Plant Molecular Biology **52:** 761–773, 2003.
© 2003 Kluwer Academic Publishers. Printed in the Netherlands.

Transgene structures in T-DNA-inserted rice plants

Sung-Ryul Kim[1,2], Jinwon Lee[1], Sung-Hoon Jun[1], Sunhee Park[1], Hong-Gyu Kang[1], Soontae Kwon[2] and Gynheung An[1,*]
[1]*National Research Laboratory of Plant Functional Genomics, Division of Molecular and Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Republic of Korea (*author for correspondence; email genean@postech.ac.kr);* [2]*Department of Horticulture and Breeding, Andong National University, Andong 760-749, Republic of Korea*

Received 19 December 2002; accepted in revised form 5 March 2003

Key words: border, flanking sequence, *Oryza sativa*, repeat, rice, T-DNA integration, vector backbone

Abstract

T-DNA is commonly used for delivery of foreign genes and as an insertional mutagen. Although ample information exists regarding T-DNA organization in dicotyledonous plants, little is known about the monocot rice. Here, we investigated the structure of T-DNA in a large number of transgenic rice plants. Analysis of the T-DNA borders revealed that more than half of the right ends were at the cleavage site, whereas the left ends were not conserved and were deleted up to 180 bp from the left border (LB) cleavage site. Three types of junctions were found between T-DNA and genomic DNA. In the first, up to seven nucleotide overlaps were present. The frequency of this type was much higher in the LB region than at the right border (RB). In the second type, which was more frequent in RB, the link was direct, without any overlaps or filler DNA. Finally, the third type showed filler DNA between T-DNA and the plant sequences. Out of 171 samples examined, 77 carried the vector backbone sequence, with the majority caused by the failure of T-strand termination at LB. However, a significant portion also resulted from co-integration of T-DNA and the vector backbone to a single locus. Most linkages between T-DNA and the vector backbone were formed between two 3′ ends or two 5′ ends of the transferred DNAs. The 3′ ends were mostly linked through 3–6 bp of the complementing sequence, whereas the 5′ ends were linked through either precise junctions or imprecise junctions with filler DNA.

Abbreviations: 3 SD/SA, triple splice donor and acceptor sequences; iPCR, inverse PCR; LB, left border; RB, right border; T-DNA, transferred DNA; Ti, tumor-inducing; T-strand, transferred DNA strand

Introduction

The phytopathogen *Agrobacterium* can induce a tumor to host wound sites by transferring a segment of DNA (T-DNA) from a Ti (tumor-inducing) plasmid into the plant's genome (Nester and Kosuge, 1981; Otten *et al.*, 1992; Galbiati *et al.*, 2000; Gelvin, 2000). In nature, most dicotyledonous species and some monocots are infected by *Agrobacterium* (De Cleene and de Ley, 1976; De Cleene, 1985; Bytebier *et al.*, 1987). Ti plasmids have now been modified to serve as vectors for introducing foreign DNA into plant chromosomes (Hoekema *et al.*, 1983; Zambryski *et al.*, 1983; Bevan, 1984; An *et al.*, 1985). Such vectors have been successfully used for DNA transfer into a large number of dicotyledonous species (Birch, 1997; Binns, 2002) as well as some monocotyledonous plants, such as asparagus (Bytebier *et al.*, 1987), maize (Ishida *et al.*, 1996), barley (Tingay *et al.*, 1997), and rice (Hiei *et al.*, 1994; Jeon *et al.*, 2000).

Sheng and Citovsky (1996) and Zupan *et al.* (2000) have studied several dicotyledonous species and have established that T-DNA transfer is mediated largely by the virulence proteins encoded by *Agrobacterium*. The *virulence* (*vir*) genes are located on Ti plasmids; their expression is inducible by plant phenolic compounds

(Stachel *et al.*, 1985). The VirD1 and VirD2 proteins recognize the border repeat sequences located at the ends of the T-DNA, and produce a single-stranded nick between the third and fourth bases on the bottom strand of each repeat (Yanofsky *et al.*, 1986). These nicks determine the initiation and termination sites of the T-strand at the right border (RB) and left border (LB), respectively. The 25 bp border sequence is directional in its mode of action (Wang *et al.*, 1984; Miranda *et al.*, 1992). VirD2 protein covalently binds to the 5′ end of the T-strand and guides the single-stranded DNA into the plant nucleus (Sheng and Citovsky, 1996; Zupan *et al.*, 2000).

It has been generally postulated that T-DNA is integrated by illegitimate recombination. The 3′ end of single-stranded T-DNA lands at a single-stranded region of genomic DNA via sequence homology-dependent annealing (Gheysen *et al.*, 1991; Mayerhofer *et al.*, 1991; Tinland, 1996). Recently, a synthesis-dependent strand-annealing mechanism has been proposed. In this model, a short double-strand region is made from single-stranded T-DNA before being ligated to a genomic double-strand break (Puchta, 1998; Salomon and Puchta, 1998; Kumar and Fladung, 2002).

In dicotyledonous plants, T-DNA integrates either at one locus or at several independent loci. In addition, multiple T-DNA copies frequently are formed at one locus, in either direct or inverted repeats (Jorgensen *et al.*, 1987; De Neve *et al.*, 1997; Krizkova and Hrouda, 1998; De Buck *et al.*, 1999; Kumar and Fladung, 2000). The T-DNA boundary is not always clearly defined in transgenic plants. Ooms *et al.* (1982) and Ursic *et al.* (1983) have observed that the sequence beyond the left T-DNA border is often present in crown gall tissues. Vector backbone sequences that reside outside LB and RB have been found in dicot plants (Cluster *et al.*, 1996; Kononov *et al.*, 1997; Wenck *et al.*, 1997; Wolters *et al.*, 1998, De Buck *et al.*, 2000). The proportion of plants with such backbone DNA varies between 15% and 75%. Kononov *et al.* (1997) have suggested that the vector backbone sequence can be integrated into the plant genome independent of T-DNA action, and Durrenberger *et al.* (1989) have observed that VirD2 can be linked not only to the 5′ end of the bottom-strand nick at RB but also to the 5′ end of the bottom-strand nick at LB. In fact, LB can serve as an initiation site for DNA transfer, generating vector backbone single-stranded DNA (Ramanathan and Veluthambi, 1995; van der Graaff *et al.*, 1996).

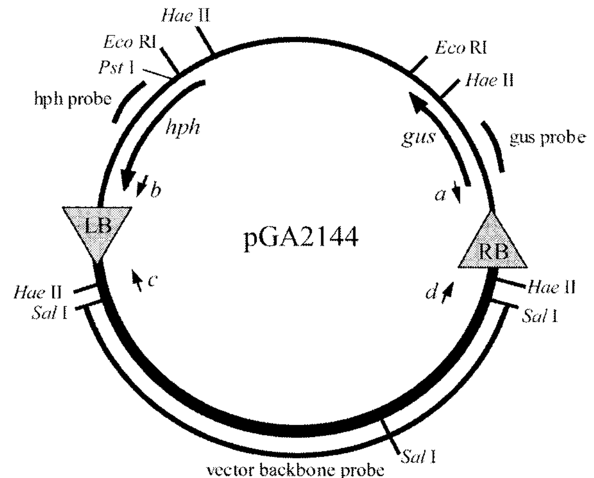


Figure 1. Map of binary vector pGA2144. LB and RB represent the left border and right border, respectively. The upper portion of the map between LB and RB is T-DNA; the lower portion, the vector backbone. T-DNA contains the promoter-less *gus* coding region followed by the transcription terminator of the nopaline synthase gene. T-DNA also carries the *hph* gene that consists of the *OsTubA1* promoter, the *hph* coding region, and the transcription termination region of Gene 7 of pTiA6. The *a*, *b*, *c*, and *d* represent the PCR primers adjacent (400–700 bp) to LB and RB. Arrows indicate direction of primers from the 5′ to the 3′ end. Lines above vector circle represent hybridization probes used in Figure 3 and 5. Restriction enzyme sites are also indicated above the vector.

Although T-DNA organization has been studied in detail in dicotyledonous plants, little information is available about its structure in monocot species. In this study, we investigated the process of T-DNA integration in transgenic rice plants.

Materials and methods

Plant materials

We used plants of transgenic rice (*Oryza sativa* L. cv. Dongjin) that had been transformed with binary Ti vector pGA2144 (Figure 1; Jeon *et al.*, 2000). Within the T-DNA, the hygromycin phosphotransferase (*hph*) gene used for selecting transformants is located next to the LB of T-DNA, and the promoterless β-glucuronidase (*gus*) gene is next to RB. Triple splice donor and acceptor sequences (3 SD/SA) are present between RB and *gus*. The vector is designed to generate insertional mutants in rice. When integrated into an active gene, the 3 SD/SA enhances the opportunity for transcriptional fusion between an endogenous gene and the *gus* reporter (Jeon *et al.*, 2000). Our vector backbone sequence was derived from pTJS75 (An

et al., 1985). The transgenic plants were generated from scutellum-derived calluses after co-cultivation with *Agrobacterium tumefaciens* LBA4404 that carried the binary vector, as previously described (Lee *et al.*, 1999).

Genomic DNA isolation and DNA gel-blot analysis

Genomic DNA was isolated from the leaf material of transgenic plants according to the modified cetyltrimethylammonium bromide method (Chen and Ronald, 1999). To analyze the primary plants, mature leaves were harvested before heading occurred. For T_2 transgenic plants, 4–10 seedlings were grown for 2–3 weeks, and their leaves were then pooled. A 2 μg portion of genomic DNA was digested with *Eco*RI, *Hae*II, or *Sal*I, separated on 0.7% agarose gel, blotted onto a nylon membrane, and hybridized with a ^{32}P-labeled probe. The *gus* and *hph* probes were obtained via PCR amplification, while the vector backbone probe was prepared by isolating two *Sal*I fragments (5 and 3 kb) from pTJS75.

PCR analysis and flanking-sequence isolation

Junction sequences between T-DNA and the vector backbone were PCR-amplified with four primers located near the borders (Figure 1). The sequences were 5′-AATATCTGCATCGGCGAACT-3′ (*a*), 5′-AGAGCCTGACCTATTGCATC-3′ (*b*), 5′-GACGCCGTTGGATACACCAA-3′ (*c*), and 5′-TC-AGTGAGGGCCAAGTTTTC-3′ (*d*). Sequences between the direct T-DNA repeats were amplified by PCR with primers *a* and *b*. Genomic DNA sequences flanking the T-DNA or vector backbone were isolated by inverse PCR (iPCR; Ochman *et al.*, 1988; Triglia *et al.*, 1988). A 1 μg portion of genomic DNA was digested with *Pst*I or *Hae*II, self-ligated with T4 DNA ligase, then amplified by nested PCR with the following primers: (1) for isolating genomic sequences flanking the RB of T-DNA, the primers were 5′-CCACAGTTTTCGCGATCCAGACTG-3′ and 5′-CCATGTAGTGTATTGACCGATTC-3′, and the nested primers were 5′-GGGTTGGGGTTTCTACA-GGACGT-3′ and 5′-TCGTCTGGCTAAGATCGG-CCGCA-3′; (2) for isolating the DNA sequences flanking the LB of T-DNA, the primers were 5′-GGTGAATGGCATCGTTTGAA-3′ and 5′-GATCG-TTATGTTTATCGGCACTT-3′, and the nested primers were 5′-ACAAGCCGTAAGTGCAAGTG-3′ and 5′-AGTGCTTGACATTGGGGAATTCAG-3′; and (3) for isolating the genomic sequences flanking the LB of the vector backbone, the primers were 5′-CATCGGTAACATGAGCAAAG-3′ and 5′-CGATC TTGAGAACTATGCCGA-3′, and the nested primers were 5′-GCCTGTATCGAGTGGTGATT-3′ and 5′-AGGAAATCGCTGGATAAAGC-3′. The 20 μl iPCR-reaction solution contained 1× Ex TaqTM buffer, 200 μM each of dNTPs, 20–100 ng of template DNA, 1 unit of Ex TaqTM (TaKaRa), and 0.5 μM of each primer. We used 35 cycles, each consisting of denaturation at 94 °C for 1 min, primer annealing at 55 °C for 1 min, and extension at 72 °C for 2 min. Nested PCR was performed under the same conditions, with the 400-fold dilution of the first PCR product as template.

DNA sequencing and analysis

PCR products were sequenced directly after gel elution using the ABI PRISM 3100 Genetic Analyzer System. The isolated sequences were then aligned to Rice GD (http://btn.genomics.org.cn/rice/index.php) (Yu *et al.*, 2002) and the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/). Junction sequences were analyzed with the BLAST2 sequences alignment tool at NCBI.

Results

Organization between T-DNA and rice genomic DNA

Genomic DNA fragments that flanked the T-DNA were amplified from transgenic plants by iPCR, with nested primers located near the borders and restriction enzyme sites. The amplified DNA contained either 165 bp of the T-DNA sequence from RB or 236 bp from LB. We analyzed a total of 53 fragments flanking RB, and 61 fragments flanking LB. Figure 2A shows the sequences at the junctions between RB and the flanking genomic region. In 29 events, junction points were observed between the third and fourth nucleotide (nt) of the 25 nt border sequence. This result was similar to previous observations that 7 out of 15 junctions in *Arabidopsis* and tobacco (Tinland, 1996), 18 out of 27 in aspen (Kumar and Fladung, 2002), and 3 out of 11 in barley (Stahl *et al.*, 2002) were located between the third and fourth nt. This site is a known cleavage site for generating single-stranded T-DNA fragments (Yanofsky *et al.*, 1986; Stachel *et al.*, 1987). In most of the remaining plants, usually 1–28 nt were deleted from the RB cleavage site, although in one sample, 127 were deleted. Unlike with RB, no cases were found in which the cleavage site remained in LB

A

```
Line number                  T-DNA                            RB cleavage site    Rice DNA
                                                                      ↓
            ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGACAGGATATATTGGCGGGTAAAC
1A16904     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAcccataaaaaccaacatagt
1A16942     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAcgtgattttgtatattatata
1A17033     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGActgctgcgcgattagcgcaag
1A17108     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGActgttggagcacgtttttttggctagaattgct
1A17125     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGActgtgtgagtggggactactg
1A17228     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAccttttctctcgactctacc
1A16924     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAggaaattctatttccttgtt
1A16925     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtcaaaacgggtggacggtga
1A16932     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAagtccttgggagaattcttca
1A16938     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAatatttccggttatcgattta
1A16944     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgcgcggagctcctaaacgtg
1A17017-1   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtgattatccactgaagagat
1A17028     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgtcagcaatgttggatactt
1A17032     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtttcaaacccatcacccttc
1A17041     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgaggtcgccgactagccgta
1A17103     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgaccatcctgtagcttttca
1A17105     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAttttgcaagccaaggacaa
1A17120     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtcaaaacgggtggacggtga
1A17126     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgcacgcttgttgcttggacc
1A17128-1   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtttccaaattgttttgggtggtgaac
1A17128-2   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgtgggctagagaggacggaa
1A17137-1   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAatgcacacttttatttacatt
1A17137-2   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtaccctctctttctaggcgccc
1A17202     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtatataacatgaactactaaa
1A17211     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAggccaaaatggccctgttga
1A17216     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAacgactggctaggtgataaag
1A17223     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtgtcgggcttttgcttctcga
1A17242     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAgaggtcgccgactagccgta
1A17306     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGAtgattgaaaaataaattacc
1A17017-2   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGtgcaaatttcagtagtgagg
1A17122     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGctggagtcaaaagttatgcc
1A17213     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGtggttgggagcctgtaaggc
1A17236     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTGggacctagcgatatcccaac
1A16922-1   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTatttcattcccaaatgccac
1A17013-2   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTttgaattgttcgcctagtgt
1A17129     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTaaaccgctcttgcttgctcg
1A17231     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTcaagtcaacaggcttatttt
1A17235     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTTtagcctttattctttgctcc
1A16922-2   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTccgctggttcgctccatcccg
1A17113     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTacggccgaagcgccacagcc
1A17217     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTaatgaaccgtggtagcaaaaaaaccctattgttcgagttctc
1A17310     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGTTaagctatatatagcttttga
1A17315     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCGTGgtagcaaactgaagcatcgatc
1A16911     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCaatgcctaatcaaacgatgtttttct
1A17245     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATCCatgtaattagcctggtcgtacgagcacga
1A17013-1   ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGATtcttcgcatgcgcttgcctg
1A17134     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTCGGctccgctgcccgtcacgacg
1A17305     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACCTggtatcttcccaatctcctca
1A17015     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACtcttacctgctggtctggaat
1A17121     ATATAAC (90 bp) ACTCACCTGGTACCTGGTACatgttctgcactttgagtaa
1A17127     ATATAAC (90 bp) ACTCACCTGGTACCggccgctggtatgcgaaagccagcag
1A17208     ATATAAC (90 bp) ACTCACCTGacaccctaggcctaagctag
1A16940     ATATAACttttgttttggtttttatag
```

Figure 2. Sequences at junction regions between the T-DNA right end and rice DNA (A), and between the T-DNA left end and rice genomic DNA (B). T-DNA sequences are presented in capitals with RB and LB sequences in bold. The 20 bp plant DNA sequences are given in lower case. Homologous sequences between the T-DNA end and the rice genomic DNA are shaded. 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.

(Figure 2B). Among those 61 samples examined, 28 retained a portion of LB, while in the remainder, up to 180 bp of T-DNA were deleted at various lengths from the LB cleavage site.

We compared the T-DNA end-point sequence with the receptive rice genomic DNA sequence at the junctions between T-DNA and plant DNA. The T-DNA integrations in rice could be divided into three groups according to the sequence found at the junction region. At RB, 15 events had homologous sequences (1–6 bp), 21 had no homology, and 17 had filler DNAs (1–22 bp) at the junction positions (Figure 2A). Because those filler DNAs were short, it was difficult to identify their origins.

Of the 61 left-border T-DNA junctions analyzed, 40 events had homologous sequences (1–7 bp), 3 had none, while the remaining 18 had filler DNA between the left T-DNA end and the rice genomic sequence (Figure 2B). The filler DNA was 1–15 bp long at most of the junctions. In line 1A17024, the 108 bp filler DNA was derived from the 3SD/SA region near the RB on the T-DNA. Our results indicated that three groups of junctions also were present in LB, and the frequency of groups having homologous sequences was higher in the LB region than in the RB region. Gheysen *et al.* (1991), Mayerhofer *et al.* (1991), and Kumar and Fladung (2002) had also observed three types of junctions between the T-DNA end and plant DNA in dicotyledonous species.

We also investigated the deleted sequences of rice chromosomal DNA at the T-DNA insertion loci from nine transgenic lines, and found that those deletions ranged from 0 to 76 bp, with the degree of deletion being similar between single- and multiple-copy

B

```
Rice DNA      LB cleavage site                T-DNA                                         Line number
                               ↓
                               TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA
    atcatcaacggaagccagagttctagttctAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16911
  gtacatggaaggcatggagccct:: 108 bp::GGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17024
           ataagggaaatttgtaatatGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17044
      cctacaaaacaagagttgagctgccggtcGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17108
            ttctagccttctacacctatGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17123
            gcactagaagcagcaagattGATATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17205
               gcggtagccgtatcgcaccgATATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17237
               ttttccaagttgctgatcttacATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17043
                 tgagtctggtccagaaatagATTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17227-1
                  tggtacagctcgatcgaaacTTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17039
                  gggcccgcaattcaatcgatTTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17134
                  ggccgtatatgaaatttaacTTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17223
                   ccgcaatgtgggttcgtcccTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16904
                  acgcggttacgggcggtatgaTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16918
                   ccaaggctcgtccatcgtacTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17102-1
                   taaaaagaaggaaggtttaaTGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17211
                    cagcgttgcttttatgccaaGTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16924
                     gtgaagactaatcggagataTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17001
               atcccttagtggatcggctcgcgaccTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17102-2
                     atggggcaaaaatatttgaaTGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17105
                      ttctttcaatgtgtgaaagaGGTGTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16942
                          acaaacaatatatagacctcTAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17012
                           tgggaaatggaacagaaatgAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16921
                           tatttatttttacaaaggaaAAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17041
                            cattagagtgaacataaaagAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17021
                    ctggtgaacccacccttcagcaatggaacAACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17023
               catcgagcggcttgttagtgcttatagggttctggACAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A16940-2
                              catcatattatgttactgatCAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17028
                  cgccatgccatgatgcttagtactgaagtagatatAAATTGACGCTTAGACAA (140 bp) TGCTAAA 1A17009
                      ctttcaatcgtgtccgttgctggtgtacttacTGACGCTTAGACAA (140 bp) TGCTAAA 1A17225
                       ccattcacctcacctcacattgcggacgtttaGACGCTTAGACAA (140 bp) TGCTAAA 1A17114
                                  aacaatcaggaccatttggtGCTTAGACAA (140 bp) TGCTAAA 1A17129
                                       tttggtatgtgaattacaggCAA (140 bp) TGCTAAA 1A16932
                                       taattttcccctcaaggcctCAA (140 bp) TGCTAAA 1A16940-1
                            ttcgagtaaaacagcttcagCTTAATAACACATTGCGGAC (120 bp) TGCTAAA 1A17017
                             aacgtaagtgtcaattcagaTTAATAACACATTGCGGAC (120 bp) TGCTAAA 1A17037
                              aagagtagtcatcaccagaaAATAACACATTGCGGAC (120 bp) TGCTAAA 1A16909
                      ccgtactgacgtacccaaatttagacagaAATAACACATTGCGGAC (120 bp) TGCTAAA 1A17034
                                tcacaatgctttccattaacTAACACATTGCGGAC (120 bp) TGCTAAA 1A17008
                                 tgaactgcgatgtgggccccAACACATTGCGGAC (120 bp) TGCTAAA 1A16945
                                 ccataaaattagtgtaaatatACACATTGCGGAC (120 bp) TGCTAAA 1A17032
                                  cttacggaagtcacttgtccCACATTGCGGAC (120 bp) TGCTAAA 1A17115
                           gacaccgcgaaggaggcgtgaaatgaaggACATTGCGGAC (120 bp) TGCTAAA 1A17229
                                    tccgatgtgattaaaaaaaaATTGCGGAC (120 bp) TGCTAAA 1A17007
                                      gggggctttgatttggcgcgGCGGAC (120 bp) TGCTAAA 1A17132
                           ctctagatataaggacctaaACGTTTTTAATGTACTGAATTC (100 bp) TGCTAAA 1A17208
                                 tacgtgtgcgctcgattttgATGTACTGAATTC (100 bp) TGCTAAA 1A17015
                                   acaggttgccttccggacgcGTACTGAATTC (100 bp) TGCTAAA 1A17315
                                    cttgcggggtggccattttaTACTGAATTC (100 bp) TGCTAAA 1A17107
                                          ggccgccaccgcggtggagcTC (100 bp) TGCTAAA 1A17103
                 cgatttacagcatcacctatTCGAGACGCGTAGTACTGAAGCTTCTAGAG ( 70 bp) TGCTAAA 1A17122
                   ttgccgtatgcgtacatatcaAGACGCGTAGTACTGAAGCTTCTAGAG ( 70 bp) TGCTAAA 1A17242
                     ctccgcccagagaccggggaACGCGTAGTACTGAAGCTTCTAGAG ( 70 bp) TGCTAAA 1A17010
                     agtagtagtattaaatggagacttGTAGTACTGAAGCTTCTAGAG ( 70 bp) TGCTAAA 1A17137-1
                         aggcaggccctagcctctacTAGTACTGAAGCTTCTAGAG ( 70 bp) TGCTAAA 1A17137-2
                       tgcgttagtttctttctatttcttGACATATATATCTATCTAA ( 30 bp) TGCTAAA 1A17228
                            ctccagctactgttacgtgtACATATATATCTATCTAA ( 30 bp) TGCTAAA 1A16922
                                  aatacctgatacagaaatagTATCTATCTAA ( 30 bp) TGCTAAA 1A17310
                                ctagtgagtgtgtgtgtgtgTGATATCAATATGAATGTGGAAAATGCTAAA 1A17239
                                  cttgcgtacctcctgtcgattacgggcctaAATGTGGAAAATGCTAAA 1A17227-2
                                                   tgtcggtgaggagtcgccagTGCTAAA 1A16934
```

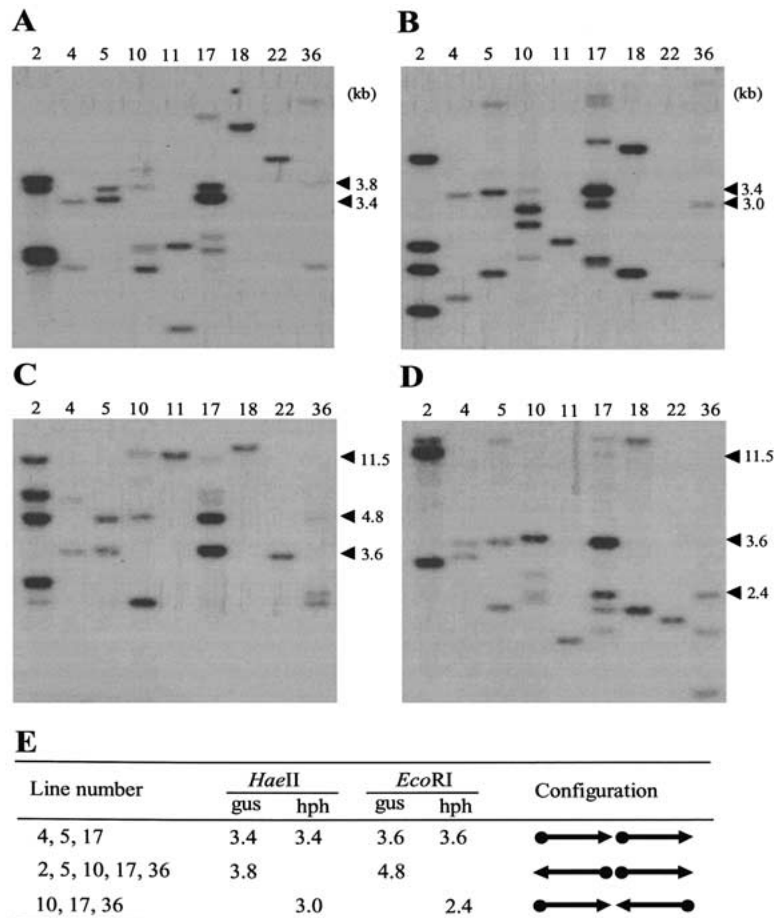
Figure 2. Continued.

insertions at a single locus (data not shown). These results agree with those reported from studies of dicotyledons, in which T-DNA integration generated target site deletions of 13–73 bp (Gheysen *et al.*, 1991; Mayerhofer *et al.*, 1991).

Configuration of T-DNA repeats

T-DNA direct and inverted repeats have been reported in both dicotyledonous plants (Jorgensen *et al.*, 1987; De Neve *et al.*, 1997; Krizkova and Hrouda, 1998; De Buck *et al.*, 1999; Kumar and Fladung, 2000) and monocot species (Jeon *et al.*, 2000; Stahl *et al.*, 2002). To investigate the T-DNA repeat configuration in rice, we analyzed 43 transformants (Figure 3), employing DNA gel-blot analysis because inverted repeats cannot be amplified by PCR (De Buck *et al.*, 1999). Our results showed that 15 lines (35%) carried a single T-DNA insert, 14 lines (33%) had T-DNA direct repeats, 11 (26%) had inverted repeats with the 5′-end junctions, and 4 (9%) had inverted repeats with the 3′-end junctions. In addition, some lines, such as lines 5, 10, 17 and 36, had more than one type of T-DNA configurations.

To further study the organization of direct repeats at the DNA-sequence level, we either PCR-amplified the nine junctions with the *a* and *b* primers, or used iPCR to capture them at the right-end T-DNA junctions. Sequence analysis revealed two types of junctions: three that were precise (Figure 4A) and six that were imprecise (Figure 4B). Whereas one T-DNA end was directly joined to the next T-DNA in the precise junctions, filler DNA sequences were present between two T-DNA ends in the imprecise junctions. Lengths



Line number	*Hae*II		*Eco*RI		Configuration
	gus	hph	gus	hph	
4, 5, 17	3.4	3.4	3.6	3.6	●⟶●⟶
2, 5, 10, 17, 36	3.8		4.8		⟵●●⟶
10, 17, 36		3.0		2.4	●⟶⟵●

Figure 3. DNA gel-blot analysis of T-DNA repeat structures. Numbers above each blot indicate independent transgenic lines generated by T-DNA insertion. Genomic DNAs were digested with *Hae*II (A and B) or *Eco*RI (C and D), and then hybridized with either the *gus* probe (A and C) or the *hph* probe (B and D). E. Summary of DNA gel-blot analysis. Three types of T-DNA repeat configurations are depicted, and the predicted sizes (kb) that hybridized with the probes are represented. Arrows indicate T-DNA direction from RB to LB.

of filler DNA ranged from 2 to 182 bp. Two of the filler DNAs were derived from T-DNA near the LB or RB; the remaining filler DNAs were short, making it difficult to characterize their origins. Whereas the left T-DNA ends were deleted 51–1655 bp from the LB cleavage site, deletions at the RB were minimal and, in fact, no deletions from the RB cleavage sites were found in three of the nine junctions. In the remaining samples, deletions ranged from 1 to 33 bp.

The vector backbone sequence is frequently co-transferred with T-DNA

The vector backbone sequence is found in dicotyledonous plants transformed with Ti vectors (Wenck *et al.*, 1997; Kononov *et al.*, 1997; Wolters *et al.*, 1998; De Buck *et al.*, 2000). Therefore, we investigated whether that backbone could also be transferred into the rice genome. Genomic DNAs were isolated from randomly selected transgenic plants, digested with *Sal*I or *Hae*II, and hybridized with the vector backbone probe (Figure 5). Of the 171 lines analyzed, the vector backbone sequence was detected in 77 (45%). The entire sequence was observed in 66 of those lines; in the remaining 11, the backbone sequence near RB was truncated in 6, the sequence near LB was truncated in 4, while both ends were deleted in the last one.

Left-border read-through is quite frequent

The vector backbone sequence found in our transgenic plants could be a consequence of the read-through product. If T-DNA is not terminated at the left border, it reads-through the vector backbone, resulting

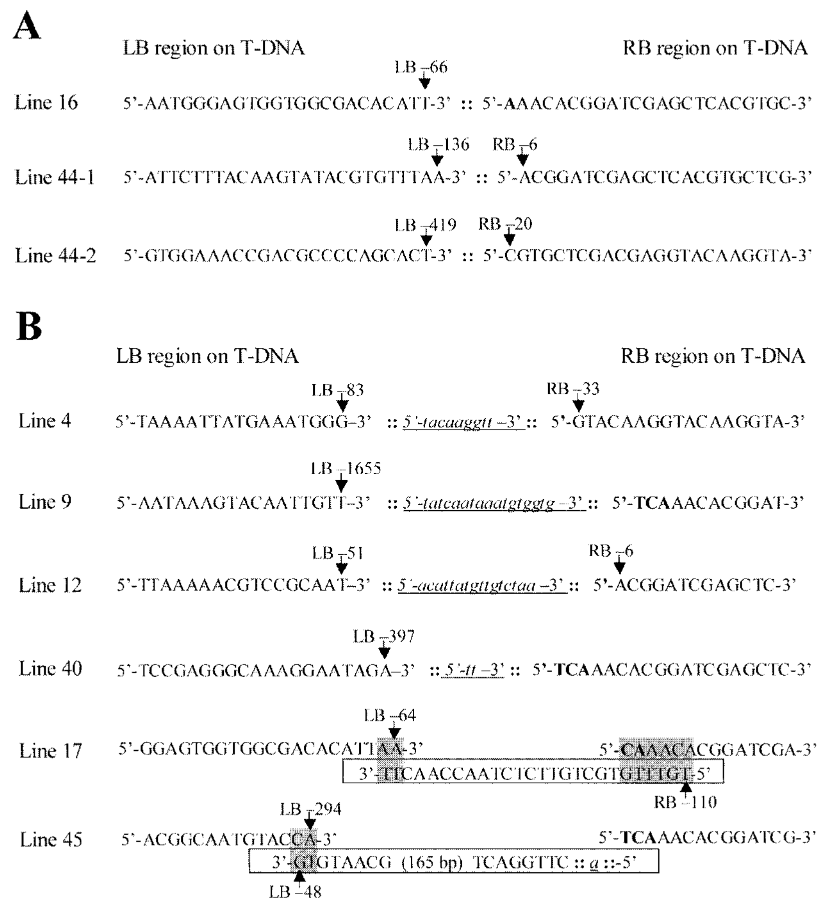


Figure 4. Sequence analysis of T-DNA direct repeats, with sequences at precise junctions (A) and imprecise junctions (B). T-DNA sequences are presented in capitals with right- and left-border sequences in bold. Filler DNAs are depicted in underlined italics; those derived from T-DNA are boxed. Arrows represent the length of the deletion from the cleavage site at the borders. Double colons (::) indicate junction positions between different DNAs.

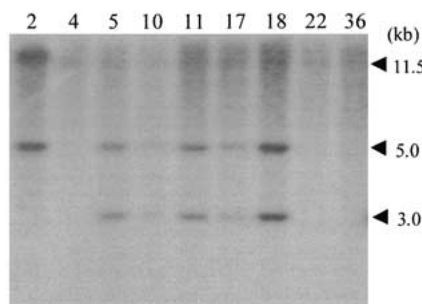


Figure 5. DNA gel-blot analysis of vector backbone sequences in transgenic rice plants. Numbers above each blot indicate independent transgenic lines generated by T-DNA insertion. Genomic DNA was digested with *Sal*I, and then hybridized with the vector backbone probe (see Figure 1). When the entire vector backbone sequence was integrated into the rice genome, the 5 kb band near the LB and the 3-kb band near the RB appeared.

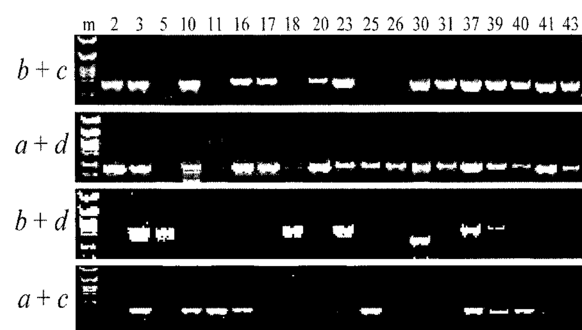


Figure 6. PCR amplification of junction regions between T-DNA and vector backbone. Numbers above each lane represent individual transgenic rice plants carrying the vector backbone. The *a*, *b*, *c*, and *d* represent PCR primers, which are located near RB or LB (see Figure 1). The right-most lane is the λ DNA size marker (m), which was digested with *Pst*I.

in co-transformation of T-DNA and vector backbone sequences into a plant's chromosome. To understand the detailed structural organization of the backbone sequence, we used four primers to amplify DNAs from the transformants (Figure 6). Two of the primers (*a* and *b*) were located within T-DNA; the other two (*c* and *d*), within the backbone (Figure 1). The read-through product at the left border could be amplified by primers *b* and *c*. Such read-through products have been reported in crown gall tumors induced on tobacco plants (Ooms *et al*., 1982; De Buck *et al*., 2000).

We used the 77 lines carrying the vector backbone sequence as templates for PCR amplification, and found that 55 (71%) contained amplified LB regions (Table 1). This indicates that LB read-through is quite frequent. The PCR fragments were sequenced to verify the read-through products (Figure 7A). In 19 of 20 cases, T-DNA was linked to the vector backbone through the intact border sequence, likewise confirming that read-through is quite common during T-DNA processing. In one sample (line 2), 73 bp sequences were deleted in the border region, and T-DNA was linked to the vector backbone through a 30 bp filler DNA. This could have been caused by two DNA molecules integrating at the same locus during the DNA transfer process.

The left border can serve as the start point of DNA transfer

In *Arabidopsis* and tobacco, the LB can serve as the initiation point for DNA transfer (Ramanathan and Veluthambi, 1995; van der Graaff *et al*., 1996). When this occurs, the read-through product can be amplified by primers *a* and *d* if the transfer does not stop at the RB. Interestingly, we found that our RB read-through frequency was quite high (Table 1), with PCR fragments being detectable with the primers in 67 lines (87%). Therefore, it was evident that a majority of the vector read-through products were not terminated in RB, thereby resulting in the transfer of T-DNA/vector backbone/T-DNA (types 1–4 in Table 1). In types 5 and 6, only RB read-through product was detected, which suggested that T-DNA transfer started at LB and continued through RB, transferring vector backbone/T-DNA. In these types, termination must have occurred at the LB of the T-DNA because the LB read-though product was not amplified.

PCR products at RB were sequenced to determine the detailed structure of the region (Figure 7B). In 22 of 23 samples, the vector backbone was connected to T-DNA through intact RB, confirming that a majority of the *a*+*d* PCR products were due to the RB read-through from the vector backbone to T-DNA. In one sample (line 750), however, 93 bp sequences were deleted (including the 25 bp RB), a result of the ligation product between two DNA fragments.

T-DNA and the vector backbone can be cross-linked

Because more than one T-DNA can be inserted into one locus, it is also possible for T-DNA and vector backbone to be integrated into one locus. PCR analysis of the 77 samples carrying the vector backbone showed that the primer sets of *b* and *d* amplified DNA fragments in 12 lines (types 1, 3, and 9; Table 1). In these types, the LB of T-DNA and the RB of the vector backbone were linked. Both borders were the 3′ ends of the single-strand transfer DNAs. Sequencing the PCR products showed that a majority were linked through 3–6 bp of complementing sequence between the 3′ ends (Figure 7C). In the remaining sample, they were linked through a 5 bp filler DNA sequence.

The link between RB of T-DNA and LB of the vector backbone also demonstrated that both ends were start points (i.e., the 5′ end) for transfer DNAs. These types could be visualized by PCR amplification with primers *a* and *c*. This category comprised 21 lines (types 1, 2, 5, and 8; Table 1). Sequencing the amplified bands revealed two groups: precise and imprecise junctions (Figure 7D), with 10 samples carrying the former junction, in which two ends were linked without filler DNA. In five other samples, 1–37 bp filler DNA sequences were inserted between the ends, forming imprecise junctions. That filler DNA had been derived from the vector backbone sequence containing 4 bp of the RB.

Isolation of genomic sequence flanking the vector backbone

To further determine whether LB could be a start point for DNA transfer, we used iPCR to isolate rice genomic DNA that flanked the vector backbone. Transgenic plants carrying that backbone sequence were then randomly selected and their DNAs digested with *Hae*II, self-ligated, and PCR-amplified. As a result, we were able to obtain two rice genomic DNA sequences linked to the vector backbone LB (Figure 8). The genomic clone of line 18 matched PAC clone P0503E05 at NCBI, while line 5 hit the contig 6649 at Rice GD. The junction points were one to

A

LB region on T-DNA | LB | LB region on vector backbone

Vector 5'-GACACA (40 bp) CAATTT**GTTTACACCACAATATATCCTGCCA**CCAGCCAGCCAACA-3'
19/20 junctions 5'-GACACA (40 bp) CAATTT**GTTTACACCACAATATATCCTGCCA**CCAGCCAGCCAACA-3'
Line 2 5'-GA :: *gcttgaaacattatcaaattgaataccata* :: CACCAGCCAGCCAACA-3'
73 bp deletion

B

RB region on vector backbone | RB | RB region on T-DNA

Vector 5'-GAACGGAT (50 bp) TTCTCTTAG**GTTTACCCGCCAATATATCCTGTCA**AACACGGATC-3'
22/23 junctions 5'-GAACGGAT (50 bp) TTCTCTTAG**GTTTACCCGCCAATATATCCTGTCA**AACACGGATC-3'
Line 750 5'-GA :: ACGGATC-3'
93 bp deletion

C

RB region on vector backbone | LB region on T-DNA

Lines 3, 5, 23, 37, 726, 744 5'-CTCTTAG**GTTTACCCGCCAATATAT**-3' / 3'-**TATATAACACCACATTTG**TTTAAC-5'
Line 39 5'-CTCTTAG**GTTTACCCGCCAATAT**-3' / 3'-**TATAACACCACATTTG**TTTAACTG-5'
Line 18 5'-ACGCTCTTTTCTCTTAGG-3' / 3'-**TCCTATATAACACCACATTTG**TTTAACTG-5'
Line 737 5'-ATCGCACGGAATGCCAAGCACTCCCGA-3' / 3'-GCTAGACTTCTAAAAAGTTT-5' (RB -126; LB -266)
Line 747 5'-GAAGTCATTATCGCACG-3' / 3'-TGCGAATCTGTTGAATTATTGTGTAACGCCT-5' (RB -145; LB -28)
Line 30 5'-TGAGCGTCGGAAAGGGTCGA :: *tgtgt* ::-3' / 3'-GCGAATCTGTTGAATTAT-5' (RB -497; LB -29)

D

i) Precise junctions

LB region on vector backbone | RB region on T-DNA

Lines 11, 40, 732 3'-ACAACCGACCGACCA**CC**-5' / 5'-**TCA**AACACGGATC-3'
Lines 750, 839 3'-ACAACCGACCGACCA**C**-5' / 5'-**TCA**AACACGGATC-3'
Line 830 3'-GCCCCTCGACAACCGACCGACCA**C**-5' / 5'-**CA**AACACGGATC-3'
Line 16 3'-CCCCTCGACAACCGACCGACCA**CC**-5' / 5'-AAACACGGATC-3'
Line 10 3'-CCCCTCGACAACCGACCGACCA**CC**-5' / 5'-AACACGGATC-3' (RB -3)
Line 3 3'-CCCCTCGACAACCGACCGA-5' / 5'-ACACGGATC-3' (LB -5; RB -4)
Line 749 3'-GCCCCTCGACAACCGACCGACCA-5' / 5'-TGTTACGTCCTGTA-3' (RB -161)

ii) Imprecise junctions

LB region on vector backbone | RB region on T-DNA

Lines 759, 37-1, 837, 25 3'-GCCCCTCGACAACCGACCGACCA**CC**-5' :: 5'- *c* / *ttgg* / *tccct* / *ctacct* -3' :: 5'-**TCA**AACACGGATC-3' (deletion range ... deletion range)
Line 37-2 3'-ACCGACCA-5' :: *5'-ccgattattctaataaacgctcttttctcttag****GTTT****-3'* :: 5'-ATATGCAGG-3' (RB -169)

Figure 7. Sequence analysis of junction regions between T-DNA and the vector backbone. T-DNA and vector backbone sequences are presented in capitals, with RB and LB sequences in bold. Homologous sequences between the T-DNA end and the vector backbone end are shaded. Filler DNAs are represented in underlined italics. Digits with arrows at the ends indicate the length of deletions from the cleavage sites to the end. Double colons (::) are placed between two different DNAs.

Table 1. Organization of T-DNA and vector backbone linkages. Seventy-seven lines carrying the vector backbone sequence were PCR-amplified using four primers (*a*, *b*, *c*, and *d*) described in Figure 1. O, positive PCR; X, negative PCR.

Type	$a+d$	$b+c$	$b+d$	$a+c$	Number of transgenic lines (%)
1	O	O	O	O	5 (6.5)
2	O	O	X	O	6 (7.8)
3	O	O	O	X	5 (6.5)
4	O	O	X	X	36 (46.8)
5	O	X	X	O	8 (10.4)
6	O	X	X	X	7 (9.1)
7	X	O	X	X	3 (3.9)
8	X	X	X	O	2 (2.6)
9	X	X	O	X	2 (2.6)
10	X	X	X	X	3 (3.9)
Total	67	55	12	21	77 (100)

three nucleotides from the LB cleavage site, respectively. These results confirm that LB can integrate host chromosomes.

Discussion

We studied T-DNA integration patterns in rice chromosomes. In 29 of the 53 RB examined (55%), the junction points were at the cleavage site for generating single-stranded T-DNA fragments. This result supports those from earlier rice experiments. For example, Hiei *et al.* (1994) reported that two out of three RB generated by pIG121Hm contained the cleavage site. Jeon *et al.* (2000) also found that 11 of 32 RB junction points were at the cleavage site. In addition, Gheysen *et al.* (1991) analyzed six *Arabidopsis* and two tobacco T-DNA inserts, and reported that three events had homologous sequences, two events had no homology, and the rest had filler DNA at the junction between T-DNA and plant DNA. Mayerhofer *et al.* (1991) also examined seven T-DNA insertions in *Arabidopsis*, finding that three events had homology, three had none, and the remaining one event had filler DNA. In aspen, Kumar and Fladung (2002) showed that two events had homologous sequences (3–7 bp), two had none, and six had filler DNA between the right T-DNA end and the plant genomic DNA sequence. Our data, with our larger sample size, confirms and extends previous observations that three types of junction are also present between RB and rice genomic DNA.

In contrast, LB junction points were not conserved. Among the 61 LB examined in this study, there were no cases where the LB cleavage site had remained. Hiei *et al.* (1994) also observed the same in their study of five junctions. Our comparison of deleted sequences showed that the amount of deletion was greater in the left T-DNA ends than in the right ends. In research on *Arabidopsis* and tobacco plants, Tinland (1996) found that only 3 of 15 LB cleavage sites were conserved, with the remaining samples carrying 5–109 nucleotide deletions, and one event having a 1500 bp deletion. In aspen, only 4 of 20 LB cleavage sites remained, and deletions ranged from 2 to 24 bp (Kumar and Fladung, 2002). In barley, 3 out of 39 LB cleavage sites were observed; the deletions were 1 to 95 nucleotides from the cleavage sites in the rest of the studied lines (Stahl *et al.*, 2002). Our results were generally similar to those observed with T-DNA junctions in other plants, which suggests that T-DNA integration mechanisms are comparable between monocotyledonous rice and dicot species.

Our analytical comparison of rice genomic DNA and T-DNA presents three distinct types of junction regions. In the first, up to seven nucleotide overlaps exist between the plant's DNA and the T-DNA. The frequency of this type is much higher in the LB than in the RB region. Tinland (1996) has proposed that the $3'$ end of T-DNA finds sequence similarity with the single-strand region of plant chromosomal DNA, thereby facilitating T-DNA integration. Brunaud *et al.* (2002) reported that the plant genome sequence upstream of the T-DNA insertion was T-rich and it was particularly striking in the cases of canonical insertions at LB. However, we did not observe the T-rich

regions at the LB junctions in rice. The difference might be due to the fact that the canonical insertion was a rare event in rice.

In the second type of junction, T-DNA and the rice genome sequence are linked directly, without any overlaps or filler DNA. This type is more frequent in RB, with 21 of 53 RB and only 3 of 61 LB belonging to this category. Because this type is so commonly found in the junctions between RB and plant DNA, it could be postulated that the VirD2 protein attached to RB may play a critical role during the synapsis of T-DNA to plant chromosomal DNA. However, Ziemienowicz *et al.* (2000) have reported that VirD2 shows no ligase activity *in vitro*. An alternative, synthesis-dependent strand-annealing model has now been proposed by Puchta (1998), Salomon and Puchta (1998), and Kumar and Fladung (2002). In that model, the 3′ end of single-stranded T-DNA anneals at the genomic double-strand break, and DNA synthesis progresses toward the 5′ end of the T-DNA. The newly synthesized sequence is then patched by sequence homology or blunt-end ligation. It is difficult to speculate which mechanism occurred at the precise junctions in our rice experiments.

Finally, in the third type, filler DNA is present between the T-DNA and the plant sequences. Because it normally is short (1–22 bp), it is usually difficult to deduce its origin. Our 108 bp filler DNA, however, was derived from the RB region on the T-DNA. In aspen, filler DNAs arise from either the host genome near the target site or T-DNA near the borders (Kumar and Fladung, 2002). Filler DNA may serve to facilitate T-DNA integration into plant chromosomes.

More than one T-DNA can be integrated into one locus. We estimated that 33% of the transformants carried direct repeats, 21% had inverted repeats with the 5′-end junctions, and 7% carried inverted repeats with the 3′-end junctions. Sequencing the direct repeats showed that both precise and imprecise junctions had occurred. The latter contained 2–181 bp filler DNAs, and were derived from T-DNA near the borders. They were inserted in a reverse orientation, suggesting that such an oriented filler might help in connecting two T-DNA fragments. Precise and imprecise junctions have also been reported in tobacco, *Arabidopsis*, and aspen (Krizkova and Hrouda, 1998; De Buck *et al.*, 1999; Kumar and Fladung, 2000). There, the lengths of filler DNAs vary from only a few to up to 300 bp, with their derivations being either T-DNA or host DNA.

In the current study, we observed that 45% of the transformants carry the vector backbone sequence, a phenomenon largely due to the failure of T-DNA termination at LB, thereby generating a read-through product. The vector DNA can be started at LB, although its efficiency is low. We also showed that RB can serve as a termination site for DNA transfer, again with only low efficiency. Therefore, the vector backbone is linked to T-DNA in the majority of the lines, and co-transfer of the backbone sequences with T-DNA complicates transformant analysis. Furthermore, bacterial sequences are undesirable when the transformants are to be developed for elite varieties. Hence, it would be useful if termination efficiency at LB could be increased and its initiation efficiency decreased.

In 11 out of 77 cases, the vector backbone sequence was partial, probably because of early termination during T-strand generation in *Agrobacterium*. Alternatively, the partial sequences could have been produced from the intact T-strand by nuclease digestion in the plant cells. Truncated vector backbones have been reported in other species, including carrot and potato (Miranda *et al.*, 1992; Wolters *et al.* 1998).

We also observed that T-DNA and the vector backbone are independently produced and linked. In this study, the frequency of linkage between their LBs was quite low and, likely, the frequency of linkage between T-DNA RB and vector backbone RB was also low. However, cross-border links were high between T-DNA RB and vector backbone LB, as were those between T-DNA LB and vector backbone RB. These results suggest that two single-stranded DNAs are efficiently linked by complementing sequences at the 3′ end of the transfer DNAs, as well as by synapsis of the 5′ ends of the transfer DNAs.

Based on our research, we believe that LB can be a start point of DNA transfer. Rice genomic DNA sequences were isolated here that flanked LB of the backbone sequence. Although this has also been reported in *Arabidopsis* (van der Graaff *et al.*, 1996), other examples are rare. Our results confirm that the

Vector backbone — LB cleavage site ↓

Vector TGTTGGCTGGCTGG**TGGCAGGATATATTGTGGTGTAAAC**

Line 18 TGTTGGCTGGCTG**GTG**tgttttttctgtcctgctttagcgcagaagatacaatactctagcagcac

Line 5 TGTTGGCTGGCTGG*acgaggtacagacttaacctcg*tctctcctccatcgtcgttggcatttatccgtgacg

Figure 8. Genomic DNA flanking LB of the vector backbone. Nucleotides of the vector backbone are presented in capitals, with LB sequences in bold. Plant DNA sequences are given in lower case. Homologous sequences between the vector backbone and genomic DNA are shaded, and filler DNAs at the junction are depicted in underlined italics. The LB and vector backbone sequences near LB in the binary vector are shown on top for comparison. The cleavage site is indicated with an arrow.

left border of the vector backbone can be inserted into the monocot rice chromosome.

Acknowledgments

We thank Sunok Moon for valuable discussion, Min-Jung Han and Kyungsook An for their production of transgenic rice plants, and Priscilla Licht for critical reading of the manuscript. J.L., S.-H.J., and H.-G.K. were recipients of a BK21 fellowship. This work was supported, in part, by grants from the Crop Functional Genomic Center (CG-1111), the 21 Century Frontier Program, and from the Biogreen 21 Program, Rural Development Administration. The study was also funded by the National Research Laboratory Program of the Korea Institute of Science and Technology Evaluation and Planning.

References

An, G., Watson, B.D., Stachel, S., Gordon, M.P. and Nester, E.W. 1985. New cloning vehicles for transformation of higher plants. EMBO J. 4: 277– 284.

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

Binns, A.N. 2002. The T-DNA of *Agrobacterium tumefaciens*: 25 years and counting. Trends Plant Sci. 7: 231–233.

Birch, R.G. 1997. Plant transformation: problems and strategies for practical application. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48: 297–326.

Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., de Rose, R., Pelletier, G., Lepiniec, L., Caboche, M. and Lecharny, A. 2002. T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. EMBO Rep. 3: 1152–1157.

Bytebier, B., Deboeck, F., De Greve, H., Van Montagu, M. and Hernalsteens, J.P. 1987. T-DNA organization in tumor cultures and transgenic plants of monocotyledon *Asparagus officianalis*. Proc. Natl. Acad. Sci. USA 84: 5345–5349.

Chen, D.H. and Ronald, P.C. 1999. A rapid DNA minipreparation method suitable for AFLP and other PCR applications. Plant Mol. Biol. Rep. 17: 53–57.

Cluster, P.D., O'Dell, M., Metzlaff, 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.

De Buck, S., Jacobs, A., Van Montagu, M. and 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 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.

De Cleene, M. 1985. The susceptibility of monocotyledons to *Agrobacterium tumefaciens*. Phytopath. Z. 113: 81–89.

De Cleene, M. and De Ley, J. 1976. The host range of crown gall. Bot. Rev. 42: 389–466.

De Neve, M., De Buck, S., Jacobs, A., Van Montagu, M. and 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.

Durrenberger, F., Crameri, A., Hohn, B. and Koukolikova-Nicola, Z. 1989. Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc. Natl. Acad. Sci. USA 86: 9154–9158.

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

Gelvin, S.B. 2000. *Agrobcaterium* and plant genes involved in T-DNA transfer and integration. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51: 223–256.

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

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.

Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. and Schilperoort, R.A. 1983. A binary plant vector strategy based on separation of *vir*- and T-regions of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303: 179–180.

Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T. and Kumashiro, T. 1996. High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. Nature Biotechnol. 14: 745–750.

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.

Jorgensen, R., Snyder, C. and Jones, J.G.D. 1987. T-DNA is organized predominantly in inverted repeat structure in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. Mol. Gen. Genet. 207: 478–485.

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.

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

Kumar, S. and Fladung, M. 2000. Transgene repeats in aspen: molecular characterization suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome. Mol. Gen. Genet. 264: 20–28.

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.

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.

Miranda, A., Janssen, G., Hodges, L., Peralta, E.G. and Ream, W. 1992. *Agrobacterium tumefaciens* transfers extremely long T-DNAs by a unidirectional mechanism. J. Bact. 174: 2288–2297.

Nester, E.W. and Kosuge, T. 1981. Plasmids specifying plant hyperplasias. Annu. Rev. Microbiol. 35: 531–561.

Ochman, H., Gerber, A.S. and Hartl, D.L. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120: 621–623.

Ooms, G., Bakker, A., Molendijk, L., Wullems, G.J., Gordon, M.P., Nester, E.W. and Schilperoort, R.A. 1982. T-DNA organization in homogeneous and heterogeneous octopine-type crown gall tissues of *Nicotiana tabacum*. Cell. 30: 589–597.

Otten, L., Canaday, J., Gerard, J.C., Fournier, P., Crouzet, P. and Paulus, F. 1992. Evolution of agrobacteria and their Ti plasmids: a review. Mol. Plant-Microbe Interact. 5: 279–287.

Puchta, H. 1998. Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. Plant J. 13: 331–339.

Ramanathan, V. and Veluthambi, K. 1995. Transfer of non-T-DNA portions of the *Agrobacterium tumefaciens* Ti plasmid pTiA6 from the left terminus of T_L-DNA. Plant Mol. Biol. 28: 1149–1154.

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

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

Stachel, S.E., Messens, E., Van Montagu, M. and Zambryski, P. 1985. Identification of the signal molecules produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. Nature 318: 624–629.

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.

Tingay, S., McElroy, D., Kalla, R., Fieg, S., Wang, M., Thornton, S. and Bretell, R. 1997. *Agrobacterium tumefaciens*-mediated barley transformation. Plant J. 11: 1369–1376.

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

Triglia, T., Peterson, M.G. and Kemp, D.J. 1988. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. Nucl. Acids Res. 16: 8186.

Ursic, D., Slightom, J.L. and Kemp, J.D. 1983. *Agrobacterium tumefaciens* T-DNA integrates into multiple sites of the sunflower crown gall genome. Mol. Gen. Genet. 190: 494–503.

van der Graaff, E., den Dulk-Ras, A. and Hooykaas, P.J.J. 1996. Deviating T-DNA transfer from *Agrobacterium tumefaciens* to plants. Plant Mol. Biol. 31: 677–681.

Wang, K., Herrera-Estrella, L., Van Montagu, M. and Zambryski, P. 1984. Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. Cell. 38: 455–462.

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.

Wolters, A.M.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.

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.

Yu, J., Hu, S.N., Wang, J., Wong, G.K.S., Li, S.G., Liu, B., Deng, Y.J., Dai, L., Zhou, Y., Zhang, X.Q., Cao, M.L., Liu, J., Sun, J.D., Tang, J.B., Chen, Y.J., Huang, X.B., Lin, W., Ye, C., Tong, W., Cong, L.J., Geng, J.N., Han, Y.J., Li, L., Li, W., Hu, G.Q., Huang, X.G., Li, W.J., Li, J., Liu, Z.W., Li, L., Liu, J.P., Qi, Q.H., Liu, J.S., Li, L., Li, T., Wang, X.G., Lu, H., Wu, T.T., Zhu, M., Ni, P.X., Han, H., Dong, W., Ren, X.Y., Feng, X.L., Cui, P., Li, X.R., Wang, H., Xu, X., Zhai, W.X., Xu, Z., Zhang, J.S., He, S.J., Zhang, J.G., Xu, J.C., Zhang, K.L., Zheng, X.W., Dong, J.H., Zeng, W.Y., Tao, L., Ye, J., Tan, J., Ren, X.D., Chen, X.W., He, J., Liu, D.F., Tian, W., Tian, C.G., Xia, H.G., Bao, Q.Y., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W.M., Li, P., Chen, W., Wang, X.D., Zhang, Y., Hu, J.F., Wang, J., Liu, S., Yang, J., Zhang, G.Y., Xiong, Y.Q., Li, Z.J., Mao, L., Zhou, C.S., Zhu, Z., Chen, R.S., Hao, B.L., Zheng, W.M., Chen, S.Y., Guo, W., Li, G.J., Liu, S.Q., Tao, M., Wang, J., Zhu, L.H., Yuan, L.P., Yang, H.M. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79–92.

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M. and Schell, J. 1983. Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. EMBO J. 2: 2143–2150.

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

Zupan, J., Muth, T.R., Draper, J. and Zambryski, P. 2000. The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. Plant J. 23: 11–23.