Characterization of integration sites and transfer DNA structures in Agrobacterium -mediated transgenic events of maize inbred B104

Anjanasree Neelakandan, Mercy Kabahuma, Qin Yang, Miriam Lopez, Randall J. Wisser, Peter Balint-Kurti, Nick Lauter

G3, 2023, 13(10), jkad166
https://doi.org/10.1093/g3journal/jkad166
Advance Access Publication Date: 31 July 2023
Plant Genetics and Genomics

Characterization of integration sites and transfer DNA structures in *Agrobacterium*-mediated transgenic events of maize inbred B104

Anjanasree K. Neelakandan,[1] Mercy Kabahuma,[1,2] Qin Yang,[3,4] Miriam Lopez,[1,5] Randall J. Wisser,[6,7] Peter Balint-Kurti,[3,8,*] Nick Lauter[1,2,5]

[1]Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA
[2]Interdisciplinary Genetics and Genomics Graduate Program, Iowa State University, Ames, IA 50011, USA
[3]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA
[4]State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University, Yangling 712100, China
[5]Corn Insects and Crop Genetics Research Unit, USDA-ARS, Ames, IA 50011, USA
[6]Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19716, USA
[7]Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux, INRAE, University of Montpellier, L'Institut Agro, Montpellier 34000, France
[8]Plant Science Research Unit, USDA-ARS, Raleigh, NC 27695, USA

*Corresponding author: Department of Entomology and Plant Pathology, North Carolina State University, 2572 Thomas Hall, 112 Derieux Place, Raleigh, NC 27695-7613, USA. Email: Peter.Balint-Kurti@USDA.GOV

Abstract

In maize, the community-standard transformant line B104 is a useful model for dissecting features of transfer DNA (T-DNA) integration due to its compatibility with *Agrobacterium*-mediated transformation and the availability of its genome sequence. Knowledge of transgene integration sites permits the analysis of the genomic environment that governs the strength of gene expression and phenotypic effects due to the disruption of an endogenous gene or regulatory element. In this study, we optimized a fusion primer and nested integrated PCR (FPNI-PCR) technique for T-DNA detection in maize to characterize the integration sites of 89 T-DNA insertions in 81 transformant lines. T-DNA insertions preferentially occurred in gene-rich regions and regions distant from centromeres. Integration junctions with and without microhomologous sequences as well as junctions with de novo sequences were detected. Sequence analysis of integration junctions indicated that T-DNA was incorporated via the error-prone repair pathways of nonhomologous (predominantly) and microhomology-mediated (minor) end-joining. This report provides a quantitative assessment of *Agrobacterium*-mediated T-DNA integration in maize with respect to insertion site features, the genomic distribution of T-DNA incorporation, and the mechanisms of integration. It also demonstrates the utility of the FPNI-PCR technique, which can be adapted to any species of interest.

Keywords: maize, T-DNA, transformation, nonhomologous end-joining, integration, *Agrobacterium*, genetic engineering, disease resistance, Plant Genetics and Genomics

Introduction

Agrobacterium tumefaciens is a plant pathogenic bacterium that is naturally capable of mediating transkingdom DNA transfer (Escobar and Dandekar 2003). This capability has been co-opted as a tool for the genetic engineering of plant genomes (Gelvin 2003; Zambryski 2013). A plant cell is transformed when a prokaryotic transfer DNA (T-DNA) element of the *A. tumefaciens* tumor-inducing (Ti) plasmid is integrated into its host's genome. During the initiation of transfer, the T-DNA, delineated by 25-bp direct repeats termed left and right borders (LB and RB), is recognized and cleaved by plasmid-encoded virulence (Vir) proteins (VirD1 and VirD2). This releases a single-stranded T-DNA that gets transported into the plant cell. VirD2 also covalently binds and protects the RB at the 5′-end, while another Vir protein, VirE2, associates and protects the T-DNA strand throughout its length. This T-DNA:Vir complex makes its way through the host cell cytoplasm and into the nucleus, where the T-DNA becomes double-stranded and integrates into the nuclear genome (Tzfira and Citovsky 2006; Gelvin 2012). While there is considerable information available on T-DNA production and transfer, little is known about the mechanism of host genome integration and the role of host proteins in facilitating this union (Singer 2018). Frequently, nucleotide bases predicted to be at the T-DNA border cleavage site are deleted during this process (Kleinboelting *et al.* 2015).

T-DNA integration is thought to take place at double-strand breaks, with the help of host endogenous repair proteins via homologous recombination (HR) or nonhomologous end-joining (NHEJ) pathways (Gelvin 2017). The HR mode of repair is precise, preferentially operates during the postreplicative cell cycle in the S/G2 phase of cell division, and depends on the availability of a homologous template (Gherbi *et al.* 2001). In contrast, NHEJ can occur at all stages of the cell cycle (but frequently occurs during G1) and is capable of ligating nonhomologous sequences; however, NHEJ is more prone to error (Waterworth *et al.* 2011). Double-strand break repair by NHEJ is considered the main DNA

Received: April 28, 2023. **Accepted:** July 20, 2023
Published by Oxford University Press on behalf of The Genetics Society of America 2023.

Table 1. Gene constructs used in maize transformation.

Construct number	Type	Gene ID	Gene name
T1	Overexpression	GRMZM2G416632	Glutathione S-transferase 23
T2	Overexpression	GRMZM2G305439	Leaf senescence–related protein
T3	Overexpression	GRMZM2G031364	Cytochrome P450 protein
T4	Silencing	GRMZM2G127328	Lysine histidine transporter 1
T5	Overexpression	GRMZM2G099363	Caffeoyl-CoA O-methyltransferase 2
T6	Overexpression	GRMZM2G021149	LSM domain–containing protein
T7	Overexpression	GRMZM2G173428	CTC-interacting domain 11
T8	Overexpression	GRMZM2G107774	Remorin 1
T9	Silencing	GRMZM2G068465	Cytochrome P450 51
T10	Overexpression	GRMZM2G168364	CHIT13—chitinase family protein precursor
T11	Silencing	GRMZM2G127342	Amino acid/auxin permease 40
T12	Overexpression	GRMZM2G074572	Stomatal cytokinesis–defective protein (SCD1)
T13	Overexpression	GRMZM2G080499	Protein kinase superfamily protein
T14	Overexpression	GRMZM2G132936	RNA recognition motif (RRM)–containing protein
T15	Overexpression	GRMZM2G061806	Hydroxycinnamoyl transferase 1
T16	Overexpression	GRMZM2G114918	Hydroxycinnamoyl transferase 2

repair pathway in plant cells (Waterworth *et al.* 2011). In the classical NHEJ (cNHEJ) model, a microhomologous region of up to 5 bp between the repaired ends facilitates annealing and recession of overhangs, leading to deletions in some cases. In the alternative NHEJ or microhomology-mediated end-joining (MMEJ) repair, longer microhomologies of 6–25 bp are required and they potentially cause longer deletions and complex rearrangements at the insertion site. Both these pathways are suggested to be involved in the genomic capture of T-DNA ends during integration (Kralemann *et al.* 2022). Alternatively, the repaired ends are extended based on a nonhomologous template by a synthesis-dependent strand annealing mechanism to generate "fillers" that lead to insertions or duplications at the target site (Mayerhofer *et al.* 1991; Windels *et al.* 2003; Kleinboelting *et al.* 2015). Recent studies suggest the role of DNA polymerase θ (PolQ) and chromatin configuration in the T-DNA integration process (van Kregten *et al.* 2016; Gelvin 2021; Nishizawa-Yokoi *et al.* 2021). Thus, the different modes of T-DNA integration leave distinct footprints at integration sites within the genome.

Using large collections of transformants, genome-wide sequence analysis of T-DNA integration junctions can help us understand which of the processes, HR, NHEJ, or MMEJ, are most important for T-DNA insertion (Gelvin 2017). Kleinboelting *et al.* (2015) analyzed the integration junctions of GABI-Kat T-DNA insertions in *Arabidopsis* and found that most integration events showed evidence of double-strand break repair via NHEJ. In some cases, they identified microhomology of up to 20 bp, suggesting that MMEJ plays a minor role in T-DNA integration.

Agrobacterium-mediated genetic transformation has been widely used for functional studies and trait manipulation in a wide variety of crops (Gelvin 2003). Transformation efficiencies vary in a genotype-specific fashion, and many genotypes within a species are not transformable with current protocols (Frame *et al.* 2006; Que *et al.* 2014). In maize, the Hi-II (high type II callus production) hybrid genotype has been the favored genotype for transformation due to its relatively high transformation efficiency (Armstrong *et al.* 1991). Additionally, a few inbred lines have been reported to be amenable to *Agrobacterium*-mediated transformation with varying efficiencies (Frame *et al.* 2006; Cho *et al.* 2014). Among these, the inbred B104 has been favored for transformation (Hallauer *et al.* 1997).

Plant transgenesis can produce a variety of different types of integrations. The ideal T-DNA insertion for optimal expression and trait engineering is a single copy at a given locus. However, the transgenic events oftentimes display complex T-DNA architecture including direct or inverted T-DNA repeats, incomplete T-DNAs, and/or the inclusion of plasmid backbones. Recent advances in long-read sequencing technologies have helped to elucidate large-scale genome structural changes including duplications, chromosomal translocations, and the exchange of chromosome arm ends associated with the T-DNA integration (Jupe *et al.* 2019; Pucker *et al.* 2021). The "position effects" of transgene integrations in maize and soybean at different chromosomal locations (noncentromeric and nontelomeric insertions) have also been reevaluated with current sequencing technologies and were found to have substantially smaller effects than previously thought (Betts *et al.* 2019). Integrations may also occur in the coding sequence or regulatory elements of an endogenous gene, causing event-specific alterations of gene expression and resulting phenotypes (Nicolia *et al.* 2017). During the commercialization process, multiple insertion events are routinely screened to identify single-copy lines with insertions that are not predicted to alter the expression of endogenous nontarget genes.

The objective of this study was to dissect the molecular structure of T-DNA insertion sites, integration dynamics, and transgene structures resulting from the *Agrobacterium*-mediated transformation of B104. Using the complete B104 genome assembly (https://www.maizegdb.org/gbrowse/maize_b104), we characterized the genomic context of transgene insertions and inferred the molecular mechanisms of the T-DNA integration process based on junction sequence features.

Methods

Generation and maintenance of transgenic maize B104

A set of 16 different transgene constructs (Table 1) in the binary plasmid pMCG1005 developed at North Carolina State University was used to transform the B104 inbred at the Iowa State University Plant Transformation Facility (Frame *et al.* 2011, 2015). The transgenic events were selected based on Basta resistance, regenerated, and propagated in the greenhouse. Transgenic T1 maize seedlings of >100 events were grown in the transgenic field nursery of Iowa State University in the summer of 2016 and screened with herbicide Liberty (glufosinate–ammonium) leaf painting assay. The Basta-resistant plants were identified and tagged for sample collection. The young leaf tissue was collected in plates, lyophilized, and homogenized with metal balls in a genogrinder (Geno/Grinder 2010; SPEX SamplePrep). Genomic DNA was isolated as described (Hsia *et al.* 2010), and the extracted

DNA was dissolved in 100 µl 1XTE. The concentration and purity of DNA were checked with a NanoDrop ND1000 spectrophotometer and diluted with molecular biology grade water (MBGW) to make 20 ng/µl stocks for use as a template for PCR.

Primer design and amplification of T-DNA insertions

A modified thermal asymmetric interlaced PCR (TAIL-PCR) approach, fusion primer and nested integrated PCR (FPNI-PCR) (Wang *et al.* 2011), was employed to identify the insertion sites and characterize the T-DNA junction sequences in the maize genome. All oligonucleotide primers (except the MFAD set; see below) were designed based on the plasmid or maize sequence in Vector NTI (Advance 11, Invitrogen). Oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA).

This study used 8 arbitrary degenerate (AD) primers that have been optimized for the maize genome (maize fusion AD, or MFAD primers) (Settles *et al.* 2004) with annealing temperatures ranging from 40 to 54°C (Supplementary Table 1). Each primer was fused to an artificial adapter oligo that serves as 1 of the binding sites in subsequent PCR steps. For PCR, the adapter-fused AD primers were paired with nonoverlapping nested primers that target different positions within and proximal to the left and right T-DNA border sequences (LP1, LP2, LP3, RP1, RP2, and RP3) so that each reaction contained 1 AD–MFAD primer and 1 specific primer. The LP and RP oligos have annealing temperatures of >62°C for high-stringency PCR. The primer sequences are outlined in Supplementary Table 1.

Primary PCR was carried out with adapter-fused AD primers (1 µM) and T-DNA-specific primer 1 (LP1 or RP1; 0.25 µM) in a 20-µl reaction volume, with 40 ng of genomic DNA template, and 0.625 U of Taq DNA Polymerase (New England Biolabs) with ThermoPol reaction buffer. The product of primary PCR (1 µl) was used as a template for secondary PCR with nested T-DNA-specific primer 2 (LP2 or RP2) and fusion-specific primer 1 (FSP1), followed by tertiary PCR after 50 times dilution of secondary PCR. The products of tertiary PCR, generated by T-DNA primer 3 (LP3 or RP3) and fusion-specific primer 2 (FSP2), were directly submitted for Sanger sequencing if they produced a single band after ExoSAP-IT treatment (Affymetrix) as per the manufacturer's recommendations.

Sequence analysis

The T-DNA insertion loci flanking sequence tags obtained after trimming the border sequences were used as queries to blast against the maize B73 RefGen_V4 genome assembly (Jiao *et al.* 2017). Upon release of the B104 genome assembly (https://www.maizegdb.org/gbrowse/maize_b104), the sequences were also compared to confirm the junction features at the chromosomal insertion sites. The web tool FSTVAL (http://bioinfo.mju.ac.kr/fstval/) (Kim *et al.* 2012) was used to annotate the flanking sequences (e.g. genic, intergenic, and repetitive).

PCR resequencing for integration site validation and genotyping

The insertion sites were confirmed based on having the expected product size from PCR amplifications using the independent insertion site-specific maize genomic primer and T-DNA-specific primer. Sanger sequencing of the amplification products was performed at the Iowa State University DNA facility. Three separate transgenes tagged by both the LB and RB flanking sequences were amplified to validate both the flanking sequence data and the T-DNA at each locus. Amplifications with Phusion High-Fidelity DNA Polymerase (New England Biolabs) were performed based on the integration junction sequence or filler sequence (Supplementary Table 1), with reaction conditions favoring larger products as recommended by the manufacturer.

Results

Sequence characterization of T-DNA transformants

Sixteen recombinant gene constructs were created in the plasmid binary vector pMCG1005 (Yang *et al.* 2017) (Fig. 1a). Each construct was designed to overexpress or silence 1 of 16 genes that had been implicated in resistance to northern leaf blight or southern leaf blight diseases of maize (Table 1). A total of 106 T0 primary transformants of B104 were either selfed or crossed to nontransgenic B104 to generate T1 plants. Each of the T1 plants was confirmed to carry a transgene using Basta herbicide screening and PCR.

To amplify flanking sequences adjacent to the LB and RB of each T-DNA insertion site, we used a FPNI-PCR with maize-customized AD primers (Supplementary Table 1) (Settles *et al.* 2004; Wang *et al.* 2011). Using degenerate random primers with a wide range of GC content as fusion partners for adapter sequences was more efficient at amplifying flanking sequences than primers with a narrower GC range (data not shown). The flanking sequence data were independently validated using distinct primer sets for 12 insertion sites, 3 of which were also validated by amplifying the T-DNA insertion sequence (Supplementary Fig. 1). In each case, the expected flanking sequence identified by FPNI-PCR or the expected T-DNA insertion was identified, confirming the accuracy of the approach.

In total, 161 flanking sequence products were produced from 97 T1 plants (9 T1 plants produced no FPNI-PCR product). Fifty-two of the flanking sequences contained only vector sequences, so the insertion site could not be mapped to the B104 genome. Half (i.e. 26) of these FPNI-PCR products were derived from 16 T1 plants for which no other flanking sequence contained plant genomic (Pl) sequences. The other 26 FPNI-PCR products without Pl sequences were detected in T1 plants, from which FPNI-PCR products with Pl sequences were also detected. Altogether, 109 FPNI-PCR products with mappable flanking sequences were generated from 81 T1 plants. These 109 sequences corresponded to 89 insertions (Supplementary File S1), which were identified from the LB flanking sequence only ($n = 38$), the RB flanking sequence only ($n = 31$), or from both the LB and RB flanking sequences ($n = 20$) (Fig. 1b). Nine of the insertions were located in repetitive regions of the genome, so they could not be confidently mapped to a single locus, leaving 80 mappable insertion sites.

Analysis of T-DNA insertion sites

To characterize the positional distribution of the 80 mapped insertion sites across chromosomes, we computed the mean relative distance, which is the ratio of the physical distance (Mb) of the insertion site from the centromere relative to the physical size of the corresponding chromosome arm (Table 2; Fig. 1c). Across all chromosome arms, T-DNA insertions had a mean relative distance of 0.70, ranging from 0.54 on chromosome 3 to 0.90 on chromosome 10. Chromosome size was partially predictive of the number of insertion events ($R^2 = 0.31$). Chromosome 5 had the most T-DNA insertion sites ($n = 17$), followed by chromosome 3 ($n = 14$), while chromosome 6 had the least T-DNA insertions ($n = 2$).

The genomic sequence context for the insertion events was also analyzed: 64% ($n = 57$) of the insertions occurred in genic regions (within 5 kb flanking the transcription start and termination

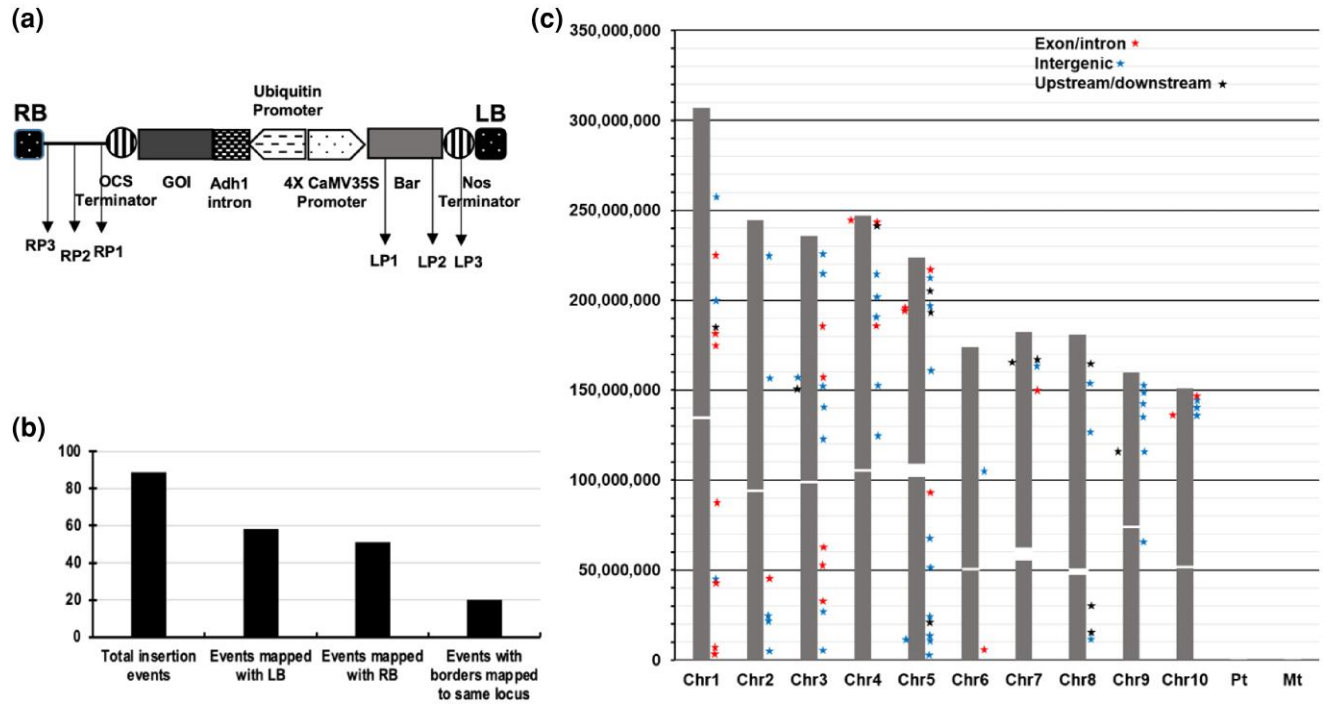

Fig. 1. a) Schematic diagram of the T-DNA region of the pMCG1005 vector showing the region from the LB to the RB. The primer locations for flanking sequence acquisition are indicated by arrows (RP1–3, LP1–3). The inserted gene sequence is represented as GOI ("gene of interest") downstream to the ubiquitin promoter. The rest of the plasmid includes the bacterial origin of replication and the *aadA* gene for spectinomycin resistance (marker for bacterial selection). Other acronyms: OCS, octopine synthase; Adh1, alcohol dehydrogenase-1; CaMV, cauliflower mosaic virus; Bar, bialaphos resistance; Nos, nopaline synthase. b) A summary of the T-DNA integration flanking sequences mapped in the maize genome. N.B. the "Events mapped with LB" category includes events mapped with both LB and RB. Likewise, for the "Events mapped with RB" category. c) Chromosomal localization of integration sites in maize. The 10 chromosomes of maize with their centromeres are depicted with the flanking sequence information mapped with asterisks. The colors of the asterisks indicate the nature of the flanking sequence as indicated.

Table 2. Chromosomal location of T-DNA insertion loci.

Chromosome	Number of insertions	Percentage of the total (%)	Mean relative distance from the centromere (±SD)[a]
Chr01	11	13.75	0.55 ± 0.26
Chr02	6	7.50	0.71 ± 0.21
Chr03	14	17.50	0.54 ± 0.25
Chr04	8	10.00	0.74 ± 0.24
Chr05	17	21.25	0.72 ± 0.23
Chr06	2	2.50	0.65 ± 0.29
Chr07	4	5.00	0.84 ± 0.07
Chr08	6	7.50	0.71 ± 0.18
Chr09	7	8.75	0.62 ± 0.30
Chr10	5	6.25	0.90 ± 0.04

[a] Mean relative distance from the centromere is the ratio of the Mb of the flanking sequence from the centromere to the full measure of the corresponding chromosome arm.

Table 3. Genomic context of T-DNA insertion loci.

Nature of the flanking sequence	Number of insertions	Percentage of the total (%)
Intergenic	23	26
Genic[a]	57	64
Exonic	16	18
Intronic	8	9
Upstream	22	25
Downstream	11	12
Repetitive region	9	10

[a] Insertion sites within 5 kb of the transcription start or termination site were considered genic.

sites) and 26% ($n = 23$) occurred in intergenic regions with few repeats (Table 3). As mentioned above, 9 insertions (10%) occurred in repeat-rich regions, which could not be mapped to a single locus. Further partitioning of the genic insertions showed that 25% occurred in 5′-upstream sequences, 18% in exons, 9% in introns, and 12% in 3′-downstream sequences.

Analysis of T-DNA integration junctions

During the transfer process, the T-DNA is cleaved from the Ti plasmid at the LB and RB. The single-stranded T-DNA molecule is then exported to the host cell as part of a complex involving several *Agrobacterium*-derived Vir proteins. T-DNA integration is often associated with the degradation of the T-DNA end or additional border sequence retained during integration, leading, respectively, to less or more border sequences than expected at the cleavage site (Tzfira *et al.* 2004).

We characterized the integrity of integration for 147 flanking sequences (FPNI-PCR products) in which a cleavage site was detected (constituting 81 LB and 66 RB integration junctions; Fig. 2). Considering the canonical cleavage site as position 0 and the number of fewer or greater than expected nucleotides as negative or positive values, respectively, the cleavage sites deviated by −108 to +283 at the LB and by −154 to +203 at the RB. The proportion of integration junctions exhibiting the canonical cleavage site at the LB was low (3.7%, 3 of 81) as compared to the RB (22.7%, 15 of 66). The distribution of integration errors was highly skewed, with the majority of cleavage sites at both the LB (91.4%, 74 of 81) and

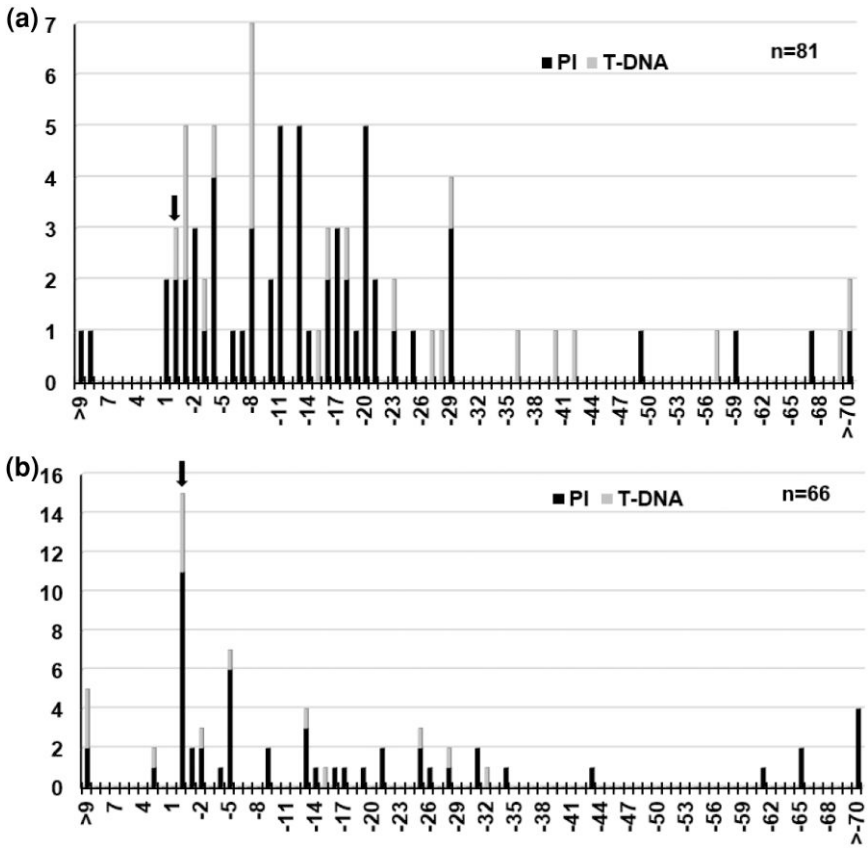

Fig. 2. Distribution of nucleotides relative to the expected T-DNA cleavage site. The number of T-DNA insertion sites (Y-axis) is plotted for the observed number of nucleotides preceding a cleavage site (X-axis). Cleavage sites are summarized for the a) LB and b) RB sequences based on the number of fewer or greater than expected nucleotides as negative or positive values, respectively. T-DNA insertions with proper cleavage are set to zero (noted by the arrow). Flanking sequences with Pl and T-DNA sequences are indicated.

RB (66.7%, 44 of 66) having end sequences that were deleted, while only 4.9% (4 of 81) of LB and 10.6% (7 of 66) of RB junctions had more than the expected number of nucleotides at the cleavage site (Fig. 2). At the LB, 85% (69 of 81) of the T-DNA ends were clustered around sites +1 to −29 (Fig. 2a). Similarly, at the RB, 80% (53 of 66) of the cleavage sites were distributed from +3 to −34 (Fig. 2b). At the LB, 28.4% (23 of 81) of cleavage sites were linked to a full-length or truncated T-DNA (concatemeric structures), and at the RB, this proportion was 22.7% (15 of 66) (Fig. 2).

By reference to the B104 genome, we were able to infer that in 19 (12.9%) cases, integration events resulted from simple direct ligation (no filler, no homology), and in 63 (42.9%) cases, junctions with microhomologous regions ranging from 1 to 12 bp were detected (Supplementary Fig. 2), and junctions with additional nucleotides or "filler" sequences were detected in 65 (44.2%) cases (Fig. 3a and b). At the LB, 39 out of 81 (48%) of integrations were microhomology-mediated compared to 24 out of 66 (36%) at the RB. Simple ligations were observed in 11 out of 66 (17%) of RB and 8 out of 81 (10%) of LB sequences, and junctions with filler sequences were observed in 31 out of 66 (47%) of RB as compared to 34 out of 81 (42%) of LB sequences (Supplementary Fig. 3). Among the microhomologies observed, 91.8% were 1–5 bp; all the >5-bp microhomologies observed were at LB junctions (Fig. 3c). Analysis of the 30-bp sequences flanking each T-DNA border (for integrations without filler sequences) revealed low levels of conservation of sequences at either border with higher levels of nucleotide conservation at the LB flanking sequence as compared to the RB (Supplementary Fig. 4).

Filler sequences from 1 bp to >400 bp were observed (Fig. 3d), of which most (23 out of 34, 68% for LB; 23 out of 31, 74% for RB) were <50 bp. The filler sequences comprised simple insertions, duplications, or homologous regions of upstream/downstream flanking sequences adjacent to the insertion loci or the T-DNA or composite/chimeric sequences originating from the T-DNA and flanking sequences.

Discussion

Molecular characterization of genetically modified plants in terms of transgene detection, monitoring, traceability, and safety evaluation is integral for plant biotechnology (Que *et al.* 2014; Li *et al.* 2017). Efficient detection of transgene insertion sites is essential for quality assessment and to deduce the extent of correlation between transgene locations and/or the number of insertions with expression levels and the phenotype.

FPNI-PCR is a modified form of TAIL-PCR and is relatively easy to adapt to any species of interest. While the same sets of AD primers may be used across species, some optimization may be advisable to account for differences in GC content across genomes. If the genome sequence is available, then further optimization is possible to select degenerate sequences that are enriched in genes of that species (Settles *et al.* 2004).

The use of FPNI-PCR with maize-customized AD primers proved to be an efficient and specific method for high-throughput recovery of sequences flanking T-DNA insertion sites. The amplification of full-length T-DNA of 3 independent single-copy events

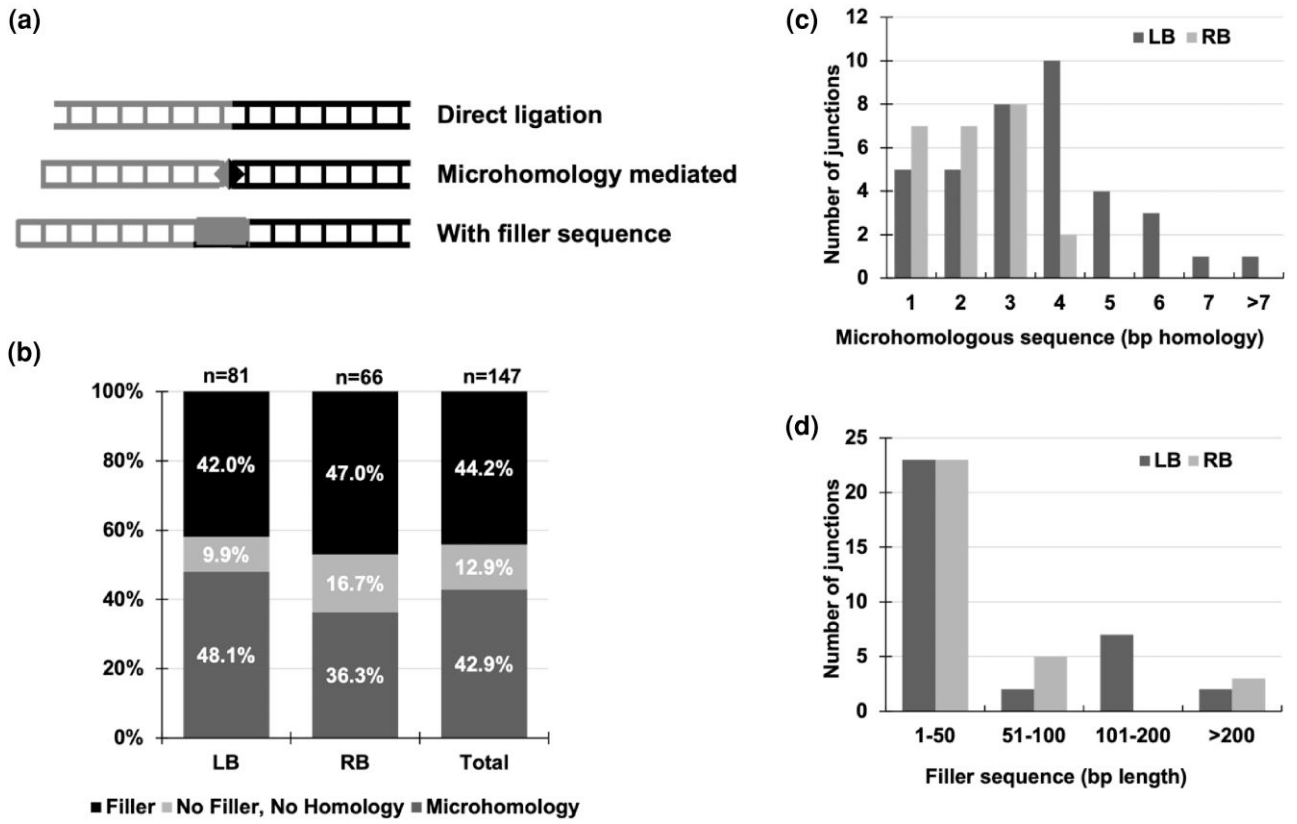

Fig. 3. Characteristics of T-DNA junction sequence features in maize. a) The types of junction sequence features of T-DNA integration in the plant genome. b) The percentage of junctions with microhomologous sequences, "filler" sequences, and direct ligation at either of the borders and the combined data for both. c) The distribution of microhomologous sequences in bp (1–7 bp and >7) among the junctions for both borders. d) The size distribution of filler sequences at the T-DNA::plant junction for both borders.

based on flanking sequence information substantiated the reliability and specificity of the sequence information generated (Supplementary Fig. 1). The frequency of T-DNA insertions was higher toward the distal ends of the chromosomes, correlating with higher gene density and transcriptionally active chromatin, which is consistent with the observed enrichment of insertions in genic regions (Table 3) as well as previous studies in other species (Szabados *et al.* 2002; An *et al.* 2003; Yang *et al.* 2011; Perez-Martın *et al.* 2017; Schouten *et al.* 2017). One contributing factor to this distribution is that gene expression may be higher in distal areas (Mizuno *et al.* 2011), leading to preferential selection of events with higher expression of the herbicide-resistant selectable marker. However, a uniform distribution of insertions across chromosomes has been observed in rice (Kim *et al.* 2007; Shilo *et al.* 2017). The relatively lower frequency of T-DNA integration in chromosome 6 could in part be due to the presence of the nucleolus organizer region that might influence the recombination process important for T-DNA integration (Anderson *et al.* 1955; Luo *et al.* 1998).

The structural features of T-DNA integration sites have been comprehensively analyzed and characterized in several model species (Kim *et al.* 2003; Kleinboelting *et al.* 2015; Nicolia *et al.* 2017), but not much information is known about maize. The T-DNA ends are subjected to different degrees of resection owing to degradation by the action of nucleases during transfer or recombination-related processes during integration that involves host proteins. The higher percentage of canonical cleavage site retention and fewer total cleavage sites that were observed at the RB compared to the LB could be attributed, in part, to its protection by the bound bacterial effector VirD2 (Tinland *et al.* 1995). Overall, the primary cleavage site seems to be conserved as expected, with limited end sequence polymorphisms owing to varying degrees of nucleotide loss or trimming, possibly due to degradation or repair machinery that mediates the integration process, consistent with other species (Kleinboelting *et al.* 2015).

We found microhomologous regions at both the LB and RB junctions in maize (Fig. 3; Supplementary Fig. 2). The large majority of these (91.8%) were 1–5 bp, suggesting a predominant role for NHEJ. Some longer microhomologies were observed at the LB, suggesting that MMEJ may also occur (Fig. 3). The observation that homology length and frequency were higher at the LB as compared to the RB could be due to the nature of the T-DNA integration process (Park *et al.* 2015; Gelvin 2017), which is often initiated at the LB of the T-DNA. It should be noted that we also observed higher levels of nucleotide conservation at the LB flanking sequence as compared to the RB (Supplementary Fig. 4). All the filler sequences shared homology to either plant sequences adjacent to the insertion junction or to the T-DNA. This suggests a template-dependent synthesis mechanism for repair following integration. In a previous study in maize, contrary to our findings, the filler sequences were observed to be nonhomologous to the plant flanking or T-DNA sequences (Yang *et al.* 2011). This could be due to the differences in the inbred background that were used in the 2 studies (in the previous study, line 18-599 was used, while we used B104). The identification of microhomologous regions and filler sequences at the T-DNA–host junction is consistent with the role of the recently characterized enzyme, PolQ, in the T-DNA integration process (van Kregten *et al.* 2016).

Multiple full-length or truncated T-DNAs in tandem at a single locus are often observed in transgenic plants; their presence needs to be scrutinized for regulatory compliance and event quality. Furthermore, the fact that 64% of insertions we characterized were in genic regions with 18% of those in exons emphasizes the fact that T-DNA insertions are relatively likely to influence the expression of endogenous genes and that transgenic events should be appropriately scrutinized with this in mind. In the future, information on good-quality, single-copy, transgene insertion sites can be correlated with expression levels to develop putative "genomic safe harbors" that are integration sites for stable and predictable expression and for trait stacking applications.

Data availability

Plasmids mentioned are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Supplemental material available at G3 online.

Acknowledgments

The authors acknowledge Dr. Kan Wang and Bronwyn Frame at the Plant Transformation Facility (Iowa State University, Ames, IA, USA) for generating the transgenic lines examined in this study. We thank Dr. Tes Posekany, Dr. Siva Chudalayandi, Grace Kuehne, and Jacob Studt for their helpful comments and technical assistance. This is dedicated to the memory of Nick Lauter (Yandeau-Nelson and Wisser 2021).

Funding

This research was funded by the US National Science Foundation (Award #1127076) with additional support from the USDA-ARS project #5030-21000-067.

Conflicts of interest

The authors declare no conflict of interest.

Literature cited

An S, Park S, Jeong D-H, Lee D-Y, Kang H-G, Yu J-H, Hur J, Kim S-R, Kim Y-H, Lee M, *et al.* Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol. 2003; 133(4):2040–2047. doi:10.1104/pp.103.030478.

Anderson EG, Kramer HH, Longley AE. Translocations in maize involving chromosome 6. Genetics. 1955;40(4):531–538. doi:10.1093/genetics/40.4.531.

Armstrong CL, Green CE, Phillips RL. Development and availability of germplasm with high type II culture formation response. Maize Genet Coop Newsl. 1991;65:92–93.

Betts SD, Basu S, Bolar J, Booth R, Chang S, Cigan AM, Farrell J, Gao H, Harkins K, Kinney A, *et al.* Uniform expression and relatively small position effects characterize sister transformants in maize and soybean. Front. Plant Sci. 2019;10:1209. doi:10.3389/fpls.2019.01209.

Cho MJ, Wu E, Kwan J, Yu M, Banh J, Linn W, Anand A, Li Z, TeRonde S, Register JC, *et al.* *Agrobacterium*-mediated high-frequency transformation of an elite commercial maize (*Zea mays* L.) inbred line. Plant Cell Rep. 2014;33(10):1767–1777. doi:10.1007/s00299-014-1656-x.

Escobar MA, Dandekar AM. *Agrobacterium tumefaciens* as an agent of disease. Trends Plant Sci. 2003;8(8):380–386. doi:10.1016/S1360-1385(03)00162-6.

Frame B, Main M, Schick R, Wang K. Genetic transformation using maize immature zygotic embryos. Methods Mol Biol. 2011;710: 327–341. doi:10.1007/978-1-61737-988-8_22.

Frame BR, McMurray JM, Fonger TM, Main ML, Taylor KW, Torney FJ, Paz MM, Wang K. Improved *Agrobacterium*-mediated transformation of three maize inbred lines using MS salts. Plant Cell Rep. 2006;25(10):1024–1034. doi:10.1007/s00299-006-0145-2.

Frame B, Warnberg K, Main M, Wang K. Maize (*Zea mays* L.). Methods Mol Biol. 2015;1223:101–117. doi:10.1007/978-1-4939-1695-5_8.

Gelvin SB. *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev. 2003; 67(1):16–37. doi:10.1128/MMBR.67.1.16-37.2003.

Gelvin SB. Traversing the cell: *Agrobacterium* T-DNA's journey to the host genome. Frontiers in Plant Sci. 2012;3:52. doi:10.3389/fpls.2012.00052.

Gelvin SB. Integration of *Agrobacterium* T-DNA into the plant genome. Annu Rev Genet. 2017;51(1):195–217. doi:10.1146/annurev-genet-120215-035320.

Gelvin SB. Plant DNA repair and *Agrobacterium* T–DNA integration. Int J Mol Sci. 2021;22(16):8458. doi:10.3390/ijms22168458.

Gherbi H, Gallego ME, Jalut N, Lucht JM, Hohn B, White CI. Homologous recombination in planta is stimulated in the absence of Rad50. EMBO Rep. 2001;2(4):287–291. doi:10.1093/embo-reports/kve069.

Hallauer AR, Lamkey KR, White PR. Registration of five inbred lines of maize: B102, B103, B104, B105, and B106. Crop Sci. 1997;37(4): 1405–1406. doi:10.2135/cropsci1997.0011183X003700040094x.

Hsia A-P, Chen HD, Ohtsu K, Schnable PS. DNA extraction from freeze-dried plant tissue with CTAB in a 96-well format. Cold Spring Harb Protoc. 2010;2010(11):pdb.prot5516. doi:10.1101/pdb.prot5516.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, *et al.* Improved maize reference genome with single-molecule technologies. Nature. 2017;546(7659):524. doi:10.1038/nature22971.

Jupe F, Rivkin AC, Michael TP, Zander M, Motley ST, Sandoval JP, Slotkin RK, Chen H, Castanon R, Nery JR, *et al.* The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS Genet. 2019;15(1):e1007819. doi:10.1371/journal.pgen.1007819.

Kim JS, Kim J, Lee T-H, Jun KM, Kim TH, Kim Y-H, Park H-M, Jeon J-S, An G, Yoon U-H, *et al.* FSTVAL: a new web tool to validate bulk flanking sequence tags. Plant Methods. 2012;8(1):19. doi:10.1186/1746-4811-8-19.

Kim SR, Lee J, Jun SH, Park S, Kang HG, Kwon S, An G. Transgene structures in T-DNA-inserted rice plants. Plant Mol Biol. 2003; 52(4):761–773. doi:10.1023/A:1025093101021.

Kim SI, Veena, Gelvin SB. Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. Plant J. 2007;51(5):779–791. doi:10.1111/j.1365-313X.2007.03183.x.

Kleinboelting N, Huep G, Appelhagen I, Viehoever P, Li Y, Weisshaar B. The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. Mol Plant. 2015;8(11):1651–1664. doi:10.1016/j.molp.2015.08.011.

Kralemann LEM, de Pater S, Shen H, Kloet SL, van Schendel R, Hooykaas PJJ, Tijsterman M. Distinct mechanisms for genomic attachment of the 5′ and 3′ ends of *Agrobacterium* T-DNA in plants. Nat Plants. 2022;8(5):526–534. doi:10.1038/s41477-022-01147-5.

Li R, Quan S, Yan X, Biswas S, Zhang D, Shi J. Molecular characterization of genetically-modified crops: challenges and strategies. Biotechnol Adv. 2017;35(2):302–309. doi:10.1016/j.biotechadv.2017.01.005.

Luo MC, Yang ZL, Dvořák J. Position effects of ribosomal RNA multigene loci on meiotic recombination in wheat. Genetics. 1998;149(2):1105–1113. doi:10.1093/genetics/149.2.1105.

Mayerhofer R, Konczkalman Z, Nawrath C, Bakkeren G, Crameri A, Angelis K, Redei GP, Schell J, Hohn B, Koncz C. T-DNA integration—a mode of illegitimate recombination in plants. EMBO J. 1991;10(3):697–704. doi:10.1002/j.1460-2075.1991.tb07999.x.

Mizuno H, Kawahara Y, Wu J, Katayose Y, Kanamori H, Ikawa H, Itoh T, Sasaki T, Matsumoto T. Asymmetric distribution of gene expression in the centromeric region of rice chromosome 5. Front Plant Sci. 2011;2:16. doi:10.3389/fpls.2011.00016.

Nicolia A, Ferradini N, Veronesi F, Rosellini D. An insight into T-DNA integration events in *Medicago sativa*. Int J Mol Sci. 2017;18(9):1951. doi:10.3390/ijms18091951.

Nishizawa-Yokoi A, Saika H, Hara N, Lee L-Y, Toki S, Gelvin SB. *Agrobacterium* T-DNA integration in somatic cells does not require the activity of DNA polymerase θ. New Phytol. 2021;229(5):2859–2872. doi:10.1111/nph.17032.

Park SY, Vaghchhipawala Z, Vasudevan B, Lee LY, Shen Y, Singer K, Waterworth WM, Zhang ZJ, West CE, Mysore KS, *et al.* *Agrobacterium* T-DNA integration into the plant genome can occur without the activity of key non-homologous end-joining proteins. Plant J. 2015;81(6):934–946. doi:10.1111/tpj.12779.

Perez-Martın Fernando, Yuste-Lisbona Fernando J., Pineda Benito, Angarita-Díaz MP, García-Sogo B, Antón T, Sánchez S, Giménez E, Atarés A, Fernández-Lozano A, *et al.* A collection of enhancer trap insertional mutants for functional genomics in tomato. Plant Biotechnol J. 2017;15(11):1439–1452. doi:10.1111/pbi.12728.

Pucker B, Kleinbölting N, Weisshaar B. Large scale genomic rearrangements in selected *Arabidopsis thaliana* T-DNA lines are caused by T-DNA insertion mutagenesis. BMC Genomics. 2021;22(1):599. doi:10.1186/s12864-021-07877-8.

Que Q, Elumalai S, Li X, Zhong H, Nalapalli S, Schweiner M, Fei X, Nuccio M, Kelliher T, Gu W, *et al.* Maize transformation technology development for commercial event generation. Front Plant Sci. 2014;5:379. doi:10.3389/fpls.2014.00379.

Schouten HJ, vande Geest H, Papadimitriou S, Bemer M, Schaart JG, Smulders MJM, Perez GS, Schijlen E. Re-sequencing transgenic plants revealed rearrangements at T-DNA inserts, and integration of a short T-DNA fragment, but no increase of small mutations elsewhere. Plant Cell Rep. 2017;36(3):493–504. doi:10.1007/s00299-017-2098-z.

Settles AM, Latshaw S, McCarty DR. Molecular analysis of high-copy insertion sites in maize. Nucleic Acids Res. 2004;32(6):e54. doi:10.1093/nar/gnh052.

Shilo S, Tripathi P, Melamed-Bessudo C, Tzfadia O, Muth TR, Levy AA. T-DNA-genome junctions form early after infection and are influenced by the chromatin state of the host genome. PLoS Genet. 2017;13(7):1–16. doi:10.1371/journal.pgen.1006875.

Singer K. The Mechanism of T-DNA Integration: Some Major Unresolved Questions. Gelvin SB, editor. Cham: Springer International Publishing; 2018.

Szabados L, Kovács I, Oberschall A, Ábrahám E, Kerekes I, Zsigmond L, Nagy R, Alvarado M, Krasovskaja I, Gál M, *et al.* Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. Plant J. 2002;32(2):233–242. doi:10.1046/j.1365-313X.2002.01417.x.

Tinland B, Schoumacher F, Gloeckler V, Bravo-Angel AM, Hohn B. The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. EMBO J. 1995;14(14):3585–3595. doi:10.1002/j.1460-2075.1995.tb07364.x.

Tzfira T, Citovsky V. *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. Curr Opin Biotechnol. 2006;17(2):147–154. doi:10.1016/j.copbio.2006.01.009.

Tzfira T, Li J, Lacroix B, Citovsky V. *Agrobacterium* T-DNA integration: molecules and models. Trends Genet. 2004;20(8):375–383. doi:10.1016/j.tig.2004.06.004.

van Kregten M, de Pater S, Romeijn R, van Schendel R, Hooykaas PJJ, Tijsterman M. T-DNA integration in plants results from polymerase-θ-mediated DNA repair. Nat Plants. 2016;2(11):16164. doi:10.1038/nplants.2016.164.

Wang Z, Ye S, Li J, Zheng B, Bao M, Ning G. Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning. BMC Biotechnol. 2011;11(1):109. doi:10.1186/1472-6750-11-109.

Waterworth WM, Drury GE, Bray CM, West CE. Repairing breaks in the plant genome: the importance of keeping it together. New Phytol. 2011;192(4):805–822. doi:10.1111/j.1469-8137.2011.03926.x.

Windels P, De Buck S, Van Bockstaele E, De Loose M, Depicker A. T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. Plant Physiol. 2003;133(4):2061–2068. doi:10.1104/pp.103.027532.

Yandeau-Nelson MD, Wisser RJ. Remembering Dr. Nick Lauter (December 13, 1972–January 7, 2021). Curr Plant Biol. 2021;27:100214. doi:10.1016/j.cpb.2021.100214.

Yang L, Fu FL, Zhang ZY, Zhou SF, She YH, Li WC. T-DNA integration patterns in transgenic maize lines mediated by *Agrobacterium tumefaciens*. African J Biotechnol. 2011;10(59):12614–12625.

Yang Q, He Y, Kabahuma M, Chaya T, Kelly A, Borrego E, Bian Y, El Kasmi F, Yang L, Teixeira P, *et al.* A gene encoding maize caffeoyl-CoA O-methyltransferase confers quantitative resistance to multiple pathogens. Nat Genet. 2017;49(9):1364–1372. doi:10.1038/ng.3919.

Zambryski P. Fundamental discoveries and simple recombination between circular plasmid DNAs led to widespread use of *Agrobacterium tumefaciens* as a generalized vector for plant genetic engineering. Int J Dev Biol. 2013;57(6–7–8):449–452. doi:10.1387/ijdb.130190pz.

Editor: L. McIntyre