T-DNA Integration in Arabidopsis Chromosomes. Presence and Origin of Filler DNA Sequences

Pieter Windels, Sylvie De Buck, Erik Van Bockstaele, Marc De Loose, and Ann Depicker*

Department of Plant Genetics and Breeding, Center of Agricultural Research-Gent, Caritasstraat 21, B–9090 Melle, Belgium (P.W., E.V.B., M.D.L.); Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology, Ghent University, Technologiepark 927, B–9052 Gent, Belgium (S.D.B., A.D.); and Department for Plant Production, Ghent University, Coupure Links 653, B–9000 Gent, Belgium (E.V.B.)

To investigate the relationship between T-DNA integration and double-stranded break (DSB) repair in Arabidopsis, we studied 67 T-DNA/plant DNA junctions and 13 T-DNA/T-DNA junctions derived from transgenic plants. Three different types of T-DNA-associated joining could be distinguished. A minority of T-DNA/plant DNA junctions were joined by a simple ligation-like mechanism, resulting in a junction without microhomology or filler DNA insertions. For about one-half of all analyzed junctions, joining of the two ends occurred without insertion of filler sequences. For these junctions, microhomology was strikingly combined with deletions of the T-DNA ends. For the remaining plant DNA/T-DNA junctions, up to 51-bp-long filler sequences were present between plant DNA and T-DNA contiguous sequences. These filler segments are built from several short sequence motifs, identical to sequence blocks that occur in the T-DNA ends and/or junctions, up to 51-bp-long filler sequences were present between plant DNA and T-DNA contiguous sequences. These filler segments are built from several short sequence motifs, identical to sequence blocks that occur in the T-DNA ends and/or the plant DNA close to the integration site. Mutual microhomologies among the sequence motifs that constitute a filler segment were frequently observed. When T-DNA integration and DSB repair were compared, the most conspicuous difference was the frequency and the structural organization of the filler insertions. In Arabidopsis, no filler insertions were found at DSB repair junctions. In maize (Zea mays) and tobacco (Nicotiana tabacum), DSB repair-associated filler was normally composed of simple, uninterrupted sequence blocks. Thus, although DSB repair and T-DNA integration are probably closely related, both mechanisms have some exclusive and specific characteristics.

The bacterial phytopathogen Agrobacterium tumefaciens provides a natural system to transfer and introduce genetic information into susceptible plant cells. The T-DNA region, a portion of the tumor-inducing plasmid delimited by two 24-bp border repeats, is transferred to the plant cell as a single-stranded DNA after vir (virulence) gene induction (Gheysen et al., 1998; Gelvin, 2003). Upon arrival in the plant cell, the T-strand is targeted to the cell nucleus, and the T-DNA becomes integrated into the host genome by the plant’s illegitimate recombination apparatus (Matsumoto et al., 1990; Gheysen et al., 1991; Mayerhofer et al., 1991; Tinland, 1996). Recently, T-DNA integration in the Arabidopsis genome has been shown to depend on the sequence of the pre-insertion site (Brunaud et al., 2002).

That T-DNA strands can be captured at double-stranded breaks (DSBs; Salomon and Puchta, 1998) in tobacco (Nicotiana tabacum) and also that T-DNA integration in yeast (Saccharomyces cerevisiae) depends on nonhomologous end-joining (NHEJ) enzymes (van Attikum et al., 2001) suggest that double-stranded break repair via NHEJ is involved as a pathway in T-DNA integration. More recently, the T-DNA integration efficiency has been found to decrease in Ku80- and DNA ligase IV-deficient plants (Friesner and Britt, 2003). The first molecular data on NHEJ in plants came from the footprint analysis of excised transposons. Repair at transposon-induced DSBs seems to generate only minor changes, such as small deletions, insertion of filler, or inversions (Gorbunova and Levy, 1999; Yan et al., 1999). Also, T-DNA integration has been studied as an example of NHEJ. Short regions of microhomology, insertions of filler DNA, and small deletions have been reported as features of T-DNA/plant DNA junctions (Matsumoto et al., 1990; Gheysen et al., 1991; Mayerhofer et al., 1991; Tinland, 1996; Fladung, 1999; Kumar and Fladung, 2002; Meza et al., 2002). More direct NHEJ analysis in plants involved the characterization of naturally occurring deletions in plant genomes (Ralston et al., 1988; Wessler et al., 1990) and the sequencing of newly formed junctions at repaired DSBs (Gorbunova and Levy, 1997; Salomon and Puchta, 1998). For five spontaneous deletions of the waxy locus in maize (Zea mays; Wessler et al., 1990), for a spontaneous deletion in the maize bz-R allele (Ralston et al., 1988), and for NHEJ-repaired junctions in tobacco cells (Gorbunova and Levy, 1997; Salomon and Puchta, 1998), the majority of repaired junctions contained uninterrupted filler insertions (Gorbunova et al., 1999; Ralston et al., 1988; Salomon et al., 1988; Tinland, 1996; Fladung, 1999; Kumar and Fladung, 2002; Meza et al., 2002). More direct NHEJ analysis in plants involved the characterization of naturally occurring deletions in plant genomes (Ralston et al., 1988; Wessler et al., 1990) and the sequencing of newly formed junctions at repaired DSBs (Gorbunova and Levy, 1997; Salomon and Puchta, 1998).
and Levy, 1999). Recently, species-specific differences in plant DSB repair have been reported by Kirik et al. (2000), who showed that DSB repair in tobacco cells is accompanied by insertions but not in Arabidopsis in which large deletions of the joined plant DNA ends are frequent. These findings have been further substantiated by Orel and Puchta (2003). These authors suggest that free DNA break ends in Arabidopsis are less stable and more prone to end degradation than free DNA ends in tobacco cells.

To investigate the relationship between DSB repair by NHEJ and T-DNA integration in Arabidopsis, we sequenced and analyzed 67 T-DNA/plant DNA junctions and 13 junctions between linked T-DNAs. We compared the outcome of the T-DNA/plant DNA and T-DNA/T-DNA joining reaction with the characteristics of NHEJ-repaired DSBs reported in Arabidopsis (Kirik et al., 2000; Orel and Puchta, 2003) and other plant species (Gorbunova and Levy, 1997; Salomon and Puchta, 1998). This comparison mainly focused on the frequency, the presence, and the origin of filler DNA sequences.

RESULTS

Experimental Setup

T-DNA junctions, derived from a population of Arabidopsis transformants (see “Materials and Methods”), were amplified and sequenced. We characterized 44 left T-DNA/plant DNA junctions, 23 right T-DNA/plant DNA junctions, eight junctions between tandemly linked T-DNAs, and five junctions connecting two T-DNAs in inverted orientation, for which one T-DNA was truncated for the T-DNA border region (see supplemental data, available in the online version of this article at http://www.plantphysiol.org). To determine the end point of the T-DNA borders, the T-DNA region was aligned against the T-DNA plasmid sequence. We will refer throughout to a correct nicking process of the T-strand when the T-DNA is cleaved between the third and fourth base of the 24-bp border repeat (Van Haaren et al., 1988).

Further sequence analysis revealed three distinct junction types. A first class included T-DNA junctions that were joined by a simple ligation-like mechanism, resulting in a junction without microhomology or filler DNA insertions. A second class had microhomology at the transition point between two adjacent DNA segments. Here, microhomology is defined as small DNA sequences, ranging from 1 to 7 bp and found at the junction point between the T-DNA and the plant DNA for plant DNA/T-DNA junctions and between two T-DNAs, when linked T-DNAs are studied (Roth et al., 1985; Roth and Wilson, 1986). A third class grouped junctions with insertion of filler DNA. These new, intermediate filler sequences were detected easily because the end point of the T-DNA border, and the starting point of the plant DNA for T-DNA/plant DNA junctions could be determined. Similarly, for linked T-DNAs, the end points of the two T-DNA borders were known. Filler sequences longer than 17 bp were screened for complete colinear sequence identity with the Arabidopsis genome to evaluate whether the observed filler sequences could be uninterrupted sequence blocks that had been copied from the plant genome. In addition, for 21 filler DNAs larger than 6 bp, we analyzed the origin of the filler DNA by looking for identical sequence motifs with: (a) the first 100 bp of the T-DNA sequence situated immediately adjacent to the transition point of the T-DNA junction, (b) a 200-bp plant DNA segment surrounding the T-DNA integration point, and (c) the complete T-DNA plasmid sequence (see “Materials and Methods” and Supplemental Data). An experimental, statistical analysis was performed to determine thresholds for reporting identities, allowing us to discriminate mechanistic relevance from by chance occurrence (see “Materials and Methods” and Supplemental Data).

Outcome of the Joining Reaction between T-DNA and Plant DNA Ends

We wanted to address the question whether T-DNA junction characteristics and the mode of T-DNA joining are related. Therefore, we characterized the outcome of the joining reaction of 67 plant DNA/T-DNA junctions, taking into account the T-DNA border type involved and the degree of T-DNA end processing (Table I). Of 44 left border T-DNA/plant DNA junctions, four had the T-DNA end joined to the plant DNA without the presence of microhomology or the formation of filler DNA sequences, 23 displayed microhomology of 1 bp up to a stretch of 6 bp, and 17 had filler DNA insertions ranging from 1 up to 48 bp (Table I; Supplemental Data). Similar results were obtained for the right border junctions: three of 23 junctions were joined by means of a simple ligation-like joining reaction, 10 had microhomology of 2 up to 7 bp, and 10 had filler DNA ranging from 1 to 51 bp (Table I; Supplemental Data). These results clearly indicate that the frequency by which simple ligation, microhomology, or filler DNA is observed at left and right border junctions is comparable.

We also analyzed whether T-DNA joining characteristics and the degree of T-DNA end processing

<p>| Table I. Features of 67 T-DNA/plant DNA junctions |
|------------------------------------------|-----------|-----------|-----------|</p>
<table>
<thead>
<tr>
<th>T-DNA end</th>
<th>Simple Ligation</th>
<th>Microhomology</th>
<th>Filler Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left T-DNA border</td>
<td>4/44</td>
<td>23/44</td>
<td>17/44</td>
</tr>
<tr>
<td>Right T-DNA border</td>
<td>3/23</td>
<td>10/23</td>
<td>10/23</td>
</tr>
<tr>
<td>Intact T-DNA ends</td>
<td>2/13</td>
<td>3/13</td>
<td>8/13</td>
</tr>
<tr>
<td>Processed T-DNA ends</td>
<td>5/54</td>
<td>30/54</td>
<td>19/54</td>
</tr>
<tr>
<td>Total</td>
<td>7/67</td>
<td>33/67</td>
<td>27/67</td>
</tr>
</tbody>
</table>
were correlated (Table I; Supplemental Data). Regarding the T-DNA end processing, our results confirm recent observations (Brunaud et al., 2002; Krysan et al., 2002; Meza et al., 2002) that, although the deletion size at the left T-DNA ends is larger than that at the right borders, processing and trimming is a ubiquitous feature of T-DNA ends. Interestingly, when the degree of T-DNA end processing is correlated with the outcome of the joining reaction, three of 13 intact T-DNA ends were joined to the plant DNA with microhomology (Table I), in contrast to 30 of 54 processed T-DNA ends (Table I). Also striking is that eight of 13 intact and only 19 of 54 processed T-DNA ends were joined with insertion of filler sequences. In conclusion, microhomology was on average 2.4-fold more associated with deleted than with intact T-DNA borders (Table I). Intact T-DNA ends were on average 1.7-fold more associated with filler DNA than were T-DNA junctions with deleted T-DNA ends (Table I). A Fisher exact test was performed that verified that the correlation between the intactness or the processing of T-DNA ends, and the presence of filler DNA or microhomology, respectively, was significant (P < 0.05; see “Materials and Methods” and Supplemental Data). However, these conclusions should be considered with caution because of the small number of analyzed junctions.

Characterization of Filler DNA

Origin of Filler DNA at T-DNA/Plant DNA Junctions

Analysis of DSB lesions repaired via NHEJ in Arabidopsis has shown that none of 40 analyzed junctions had filler insertions (Kirik et al., 2000). Because of the striking difference with our observation that in Arabidopsis 27 of 67 T-DNA junctions (approximately 40%) harbor filler insertions (Table I), we investigated the origin and structure of the filler sequences found at the T-DNA junctions.

The origin of the filler DNA sequences was determined by comparative sequence analysis. The results of this analysis are presented in Figure 1A. For a full overview, containing all filler sequences and their origin analysis, see Supplemental Data. None of the filler sequences larger than 17 bp had a complete colinear sequence identity with the Arabidopsis genome sequence, indicating that the observed filler insertions did not consist of uninterrupted sequence blocks duplicated from the Arabidopsis genome. For 15 of the 17 T-DNA/plant DNA junctions with intermediate filler DNA of 6 bp or more, the filler DNA was made up of short duplicated sequences, identical to sequence motifs occurring in either the plant DNA surrounding the T-DNA insertion point and/or the T-DNA plasmid sequence (Fig. 1; Supplemental Data; plant DNA and T-DNA plasmid-derived repeats are indicated above or below the junction sequence, respectively). The remaining two filler insertions of 8 and 11 bp did not contain a sequence motif longer than the determined thresholds for reporting sequence identities. Therefore, in these two cases, we were unable to discern mechanistic relevance and by chance occurrence (see “Materials and Methods”). Of the 15 analyzed filler DNAs for which the origin could be attributed, four contained sequence motifs, identical to the plant DNA target or the T-DNA end, four had sequence blocks identical to the plant DNA target only, and seven consisted of sequence identities with the T-DNA end only (Fig. 1; Supplemental Data).

The first class of filler DNAs, with both plant DNA- and T-DNA-derived identical sequence blocks, is clearly illustrated by junction kg353LB-21 (Fig. 1A-3). This T-DNA/plant DNA junction had a 13-bp filler sequence, made up of five identified, different, and overlapping sequence motifs. Two sequence blocks of 9 bp (Fig. 1A-3, motifs 9 and 10) were identical to sequence motifs present in the plant target, whereas the other motifs, one of 7 bp (Fig. 1A-3, motif 13) and two of 6 bp (Fig. 1A-3, motifs 11 and 12), were identical to sequence blocks occurring in the T-DNA border sequence. The sequence identities with the plant genome corresponded either with the plant DNA immediately adjacent to the left border end of the T-DNA (Fig. 1A-3, motif 9) or with the plant DNA adjacent to the opposite T-DNA end of the same T-DNA insert (Fig. 1A-3, motif 10). The different sequence motifs, which make up the filler, shared small stretches of microhomology of up to a few bases, and microhomology was usually restricted to the ends of adjacent motifs. Moreover, microhomology was also observed between the filler DNA and the T-DNA end and/or the plant DNA end (Fig. 1A-3, motif 13). On the other hand, sequence blocks with a full overlap were also detected. For junction kg353LB-21, motif 12 lies entirely within motif 9 (Fig. 1A-3). Similar results were obtained for the other filler DNA sequences (see Supplemental Data), and mutual microhomology seemed to be a general characteristic of all sequence identities that make up a filler sequence at T-DNA/plant DNA junctions. Furthermore, junction kg353LB-21 also nicely illustrated that sequence identities that make up a filler can be in direct or inverted orientation relative to their template DNAs (Fig. 1, solid or arrowed line, respectively). One 9-bp sequence block, identical to a sequence motif present in the plant target (Fig. 1A-3, motif 9), was in inverted orientation, whereas the other four sequence identities were in direct orientation relative to their template DNAs. In general, 20% of all sequence identities in the filler had an inverted orientation, but the majority was in direct orientation relative to their template DNA.

The second class of filler sequences, containing sequences identical with the plant target only, is illustrated by junction kg353RB-8 (Fig. 1A-1). This right border T-DNA/plant DNA junction harbored a 51-bp filler segment composed of three sequence motifs, identical to sequence blocks occurring in the
Figure 1. Origin of T-DNA at plant DNA/T-DNA junctions derived from the same T-DNA loci.

(A) Origin of T-DNA at plant DNA/T-DNA junctions derived from the same T-DNA loci.

1) kg301B-19: plant/LB-5

2) CK1761B-6: plant/M-7

3) kg501B-21: plant/D3R-29

4) kg501B-8: plant/KL-8

(B) Origin of T-DNA at randomly linked T-DNA junctions.

1) kg714-8: RE-745LL-8B

2) kg509-7: RE-123181B-425

(C) Origin of T-DNA at inserted repeat T-DNA junction.

1) kg65-12: RE-1049RB-2679

2) kg621-18: RE-31251B-2136

(Figure legend continues on following page)
Origin of Filler DNA at T-DNA/T-DNA Junctions

Because the frequency of filler insertions differed in T-DNA/plant DNA junctions and DSB-repaired plant DNA/plant DNA joints, we also analyzed the frequency and origin of filler insertions at T-DNA/T-DNA junctions to investigate the influence of the plant DNA end in a T-DNA/plant DNA joining re-

action. Of eight junctions between tandemly linked T-DNAs, three contained filler DNA sequences of 1, 21, and 38 bp; of three junctions between T-DNA-inverted repeats over the left border, one junction harbored an intermediate 21-bp filler DNA, and of two junctions between two right border T-DNA ends, both contained filler sequences of 4 and 40 bp. Taken together, six of 13 linked T-DNAs harbor filler insertions at their junction (see Supplemental Data).

For the two junctions between tandemly linked T-DNAs with filler DNA of 21 and 38 bp, the origin of the filler sequences was determined (Fig. 1B). For these two junctions, the origin of the filler was attributed to T-DNA sequences. The filler DNA at junction kd73-4 (Fig. 1B-1) was identical to a sequence motif found in the T-DNA sequence 1,181 bp distant from the T-DNA border end. It is the only filler reported in this study that consisted of an uninterrupted sequence motif. In contrast, the 38-bp filler sequence at junction kg150-7 (Fig. 1B-2) is a nice example of how alternating sequence motifs identical to different T-DNA sequences make up a filler DNA. Similar results were observed for the filler DNAs of two inverted repeat junctions (Fig. 1, C-1 and C-2).

In addition, we analyzed whether the observed filler DNAs at linked T-DNAs could be attributed to Arabidopsis genomic sequences. In none of the filler DNAs could an identical sequence longer than 15 bp be detected with 100% identity with the plant genome. Only for junction kd12-1, a 16-bp plant DNA sequence was found that overlapped with two T-DNA-located sequence identities (Fig. 1C-2, italicized sequence). Nevertheless, the statistical significance of this 16-bp sequence motif, identical to a sequence block occurring in the plant genome, is difficult to evaluate because the surrounding plant DNA for the kd12 T-DNA locus had not been characterized. In conclusion, our findings with regard to the origin of the filler sequences at T-DNA/T-DNA junctions resemble those for T-DNA/plant DNA junctions.
DISCUSSION

That T-DNAs can be captured at DSB sites in tobacco (Salomon and Puchta, 1998), that NHEJ enzymes in yeast are involved in T-DNA integration (van Attikum et al., 2001), and that Ku80- and DNA ligase IV-deficient plants exhibit a reduced T-DNA integration efficiency (Friesner and Britt, 2003) have been taken as evidence for the role played by NHEJ mechanisms in a pathway for T-DNA integration. However, when we compare the outcome of the joining reaction at the T-DNA junctions with junction characteristics of repaired DSBs in plant cells, some remarkable differences appear. First, NHEJ-repaired junctions in Arabidopsis are characterized by the occurrence of large deletions ranging up to 2,207 bp (Kirik et al., 2000). T-DNA insertions in Arabidopsis are normally associated with much smaller target site deletions. We characterized target site deletions for eight single-copy CK2L-transgenic lines, and we found that on average, 48 bp of the genomic pre-insertion target site are deleted (data not shown), confirming earlier findings that on average the target site deletion in the host genome is small after T-DNA integration (Krysan et al., 2002; Meza et al., 2002). However, the most conspicuous difference is that approximately 40% of all T-DNA junctions harbor filler, whereas filler insertions are absent at NHEJ-repaired junctions in Arabidopsis (Kirik et al., 2000). Moreover, not only the frequency of filler insertions seems different, but also the composition of T-DNA-associated filler can be discriminated from filler insertions found at repaired DSBs in plant species other than Arabidopsis (Ralston et al., 1988; Wessler et al., 1990; Gorbunova and Levy, 1997; Salomon and Puchta, 1998; Yan et al., 1999). Filler insertions at T-DNA/plant DNA junctions consist of several, predominantly short, sequence motifs that are identical to sequence blocks occurring in: (a) the immediate surrounding of the plant T-DNA target site, (b) the T-DNA sequence, or (c) both. The T-DNA target site-derived motifs were identical to the pre-insertion site deletion or to the plant DNA adjacent to either site of the T-DNA, whereas T-DNA-derived sequence identities were scattered along the whole T-DNA sequence. Alternatively, chimeric sequence motifs that form internal repeat structures in the filler DNA sequence are observed. The majority of the sequence motifs are in direct orientation relative to their template DNA (Fig. 1; Supplemental Data). In contrast, the majority of the filler insertions found at DSB-repaired junctions in tobacco and maize are made up of simple, uninterrupted sequence blocks identical to sequences of the plant genome. Analogously, in yeast, in which illegitimate recombination is normally not accompanied by insertion of filler DNA, filler sequences are observed at the junctions upon T-DNA integration into the yeast genome (Bundock and Hooykaas, 1996). Moreover, DNA ligase IV and Ku80, which are required for repair of DNA damage in Arabidopsis, seem not to be involved in the integration of T-DNA into plants (Gallego et al., 2003; van Attikum et al., 2003). Taken together, although DSB repair and T-DNA integration might be associated, both mechanisms probably have some exclusive and specific characteristics. The observed differences are seemingly linked to the T-DNA integration process, meaning that the T-DNA may follow a pathway of its own during T-DNA integration rather than an entirely plant cell-determined route.

The differences between T-DNA integration and DSB repair might be explained in several ways. First, the T-DNA is transferred to the plant nucleus as a nucleoprotein complex, and T-DNA-associated factors or cotransferred virulence proteins might influence the outcome of the joining reaction. After transposon excision, transposition-specific proteins probably play a role in protecting and repairing the broken sites (Rinehart et al., 1997; Gorbunova and Levy, 1999). We show that right and left T-DNA borders are joined to the plant DNA in a similar manner (Table I), implying that the right border-associated VirD2 does not play a major role in it, confirming the findings of Ziemienowicz et al. (2000). However, other—yet unidentified—virulence factors might take part in the T-DNA joining process. Second, subtle differences between T-DNA integration and DSB repair may affect the NHEJ process. In our opinion, complex filler at T-DNA junctions could result from unstable interactions between the plant DNA target site and the invading T-DNA end. DNA repair via NHEJ has been shown to involve DNA end-stabilizing factors, such as DNA end-binding Ku proteins, a DNA-dependent protein kinase, the XRCC4/DNA ligase IV complex, and the Mre11/Rad50/Xrs2 complex (for review, see Tsukamoto and Ikeda, 1998; Pfeiffer et al., 2000), for which analogs in Arabidopsis have been identified (Hartung and Puchta, 1999; West et al., 2000; Gallego et al., 2001; Daudal-Cotterell et al., 2002; Tamura et al., 2002). This DNA end-joining protein complex might protect the DNA ends against nuclease-mediated degradation and also stabilize the intramolecular association of the interacting DNA ends (Tsukamoto and Ikeda, 1998). Our results suggest that when filler insertions are observed at T-DNA/plant DNA junctions, the initial interaction between the invading T-DNA end and the plant target is not stabilized immediately by this kind of protein complexes. Before stabilization, a free 3′ protruding T-DNA end either from the left or right T-DNA end lands and screens for microcomplementarity. Upon landing, the free 3′ end is taken as primer for simultaneous template-based DNA synthesis. Because at this stage the initial interaction is not stabilized by host-dependent NHEJ-associated protein complexes, the T-DNA takes off and lands in the neighborhood. Landing and taking off is repeated regularly. Different regions, such as the surrounding of the T-DNA target site and the T-DNA plasmid
sequence, can be invaded, resulting in a patchwork-like filler sequence. It is tempting to speculate that at least during this initial interaction, DNA end-stabilizing protein complexes are absent, whereas a repair-polymerase complex might be involved. The presence of polymerase activity and the absence of exonuclease activity during this initial interaction might explain why filler DNA has been associated more frequently with intact T-DNA ends than with T-DNA ends that are joined via a microhomology-based mechanism. In contrast, DSB repair sites might not recruit DNA replication enzyme activities in combination with the plant’s NHEJ machinery, which might result in a faster and stable interaction between both DSB ends.

MATERIALS AND METHODS

Transgenic Lines

Two populations of transgenic Arabidopsis C24 lines were used for amplification of T-DNA junction fragments. A first population included 21 transgenic lines that were cotransformed with two Agrobacterium tumefaciens strains containing each a different T-DNA, of which six transgenic lines (the kd lines) were cotransformed with the T-DNAs from plasmids pAK1202 and pAD120l (De Neve et al., 1997). 14 lines (the kg lines) with those from plasmids pAK1202 and pAG1201 (De Neve et al., 1997), and one (Hsb-K2L 67/1) with those from plasmids pHsb and pK2L610 (De Buck et al., 1999). A second population included 12 CK2L-transgenic lines that were transformed with the A. tumefaciens strain C58C1RIP (pGV2260 and pK2L610; De Buck et al., 2000) and subsequently screened for single-copy inserts by T-DNA fingerprinting (Theuns et al., 2002).

Amplification of T-DNA Junctions

All T-DNA junctions were amplified with T-DNA fingerprinting (Theuns et al., 2002). From each transgenic line, a T-DNA fingerprint was generated either with Msel- or with BflI-digested DNA as template DNA. Subsequently, the sequence of these fragments was determined after amplification upon elution (Windels et al., 2001). All transgenic lines were fingerprinted by using a right and left border-specific anchor primer, situated 129 and 277 bp internally to the right and left border repeats, respectively.

Characterization of the T-DNA Junctions and Filler Sequences

Pair-wise sequence alignment was used to screen for sequence identity between the amplified junctions and the T-DNA plasmid. Sequence identity between the plant region of the T-DNA junction fragment and the Arabidopsis genome was determined with the BLAST algorithm (Altschul et al., 1990) against the Arabidopsis database (http://www.arabidopsis.org/BLAST, Huala et al., 2003). We first screened the complete Arabidopsis genome sequence with the BLAST algorithm for the filler sequences larger than 17 bp. In this manner, we could determine whether the observed filler DNAs were uninterrupted sequence blocks that had been copied from the Arabidopsis genome. A threshold was set at 17 bp because the probability of finding a 17-bp identical sequence motif in a 125-Mb sequence, which is the length of the complete Arabidopsis genome sequence (The Arabidopsis Genome Initiative, 2000), is less than 0.01.

Subsequently, for those junctions for which a filler DNA segment of 6 bp or more was detected, their origin was analyzed as follows. First, we screened the filler DNA at plant DNA/T-DNA junctions for sequence motifs, identical either in direct or inverted orientation to sequence blocks occurring in the 200-bp plant DNA region surrounding the T-DNA integration point. For those T-DNA inserts for which both T-DNA junctions were amplified, the filler sequence was easily compared with the target site deletion and the plant DNA adjacent to the opposite T-DNA junction. For the filler in T-DNA junctions for which the corresponding, opposite T-DNA junction had not been amplified, sequence identities with the target site deletion or the plant DNA adjacent to the opposite T-DNA border could not be distinguished. We chose to evaluate the statistical relevance of the reported sequence identities by means of an experimental statistical test. To perform a numerical, statistical analysis for each of the observed identical sequence motifs is extremely difficult because the statistical relevance is influenced by several parameters, such as length of the sequence identity, distance between the reported identity and the originating sequence, number of identities observed per filler insertion, and length of the filler sequence itself. Therefore, we compared the actual, observed filler sequences with the 200-bp plant target regions of other, characterized T-DNA inserts. Because this analysis revealed that sequence identities of 8 bp and shorter could occur simply by chance when a 200-bp plant region is screened (see Supplemental Data), only sequence identities of at least 9 bp with the plant target site were reported. Second, we compared the filler DNA with the first 100 bp of the T-DNA sequence that are situated immediately adjacent to the filler DNA. To determine a threshold for statistically relevant sequence identities when this 100-bp T-DNA border region was taken into account, the actual filler sequence was shuffled and compared with 100-bp T-DNA border region. In that case, identities of at least 6 bp in length were found to be relevant. To detect identical sequences in the filler with T-DNA plasmid sequences that were not situated immediately adjacent to the filler DNA, we performed a pair-wise comparison between the T-DNA plasmid and filler sequences. Here, only sequence identities of at least 11 bp are reported. The probability of finding an 11-bp repeat in a 12-kb sequence (which is the average length of the T-DNA plasmids used in this study) is less than 0.01.

Fisher Exact Test for Statistical Significance

To test whether the differences observed with regard to the degree of T-DNA processing and the frequency of either filler DNA or microhomology at the T-DNA junction were significant, a Fisher exact test was performed. A two-by-two contingency table was constructed with the following values: A1, 3; A2, 30; B1, 8; and B2, 19 (see Supplemental Data). A Fisher two-sided exact P value of 0.043 was obtained.

ACKNOWLEDGMENTS

The authors thank Isabel Roldán-Ruiz, Rosalinde van Lipzig, and Sergei Kushnir for critical reading of the manuscript and helpful comments, and Martine De Cock for layout.

Received May 26, 2003; returned for revision July 20, 2003; accepted August 21, 2003.

LITERATURE CITED


