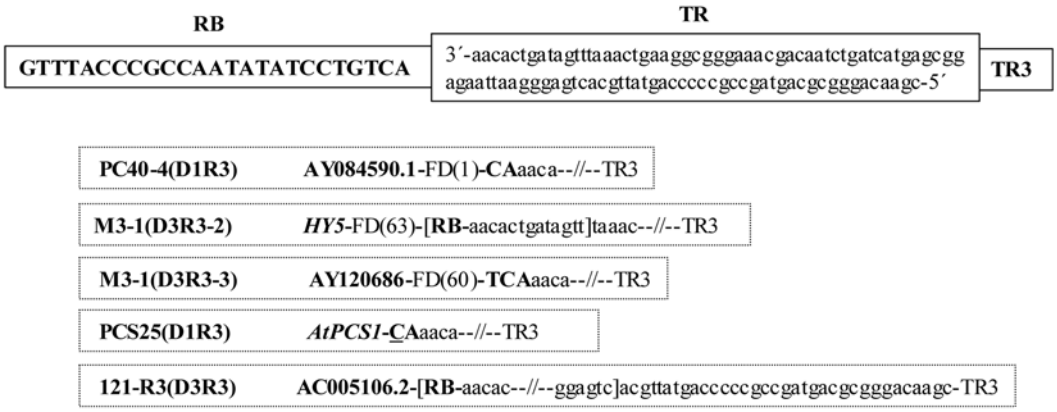

Figure 2. Analysis of T-DNA junctions between plant genome and RB regions. RB, right border; TR, T-DNA region close to RB; TR3, sequence of primer located in TR; FD, filler DNA. Numbers in parentheses indicate total number of nucleotides, and nucleotide sequence in brackets means deletion. Underlined nucleotide belongs to each side of junction, but is considered part of RB.

Figure 3. Analysis of T-DNA junctions between plant genome and LB regions. LB, left border; TL, T-DNA region close to LB; TL3, sequence of primer located in TL; FD, filler DNA. Numbers in parentheses indicate total number of nucleotides, and nucleotide sequence in brackets means deletion. Underlined nucleotides belong to each side of junction, but are considered part of T-DNA.

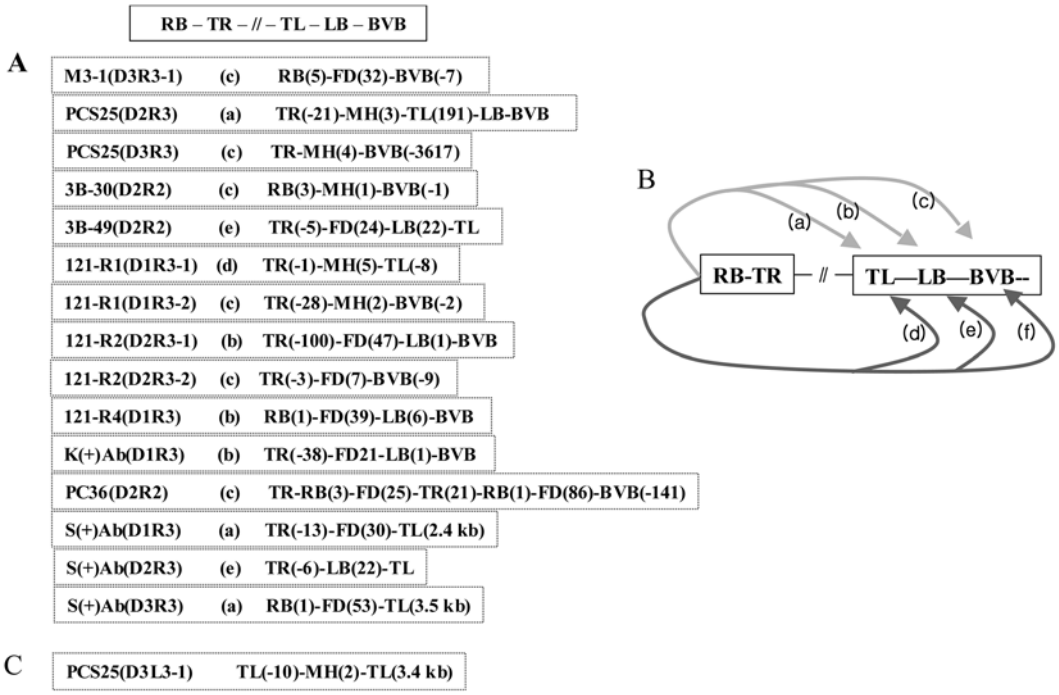

Figure 4. Analysis of LB and RB regions where T-DNA integration occurs in plant expression construct. (**A**) Integration of RB regions into plant expression vector, (**B**) types of RB integrations classified by position and orientation, and (**C**) integration of LB region into plant expression vector. LB, left border; RB, right border; TL, T-DNA region close to LB; TR, T-DNA region close to RB; FD, filler DNA; BVB, binary vector backbone; MH, microhomology. Positive and negative numbers in parentheses indicate total remaining (or position from LB) and deleted number of nucleotides, respectively.

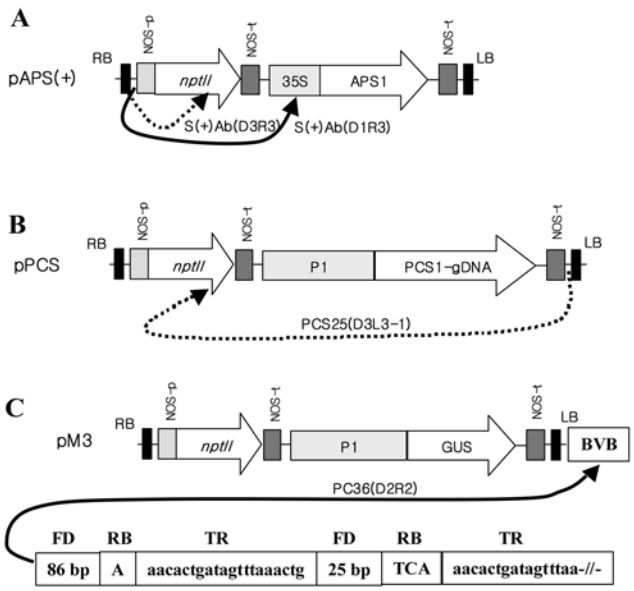

Figure 5. Extraordinary examples of T-DNA integration into plant expression construct. (**A**) Integration of RB region into middle of T-DNA, (**B**) integration of LB region into middle of T-DNA, and (**C**) integration of RB region into middle of BVB. LB, left border; RB, right border; TR, T-DNA region close to RB; FD, filler DNA; BVB, binary vector backbone.

DNA region (Fig. 4). Figure 5 presents extraordinary examples of integration into our plant expression construct. For instance, in the S(+)Ab line, two PCR products (D3R3 and D1R3) showed that the RB region was linked to the middle of the T-DNA (Fig. 5A). In Line PCS25, the LB region was linked to the T-DNA region close to the RB (Fig. 5B) while Line PC36 had two repeats of the RB region that were linked to the BVB (Fig. 5C).

A unified model has been configured for T-DNA integration into double-strand breaks (DSBs) in plants and yeast cells (Tzfira et al., 2004). There, the ss T-DNA is transported into the nucleus and converted into ds DNA by the random-priming nature of the complementation process in the 3′-end. This ds T-DNA then integrates into the host DNA through either non-homologous recombination (NHR) or homologous recombination (HR). However, our results suggest that the conversion of ss T-DNA into ds T-DNA may initiate not at the 3′-end but at the 5′-end of the T-DNA. We believe this because many of the RBs here were linked to either T-DNA or the BVB. Furthermore, the RBs were more severely deleted than the LBs, and most of the latter were linked to the plant genome. We also suggest that the BVB in the plant genome may have originated not only through a mistake in the LBs by the VirD1/VirD2 proteins but also by the linkage of the RBs to either the BVB or T-DNA.

ACKNOWLEDGEMENT

This work was supported by a grant from the Korea Research Foundation (KRF-2004-042-F00006).

Received March 2, 2007; accepted June 13, 2007.

LITERATURE CITED

Buck SD, Wilde CD, van Montagu M, 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

Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16: 735-743

de Groot MJ, Bundock P, Hooykaas PJ, Beijersbergen AG (1998) *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat Biotechnol 16: 839-842

Gheysen G, Villarroel R, van Montagu M (1991) Illegitimate recombination in plants: A model for T-DNA integration. Gene Dev 5: 287-297

Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucl Acids Res 25: 4560-4657

Jain A, Leustek T (1994) cDNA clone for 5′-adenylphosphosulfate kinase from *Arabidopsis thaliana*. Plant Physiol 105: 771-772

Kelly BA, Kado CI (2002) *Agrobacterium*-mediated T-DNA transfer and integration into the chromosome of *Streptomyces lividans*. Mol Plant Pathol 3: 125-134

Kim JH, Lee S (2007) Overexpression of *Arabidopsis* phytochelatin synthase (*AtPCS1*) does not change the maximum capacity for non-protein thiol production induced by cadmium. J Plant Biol 50: 220-223

Kononov ME, Bassuner B, Gelvin SB (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

Krizkova L, Hrouda M (1998) Direct repeats of T-DNA integration in tobacco chromosome: Characterization of junction regions. Plant J 16: 673-680

Ku JJ, Park YH, Park YD (2006) A non-biotic selection system uses the phosphomannose isomerase (PMI) gene for *Agrobacterium*-mediated transformation of Chinese cabbage. J Plant Biol 49: 115-122

Kunik T, Tzfira T, Kapulnik Y, Dingwall C, Citovsky V (2001) Genetic transformation of HeLa cells by *Agrobacterium*. Proc Natl Acad Sci USA 98: 1871-1876

Lee S, Moon JS, Domier LL, Korban SS (2002) Molecular characterization of phytochelatin synthase expression in transgenic *Arabidopsis*. Plant Physiol Biochem 40: 727-733

Lee S, Moon JS, Ko TS, Petros D, Goldsbrough PB, Korban SS (2003) Overexpression of *Arabidopsis* phytochelatin synthase paradoxically leads to hypersensitive to cadmium stress. Plant Physiol 131: 656-663

Leustek T, Murillo M, Cervantes M (1994) Cloning of a cDNA encoding ATP sulfurylase from *Arabidopsis thaliana* by functional expression in *Saccharomyces cerevisiae*. Plant Physiol 105: 897-902

Liu YG, Whittier RF (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

Mayerhofer R, Koncz-Kalman Z, Nawrath C, Bakkeren G, Crameri A, Angelis K, Redei GP, Schell J, Hohn B, Koncz C (1991) T-DNA integration: A mode of illegitimate recombination in plants. EMBO J 10: 697-704

Meza TJ, Stangleand B, Mercy IS, Skarn M, Nymoen DA, Berg A, Butenko MA, Hakelien AM, Haslekas C, Meza-Zepeda LA, Aalen RB (2002) Analysis of single-copy *Arabidopsis* T-DNA transformed lines shows that the presence of vector backbone sequences, short inverted repeats and DNA methylation is not sufficient or necessary for the induction of transgene silencing. Nucl Acids Res 30: 4556-4566

Murashige T, Skoog T (1962) A revised medium for growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 473-479

Piers KL, Heath JD, Liang X, Stephens KM, Nester EW (1996) *Agrobacterium tumefaciens* mediated transformation of yeast. Proc Natl Acad Sci USA 93: 1613-1618

Tzfira T, Li J, Lacroix B, Citovsky V (2004) *Agrobacterium* T-DNA integration: Molecules and models. Trends Genet 20: 375-383