Genes Genet. Syst. (2020) **95**, p. 173–182

Effects of *Arabidopsis* Ku80 deletion on the integration of the left border of T-DNA into plant chromosomal DNA via *Agrobacterium tumefaciens*

Ryouhei Yoshihara[1*], Yuka Mitomi[1], Maki Okada[1], Hanako Shibata[1], Mai Tanokami[1†], Yurie Nakajima[1], Hideyuki Inui[2], Yutaka Oono[3], Hiroyuki Furudate[1] and Shuuitsu Tanaka[1]

[1]*Department of Regulatory Biology, Faculty of Science, Saitama University, Saitama, Saitama 338-8570, Japan*
[2]*Biosignal Research Center, Kobe University, Kobe, Hyogo 657-8501, Japan*
[3]*Takasaki Advanced Radiation Research Institute, National Institutes for Quantum and Radiological Science and Technology, Takasaki, Gunma 370-1292, Japan*

(Received 5 December 2019, accepted 19 March 2020; J-STAGE Advance published date: 26 August 2020)

T-DNA integration into plant chromosomal DNA via *Agrobacterium tumefaciens* can be achieved by exploiting the double-strand break repair system of the host's DNA. However, the detailed mechanism of T-DNA integration remains unclear. Here, a sequence analysis of the junction sequences of T-DNA and chromosomal DNA was performed to assess the mechanism of T-DNA integration. T-DNA was introduced into *Arabidopsis* wild-type and NHEJ-deficient *ku80* mutant plants using the floral dip method; the junctions of the left border (LB) of T-DNA were subsequently analyzed by adapter PCR. The most frequent junction of the LB of T-DNA with chromosomal DNA was of the filler DNA type in both lines. The lengths of direct or inverted repeat sequences within or around the filler DNA sequence were greater in the *ku80* mutant. In addition, the frequency of T-DNA integration near a transcription start site was significantly higher in the *ku80* mutant. Our observations suggest that the presence of the Ku80 protein affects the location of the integration of T-DNA and the pattern of formation of repeat sequences within or around the filler DNA during LB integration into chromosomal DNA.

Key words: *Agrobacterium tumefaciens*, double-strand break, repair pathway, T-DNA integration

INTRODUCTION

Agrobacterium tumefaciens is a gram-negative phytopathogenic soil bacterium (Young et al., 2001) that causes crown gall disease in a wide range of host plants. As this pathogenic bacterium has the ability to naturally introduce genes into host plant cells, it has been widely used for plant transformation. Pathogenic *A. tumefaciens* harbors a tumor-inducing (Ti) plasmid that carries transfer DNA (T-DNA) and virulence (*vir*) genes that are required for gene transfer (Komari et al., 1986). T-DNA is a DNA region located between two direct repeat sequences, the left border (LB) and the right border (RB). The single-stranded T-strand DNA is processed from T-DNA and transmitted into the host plant cells. Proteins encoded by the *vir* genes are involved in T-DNA transfer to the host plant cells (Gelvin, 2017). The VirD2 protein covalently attaches to the 5′ end of the T-strand. Subsequently, this DNA–protein complex is coated with VirE2 proteins to form a T-complex, which is transported to the plant host nucleus, where the T-DNA is integrated into plant chromosomal DNA. T-DNA integration is a complex process that involves not only single T-DNA integration, but also multi-locus integration, or T-DNA-to-T-DNA linkages (Gelvin, 2017). Although the details of the mechanism of T-DNA integration into the plant chromosomal DNA have not been elucidated fully, it is suggested that DNA double-strand break (DSB) repair systems are involved in

Edited by Tetsu Kinoshita
* Corresponding author. ryoshihara@mail.saitama-u.ac.jp
† Present address: Tochigi Prefectural Agricultural Experiment Station, Utsunomiya, Tochigi 320-0002, Japan
DOI: http://doi.org/10.1266/ggs.19-00064

this process (Tzfira et al., 2004).

Non-homologous end joining (NHEJ), homologous recombination (HR), and alternative end joining (alt-EJ) mechanisms such as microhomology-mediated end joining (MMEJ), are candidate pathways for T-DNA integration (Tzfira et al., 2004; Citovsky et al., 2007; Bourras et al., 2015). NHEJ is a repair pathway in which DSBs are ligated directly without the need for homologous sequences. The Ku70/Ku80 heterodimer (referred to here as the Ku protein), which acts as a loading protein for other NHEJ factors, is known to bind to both ends of the DSB and therefore aligns them. After this processing, the ends of the DSB are joined by the action of the DNA ligase IV/XLF/XRCC4 complex (Davis and Chen, 2013).

HR is an error-free DSB repair pathway in which a single-stranded DNA (ssDNA) region is produced at the ends of the DSB via end resection. Subsequently, this ssDNA region is inserted into a homologous sequence on a sister chromatid via the action of recombination proteins, to form a D-loop structure. After the formation of the D-loop, the DSB is repaired by break-induced replication, synthesis-dependent strand annealing, or double-strand break repair (Li and Heyer, 2008; Nimonkar et al., 2011; Shibata et al., 2014).

Alt-EJ is an error-prone repair pathway that can bind a DSB without the Ku protein (Dueva and Iliakis, 2013). In particular, MMEJ exploits the microhomology that is present in both strands of the DSB (Ma et al., 2003; Yu and Gabriel, 2003). In this process, a DNA synapse is formed between the microhomology regions of two 3′-overhanging DSB ends. DNA polymerase θ (Polθ) synthesizes a DNA strand at the ssDNA gap regions in the DNA synapse (Black et al., 2016). In some cases, microhomology is also used in the process of NHEJ. The lengths of the microhomology regions often observed at the DNA junction site and that are joined by NHEJ or alt-EJ are 0–4 bp or 2–20 bp, respectively (Chang et al., 2017; Pannunzio et al., 2018). In the process of alt-EJ, formation of filler DNA, i.e., an insert with several nucleotides, is also observed at the repaired junction of the DSB. The formation of the filler DNA results from the extension of the 3′-ssDNA tail of the DSB by Polθ-mediated terminal transferase activity during end joining (Black et al., 2016).

Short homologous sequences are often observed between T-DNA and plant chromosomal DNA at the T-DNA insertion site (Bechtold and Pelletier, 1998; Brunaud et al., 2002). Conversely, other studies report that T-DNA is randomly integrated into the plant chromosomal DNA and that short homologous sequences between T-DNA and the site of integration are not required for the process (Bourras et al., 2015; Gelvin, 2017). Filler DNA has also been observed at the T-DNA and chromosomal DNA junctions (Windels et al., 2003). The frequency of gene targeting via HR is extremely low, even when T-DNA with a long homologous sequence to a target site is used for transformation (Iida and Terada, 2005). These facts suggest that NHEJ and alt-EJ, rather than HR, are the major pathways of T-DNA insertion. However, the role of NHEJ and of its backup system, alt-EJ, in T-DNA integration into plant chromosomal DNA remains the subject of much debate. For example, based on the analysis of the transformation frequency in NHEJ-deficient mutants, some studies have demonstrated the involvement of NHEJ in T-DNA integration (Friesner and Britt, 2003; Li et al., 2005; Mestiri et al., 2014), whereas others have indicated that NHEJ deficiency had little effect on transformation frequency (Gallego et al., 2003; van Attikum et al., 2003; Park et al., 2015). Recently, Polθ was reported to play an essential role in T-DNA integration (van Kregten et al., 2016); however, the detailed mechanism of T-DNA integration into chromosomal DNA remains unclear. Current models of T-DNA integration into chromosomal DNA suggest that T-DNA insertion is initiated by homology-dependent annealing of the LB or its adjacent sequences of the T-strand (Brunaud et al., 2002; van Kregten et al., 2016). As the first step of elucidating the mechanism of T-DNA integration via *A. tumefaciens*, in the current study, T-DNA was introduced into the chromosomal DNA of *Arabidopsis thaliana* wild-type and Ku80-deficient mutant plants via *A. tumefaciens* and the DNA sequences of the LB-side junctions were analyzed to evaluate the involvement of NHEJ in T-DNA integration.

MATERIALS AND METHODS

Plants and bacterial strains *Arabidopsis thaliana* ecotype Columbia was used as the wild-type line. The *AtKu80* knockout mutant (SALK_016627) was obtained from the SALK T-DNA collection (Alonso et al., 2003; http://signal.salk.edu). In this mutant, T-DNA was inserted into the 10th intron of the *Ku80* gene, resulting in disruption of the NHEJ pathway (Jia et al., 2012). Confirmation of the T-DNA insertion in the *Ku80* gene and selection of homozygous mutants for the T-DNA were performed by PCR analysis. Primer sets LBb1/AtKu80-1F and AtKu80-1F/AtKu80-1R were used to confirm the presence and absence, respectively, of T-DNA in *Ku80* sequences (Supplementary Fig. S1A, Supplementary Table S1). The sensitivity of the mutant to DSB induction was determined by evaluating its sensitivity to γ-rays (Supplementary Fig. S2). Plants were grown in soil or in half-strength Gamborg's B5 basal medium containing 1% sucrose, 1/1,000 Hyponex (Type: 5-10-5, Hyponex), and 0.7% agar at 25 °C under 14 h/10 h light/dark conditions. *Agrobacterium tumefaciens* GV3101 (pMP90R) was used for transformation of *A. thaliana*.

Construct for plant transformation The binary vec-

tor pAH1GB-GT2 was constructed as follows. Approximately 1 kb upstream of the stop codon of *A. thaliana* histone H1 (AT1G06760) genomic sequence was amplified using primers AthH1-F1/R1 (Supplementary Table S1) by PCR. This PCR product contained 74 bp of the 5′ untranslated region (UTR) from the start codon and the full-length histone H1 sequence except for the stop codon, but did not contain the promoter region. The C-terminal end of this promoterless histone H1 was ligated in-frame to the synthetic green fluorescent protein (sGFP)-NOS terminator sequences from the 35SΩ-sGFP(S65T) plasmid (Chiu et al., 1996) via a BamHI site and inserted into pBluescript II SK(+) to construct a promoterless histone H1-GFP fusion sequence (H1-GFP). The bialaphos resistance gene (*Bar*) cassette from pARK22 (Sawasaki et al., 1994) was ligated to the outside of the NOS terminator of the promoterless histone H1-GFP fusion cassette via an EcoRI site. A 1-kb DNA fragment containing the 3′-UTR region from the histone H1 stop codon and its downstream genome DNA region (H1DSR) was amplified with the AthH1-F8/AthH1-R8 primer pair (Supplementary Table S1) and inserted into a XhoI site with the In-Fusion HD Cloning Kit (Takara Bio). The resulting H1-GFP/*Bar*/H1DSR DNA fragment was amplified with the AthH1-F10/AthH1-R2 primer pair (Supplementary Table S1), and inserted into pBI121 (Jefferson, 1987) treated with EcoRI and NheI using the In-Fusion HD Cloning Kit, resulting in the final construct pAH1GB-GT2. The annealing positions of all primers are summarized in Supplementary Fig. S1B.

Transformation of *A. thaliana* Transformation of *A. thaliana* was conducted by the floral dip method (Clough and Bent, 1998). An aliquot (500 µl) of the preculture of *A. tumefaciens* carrying pAH1GB-GT2 grown overnight in Lennox broth (10 g/l Bacto tryptone, 5 g/l yeast extract, 10 g/l NaCl) liquid medium was used to inoculate a fresh 250-ml volume of Lennox broth liquid medium, which was then incubated at 30 °C overnight. The bacterial cells were collected by centrifugation and the pellet was resuspended in 250 ml of 5% sucrose. *Arabidopsis* inflorescences were dipped into the bacterial suspension containing 12.5 µl of Silwet L-77 (Momentive Performance Materials). We conducted three independent dipping experiments. Nine individual wild-type and *ku80* plants were immersed in the same bacterial suspension in each dipping experiment. Transformants were selected on basal medium containing 12 µg/ml bialaphos. Transformation frequency was expressed as the ratio of the number of bialaphos-resistant plants to the total number of germinated T_1 plants from selected T_0 lines.

Observation of *Arabidopsis* histone H1-GFP fusion protein T_2 seeds from transformed wild-type or *ku80* plants were aseptically sown on Gamborg's B5 basal medium containing 12 µg/ml bialaphos. Three individual bialaphos-resistant plants were selected from each transformant and placed on a glass slide with the roots covered with a cover slip. GFP fluorescence in the root cells was observed with a FLUOVIEW FV1000-F confocal laser microscope (Olympus).

Analysis of the T-DNA insertion site To analyze the location of the T-DNA integration site in plant chromosomal DNA and the features of the junction sequences, adapter PCR was conducted using chromosomal DNA from the T_2 plant as a template. The procedure of adapter PCR used was based on previously published methods (Siebert et al., 1995; Kasajima et al., 2010). The adapter sequence was attached to a restriction endonuclease site upstream of the LB junction site, and the region of the junction sequence was cloned by PCR using primers for the adapter and T-DNA sequences. For preparation of the adapter, 1 mM each of two oligonucleotides, namely 5′-CCACGCTTAGGGTAAGTAACCCTCACTGTCGCCGCCACAGGGGCTGGT-3′ and 5′-[PHO]CTAGACCAGCCC[AmC7]-3′, were mixed (Supplementary Fig. S1C). This mixture was incubated at 95 °C for 5 min and the temperature was gradually decreased by 5 °C every 5 min to 40 °C with a thermal cycler. Chromosomal DNA, digested with XbaI, NheI, SpeI or AvrII, was purified by phenol/chloroform extraction and ethanol precipitation, and the adapter was ligated to purified DNA in a 20-µl reaction at 16 °C for 2 h with 5 U of T4 DNA ligase (Nippon Gene). The first PCR cycle was conducted in a 25-µl reaction containing 3 µl of adapter-ligated DNA solution with primers NCU02695-FW9/AthH1-R4 (Supplementary Table S1). For the second PCR cycle, 1 µl of the first PCR reaction mixture was added and primers NCU02695-FW12/AthH1-R10 (Supplementary Table S1) were used for amplification of the target sequence. Primers AthH1-F11 and GFP-RV3 (Supplementary Table S1), which amplify from 216 bp upstream of the histone H1 transcription start site to the 381-bp position of GFP when T-DNA was integrated by HR, were used for detection of gene targeting.

PCR products were analyzed by agarose gel electrophoresis. The amplified target products purified from the agarose gel were sequenced directly using LBb1 primer. The NCBI BLAST website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was exploited for analysis of the position and sequence features of the insertion site. The annealing positions of all primers are summarized in Supplementary Fig. S1B and S1C.

RESULTS

Effects of the host NHEJ pathway on T-DNA introduction via *A. tumefaciens* The effects of plant DSB repair pathways on T-DNA integration into plant chromo-

somal DNA via *A. tumefaciens* were analyzed in the current study. The pAH1GB-GT2 plasmid was constructed for these analyses (Fig. 1A). This construct carried the promoterless histone H1-GFP fusion sequence at the LB side (the distance from the LB to the 5′-end of the promoterless histone H1 was 638 bases), whereas the 3′-UTR plus the untranscribed sequence of the histone H1 gene were placed at the RB side (the distance between the RB and the 3′-end of the homologous sequence was 242 bases). The transportation of this T-DNA segment (Fig. 1) into the plant cell nucleus and its integration at the host's histone H1 locus by HR would lead to the detection of GFP fluorescence. The insertion of the LB side of the promoterless histone H1-GFP fusion gene downstream of any active promoter would also lead to the detection of nuclear GFP fluorescence.

To elucidate the contribution of NHEJ and HR to T-DNA integration, *A. thaliana* wild-type and NHEJ-deficient *ku80* mutant plants were transformed with *A. tumefaciens* harboring pAH1GB-GT2 via the floral dip method. The transformation frequency was slightly lower in the *ku80* mutant, but the difference was not statistically significant (Fig. 1B, Supplementary Table S2). A total of 92 and 80 T_2 plants were obtained from the wild-type and the *ku80* mutant plants, respectively. Confocal laser microscopic analysis of the root cells of the transformants showed the presence of GFP fluorescence in 21 plants out of the 36 *ku80* background T_2 plants tested, whereas no plants with GFP fluorescence were obtained from the 42 wild-type background T_2 plants (Fig. 1C and Supplementary Fig. S3).

To investigate whether GFP fluorescence occurred as a consequence of HR, a PCR analysis of DNA extracted from GFP-positive plants was performed using primers that amplified the DNA region encompassing the segment from the host histone H1 promoter to the GFP

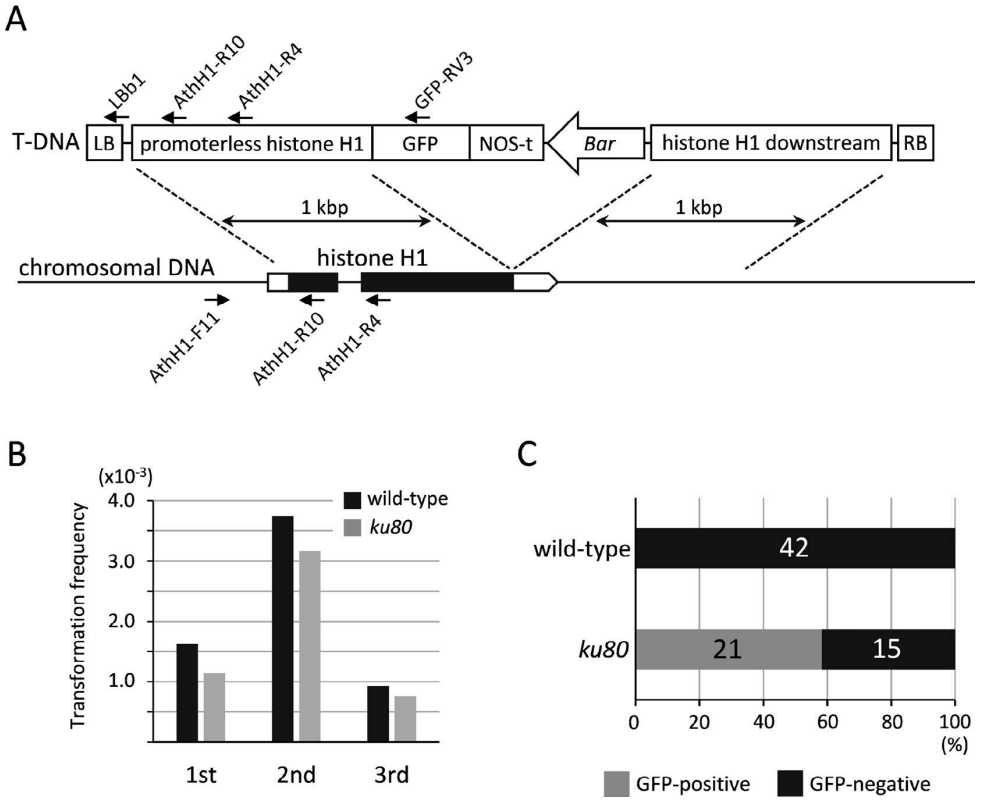

Fig. 1. Construct used in this study for transformation of wild-type and *ku80 Arabidopsis thaliana*. (A) The structure of the T-DNA region on plasmid pAH1GB-GT2 and of the *Arabidopsis* histone H1 gene region. The construct has 1 kb of promoterless *Arabidopsis* histone H1 genomic sequence fused with the green fluorescent protein (GFP) gene at the C-terminal end and 1 kb of sequence downstream of the histone H1 coding region at the left border (LB) and right border (RB) sides, respectively. *Bar* is the bialaphos resistance gene. Black and white boxes on the chromosomal DNA denote the exons and untranslated regions, respectively. The annealing positions of primers and amplification directions are indicated by arrows. LBb1 was used for sequencing of junction sequences. AthH1-R4 and AthH1-R10 were primers for adapter PCR. The primer pair AthH1-F11/GFP-RV3 was used to check for T-DNA integration by HR. (B) Transformation frequency in three independent experiments. For each experiment, statistical analysis was performed by the χ^2 test. The threshold *P*-value of the 2-way χ^2 test was approximately 7.0×10^{-9}. (C) Percentage of transformants expressing or not expressing GFP localized in the nucleus. Numbers shown within the bars are the numbers of each transformant.

gene. However, no PCR product was obtained in any of the GFP-positive plants (data not shown). This result indicates that the nuclear GFP fluorescence observed for the *ku80* mutant was not the result of T-DNA integration via HR.

LB side of T-DNA is preferentially integrated near transcription start sites in the *ku80* mutant

Although no contribution of HR to T-DNA integration was observed under our experimental conditions, GFP fluorescence was detected in approximately 60% of the *ku80* mutants. To characterize the DNA sequence at the T-DNA integration site and to elucidate the origin of the GFP fluorescence detected in the *ku80* background transformants, cloning of the LB flanking sequences of the T-DNA insertion site into plant chromosomal DNA was attempted by adapter PCR. Sixty-seven wild-type and *ku80* background T_2 plants, regardless of the presence of GFP fluorescence, were subjected to adapter PCR. Twenty-five and 24 sequences of the LB-side junctions with chromosomal DNA from wild-type and *ku80* mutant plants, respectively, were successfully analyzed (Supplementary Table S3). In the *ku80* mutant, the LBs tended to be integrated near a transcription start site of a gene other than the histone H1 gene. Approximately 24% and 63% of the LBs were integrated < 500 bp upstream or downstream from the transcription start site in the wild-type and *ku80* mutant plants, respectively (Fig. 2A, Tables 1 and 2). In this study, 75% (12/16) of GFP-expressing *ku80* mutants exhibited T-DNA integra-

Table 1. Length of filler DNA and of direct or inverted repeats, and distance of T-DNA insertion site from the nearest transcription start site for filler DNA type junctions

T_2 plant	Length of filler DNA (bp)	Direct or inverted repeat (bp)	≤ 500 bp from transcription start site	Nuclear GFP
WT-43	1	0	×	n
WT-44	3	0	×	n
WT-12	2	5	×	n
WT-66	4	6	○	n
WT-18	29	5	×	n
WT-28	12	5	×	n
WT-63	35	5	×	n
WT-22	4	6	×	n
WT-10	38	7	×	n
WT-11	7	7	×	n
WT-59	17	8	×	n
WT-26	6	9	×	n
WT-55	29	10	○	n
WT-51	16	12	○	n
ku-15	2	0	○	p
ku-27	2	0	×	n
ku-54	6	0	○	p
ku-11	10	6	×	n
ku-45	29	6	×	p
ku-8	19	7	○	p
ku-47	18	7	×	p
ku-51	29	11	○	p
ku-38	8	12	○	p
ku-12	20	15	○	n
ku-49	27	15	○	p
ku-34	67	17	○	n
ku-4	14	21	○	p
ku-53	71	28	×	p

WT- and ku- indicate wild-type and *ku80* background plants, respectively. Insertions occurring ≤ 500 bp from a transcription start site are shown as "○" and others are "×". "p" and "n" represent GFP-positive and -negative, respectively.

Table 2. Length of microhomology and distance of T-DNA insertion site from the nearest transcription start site for 0–1-bp or ≥ 2-bp microhomology type junctions

T_2 plant	Length of microhomology (bp)	≤ 500 bp from transcription start site	Nuclear GFP
WT-13	0	×	n
WT-58-1	0	×	n
WT-64	0	○	n
WT-3	1	×	n
WT-15	1	○	n
WT-29	2	×	n
WT-57	2	×	n
WT-49	3	×	n
WT-30	4	○	n
WT-54	4	×	n
WT-56	4	×	n
ku-16	0	○	p
ku-20	0	○	p
ku-10	0	○	p
ku-63-2	1	×	p
ku-65	2	○	p
ku-60	3	○	n
ku-9	4	○	p
ku-13	6	×	n
ku-41	6	×	n
ku-21	9	×	n

WT- and ku- indicate wild-type and *ku80* background plants, respectively. Insertions occurring ≤ 500 bp from a transcription start site are shown as "○" and others are "×". "p" and "n" represent GFP-positive and -negative, respectively.

tion near the transcription start site (Tables 1 and 2).

The presence of AtKu80 affects the joining of the LB of T-DNA with chromosomal DNA The deletion observed at the repair junctions in Ku-deficient human somatic cells is longer than that detected in wild-type cells (Fattah et al., 2010). In addition, the Ku protein prevents extensive DNA end resection by exonuclease (Mimitou and Symington, 2010). Therefore, Ku proteins are considered to be important for the protection of the DNA ends. However, in the current work, we observed deletion of T-DNA from most of the LB junction sequences from both wild-type (25/25) and *ku80* mutant (23/24) plants (Fig. 2B). The average size of the deletion of the LB region integrated into chromosomal DNA was 15 bp in both the wild-type and *ku80* mutant plants, with no substantial differences observed between the size profiles (Fig. 2B). These findings indicate that the Ku80 protein does not participate in the protection of the T-DNA LB against degradation.

It has been reported that, in addition to blunt ligated junctions having 0–1-bp microhomology, junction sequences of DSBs repaired by NHEJ and its backup system, alt-EJ, have insertion sequences (called filler DNA in the present study) or ≥ 2-bp microhomology (Chang et al., 2017). Therefore, as shown in Fig. 2C, the junction sequences analyzed were classified into three types (the filler DNA type, the 0–1-bp microhomology type and the ≥ 2-bp microhomology type) to evaluate the contribution of NHEJ and alt-EJ to LB-side integration into the chromosomal DNA (LB-to-ChrDNA). The frequency of these junction sequences of LB-to-ChrDNA was simi-

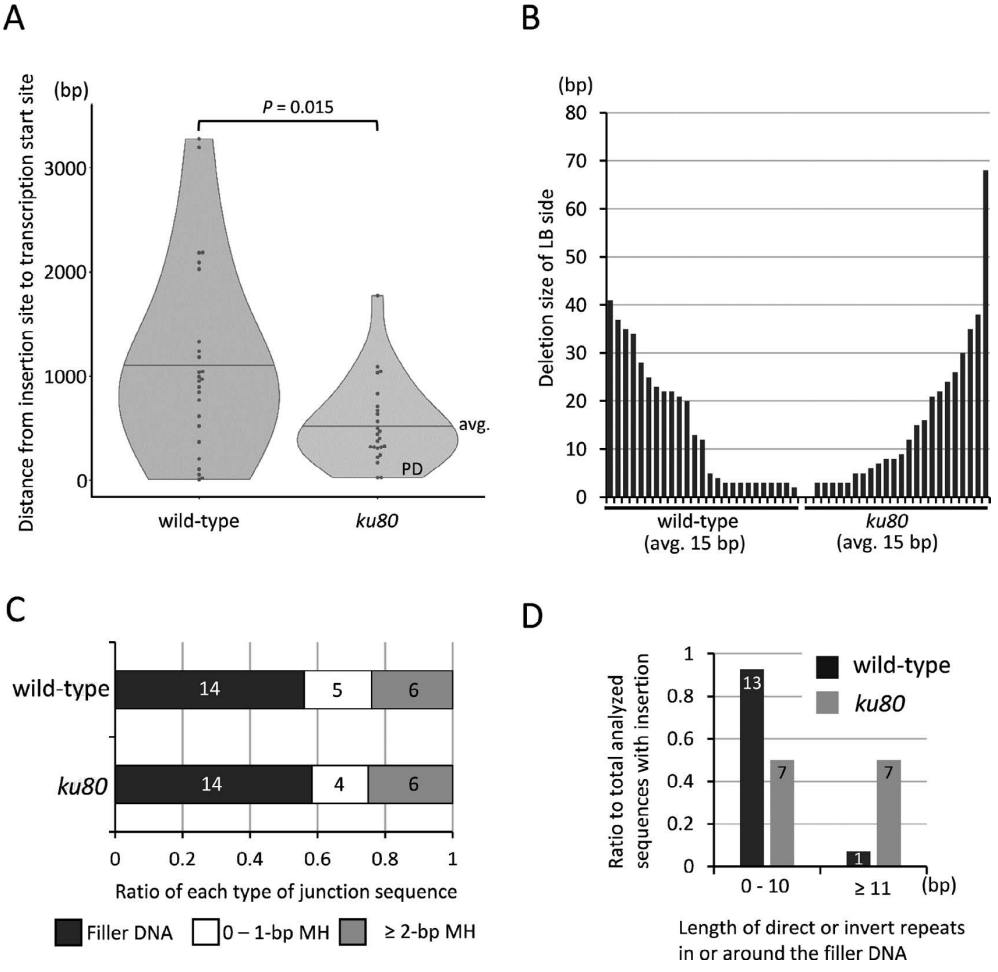

Fig. 2. Sequence analysis of the left-border (LB)-to-chromosomal DNA junction. (A) Distance from the T-DNA insertion site to the nearest transcription start site. The 5′ end of an mRNA confirmed in the NCBI database was defined as a transcription start site in this study. The Wilcoxon test was exploited for statistical analysis. "avg." and "PD" are average and probability density, respectively. (B) Deletion sizes of the LB side of T-DNA inserted into chromosomal DNA. "avg." is the average size of the deletion for each line. (C) Frequency of filler DNA, 0–1-bp microhomology (MH) and ≥ 2-bp MH observed at junction sites. (D) Comparison of the length of the direct or inverted repeats found in or around the filler DNA. The *P*-value of Fisher's exact test was 0.033. Numbers shown in the bars are the numbers of each transformant.

lar between the wild-type and the *ku80* mutant (Fig. 2C). Because NHEJ is able to join DSB ends without a homologous sequence, we expected that the frequency of 0–1-bp microhomology would decrease in the *ku80* mutant. However, almost the same frequency of the 0–1-bp microhomology was observed in the wild-type as in the *ku80* mutant, suggesting that NHEJ is not essential to produce junctions with 0–1-bp microhomology in LB-to-ChrDNA linkages. Junctions with filler DNA were the predominant type in both the wild-type and the *ku80* mutant plants. Although it depended on the structure of the DSB ends, the length of the filler DNA at the junction site is relatively short (i.e., ≤ 5 bp) in NHEJ-proficient mammalian cells (Black et al., 2016; Wyatt et al., 2016). Here, LB-to-ChrDNA junctions with a short filler DNA were minor types in both the wild-type and the *ku80* mutant plants (Table 1). Moreover, the average length of the filler DNA was similar between the wild-type and the *ku80* mutant plants, indicating that the Ku protein does not participate in the process of filler DNA formation primarily. Interestingly, direct or inverted repeat sequences were found within or around the filler DNA; moreover, the frequency of filler DNAs that were accompanied by repeat sequences > 10 bp was higher in the *ku80* mutant than in the wild-type plants (Fig. 2D, Table 1 and Supplementary Table S3). Although the number of junctions analyzed was small, T-DNA junctions with repeat sequences ≥ 10 bp tended to occur near a transcription start site (Table 1). These observations suggest that NHEJ affects the formation of filler DNA.

DISCUSSION

Although efforts to elucidate the sequences of T-DNA-to-chromosomal DNA junctions and debates regarding the mechanisms of T-DNA integration into chromosomal DNA have long been undertaken, the details of how T-DNA is integrated into chromosomal DNA are still not fully understood (Gelvin, 2017). In the present study, we have investigated the role of NHEJ in T-DNA integration, using a NHEJ-deficient *ku80* mutant to dissect the T-DNA integration process from the aspect of DSB repair pathways.

NHEJ contributes to the integration of the LB of T-DNA into chromosomal DNA The LB region of T-DNA tended to be inserted near transcription start sites in the *ku80* mutant. Moreover, filler DNA of the LB-to-ChrDNA junctions containing repeats with a length ≥ 10 bp tended to occur near transcription start sites. A larger number of T-DNA junctions containing the ≥ 10-bp repeat was detected in the *ku80* mutant than in wild-type plants. Regarding the 0–1-bp or ≥ 2-bp microhomology type of junctions, there seemed to be no correlation between the length of the microhomology and the insertion sites in wild-type plants; in contrast, in the *ku80* mutant, most of the LB-to-ChrDNA junctions with 0–4-bp, but not 6–9-bp, microhomologies were detected near a transcription start site (Table 2, Supplementary Table S3). This suggests that, in the absence of the Ku protein, longer microhomologies are required for joining the LB of T-DNA with a DSB end of chromosomal DNA that is located far from a transcription start site. Because regions that are associated with actively expressed genes and are involved in transcriptional regulation form open chromatin structures in the plant genome (Kodama et al., 2007; Zhang et al., 2012, 2014), we propose that, in the absence of the Ku80 protein, access of the LB region to DSBs occurs mainly in open chromatin regions. In mammalian cells, DSBs occurring in transcriptionally active regions are thought to be preferentially repaired by HR, while NHEJ can efficiently repair DSBs at untranscribed loci (Aymard et al., 2014). If NHEJ also shows a similar bias in plant cells, it may be difficult for the LB side of T-DNA to access transcriptionally silenced sites in the *ku80* mutant. The Ku80 protein may enhance T-DNA integration at DSB sites regardless of chromatin structure, resulting in random T-DNA insertions into chromosomal DNA. The majority (75%) of GFP-expressing *ku80* mutants exhibited T-DNA integration near a transcription start site (Tables 1 and 2) in this study. Although we could not eliminate the possibility that GFP was expressed from a segment other than the sequenced T-DNA, because multilocus T-DNA integration occurs frequently in the process of T-DNA integration, the higher frequencies of GFP expression observed in *ku80* mutant plants may be attributed to the preferential integration of T-DNA near a transcription start site in these plants. On the other hand, it is not possible to ignore the difference in the epigenetic state of the chromosomal location that may exist between the wild-type and the *ku80* mutant. Since epigenetic changes are closely related to chromosomal structure and gene expression, T-DNA insertion and expression of the introduced gene may be affected. To explain the precise mechanism of T-DNA integration, it would be necessary to consider such parameters.

HR is not a major pathway for T-DNA integration We initially expected that HR would contribute to T-DNA integration, at least to some extent, in the *ku80* mutant. However, no transformant possessing T-DNA integrated by HR was obtained in the present study for either the wild-type or the *ku80* mutant plants, suggesting that HR is not a major pathway, or not even an alternative pathway to NHEJ, regarding T-DNA integration. The frequency of gene targeting via HR is extremely low, even when T-DNAs with a long homologous sequence are used for transformation (Iida and Terada, 2005). Therefore, further analysis of transformants is needed to assess the frequency of HR-mediated

T-DNA integration. We cannot eliminate the possibility that our experimental system did not detect HR effectively because our construct carried a non-homologous vector sequence of ~600 bp between the 5′-end of histone H1 and the LB sequence. An experiment using this construct carrying the LB sequence adjacent to the homologous sequence may also be necessary to evaluate more precisely the contribution of HR to T-DNA integration.

Generation of filler DNA with long repeat sequences The most frequent junction type in LB-to-ChrDNA linkages was the filler DNA type, with most of the filler DNAs being > 5 bp in both the wild-type and the *ku80* mutant plants. Because NHEJ is expected to generate no filler DNA or short filler DNA at the junction, the detection of long filler DNA sequences suggests that DSB repair pathways other than NHEJ participate in the integration of the LB site into plant chromosomal DNA. The LB ends of the T-strand are 3′-ssDNA when the T-complex enters the nucleus (Gelvin, 2017). In addition, long 3′-ssDNA is a poor substrate for the Ku protein in mammalian cells (Yousefzadeh et al., 2014; Wyatt et al., 2016). Those reports suggest that the Ku protein is not likely to participate in the early steps of T-DNA integration.

Polθ is a candidate for the generation of long filler DNA during DSB repair. Polθ is the core factor of alt-EJ; in this pathway, Polθ synthesizes the DNA strand at the ssDNA gap regions of the DNA synapse formed between two 3′-overhanging DSB ends, via microhomology (Black et al., 2016). In addition, Polθ extends the 3′-ssDNA tail via its terminal transferase activity during end joining and forms longer filler DNA sequences (Black et al., 2016). The generation of longer filler DNA by these processes can be explained by the loop-out synthesis-dependent MMEJ, snap-back MMEJ, or the model of intermolecular templating found in fly and mammalian cells (Yu and McVey, 2010; Yousefzadeh et al., 2014). In mammalian cells, this Polθ-mediated end joining can extend the 3′-ssDNA tail independently of the Ku protein. Direct or inverted repeat sequences, which are known as templated insertions, are formed in this way (Yousefzadeh et al., 2014). A previous report showed that T-DNA integration into chromosomal DNA via *A. tumefaciens* did not occur in an *Arabidopsis* mutant deficient in Polθ (van Kregten et al., 2016). Those findings together with the results of the current study indicate that Polθ-mediated pathways are important for LB-to-ChrDNA joining in the wild-type and the *ku80*-mutant plants. In particular, because slippage and repriming were repeated, as reported by Yousefzadeh et al. (2014), it is possible that direct or inverted repeats became longer in the *ku80* mutant, in which Polθ-mediated end joining may function predominantly.

Currently, the interaction between the NHEJ and Polθ-mediated pathways in plant cells is not fully understood, and it is difficult to explain the role of NHEJ in the formation of filler DNA with repeat sequences ≤ 10 bp. The T-strand may become double-stranded DNA (dsDNA) before integration into chromosomal DNA (Liang and Tzfira, 2013). Blunt ends and short 3′- or 5′-overhanging dsDNA ends may act as a substrate for the Ku protein (Chang et al., 2017; Pannunzio et al., 2018). If such end structures are generated in the process of dsDNA formation, the Ku protein might contribute to LB-to-ChrDNA joining. Further studies are needed to address this hypothesis in detail.

In the current study, we observed the participation of the NHEJ pathway, at least partly, in T-DNA integration via *A. tumefaciens*. Our study raises the possibility that several plant host DSB repair pathways interact and contribute to T-DNA integration. The precise mechanism of T-DNA integration will be elucidated in future studies.

CONCLUSIONS

To evaluate the contribution of plant host DSB repair pathways to T-DNA integration via *A. tumefaciens*, the junction sequences of LB with chromosomal DNA were analyzed. Our results suggested that NHEJ functions to achieve LB-to-ChrDNA joining partially. In the absence of the Ku protein, direct or inverted repeated sequences found within or around the filler DNA were long and the frequency of T-DNA insertion near a transcription start site was high, suggesting that the Ku protein influences these processes.

We express our appreciation to Dr. Shin Hatakeyama and colleagues in our laboratory for their kind help and valuable advice. This work was carried out under the joint research program of the Biosignal Research Center, Kobe University.

REFERENCES

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301**, 653–657.

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., et al. (2014) Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat. Struct. Mol. Biol. **21**, 366–374.

Bechtold, N., and Pelletier, G. (1998) *In planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. Methods in Mol. Biol. **82**, 259–266.

Black, S. J., Kashkina, E., Kent, T., and Pomerantz, R. T. (2016) DNA polymerase θ: a unique multifunctional end-joining machine. Genes (Basel) **7**, 67.

Bourras, S., Rouxel, T., and Meyer, M. (2015) *Agrobacterium tumefaciens* gene transfer: how a plant pathogen hacks the nuclei of plant and nonplant organisms. Phytopathology **105**, 1288–1301.

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.

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat. Rev. Mol. Cell Biol. **18**, 495–506.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., and Sheen, J. (1996) Engineered GFP as a vital reporter in plants. Curr. Biol. **6**, 325–330.

Citovsky, V., Kozlovsky, S. V., Lacroix, B., Zaltsman, A., Dafny-Yelin, M., Vyas, S., Tovkach, A., and Tzfira, T. (2007) Biological systems of the host cell involved in *Agrobacterium* infection. Cell. Microbiol. **9**, 9–20.

Clough, S. J., and Bent, A. F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16**, 735–743.

Davis, A. J., and Chen, D. J. (2013) DNA double strand break repair via non-homologous end-joining. Transl. Cancer Res. **2**, 130–143.

Dueva, R., and Iliakis, G. (2013) Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. Transl. Cancer Res. **2**, 163–177.

Fattah, F., Lee, E. H., Weisensel, N., Wang, Y., Lichter, N., and Hendrickson, E. A. (2010) Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS Genet. **6**, e1000855.

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

Gallego, M. E., Bleuyard, J.-Y., Daoudal-Cotterell, S., Jallut, N., and White, C. I. (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in *Arabidopsis*. Plant J. **35**, 557–565.

Gelvin, S. B. (2017) Integration of *Agrobacterium* T-DNA into the plant genome. Annu. Rev. Genet. **51**, 195–217.

Iida, S., and Terada, R. (2005) Modification of endogenous natural genes by gene targeting in rice and other higher plants. Plant Mol. Biol. **59**, 205–219.

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

Jia, Q., Bundock, P., Hooykaas, P. J. J., and de Pater, S. (2012) *Agrobacterium tumefaciens* T-DNA integration and gene-targeting in *Arabidopsis thaliana* non-homologous end-joining mutants. J. Bot. **2012**, 989272.

Kasajima, I., Yoshizumi, T., Ichikawa, T., Matsui, M., and Fujiwara, T. (2010) Possible involvement of ploidy in tolerance to boron deficiency in *Arabidopsis thaliana*. Plant Biotechnol (Tokyo). **27**, 435–445.

Kodama, Y., Nagaya, S., Shinmyo, A., and Kato, K. (2007) Mapping and characterization of DNase I hypersensitive sites in *Arabidopsis* chromatin. Plant Cell Physiol. **48**, 459–470.

Komari, T., Halperin, W., and Nester, E. W. (1986) Physical and functional map of supervirulent *Agrobacterium tumefaciens* tumor-inducing plasmid pTiBo542. J. Bacteriol. **166**, 88–94.

Li, J., Vaidya, M., White, C., Vainstein, A., Citovsky, V., and Tzfira, T. (2005) Involvement of KU80 in T-DNA integration in plant cells. Proc. Natl. Acad. Sci. USA **102**, 19231–19236.

Li, X., and Heyer, W. D. (2008) Homologous recombination in DNA repair and DNA damage tolerance. Cell Res. **18**, 99–113.

Liang, Z., and Tzfira, T. (2013) *In vivo* formation of double-stranded T-DNA molecules by T-strand priming. Nat. Commun. **4**, 2253.

Ma, J.-L., Kim, E. M., Haber, J. E., and Lee, S. E. (2003) Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol. Cell. Biol. **23**, 8820–8828.

Mestiri, I., Norre, F., Gallego, M. E., and White, C. I. (2014) Multiple host-cell recombination pathways act in *Agrobacterium*-mediated transformation of plant cells. Plant J. **77**, 511–520.

Mimitou, E. P., and Symington, L. S. (2010) Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. EMBO J. **29**, 3358–3369.

Nimonkar, A. V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J. L., Wyman, C., Modrich, P., and Kowalczykowski, S. C. (2011) BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. Genes Dev. **25**, 350–362.

Pannunzio, N. R., Watanabe, G., and Lieber, M. R. (2018) Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J. Biol. Chem. **293**, 10512–10523.

Park, S.-Y., Vaghchhipawala, Z., Vasudevan, B., Lee, L.-Y., Shen, Y., Singer, K., Waterworth, W. M., Zhang, Z. J., West, C. E., Mysore, K. S., et al. (2015) *Agrobacterium* T-DNA integration into the plant genome can occur without the activity of key non-homologous end-joining proteins. Plant J. **81**, 934–946.

Sawasaki, T., Seki, M., Anzai, H., Irifune, K., and Morikawa, H. (1994) Stable transformation of *Arabidopsis* with the *bar* gene using particle bombardment. Transgenic Res. **3**, 279–286.

Shibata, A., Moiani, D., Arvai, A. S., Perry, J., Harding, S. M., Genois, M.-M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., et al. (2014) DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. Mol. Cell **53**, 7–18.

Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., and Lukyanov, S. A. (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. **23**, 1087–1088.

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

van Attikum, H., Bundock, P., Overmeer, R. M., Lee, L.-Y., Gelvin, S. B., and Hooykaas, P. J. J. (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 Kregten, M., de Pater, S., Romeijn, R., van Schendel, R., Hooykaas, P. J. J., and Tijsterman, M. (2016) T-DNA integration in plants results from polymerase-θ-mediated DNA repair. Nat. Plants **2**, 16164.

Windels, P., De Buck, S., Van Bockstaele, E., De Loose, M., and Depicker, A. (2003) T-DNA integration in Arabidopsis chromosomes. Presence and origin of filler DNA sequences. Plant Physiol. **133**, 2061–2068.

Wyatt, D. W., Feng, W., Conlin, M. P., Yousefzadeh, M. J., Roberts, S. A., Mieczkowski, P., Wood, R. D., Gupta, G. P., and Ramsden, D. A. (2016) Essential roles for polymerase θ-mediated end joining in the repair of chromosome breaks. Mol. Cell **63**, 662–673.

Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A., and Sawada, H. (2001) A revision of *Rhizobium* Frank

1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie *et al*. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. Int. J. Syst. Evol. Microbiol. **51**, 89–103.

Yousefzadeh, M. J., Wyatt, D. W., Takata, K., Mu, Y., Hensley, S. C., Tomida, J., Bylund, G. O., Doublié, S., Johansson, E., Ramsden, D. A., et al. (2014) Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. PLoS Genet. **10**, e1004654.

Yu, A. M., and McVey, M. (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res. **38**, 5706–5717.

Yu, X., and Gabriel, A. (2003) Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*. Genetics **163**, 843–856.

Zhang, W., Wu, Y., Schnable, J. C., Zeng, Z., Freeling, M., Crawford, G. E., and Jiang, J. (2012) High-resolution mapping of open chromatin in the rice genome. Genome Res. **22**, 151–162.

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2014) Open chromatin in plant genomes. Cytogenet. Genome Res. **143**, 18–27.