Plant Physiology

Focus Issue: The Agrobacterium-Plant Cell Interaction

Site-Specific Integration of *Agrobacterium tumefaciens* T-DNA via Double-Stranded Intermediates[1]

Tzvi Tzfira*, Leah Renée Frankman, Manjusha Vaidya, and Vitaly Citovsky

Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, New York 11794 (T.T., M.V., V.C.); and Westhampton Beach High School, 340 Mill Road, Westhampton Beach, New York 11978 (L.R.F.)

Agrobacterium tumefaciens-mediated genetic transformation involves transfer of a single-stranded T-DNA molecule (T strand) into the host cell, followed by its integration into the plant genome. The molecular mechanism of T-DNA integration, the culmination point of the entire transformation process, remains largely obscure. Here, we studied the roles of double-stranded breaks (DSBs) and double-stranded T-DNA intermediates in the integration process. We produced transgenic tobacco (*Nicotiana tabacum*) plants carrying an I-*Sce*I endonuclease recognition site that, upon cleavage with I-*Sce*I, generates DSB. Then, we retransformed these plants with two *A. tumefaciens* strains: one that allows transient expression of I-*Sce*I to induce DSB and the other that carries a T-DNA with the I-*Sce*I site and an integration selection marker. Integration of this latter T-DNA as full-length and I-*Sce*I-digested molecules into the DSB site was analyzed in the resulting plants. Of 620 transgenic plants, 16 plants integrated T-DNA into DSB at their I-*Sce*I sites; because DSB induces DNA repair, these results suggest that the invading T-DNA molecules target to the DNA repair sites for integration. Furthermore, of these 16 plants, seven plants incorporated T-DNA digested with I-*Sce*I, which cleaves only double-stranded DNA. Thus, T-strand molecules can be converted into double-stranded intermediates before their integration into the DSB sites within the host cell genome.

Agrobacterium tumefaciens is the only known organism capable of inter-kingdom DNA transfer (Stachel and Zambryski, 1989). In nature, this bacterium causes crown gall disease in many dicotyledonous plant species by transferring a specific DNA fragment, the transferred DNA (T-DNA), from its tumor-inducing (Ti) plasmid into the host cell (for review, see Gelvin, 2000; Tzfira et al., 2000; Zupan et al., 2000; Gelvin, 2003). The wild-type T-DNA carries two types of genes: oncogenes and genes for biosynthesis of amino acid derivatives, opines. Expression of the integrated T-DNA, therefore, results in uncontrolled cell division and formation of tumors (Gaudin et al., 1994; Das, 1998) that produce and secrete opines that are utilized by *A. tumefaciens* and several other microorganisms as the source of carbon and nitrogen (for review, see Savka et al., 2002). The T-DNA itself does not encode genes required for its transfer and integration and is defined only by two 25-bp direct repeats, termed T-DNA left and right borders (Zambryski et al., 1982). Thus, the entire wild-type T-DNA sequence between the borders can be replaced with a gene of interest that will be transferred to plants and integrated in the plant genome, representing the molecular basis for plant genetic engineering (for review, see Gelvin, 1998b, 2000, 2003; Potrykus et al., 1998).

Several *A. tumefaciens chv* (chromosomal virulence) genes and a set of *vir* (virulence) genes were contained on the Ti plasmid code for the protein machinery of the T-DNA transport (for review, see Sheng and Citovsky, 1996; Gelvin, 2000, 2003; Tzfira and Citovsky, 2000). The VirA/VirG two-component signal transduction system senses secretion of phenolics and specific sugar compounds from the wounded susceptible plant tissues and induces expression of other *vir* genes (Winans et al., 1994). The VirD2/VirD1 complex nicks the bottom strand of the Ti plasmid at the T-DNA borders and, with the assistance of the bacterial DNA synthesis and repair machinery, a single-stranded copy of the T-DNA, the T strand, is liberated from the Ti plasmid (Wang et al., 1987; Filichkin and Gelvin, 1993; Scheiffele et al., 1995). VirD1 is then released, whereas VirD2 remains covalently attached to the 5' end of the T strand (Ward and Barnes, 1988; Young and Nester, 1988; Howard et al., 1989). The resulting protein-DNA complex is exported to the host cell through the type IV secretion system encoded by the *virD4* gene and *virB* operon, with 11 open reading frames (for review, see Christie, 1997; Christie and Vogel, 2000). The VirD4/VirB channel also mediates export of VirE2 (Vergunst et al., 2000), a single-stranded DNA-binding protein (Gietl et al., 1987; Christie et al., 1988; Citovsky et al., 1988, 1989; Das, 1988; Sen et al., 1989)

thought to associate cooperatively with the T-strand in the plant cell cytoplasm (Citovsky et al., 1992; Gelvin, 1998a), producing T-complex (Howard et al., 1990; Zupan and Zambryski, 1997), a semirigid, hollow, cylindrical filament composed of a T-strand molecule attached to one molecule of VirD2 and coated by multiple VirE2 monomers (Citovsky et al., 1997). Both VirD2 and VirE2 are thought to facilitate import of the T-complex into the host cell nucleus (Howard et al., 1992; Rossi et al., 1993; Citovsky et al., 1994; Ziemienowicz et al., 2001), after which the T-DNA is integrated into the plant genome by illegitimate recombination (Gheysen et al., 1991; Mayerhofer et al., 1991; Bundock and Hooykaas, 1996; Tinland, 1996).

Initially, two different models have been suggested for T-DNA integration, double-stranded break (DSB) repair and single-stranded gap repair (Mayerhofer et al., 1991). The first model predicts that unwound ends of a double-stranded T-DNA molecule anneal with single-stranded overhangs of a DSB in the plant DNA, the residual 5' and 3' overhangs are removed, and the inserted T-DNA is ligated. The second model suggests that integration initiates with a nick, which leads to a gap, in the plant DNA, both ends of a single-stranded T-DNA molecule anneal to this gap, the residual 5' and 3' overhangs are removed, and the integration is completed by repair synthesis of the second T-DNA stand (Mayerhofer et al., 1991). Both models also suggest a role for VirD2 in recognition of nicks or gaps by interaction with plant factors (Mayerhofer et al., 1991).

The single-stranded gap repair model was extensively revised (Tinland and Hohn, 1995; Tinland et al., 1995) to suggest that the 3' end of the T strand acts as a driving force in the selection of the integration site. According to this current model, the integration process is initiated by microhomologies between the 3' end of the T strand and the upper strand of the plant DNA. The lower plant DNA strand is then nicked, and VirD2 ligates the 5' end of the T strand into this nick. The 3' overhang of the T strand is removed, and the second strand of the integrated T-DNA is filled in by the DNA repair machinery of the host cell (Tinland and Hohn, 1995; Tinland et al., 1995). This model is based on four lines of circumstantial evidence: T-DNA is transported into the plant cell as a single-stranded molecule (Tinland et al., 1994; Yusibov et al., 1994), nucleotide sequence comparison between the 3' ends of the T-DNA and the integration sites identifies regions of microhomology (Tinland et al., 1995; Brunaud et al., 2002), VirD2 associates with the 5' end of the T-DNA (Ward and Barnes, 1988; Young and Nester, 1988; Howard et al., 1989) and cleaves the T-DNA left border sequence and ligates the adjoining DNA strands in vitro (Pansegrau et al., 1993; Jasper et al., 1994), and mutations in VirD2 result in imprecise integration of the 5' end of the T-DNA (Tinland et al., 1995). Conversely, this model cannot account for the formation of several complex T-DNA integration patterns (e.g. Krizkova and Hrouda, 1998; McCormac et al., 2001), it relies on the ligase activity of VirD2 that may be dispensable for integration in vitro (Ziemienowicz et al., 2000), and it does not explain the fast kinetics of transient T-DNA expression (Janssen and Gardner, 1990; Tinland et al., 1994; Narasimhulu et al., 1996; De Neve et al., 1997) or rapid expression of T strands corresponding to the coding strand of the T-DNA (Narasimhulu et al., 1996), all of which imply that T strands can be converted to double-stranded molecules already early in the infection process.

Recent studies further supported the notion that the T-strand molecule becomes double stranded before its integration (De Neve et al., 1997; Salomon and Puchta, 1998; De Buck et al., 1999). For example, because 5' ends of single-stranded DNA molecules cannot be ligated to each other, two T-DNA right borders integrated "head to head" suggested that the T strands may have been converted to a double-stranded form before integration; these observations, however, did not rule out the possibility of this conversion occurring during the first integration event followed by immediate integration of the second T-strand molecule (De Neve et al., 1997; De Buck et al., 1999). A more direct indication of the involvement of double-stranded integration intermediates came from detection of T-DNA inserts within the sites of DSB repair in the host DNA that suggested ligation of double-stranded T-DNA molecules to both sides of the break (Salomon and Puchta, 1998).

Here, we present direct evidence for the role of DSBs and double-stranded T-DNA intermediates in T-DNA integration. Specifically, we used transient expression of an intron-encoded endonuclease I-*Sce*I (Monteilhet et al., 1990) to create DSBs in the genomes of plants transgenic for the I-*Sce*I recognition site. Our data indicate that T-DNA is preferentially targeted and integrated into these DSB sites. Furthermore, the invading T-DNA that itself contained the I-*Sce*I recognition site was often integrated after being digested by I-*Sce*I, indicating its conversion to a double-stranded form before integration.

RESULTS

Experimental System

To investigate a role that DSB may play in T-DNA integration and to determine whether T-DNA integrates as a single- or double-stranded molecule, we utilized a genetic system for in vivo induction of DSBs (Salomon and Puchta, 1998). In this approach, DSB is induced by transient expression of the I-*Sce*I endonuclease in plants carrying a 35S-I-*Sce*I-*codA* transgene composed of the 35S promoter, the I-*Sce*I recognition site, and the cytosine deaminase gene *codA* (Stougaard, 1993; Salomon and Puchta, 1998) encoded by the pBNE3I binary vector (Fig. 1A). Im-
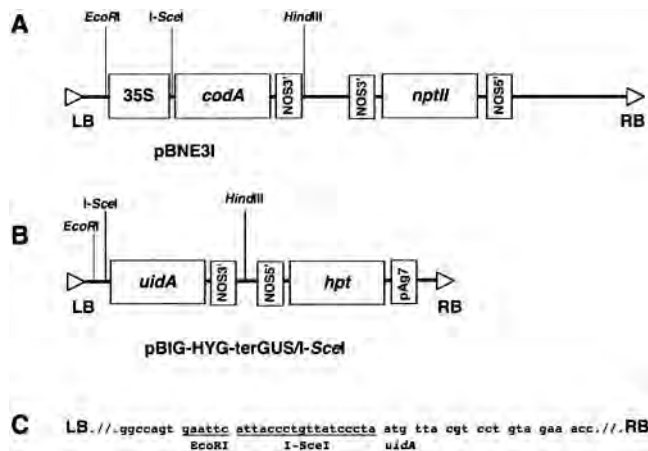
Figure 1. Maps of the T-DNA regions. A, pBNE3I. B, pBIG-HYG-terGUS/I-SceI. C, Nucleotide sequence of the I-SceI site between the T-DNA left border and the *uidA* gene in pBIG-HYG-terGUS/I-SceI. LB, T-DNA left border; RB, T-DNA right border; 35S, cauliflower mosaic virus 35S RNA promoter; *nptII*, kanamycin resistance gene; *codA*, cytosine deaminase gene, *uidA*; β-glucoronidase (GUS) gene; *hpt*, hygromycin resistance gene;NOS5', nopaline synthase promoter;NOS3', nopaline synthase terminator; pAg7, agropine synthase terminator.

proper repair of such DSBs, which may result from deletions or insertions of endogenous DNA segments and from capture of integrating T-DNA molecules, is detected by selection for loss of function of the cytosine deaminase transgene (Salomon and Puchta, 1998). In our experimental system, however, we did not select for the absence of the cytosine deaminase activity; instead, we selected for hygromycin resistance, which is encoded by the integrating I-SceI-*uidA-hpt* T-DNA derived from the pBIG-HYG-terGUS/I-SceI binary vector (Fig. 1B). This approach allowed us to detect all transformation events rather than only those that occurred within the DSB sites. The initial selection was followed by screening of the resulting population of hygromycin-resistant plants for the cytosine deaminase activity to determine the frequency of specific insertions into DSB within the 35S-I-SceI-*codA* transgene.

To determine the structure of the integrating T-DNA molecules, we designed the T-DNA to carry an I-SceI recognition sequence between the promoterless *uidA* gene and the T-DNA left border (Fig. 1B). If this T-DNA becomes double stranded to allow its cleavage with I-SceI before integration, and if the resulting fragment carrying the promoterless *uidA* integrates into the DSB created by I-SceI cleavage between the 35S promoter and the I-SceI site in the 35S-I-SceI-*codA* transgene, we expect to detect the *uidA* gene expression, i.e. GUS activity, in the transgenic plants. Thus, our experimental design aims to examine concurrently two important characteristics of the integration process: frequency of targeted integration into the specific DSB sites out of the total number of transformation events and frequency of formation of double-stranded T-DNA intermediates before integration.

Using this experimental design, we generated a total of 24 independent transgenic lines transformed with pBNE3I containing the 35S-I-SceI-*codA* T-DNA. Fifteen lines, showing a 3:1 kanamycin resistance segregation ratio, were analyzed by PCR to identify plants that do not contain tandem insertions of the 35S-I-SceI-*codA* transgene (Table I). In addition, sequence analysis of the amplified PCR fragments revealed that, in all 15 lines, the sequence of the integrated 35S-I-SceI-*codA* transgene was identical to the original sequence of this transgene within pBNE3I (data not shown). Plants were then tested for their cytosine deaminase activity, which indicates the degree of expression of the 35S-I-SceI-*codA* transgene, using leaf disc regeneration assay in the presence of 5-fluorocytosine (5-FC); 5-FC is known to select against cells that express cytosine deaminase (Stougaard, 1993). Five heterozygous lines (Sc1, Sc4, Sc33, Sc59, and Sc70) showing the highest sensitivity to 5-FC, i.e. very low or abolished ability to regenerate on a medium containing 5-FC, were selected for further studies.

Next, these 35S-I-SceI-*codA* transgenic plants were transformed with a mixture of *A. tumefaciens* strains harboring the binary vectors pCISceI, with the I-SceIORF T-DNA, and pBIG-HYG-terGUS/I-SceI,

Table I. *Primer combinations used for amplification of T-DNA/DSB junctions[a]*

Primer Set	Amplified Junction	Expected Size (bp)
35S-F2, *codA*-R	35S-I-SceI-*codA* transgene	509
35S-F2 *codA*-R	Disrupted 35S-I-SceI-*codA* (no T-DNA integration)	Unpredicted
35S-F2, T-DNA-RB	35S/T-DNA right border	Unpredicted
codA-R, T-DNA-RB	T-DNA right border/*codA*	Unpredicted
35S-F2, T-DNA-LB	35S/T-DNA left border	Unpredicted
codA-R, T-DNA-LB	T-DNA left border/*codA*	Unpredicted
35S-F2, GUS-R	35S / I-SceI-digested T-DNA left border	350
codA-R, GUS-R	I-SceI-digested T-DNA left border/*codA*	Approximately 672

[a] Primer sequences: 35S-F2, 5'CGTTCCAACCACGTCTTCAAAGC3'; *codA*-R, 5'GCCTTCAAACAG CGTGCCGG; T-DNA-RB, 5'CCCAGTCATAGCCGAATAGCC3'; T-DNA-LB, 5'GGCGTAATAGCGAA-GAGGCCC3'; GUS-R, 5'CCTGATTATTGACC CACACTTTGCCG3'.

with the I-*Sce*I-*uidA*-*hpt* T-DNA; these vectors encode the transiently expressed I-*Sce*I endonuclease and provide the integrating I-*Sce*I-*uidA*-*hpt* T-DNA, respectively. A total of 620 hygromycin-resistant transgenic lines were recovered and analyzed for their sensitivity to 5-FC; the 5-FC-resistant plants were then tested for GUS activity and the nucleotide sequence of their integration junctions was determined.

Efficiency of T-DNA Integration during DSB Repair

From 620 hygromycin-resistant plants, 82 plants exhibited significant 5-FC resistance. We then PCR amplified the putative integration junctions around the DSB site in 68 of these plants and examined them for small deletions or insertions in the DSB site. To this end, the 35S-I-*Sce*I-*codA* transgene was amplified using the 35S-F2 and *codA*-R primers (Table I) and a short extension time (1 min). All 68 plants exhibited sequence alterations within the *codA* gene, indicating approximately 11% frequency of DSB induction and improper repair and demonstrating that the 5-FC resistance of these lines was not simply and solely due to silencing of this transgene. Forty-five of 620 plants (7.25%) yielded short PCR fragments that ranged from 300 to 600 bp and, thus, could not represent the true T-DNA integration (Table II). Sequence analysis of five randomly chosen PCR fragments revealed deletions and/or insertions of short filler DNA that were most likely responsible for loss of function of the cytosine deaminase gene (data not shown).

Twenty-three of 620 plants (3.71%) did not show any visible PCR bands after short amplification, but four of them yielded longer PCR fragments, 2.2 kb, 2.8 kb, and two bands of 2.9 kb each (Table II), when longer PCR extension time (3 min) was used. Sequence analysis of these fragments revealed the integration of the full-length or truncated I-*Sce*IORF T-DNA into DSB sites. The remaining 19 plants were subjected to a series of PCR analyses using different primer combinations shown in Table I. Table II shows that five plants contained the full-length or truncated I-*Sce*I-*uidA*-*hpt* T-DNA integrated in their

DSBs, whereas in seven plants, the integrated I-*Sce*I-*uidA*-*hpt* T-DNA was found to have been digested by I-*Sce*I before integration. The last seven plants did not yield discrete PCR products, potentially due to large deletions or insertions (Table II). Thus, out of 620 transgenic lines examined in this study, we observed 16 site-specific T-DNA integration events, including those of the I-*Sce*IORF and I-*Sce*I-*uidA*-*hpt* transgenes, corresponding to 2.58% frequency of targeted integration. Furthermore, this is the minimal estimate of the frequency of insertions into DSB sites because it assumes that transient expression of I-*Sce*I created DSBs in all 620 transgenic lines.

Orientation of Integrated T-DNA within DSB Sites

Determination of the nucleotide sequence of T-DNA/DSB site integration junctions allowed their classification based on the orientation of the inserted T-DNA within the DSB site (Fig. 2). For the junctions in which the T-DNA left border was found ligated to the left side of DSB (DSB-L/LB, Fig. 2A) and the T-DNA right border was found ligated to the right side of DSB (RB/DSB-R, Fig. 2A), the sequences of the DSB site and the T-DNA were presented in sense orientation in Figures 3 to 5. When the T-DNA integrated in reverse orientation, we classified the junctions as DSB-L/RB and LB/DSB-R (Fig. 2B); in this case, to show the sense strand of the junction, we reversed and complemented the T-DNA sequence in Figures 3 to 5. Finally, when the I-*Sce*I-digested T-DNA integrated into DSB, recreating the I-*Sce*I site, we classified two additional types of junctions, DSB-L/I-*Sce*I (Fig. 2C) and I-*Sce*I/DSB-R (Fig. 2D). The T-DNA sequences of these junctions were presented in Figures 3 to 5 as sense (for DSB-L/I-*Sce*I) or reversed and complemented sequences (for I-*Sce*I/DSB-R).

Integration of I-*Sce*IORF T-DNA

In four plant lines, the full-length or truncated I-*Sce*IORF T-DNA integrated into the DSB site (Table II). In three of these lines (Sc4-2, Sc59-5, and Sc59-9),

Table II. *Frequency of site-specific T-DNA integration*

| Experimental Step	No. of Plants	Frequency of Transformation Events % of Total
Hygromycin	620	100
5-FC-resistant plants	82	13.22
Plants analyzed by PCR	68	10.96
PCR fragment > 600 bp or no PCR fragment	23	3.71
All cases of T-DNA integration	16	2.58
Integration of I-*Sce*IORF T-DNA	4	0.64
Integration of full-length or slightly truncated T-DNA	5	0.81
Integration of I-*Sce*I-digested *uidA*-*hpt* T-DNA	7	1.13
Large deletion/insertion	7	1.13
PCR fragment < 600 bp	45	7.25
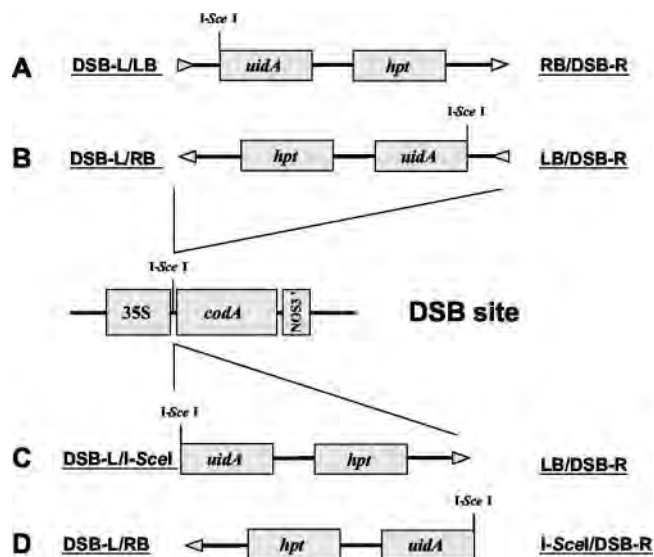
Figure 2. Orientation of the integrated T-DNA within DSB sites. A and B, Two possible orientations for integration of the I-*Sce*I-*uidA*-*hpt* T-DNA. C and D, Two possible orientations for integration of the I-*Sce*I-digested I-*Sce*I-*uidA*-*hpt* T-DNA. I-*Sce*I sites are shown within the integrating T-DNA and in the DSB site. Arrowheads indicate the left-to-right border direction of T-DNA. Designations of each T-DNA border/DSB site integration junction are indicated.

almost the entire I-*Sce*IORF T-DNA integrated into DSB, whereas in the fourth line (Sc70-3), a deletion of 797 bp was observed (Table III; Fig. 3). Sequence analysis of the DSB-L/LB junctions revealed that small deletions of 13 to 62 bp had occurred at the T-DNA left border. These deletions in the integrated T-DNA were accompanied by deletions, insertions, or no changes in the DSB site sequence (Sc4-2, Sc59-5, Sc70-3, and Sc59-9, Table III). The RB/DSB-R junctions were more conserved, showing no T-DNA right border alterations in three out of four cases (Sc4-2, Sc59-5, and Sc59-9) and were accompanied by small deletions or insertions in the DSB site (Table III). In the Sc59-5 line, insertion of 4 bp led to restoration of the TTAT fragment, originally located within the cleaved I-*Sce*I sequence of DSB (Fig. 3). In the Sc59-9 line, insertion of TA nucleotides was found at the LB/DSB-R junction (Fig. 3). In the Sc70-3 line, a large T-DNA right border deletion of 797 bp occurred together with a 10-bp deletion in the flanking DSB site (Table III; Fig. 3). Sequence homologies between the integrated T-DNA and the DSB site—analyzed within 10 bp from the integration point—were small, ranging between 3 and 10 bp at the left border integration point to 0 and 2 bp at the right border integration point (Table III; Fig. 3).

Integration of Full-Length I-*Sce*I-*uidA*-*hpt* T-DNA

In five lines (Sc1-10, Sc33-4, Sc33-7, Sc4-6, and Sc1-5), the full-length or only slightly (≤29 bp) truncated I-*Sce*I-*uidA*-*hpt* T-DNA integrated into the DSB site (Table II). Similar to integration of the I-*Sce*IORF

T-DNA, the RB/DSB-R integration junctions of the I-*Sce*I-*uidA*-*hpt* T-DNA were conserved and exhibited no alterations in the sequence of the T-DNA right border in four lines (Sc1-10, Sc33-4, and Sc1-5) and a small 18-bp deletion in one line (Sc4-6; Table III; Fig. 4). These deletions coincided with 0- to 24-bp deletions in the flanking DSB sequences (Table III). The DSB-L/LB junctions were less conserved, exhibiting deletions of 4 to 29 bp in the T-DNA left border and 1 to 3 bp in the DSB site in four lines (Sc1-10, Sc33-4, Sc4-6, and Sc1-5; Table III). In the fifth line, Sc33-4, GT nucleotides were inserted between the T-DNA left border and the DSB site (Fig. 4); however, unlike an insertion in the Sc59-5 line (Fig. 3) that recreated part of the I-*Sce*I recognition sequence, the junction of the Sc33-4 line did not restore this restriction site. Finally, in the Sc33-7 line, the T-DNA integrated in its entirety and without alterations, but the right and left sides of the DSB site underwent deletions of 22 and 26 bp, respectively (Fig. 4; Table III). Homologies between the integrated T-DNA and the DSB site ranged from 0 to 6 bp at the left side of DSB to 0 to 5 bp at the right side of DSB (Table III; Fig. 4).

Integration of I-*Sce*I-Digested I-*Sce*I-*uidA*-*hpt* T-DNA

Seven transgenic lines (Sc4-5, Sc4-8, Sc1-3, Sc4-12, Sc59-2, Sc33-9, and Sc70-9) contained only a part of the I-*Sce*I-*uidA*-*hpt* T-DNA integrated into the DSB site. In all seven lines, the integrated portion of the T-DNA lacked the sequences from the T-DNA left border to the I-*Sce*I site, probably due to cleavage of the T-DNA itself with I-*Sce*I (Table III; Fig. 5). Also, in six of these lines, a portion of the I-*Sce*I recognition sequence in the I-*Sce*I-*uidA*-*hpt* T-DNA was identified within the integration junction (lines Sc4-5, Sc4-8, Sc1-3, Sc59-2, Sc33-9, and Sc70-9 in Fig. 5), whereas in the remaining Sc4-12 line, the digested I-*Sce*I sequence was most likely deleted (see below).

In lines Sc4-5, Sc4-8, and Sc1-3, the T-DNA was inserted in sense orientation to the I-*Sce*I restriction site, producing the DSB-L/I-*Sce*I type of the integration junction (see Fig. 2C). This resulted in a precise match between the digested I-*Sce*I sequence of DSB and the digested I-*Sce*I sequence of the I-*Sce*I-*uidA*-*hpt* T-DNA, recreating the intact restriction site (Fig. 5). PCR fragments obtained from amplification of DSB-L/I-*Sce*I junctions of all three lines were cleavable by a recombinant I-*Sce*I enzyme (New England Biolabs, Beverly, MA) in vitro (data not shown). No homologies were observed between the I-*Sce*I-digested end of the integrating T-DNA and the DSB site sequences in these lines (Table III). In lines Sc59-2 and Sc33-9, in which the T-DNA also integrated in sense orientation to the I-*Sce*I site, the integration event did not recreate the restriction site due to deletions of 1 to 2 bp and 2 to 10 bp in the T-DNA and in the DSB site, respectively (Table III; Fig. 5); the amplified junction fragments from these lines failed to be digested with
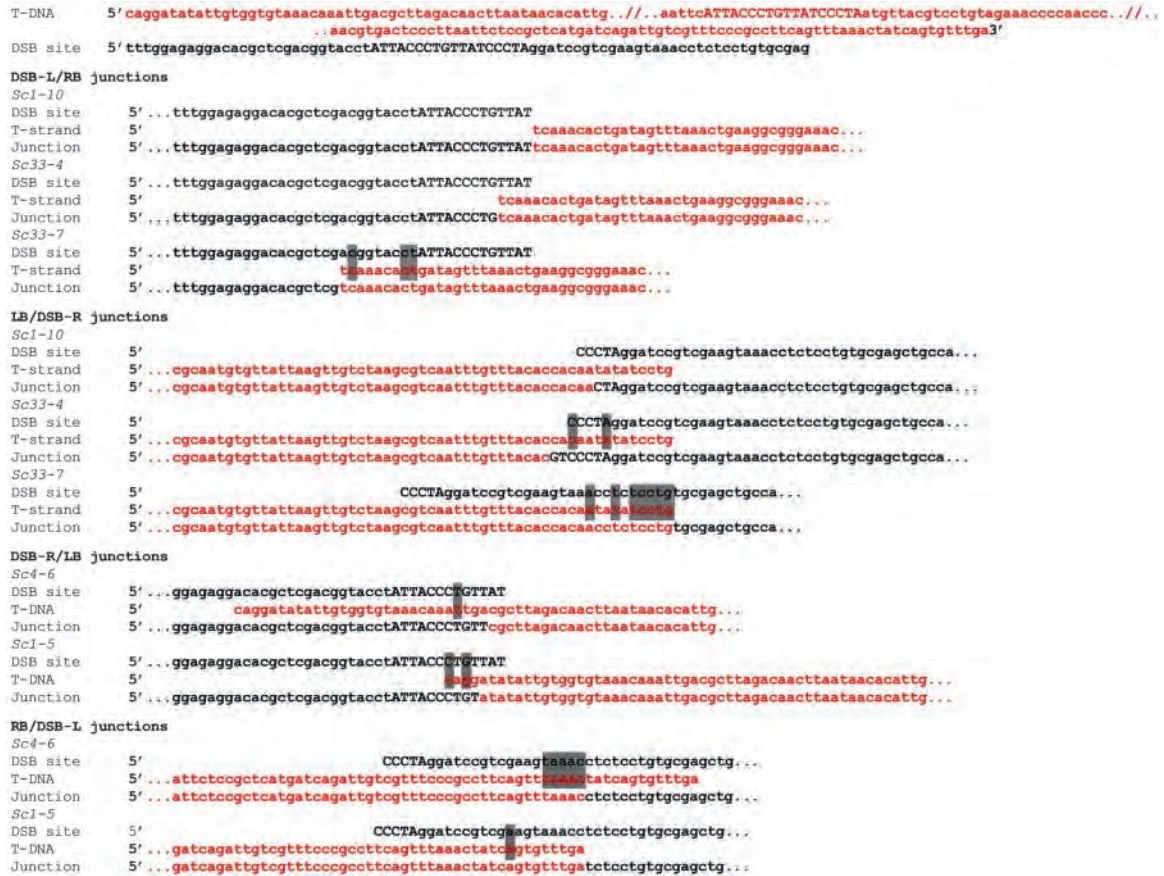
Figure 3. Nucleotide sequence of integration junctions between the I-*Sce*IORF T-DNA and DSB. Upper strands of the T-DNA and the DSB site are shown. The I-*Sce*I site is indicated in capital letters, and homologous base pairs are highlighted in gray. Numbers within the nucleotide sequences indicate the size of large deletions (see also Table III).

I-*Sce*I in vitro (data not shown). No homology between DSB and the I-*Sce*I-digested end of the T-DNA was observed in line Sc33-9, whereas a 5-bp homology was found in line Sc59-2 (Table III; Fig. 5).

Integration of the I-*Sce*I-digested I-*Sce*I-*uidA-hpt* T-DNA in sense orientation to the I-*Sce*I site of DSB should place the promoterless *uidA* gene in the T-DNA directly downstream of the 35S promoter in the DSB site (Fig. 2C). All five lines exhibiting such integration (Sc4-5, Sc4-8, Sc1-3, Sc59-2, and Sc33-9) expressed GUS activity in their tissues as detected by histochemical staining (data not shown). GUS expression was not detected in plants in which the full-length I-*Sce*I-*uidA-hpt* integrated downstream of the 35S promoter (see above) because the distance between the T-DNA left border and the *uidA* gene (see Fig. 1) was too large to allow promoter activity.

In lines Sc70-9 and Sc4-12, the T-DNA integrated in reverse orientation to the I-*Sce*I site of DSB, producing the I-*Sce*I/DSB-R type of the integration junction (see Fig. 2D). In Sc70-9, the integration junction contained only 2 bp of the I-*Sce*I site located in the T-DNA and 3 bp of the I-*Sce*I site derived from DSB. Sequence analysis of the integration junction from the Sc4-12 line showed a 17-bp deletion in the T-DNA, resulting in elimination of its entire I-*Sce*I site and a 4-bp deletion in DSB. DSB and the I-*Sce*I-digested end of the T-DNA showed no homologies to

each other in line Sc70-9, whereas a 3-bp homology was found in line Sc4-12 (Table III; Fig. 5).

The T-DNA right border of the I-*Sce*I-digested I-*Sce*I-*uidA-hpt*, which had no I-*Sce*I sequences (see Fig. 2, C and D), integrated in a relatively conserved fashion in all lines, exhibiting deletions of 0 to 4 bp in the T-DNA (with the exception of line Sc33-9, which had a 52-bp deletion) and 7- to 36-bp deletions or 1- to 12-bp insertions in the DSB site (Table III; Fig. 5). In Sc70-9, the 30-bp deletion in DSB was also accompanied by an insertion of a 12-bp DNA fragment that did not originate from the T-DNA sequences. Homologies between the integrated T-DNA right border and the DSB site ranged from 0 to 5 bp (Table III; Fig. 5).

Our sequence analysis of the RB/DSB-R integration junctions shown in Figure 5 also helped to address the likelihood of T-DNA integration in the vicinity of, rather than within, the I-*Sce*I site followed by subsequent digestion and religation of the resulting two I-*Sce*I sites, one within the plant DNA and the other near the T-DNA left border. In this scenario, which would give an appearance of I-SceI digestion before integration, the right border-host DNA junction is not expected to coincide with the original DSB. Our sequence data, however, did not detect such integration junctions, thus arguing against this possibility.

Figure 4. Nucleotide sequence of integration junctions between the I-*Sce*I-*uidA-hpt* T-DNA and DSB. Upper strands of the T-DNA and the DSB site are shown. The I-*Sce*I site is indicated in capital letters, and homologous base pairs are highlighted in gray.

DISCUSSION

T-DNA Molecules Preferentially Integrate into DSBs

Our findings indicate that DSBs in the host cell genome play an important role in T-DNA integration, representing the preferred sites for integration. This idea is based on the observations that, from a total number of 620 transgenic lines analyzed in this study, 16 lines contained T-DNA within the induced DSB site, corresponding to the 2.58% frequency of site-specific T-DNA integration. In reality, this number could be even higher because a sizable proportion of the 620 transgenic plants may not have contained DSBs, for example, due to insufficient levels of transient I-*Sce*I expression and/or activity. For comparison, the probability of random T-DNA integration into a specific single DSB within a plant genome is infinitesimally small. Based on the 4.5-Gb genome size of tobacco, the plant species used in our study, the calculated frequency of random T-DNA integration into a 18-bp-long site, corresponding to the I-*Sce*I recognition sequence, is 4×10^{-9}. Importantly, the 2.58% frequency was observed for integration into the specific DSB within the 35S-I-*Sce*I-*codA* transgene. Potentially, a much higher percentage of the 620

transgenic plants had derived from T-DNA integration into other DSBs, naturally occurring in the plant genome.

We also calculated the frequency of a DSB accepting a T-DNA insert instead of undergoing improper repair. From 63 lines identified to have a disrupted DSB, 16 lines verifiably contained T-DNA within their DSB site, corresponding to a 25% frequency. Previously, direct selection for a transgene insertion into DSBs was reported to result in 14% frequency of T-DNA integration in plants (Salomon and Puchta, 1998) and 18% frequency of insertion of a plasmid DNA in mammals (Pipiras et al., 1998). That x-ray irradiation, known to cause DSBs (Leskov et al., 2001), enhances transgene integration in plants (Kohler et al., 1989) lends additional support to the role of DSBs in T-DNA integration. We suggest that I-*Sce*I-induced DSBs mimic the naturally occurring DSBs. Such DSBs and/or reactions of DNA repair associated with DSBs may act as molecular "baits," attracting incoming molecules of foreign DNA, which then integrate into the break using the cellular DNA repair machinery. This notion helps to explain the well-known phenomenon of integration of multiple T-DNA molecules, even those originating from

Figure 5. Nucleotide sequence of integration junctions between the I-*Sce*I-digested I-*Sce*I-*uidA-hpt* T-DNA and DSB. Upper strands of the T-DNA and the DSB site are shown. The I-*Sce*I site is indicated in capital letters, and homologous base pairs are highlighted in gray.

different *A. tumefaciens* strains, into a single site in the host genome (De Block and Debrouwer, 1991; De Neve et al., 1997; Krizkova and Hrouda, 1998).

It remains unknown how T strands are targeted to their integration sites. One obvious candidate for this "pilot" function is VirD2, which is attached to the 5' end of the T strand (Ward and Barnes, 1988; Young and Nester, 1988; Howard et al., 1989) and has been proposed to recognize nicks and gaps in the plant DNA (Mayerhofer et al., 1991). This function of VirD2 in the integration process is supported by the

recent observations that VirD2 interacts with the plant transcription machinery in vivo (Bako et al., 2003). Another protein proposed to participate in targeting of the T strands to the host genome is VIP1, a plant protein that interacts with VirE2 and VirE2-single-stranded DNA complexes and facilitates nuclear import of VirE2 in the host cells (Tzfira et al., 2001). Because VIP1 contains a basic zipper motif, a hallmark of many transcription factors (van der Krol and Chua, 1991), it may also target VirE2 and its cognate T strand to chromosomal regions of transcrip-

Table III. *Sequence alterations and homology at the T-DNA integration sites*

Plant Line	Junction Type[a]	DSB Site Deletion/ Insertion (bp)	T-DNA Deletion/ Insertion (bp)	Homology[b] (bp)
Sc4-2	DSB-L/LB	−22	−13	12
Sc59-5	DSB-L/LB	0	−21	4
Sc70-3	DSB-L/LB	−30	−72	5
Sc4-2	RB/DSB-R	−7	0	2
Sc59-5	RB/DSB-R	+2	0	0
Sc70-3	RB/DSB-R	−10	−797	2
Sc59-9	DSB-L/RB	0	0	0
Sc59-9	LB/DSB-R	+2	−62	3
Sc1-10	DSB-L/RB	0	0	0
Sc33-4	DSB-L/RB	−4	0	0
Sc33-7	DSB-L/RB	−22	0	3
Sc1-10	LB/DSB-Rt	−2	−9	0
Sc33-4	LB/DSB-R	+2	−12	2
Sc33-7	LB/DSB-R	−26	0	6
Sc4-6	DSB-L/LB	−2	−29	1
Sc1-5	DSB-L/LB	−3	−4	2
Sc4-6	RB/DSB-R	−13	−18	5
Sc1-5	RB/DSB-R	−24	0	1
Sc4-5	DSB-L/I-*Sce*I	0	0	0
Sc4-8	DSB-L/I-*Sce*I	0	0	0
Sc1-3	DSB-L/I-*Sce*I	0	0	0
Sc59-2	DSB-L/I-*Sce*I	−10	−2	5
Sc33-9	DSB-L/I-*Sce*I	−2	−1	0
Sc4-5	RB/DSB-R	+2	0	0
Sc4-8	RB/DSB-R	−11	−1	0
Sc1-3	RB/DSB-R	−7	0	2
Sc59-2	RB/DSB-R	−36	0	5
Sc33-9	RB/DSB-R	+1	−52	3
Sc70-9	DSB-L/RB	−30	0	1
Sc4-12	DSB-L/RB	−17	−4	2
Sc70-9	I-*Sce*I/DSB-R	+3	−7	0
Sc4-12	I-*Sce*I/DSB-R	−4	−17	3

[a] Classification of junction types is explained in text. [b] Homology was determined within 10 bp from the integration point.

tional regulation, where the host DNA is more exposed and, thus, better suitable for T-DNA integration (Tzfira et al., 2001; Tzfira and Citovsky, 2002). In fact, a recent analysis of more than 80,000 T-DNA insertion site sequences indicated that the density of T-DNA insertion events closely parallels gene density along the Arabidopsis chromosomes, further supporting the idea that T-DNA molecules are targeted to "hot spots" of transcription (Alonso et al., 2003).

If the invading T strands are targeted to DSBs for conversion to double strands and integration, it is the frequency of DSBs in the host cell genome and the genome stability, therefore, that may represent a limiting factor for the efficiency of plant genetic transformation by *A. tumefaciens*. A statistical analysis of T-DNA integration sites in Arabidopsis revealed that T-DNA molecules mainly integrate in the vicinity of T-rich genomic regions (Brunaud et al., 2002), and several reports demonstrated that, in various plant species, genomic sequences flanking T-DNA insertions are AT rich (Gheysen et al., 1987; Takano et al., 1997; Kumar and Fladung, 2001). Such AT-rich regions, characterized by low DNA duplex stability (Breslauer et al., 1986) and strong bending (Bolshoy et al., 1991; Muller et al., 1999), are prefect candidates for naturally occurring DSBs. Also, DSBs are likely to occur during increased DNA metabolism in regenerating cells within plant wounds, contributing to the well-known susceptibility of wounded tissues to *A. tumefaciens* infection (Kahl, 1982).

Microhomology between the T strand and the plant DNA plays a critical role in the single-stranded T-DNA integration model (Tinland and Hohn, 1995; Tinland et al., 1995). However, our sequence analysis of both T-DNA right and left border integration junctions did not detect any consistent patterns of such homology; moreover, in nearly one-half of all analyzed junctions, no homology was observed at all. On

the other hand, previous studies indicated that even an extensive 500-bp homology between T-DNA and the host genome did not elicit homologous recombination-like events during T-DNA integration into DSBs (Puchta, 1998). Thus, microhomology may play only a minor role, if any, during T-DNA integration into DSBs; instead, this process most likely resembles a ligation reaction between two double-stranded DNA molecules, probably by a non-homologous end joining (NHEJ) mechanism. This notion is consistent with the recently reported involvement of the NHEJ machinery in T-DNA integration in yeast cells (van Attikum et al., 2001; van Attikum and Hooykaas, 2003). It remains unclear, however, which components of the NHEJ machinery may be involved in the T-DNA integration in plants; for example, whereas one study suggested that the DNA ligase IV of Arabidopsis is not required for this process (van Attikum et al., 2003), another study indicated a role for both Arabidopsis ligase IV and Ku80 in the T-DNA integration (Friesner and Britt, 2003).

T Strands Are Converted to a Double-Stranded Form before Integration into DSB

T strands corresponding to the coding strand of a reporter gene are rapidly expressed, which requires their conversion to a double-stranded form already early in the infection process (Narasimhulu et al., 1996). Here, we present the first direct evidence, to our knowledge, that such double-stranded forms of T-DNA, previously suggested to represent a potential dead end in the integration process (Gelvin, 1998b, 2000), are in fact the essential intermediates of T-DNA integration into DSBs. We showed that nearly one-half of all integration events within DSBs involved insertion of the I-SceI-digested T-DNA. Because this restriction endonuclease cleaves only double-stranded DNA (Monteilhet et al., 1990; Jasin, 1996), the T strands must have become double stranded before their cleavage. Moreover, that the I-SceI-cleaved T-DNA molecules integrated precisely into the I-SceI-cleaved DSB indicates that they could not have been digested after integration because, in such a case, the I-SceI site within DSB would no longer have been present. Also, the possibility that the T-DNA first integrated in the vicinity of DSB and only later became double-stranded, digested, and re-ligated to the I-SceI-cleaved DSB was ruled out because the sequences of the right border-host DNA junctions indicated that the integration occurred specifically within DSB itself. Thus, conversion of the T strand to a double-stranded molecule may be a pre-requisite for its integration into DSBs.

Formation of double-stranded intermediates before integration may also explain the well-known conservation of the 5′ end sequences of the integrated T-DNA and frequent deletions of its 3′ end sequences

(Matsumoto et al., 1990; Gheysen et al., 1991; Kumar and Fladung, 2002). The second strand synthesis must begin from the 3′ end of the T-strand template, most probably by random priming. Such priming may not occur precisely at the 3′ end, allowing completion of the synthesis to faithfully replicate the 5′ end of the T strand, but leaving the 3′ end uncomplemented and, thus, susceptible to removal by exonucleases. In an alternative pathway, the 3′ end of the T strand may be ligated to the 5′ end of one side of DSB, but the 5′ end of the T strand may still remain free. This partially integrated T strand may then be converted to a double-stranded form using the free 3′ end of the same DSB side as primer. Finally, the resulting the double-stranded T-DNA molecule is digested by I-SceI and ligated to the second side of DSB.

It is important to emphasize that our data indicate the importance of conversion of T strands into double-stranded molecules before site-specific integration into DSBs. Although integration via double-stranded intermediates into DSBs most likely represents one of the major mechanisms of T-DNA integration, an alternative integration pathway via single-stranded intermediates into microhomology regions (Brunaud et al., 2002) is certainly possible.

MATERIALS AND METHODS

DNA Constructs

Binary vectors pBNE3I and pCISceI, carrying within their T-DNA regions the cytosine deaminase (codA) and I-SceI genes, respectively, each driven by the cauliflower mosaic virus 35S promoter, were kindly provided by Dr. Holger Puchta (Puchta, 1998). The T-DNA region of pBNE3I also contained an I-SceI recognition site between the 35S promoter and codA (Fig. 1A) and carried the kanamycin resistance gene as selectable marker in transgenic plants. The T-DNA region of pCISceI, on the other hand, contained no I-SceI sites and no plant selection marker. To construct pBIG-HYG-terGUS/I-SceI, the 35S-uidA cassette of pBIG-HYG-GUS (Tzfira et al., 1997), was removed by EcoRI-HindIII digestion and replaced with PCR-amplified promoterless uidA-NOS-terminator fragment from pBI121 (Jefferson, 1987; Jefferson et al., 1987; Fig. 1B). To place this fragment in sense orientation to the T-DNA left border, the orientation of the EcoRI and HindIII sites flanking the uidA-NOS-terminator cassette were reversed using the forward 5′CGGAATTCATTACCCTGTTATCCCTAAT-GTTACGTCCTGTAGAAACCCA3′ and reverse 5′CCCAAGCTTCT-TCTAGTAACATAGATGACACC3′ primers. In addition, a new I-SceI site was incorporated between uidA and the EcoRI site in the forward primer (Fig. 1B). The pBIG-HYG-terGUS/I-SceI T-DNA also contained the hpt (hygromycin resistance) gene as a selectable marker in transgenic plants. All binary plasmids were introduced into Agrobacterium tumefaciens strain EHA105 using the calcium chloride transformation method (Tzfira et al., 1997).

Transgenic Plants

Transgenic tobacco (Nicotiana tabacum cv Turk) plants carrying the I-SceI site were generated using pBNE3I and the leaf disc transformation protocol as previously described (Horsch et al., 1985), selected on 100 μg mL⁻¹ kanamycin, and maintained in tissue culture on basal Murashige and Skoog medium (Murashige and Skoog, 1962) with no exogenous growth regulators. Plants were then transferred to soil in a greenhouse, allowed to set seed, and the transgenic progeny was selected by germinating the seeds on Murashige and Skoog agar in the presence of kanamycin. Transgenic lines showing 3:1 segregation ratio of the kanamycin resistance gene were ana-

lyzed by PCR for the presence of 35S-I-*Sce*I-*codA* as described below to identify plants without tandem T-DNA repeats. The resulting plant lines were propagated vegetatively to maintain their heterozygosity for the single 35S-I-*Sce*I-*codA* cassette; having only one copy of this DSB site helped to reduce the complexity of subsequent integration events and facilitate interpretation of the resulting data.

To identify transgenic lines with high cytosine deaminase activity, leaf explants from the 35S-I-*Sce*I-*codA* plants were placed on regeneration medium (Horsch et al., 1985) with or without 250 μg mL^{-1} 5-FC, which is toxic to cells expressing cytosine deaminase (Stougaard, 1993). Transgenic lines with high sensitivity to 5-FC and, by implication, high levels of cytosine deaminase activity (lines Sc1, Sc4, Sc33, Sc59 and Sc70) were transformed, using the leaf disc transformation protocol (Horsch et al., 1985), with a mixture of *A. tumefaciens* strains carrying the pCI*Sce*I and pBIG-HYG-terGUS/I-*Sce*I vectors. For efficient DSB induction, a ratio of 3:1 between pCI*Sce*I and pBIG-HYG-terGUS/I-*Sce*I was used in this second transformation cycle. To select for pBIG-HYG-terGUS/I-*Sce*I integration, transgenic plants were grown in the presence of 25 μg mL^{-1} hygromycin. A total of 620 hygromycin-resistant transgenic plants were obtained and maintained in tissue culture on basal Murashige and Skoog medium (Murashige and Skoog, 1962) without exogenous growth regulators.

Hygromycin-resistant transgenic lines with disrupted cytosine deaminase activity were identified by culturing their leaf explants in the presence or absence of 250 μg mL^{-1} 5-FC as described above. To identify transgenic plants with targeted integration of the functional I-*Sce*I-*uidA*-*hpt* T-DNA, leaf explants from lines resistant both to hygromycin and 5-FC were assayed for GUS activity by histochemical staining with chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-glucuronic acid as described by Jefferson (1987).

PCR and DNA Sequence Analyses of Integration Sites

Genomic DNA was extracted from transgenic leaf tissues using the DNeasy Plant Mini DNA extraction kit (Qiagen USA, Valencia, CA) and analyzed by PCR using primer sets shown in Table I. To amplify the 35S-I-*Sce*I-*codA* insertion for identification of transgenic plants from the first transformation cycle and to analyze hygromycin-resistant transgenic plants with disrupted cytosine deaminase activity from the second transformation cycle, we used the 35S-F2 and *codA*-R primers (Table I); amplification of undisrupted 35S-I-*Sce*I-*codA* was predicted to produce a 509-kb DNA fragment (Table I). To amplify the right border integration junction of pBIG-HYG-terGUS/I-*Sce*I T-DNA in the I-*Sce*I cleavage site, the primer T-DNA-RB located 431 bp downstream of the T-DNA right border of pBIG-HYG-terGUS/I-*Sce*I was used in combination with either the 35S-F2 or *codA*-R primers specific for the 35S promoter or the *codA* gene, respectively (Table I). To amplify the left border integration junction of pBIG-HYG-terGUS/I-*Sce*I T-DNA in the I-*Sce*I cleavage site, the primer T-DNA-LB, located 568 bp upstream of the T-DNA left border of pBIG-HYG-terGUS/I-*Sce*I, was used in combination with either 35S-F2 or *codA*-R (Table I). To amplify the integration junction of I-*Sce*I-digested T-DNA of pBIG-HYG-terGUS/I-*Sce*I in the I-*Sce*I cleavage site, the primer GUS-R located 310 bp upstream of the first codon of *uidA* in the T-DNA of pBIG-HYG-terGUS/I-*Sce*I was used in combination of either 35S-F2 or *codA*-R; these combinations of primers were predicted to produce 350-bp and approximately 672-bp PCR fragments, respectively (Table I).

All PCR reactions were carried out in a 50-μL volume containing 20 ng of DNA, 0.1 nM dNTP, 2.5 mM of each primer, and 2 units of TKARA EX-*Taq* polymerase (Pan Vera Corp., Madison, WI) with 5 μL of EX-*Taq* 10× reaction buffer. After 5 min of denaturation at 94°C, the junction fragment was amplified for 36 cycles of 60 s each at 94°C, 60 s at 55°C, and 1 or 3 min at 72°C. The resulting amplified junction fragments were either directly sequenced using the corresponding primers or first subcloned into pSL301 and then sequenced.

Distribution of Materials

Upon request, all novel materials described in this publication will be made available in a timely manner for noncommercial research purposes, subject to the requisite permission from any third party owners of all or parts of the material. Obtaining any permissions will be the responsibility of the requestor.

LITERATURE CITED

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301**: 653–657

Bako L, Umeda M, Tiburcio AF, Schell J, Koncz C (2003) The VirD2 pilot protein of *Agrobacterium*-transferred DNA interacts with the TATA box-binding protein and a nuclear protein kinase in plants. Proc Natl Acad Sci USA **100**: 10108–10113

Bolshoy A, McNamara P, Harrington RE, Trifonov EN (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. Proc Natl Acad Sci USA **88**: 2312–2316

Breslauer KJ, Frank R, Blocker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci USA **83**: 3746–3750

Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S, Bechtold N, Cruaud C, DeRose R, Pelletier G et al. (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. EMBO Rep **3**: 1152–1157

Bundock P, Hooykaas PJJ (1996) Integration of *Agrobacterium tumefaciens* T-DNA in the *Saccharomyces cerevisiae* genome by illegitimate recombination. Proc Natl Acad Sci USA **93**: 15272–15275

Christie PJ (1997) *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. J Bacteriol **179**: 3085–3094

Christie PJ, Vogel JP (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. Trends Microbiol **8**: 354–360

Christie PJ, Ward JE, Winans SC, Nester EW (1988) The *Agrobacterium tumefaciens virE2* gene product is a single-stranded-DNA-binding protein that associates with T-DNA. J Bacteriol **170**: 2659–2667

Citovsky V, De Vos G, Zambryski PC (1988) Single-stranded DNA binding protein encoded by the *virE* locus of *Agrobacterium tumefaciens*. Science **240**: 501–504

Citovsky V, Guralnick B, Simon MN, Wall JS (1997) The molecular structure of *Agrobacterium* VirE2-single stranded DNA complexes involved in nuclear import. J Mol Biol **271**: 718–727

Citovsky V, Warnick D, Zambryski PC (1994) Nuclear import of *Agrobacterium* VirD2 and VirE2 proteins in maize and tobacco. Proc Natl Acad Sci USA **91**: 3210–3214

Citovsky V, Wong ML, Zambryski PC (1989) Cooperative interaction of *Agrobacterium* VirE2 protein with single stranded DNA: implications for the T-DNA transfer process. Proc Natl Acad Sci USA **86**: 1193–1197

Citovsky V, Zupan J, Warnick D, Zambryski PC (1992) Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. Science **256**: 1802–1805

Das A (1988) *Agrobacterium tumefaciens virE* operon encodes a single-stranded DNA-binding protein. Proc Natl Acad Sci USA **85**: 2909–2913

Das A (1998) DNA transfer from *Agrobacterium* to plant cells in crown gall tumor disease. Subcell Biochem **29**: 343–363

De Block M, Debrouwer D (1991) Two T-DNAs co-transformed into *Brassica napus* by a double *Agrobacterium tumefaciens* infection are mainly integrated at the same locus. Theor Appl Genet **82**: 257–263

De Buck S, Jacobs A, Van Montagu M, Depicker A (1999) The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. Plant J **20**: 295–304

De Neve M, De Buck S, Jacobs A, Van Montagu M, Depicker A (1997) T-DNA integration patterns in co-transformed plant cells suggest that T-DNA repeats originate from co-integration of separate T-DNAs. Plant J **11**: 15–29

Filichkin SA, Gelvin SB (1993) Formation of a putative relaxation intermediate during T-DNA processing directed by the *Agrobacterium tumefaciens* VirD1, D2 endonuclease. Mol Microbiol **8**: 915–926

Friesner J, Britt AB (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. Plant J **34**: 427–440

Gaudin V, Vrain T, Jouanin L (1994) Bacterial genes modifying hormonal balances in plants. Plant Physiol Biochem **32:** 11–29

Gelvin SB (1998a) *Agrobacterium* VirE2 proteins can form a complex with T strands in the plant cytoplasm. J Bacteriol **180:** 4300–4302

Gelvin SB (1998b) The introduction and expression of transgenes in plants. Curr Opin Biotechnol **9:** 227–232

Gelvin SB (2000) *Agrobacterium* and plant genes involved in T-DNA transfer and integration. Annu Rev Plant Physiol Plant Mol Biol **51:** 223–256

Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev **67:** 16–37

Gheysen G, Van Montagu M, Zambryski P (1987) Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA. Proc Natl Acad Sci USA **84:** 6169–6173

Gheysen G, Villarroel R, Van Montagu M (1991) Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev **5:** 287–297

Gietl C, Koukolikova-Nicola Z, Hohn B (1987) Mobilization of T-DNA from *Agrobacterium* to plant cells involves a protein that binds single-stranded DNA. Proc Natl Acad Sci USA **84:** 9006–9010

Horsch RB, Fry JE, Hoffman NL, Eichholtz D, Rogers SG, Fraley RT (1985) A simple and general method for transferring genes into plants. Science **227:** 1229–1231

Howard E, Zupan J, Citovsky V, Zambryski PC (1992) The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. Cell **68:** 109–118

Howard EA, Citovsky V, Zambryski PC (1990) The T-complex of *Agrobacterium tumefaciens*. UCLA Symp Mol Cell Biol **129:** 1–11

Howard EA, Winsor BA, De Vos G, Zambryski PC (1989) Activation of the T-DNA transfer process in *Agrobacterium* results in the generation of a T-strand protein complex: tight association of VirD2 with the 5′ ends of T-strands. Proc Natl Acad Sci USA **86:** 4017–4021

Janssen BJ, Gardner RC (1990) Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. Plant Mol Biol **14:** 61–72

Jasin M (1996) Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet **12:** 224–228

Jasper F, Koncz C, Schell J, Steinbiss H-H (1994) *Agrobacterium* T-strand production *in vitro*: sequence-specific cleavage and 5′ protection of single-stranded DNA templates by purified VirD2 protein. Proc Natl Acad Sci USA **91:** 694–698

Jefferson RA (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep **5:** 387–405

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J **6:** 3901–3907

Kahl G (1982) Molecular biology of wound healing: the conditioning phenomenon. *In* G Kahl, J Schell, eds, Molecular Biology of Plant Tumors. Academic Press, New York, pp 211–268

Kohler F, Cardon G, Pohlman M, Gill R, Schider O (1989) Enhancement of transformation rates in higher plants by low-does irradiation: are DNA repair systems involved in incorporation of exogenous DNA into the plant genome? Plant Mol. Biol **12:** 189–199

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

Kumar S, Fladung M (2001) Gene stability in transgenic aspen (*Populus*): II. Molecular characterization of variable expression of transgene in wild and hybrid aspen. Planta **213:** 731–740

Kumar S, Fladung M (2002) Transgene integratin in aspen: structures of integration sites and mechanism of T-DNA integration. Plant J **31:** 543–551

Leskov KS, Criswell T, Antonio S, Li J, Yang CR, Kinsella TJ, Boothman DA (2001) When X-ray-inducible proteins meet DNA double strand break repair. Semin Radiat Oncol **11:** 352–372

Matsumoto M, Ito Y, Hosoi T, Takahashi Y, Machida Y (1990) Integration of *Agrobacterium* T-DNA into a tobacco chromosome: possible involvement of DNA homology between T-DNA and plant DNA. Mol Gen Genet **224:** 309–316

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

McCormac AC, Fowler MR, Chen DF, Elliott MC (2001) Efficient co-transformation of *Nicotiana tabacum* by two independent T-DNAs, the effect of T-DNA size and implications for genetic separation. Transgenic Res **10:** 143–155

Monteilhet C, Perrin A, Thierry A, Colleaux L, Dujon B (1990) Purification and characterization of the *in vitro* activity of I-*Sce* I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Res **18:** 1407–1413

Muller AE, Kamisugi Y, Gruneberg R, Niedenhof I, Horold RJ, Meyer P (1999) Palindromic sequences and A+T-rich DNA elements promote illegitimate recombination in *Nicotiana tabacum*. J Mol Biol **291:** 29–46

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant **15:** 473–497

Narasimhulu SB, Deng X-B, Sarria R, Gelvin SB (1996) Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. Plant Cell **8:** 873–886

Pansegrau W, Schoumacher F, Hohn B, Lanka E (1993) Site-specific cleavage and joining of single-stranded DNA by VirD2 protein of *Agrobacterium tumefaciens* Ti plasmids: analogy to bacterial conjugation. Proc Natl Acad Sci USA **90:** 11538–11542

Pipiras E, Coquelle A, Bieth A, Debatisse M (1998) Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. EMBO J **17:** 325–333

Potrykus I, Bilang R, Futterer J, Sautter C, Schrott M, Spangenberg G (1998) Genetic Engineering of Crop Plants, Ed 1. Marcel Dekker, New York

Puchta H (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. Plant J **13:** 77–78

Rossi L, Hohn B, Tinland B (1993) The VirD2 protein of *Agrobacterium tumefaciens* carries nuclear localization signals important for transfer of T-DNA to plant. Mol Gen Genet **239:** 345–353

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J **17:** 6086–6095

Savka MA, Dessaux Y, Oger P, Rossbach S (2002) Engineering bacterial competitiveness and persistence in the phytosphere. Mol Plant-Microbe Interact **15:** 866–874

Scheiffele P, Pansegrau W, Lanka E (1995) Initiation of *Agrobacterium tumefaciens* T-DNA processing: purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of superhelical T-border DNA *in vitro*. J Biol Chem **270:** 1269–1276

Sen P, Pazour GJ, Anderson D, Das A (1989) Cooperative binding of *Agrobacterium tumefaciens* VirE2 protein to single-stranded DNA. J Bacteriol **171:** 2573–2580

Sheng J, Citovsky V (1996) *Agrobacterium*-plant cell interaction: have virulence proteins, will travel. Plant Cell **8:** 1699–1710

Stachel SE, Zambryski PC (1989) Generic trans-kingdom sex? Nature **340:** 190–191

Stougaard J (1993) Substrate-dependent negative selection in plants using a bacterial cytosine deaminase gene. Plant J **3:** 755–761

Takano M, Egawa H, Ikeda JE, Wakasa K (1997) The structures of integration sites in transgenic rice. Plant J **11:** 353–361

Tinland B (1996) The integration of T-DNA into plant genomes. Trends Plant Sci **1:** 178–184

Tinland B, Hohn B (1995) Recombination between prokaryotic and eukaryotic DNA: integration of *Agrobacterium tumefaciens* T-DNA into the plant genome. Genet Eng **17:** 209–229

Tinland B, Hohn B, Puchta H (1994) *Agrobacterium tumefaciens* transfers single-stranded transferred DNA (T-DNA) into the plant cell nucleus. Proc Natl Acad Sci USA **91:** 8000–8004

Tinland B, Schoumacher F, Gloeckler V, Bravo-Angel AM, Hohn B (1995) The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. EMBO J **14:** 3585–3595

Tzfira T, Citovsky V (2000) From host recognition to T-DNA integration: the function of bacterial and plant genes in the *Agrobacterium*-plant cell interaction. Mol Plant Pathol **1:** 201–212

Tzfira T, Citovsky V (2002) Partners-in-infection: host proteins involved in the transformation of plant cells by *Agrobacterium*. Trends Cell Biol **12:** 121–129

Tzfira T, Jensen CS, Wangxia W, Zuker A, Altman A, Vainstein A (1997) Transgenic *Populus*: a step-by-step protocol for its *Agrobacterium*-mediated transformation. Plant Mol Biol Rep **15:** 219–235

Tzfira T, Rhee Y, Chen M-H, Citovsky V (2000) Nucleic acid transport in plant-microbe interactions: the molecules that walk through the walls. Annu Rev Microbiol **54:** 187–219

Tzfira T, Vaidya M, Citovsky V (2001) VIP1, an *Arabidopsis* protein that interacts with *Agrobacterium* VirE2, is involved in VirE2 nuclear import and *Agrobacterium* infectivity. EMBO J **20:** 3596–3607

van Attikum H, Bundock P, Hooykaas PJJ (2001) Non-homologous end-joining proteins are required for *Agrobacterium* T-DNA integration. EMBO J **20:** 6550–6558

van Attikum H, Bundock P, Overmeer RM, Lee LY, Gelvin SB, Hooykaas PJ (2003) The *Arabidopsis AtLIG4* gene is required for the repair of DNA damage, but not for the integration of *Agrobacterium* T-DNA. Nucleic Acids Res **31:** 4247–4255

van Attikum H, Hooykaas PJJ (2003) Genetic requirements for the targeted integration of *Agrobacterium* T-DNA in *Saccharomyces cerevisiae*. Nucleic Acids Res **31:** 826–832

van der Krol AR, Chua N-H (1991) The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. Plant Cell **3:** 667–675

Vergunst AC, Schrammeijer B, den Dulk-Ras A, de Vlaam CMT, Regensburg-Tuink TJ, Hooykaas PJJ (2000) VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. Science **290:** 979–982

Wang K, Stachel SE, Timmerman B, Van Montagu M, Zambryski PC (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

Ward E, Barnes W (1988) VirD2 protein of *Agrobacterium tumefaciens* very tightly linked to the 5′ end of T-strand DNA. Science **242:** 927–930

Winans SC, Mantis NJ, Chen CY, Chang CH, Han DC (1994) Host recognition by the VirA, VirG two-component regulatory proteins of *Agrobacterium tumefaciens*. Res Microbiol **145:** 461–473

Young C, Nester EW (1988) Association of the VirD2 protein with the 5′ end of T-strands in *Agrobacterium tumefaciens*. J Bacteriol **170:** 3367–3374

Yusibov VM, Steck TR, Gupta V, Gelvin SB (1994) Association of single-stranded transferred DNA from *Agrobacterium tumefaciens* with tobacco cells. Proc Natl Acad Sci USA **91:** 2994–2998

Zambryski PC, Depicker A, Kruger K, Goodman HM (1982) Tumor induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. J Mol Appl Genet **1:** 361–370

Ziemienowicz A, Merkle T, Schoumacher F, Hohn B, Rossi L (2001) Import of *Agrobacterium* T-DNA into plant nuclei: Two distinct functions of VirD2 and VirE2 proteins. Plant Cell **13:** 369–384

Ziemienowicz A, Tinland B, Bryant J, Gloeckler V, Hohn B (2000) Plant enzymes but not *Agrobacterium* VirD2 mediate T-DNA ligation *in vitro*. Mol Cell Biol **20:** 6317–6322

Zupan J, Muth TR, Draper O, Zambryski PC (2000) The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. Plant J **23:** 11–28

Zupan J, Zambryski PC (1997) The *Agrobacterium* DNA transfer complex. Crit Rev Plant Sci **16:** 279–295