T-DNA integration in plants requires MRE11- or TDP2-mediated removal of the 5' bound Agrobacterium protein VirD2

Lejon Kralemann
Institute of Biology Leiden, Leiden University https://orcid.org/0000-0002-8154-9315
Sylvia de Pater
leiden university https://orcid.org/0000-0003-2626-3661
Hexi Shen
Shandong Jianzhu University
Susan Kloet
Leiden University Medical Center https://orcid.org/0000-0001-9189-0640
Robin van Schendel
Leiden University Medical Center https://orcid.org/0000-0001-7068-0679
Paul Hooykaas
Leiden University; Institute of Biology https://orcid.org/0000-0002-9736-6927
Marcel Tijsterman (m.tijsterman@lumc.nl)
Leiden University Medical Center https://orcid.org/0000-0001-8465-9002

Letter

Keywords:

Posted Date: December 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1144888/v1

Version of Record: A version of this preprint was published at Nature Plants on May 9th, 2022. See the published version at https://doi.org/10.1038/s41477-022-01147-5.

Abstract

Agrobacterium tumefaciens, a pathogenic bacterium capable of transforming plants through horizontal gene transfer, is nowadays the preferred vector for plant genetic engineering. The vehicle for transfer is the T-strand, a single-stranded DNA molecule bound by the bacterial protein VirD2, which guides T-DNA into the plants nucleus where it integrates. How VirD2 is removed from T-DNA, and which mechanism acts to attach the liberated end to the plant genome is currently unknown. Here, using newly developed technology that yields hundreds of T-DNA integrations in somatic tissue of *Arabidopsis thaliana*, we uncover two redundant mechanisms for the genomic capture of the T-DNA's 5' end. Different from capture of the 3' end of the T-DNA, which is the exclusive action of polymerase theta-mediated end joining (TMEJ), 5' attachment is accomplished either by TMEJ or by canonical non-homologous end joining (cNHEJ). We further find that TMEJ needs MRE11, whereas cNHEJ requires TDP2 to remove the 5'-end blocking protein VirD2. As a consequence, T-DNA integration is severely impaired in plants deficient for both MRE11 and TDP2 (or other cNHEJ factors). In support of MRE11 and cNHEJ specifically acting on the 5' end, we demonstrate rescue of the integration defect of double-deficient plants by using T-DNAs that are capable of forming telomeres upon 3' capture. Our study provides a mechanistic model for how Agrobacterium exploits the plant's own DNA repair machineries to transform them.

Main Text

Agrobacterium tumefaciens-mediated transformation (AMT) is the most widely used method for generating transgenic plants. In nature, this soil bacterium transforms dicotyledonous plants by translocating part of its DNA, the transferred (T)-DNA, into plant cells, where it integrates into the plant's genome[1]. Subsequent expression of the *Agrobacterium* genes causes crown gall disease. Within *Agrobacterium* the T-DNA is located on a Tumor-inducing (Ti) plasmid flanked by a repeated sequence of 25 bp, the Left and Right Border repeat (LB and RB). These sequences are recognition sites for the virulence proteins VirD1 and VirD2, which generate ssDNA breaks required to liberate T-DNA as a single-stranded DNA molecule, the T-strand[2]. The VirD2 protein remains covalently bound to the 5' end of the T-strand[3,4] and pilots it into the plant cell through the type 4 secretion system[5] that is created by the *Agrobacterium* virulence program upon detection of wounded plant cells[6]. The T-DNA is subsequently imported into the nucleus[7] where it integrates at a random position in the genome[8]. The molecular mechanism by which the T-DNA is integrated into the plant genome has remained enigmatic until recently when it was found for *Arabidopsis thaliana* that this process critically depends on polymerase theta (Pol θ)[9], a host protein that acts in the repair of DNA double-strand breaks (DSBs) via end joining. Abundant genetic and biochemical research performed over the last few years has established that Pol θ facilitates repair of DSBs in a multitude of species by using (few) complementary bases in 3' protruding ssDNA break ends to carry out DNA extension on one break end using the other end as a template[10,11]. This biochemical property of the enzyme combined with occasionally occurring primer-template switching provides an explanation for two characteristic features that are observed at sites where Pol theta-mediated end joining (TMEJ) of genomic DSBs takes place, *i.e.* micro-homology and so-called templated

insertions. These features are also prevalent at the junctions of T-DNA integration sites[9,12] – in plant transgenesis, templated insertions have also been described as "filler" sequences[13,14]. However, while TMEJ presents a logical model for connecting the 3' end of a T-DNA to a potentially resected genomic break (Fig. 1a), the biochemistry of capture of its 5' end has not yet been elucidated. It is currently also unknown how plant cells remove the covalently attached VirD2 from 5' T-DNA ends to allow integration.

To study the capture of T-DNA by the Arabidopsis genome, and in particular the attachment of the RB, we developed an NGS-based method, which we termed TRANSGUIDE (T-DNA random integration site genome-wide unbiased identification), that allows us to identify hundreds of T-DNA-genome junctions (both LB and RB) in pools of root-transformed *Arabidopsis* cells (Fig. 1b). We employ custom-made software to filter for high-quality, reliable outcomes and annotate individual T-DNA integration junctions with respect to potentially relevant features, such as genomic position, loss of T-DNA or genomic DNA sequences, degree of microhomology, and absence or presence of filler DNA. The outcomes of this pipeline reliably represents *in vivo* biology: using PCR and Sanger sequencing on the input material we validated 23 out of 24 predictions (Supplementary Data 1).

Using this technology, we obtained a collection consisting of ~2200 RB-genome junctions and ~5100 LB junctions upon transformation of the Col-0 ecotype (Supplementary Data 2). Consistent with earlier findings[15], these junctions are scattered across the entire genome, with the exception of the pericentromeric regions (Supplementary Fig. 1). Arguing for a prominent role for Pol θ in integration, we found (very similar) filler DNAs to be abundantly present at both RB- and LB-genome junctions (Fig. 1c), however, the percentages were not identical: 32 % fillers at RB-genome junctions versus 39 % at LB-genome junctions. Also, the degree of junctional microhomology (the median being 1 bp for RB- versus 3 bp for LB-genome junctions) and loss of terminal nucleotides was different between RB and LB (Fig. 1d, 1e and Supplementary Fig. 2). As microhomology usage and filler formation are hallmark features of Pol θ activity, these data suggest an unequal involvement of this enzyme in the attachment of the two T-DNA ends.

We previously found that *Arabidopsis* plants deficient for Pol θ (*teb* mutants) are completely recalcitrant to AMT, arguing for an essential role for Pol θ in genomic capture of T-DNA. This conclusion is here further substantiated by demonstrating an almost complete absence of T-DNA-genome junctions in DNA isolated from root-transformed Pol θ deficient plants: instead of finding a few hundred T-DNA integrations, we obtained only few cases in pools of *teb* calli (Fig. 1f). To exclude potential methodological distortions, *e.g.* resulting from PCR steps within TRANSGUIDE, we also performed AMT competition experiments: we mixed DNA from wild type and *teb* that were transformed with nearly identical yet bar-coded T-DNA constructs and attributed T-DNA junctions to the appropriate genotype afterwards. These internally controlled experiments corroborate our finding that genomic T-DNA capture is Pol θ dependent (Fig. 1f, Supplementary Data 3). Of note, while the almost complete absence of T-DNA junctions in *teb* material unequivocally demonstrates that TRANSGUIDE outcomes for wild type plants represent *bona fide* biology, we cannot conclude that the residual T-DNA-genome junctions found in *teb* samples represent completed T-DNA integration, as opposed to *e.g.* one-sided capture, *in*

vivo recombination, or PCR artifacts. Interestingly, however, and in agreement with a recent report[16], we find the molecules representing genomic capture in *teb* to be almost exclusively RB-to-genome junctions (Fig. 1f). Together with the notion of a reduced signature of Pol θ activity at RB-genome junctions in Pol θ proficient plants, as compared to LB-genome junctions, this result may point to another, redundant, molecular mechanism capable of attaching the 5' end of T-DNA to the plant genome.

The obvious candidate for end joining activity other than TMEJ is canonical non-homologous end joining (cNHEJ), another pathway to repair genomic DNA breaks. Previous analysis of AMT in cNHEJ deficient *Arabidopsis* led to conflicting results: whereas some labs reported reduced T-DNA integration[17-20], others found no effects[21-23] or even elevated frequencies[23,24]. We investigated a potential involvement of cNHEJ in T-DNA capture by monitoring shoot development and performing TRANSGUIDE upon root transformation of cNHEJ deficient *ku70* and *lig4* Arabidopsis mutants. We a reduced number of shoots in cNHEJ deficient plants (Fig. 2a +2b), arguing that NHEJ action affects stable transformation but is not essential. TRANSGUIDE of calli subsequently revealed a profound effect on the composition of T-DNA-genome junctions, specifically at the RB side (Fig. 2c-e): whereas LB-genome junctions found in *ku70* and *lig4* mutant roots are indistinguishable from those found in wild type, RB-genome junctions isolated from NHEJ mutant plants were characterised by an increased degree of microhomology (median of 3 bp in *ku70* and *lig4*, versus 1 bp in wild type). In fact, when plotted for the degree of microhomology, the distribution of RB-genome junctions in NHEJ mutant conditions is similar to that of the LB-genome junction, in both NHEJ deficient and proficient contexts (Supplementary Fig. 3). This increased usage of microhomology is accompanied by increased loss of T-DNA sequence at the RB end, as well as an increased percentage of junctions containing fillers (Supplementary Fig. 3), which were of similar length as those observed in wildtype (Fig. 2e). We conclude that capture of the T-DNA 3' end critically depends on intrinsically mutagenic TMEJ, whereas the 5' end can be attached to the genome via two redundant activities, *i.e.* TMEJ and cNHEJ.

The identification of two end joining pathways capable of attaching the T-DNA 5' end to the plant genome stirs the question: which enzymatic activity removes the bacterial VirD2 protein that is covalently bound to the outermost 5' nucleotide of T-DNA? Although the sequence of events leading to completed T-DNA integration is unknown, one can envisage a scenario where Pol θ-mediated genomic capture of the T-DNA 3' end leads, simply by DNA synthesis using the T-DNA as a template, to conversion of the single stranded T-DNA into dsDNA (see Fig. 1a). The resulting structure would have a striking resemblance to DSB ends that occur during meiotic recombination (by SPO11), or follow from some types of chemotherapy (TOP2 poisons), both of which have proteins covalently attached to their 5' termini[25,26]. Removal of these end-blocking proteins is a prerequisite to DSB repair and one demonstrated mechanism for their removal involves MRE11-catalyzed nicking of the protein-linked strand distal to the DSB terminus[27]. Arabidopsis *MRE11* null mutant plants are sterile, hampering their analysis[28], however, an *mre11* hypomorphic allele (*mre11-2*) exists, which in a homozygous state confers sensitivity towards DNA damaging agents yet supports plant development[29]. We inspected T-DNA integration in this mutant background and found the RB-genome junction spectrum altered but inversely to what was observed in cNHEJ mutants: instead of a

more profound TMEJ signature we observed a clear depletion of TMEJ hallmarks in *mre11-2*: less microhomology at the junctions and reduced filler size (Fig. 2f + 2g). We conclude that MRE11 functionality is needed for Pol θ-mediated capture of the T-DNA 5' end – when impaired, only cNHEJ can perform this function. Interestingly, we find a wild-type profile for LB-genome junctions in *mre11* mutant plants (Supplementary Fig. 4), which could either mean that MRE11 is not needed to process genomic breaks for capturing a T-DNA, or that the hypomorphic *mre11-2* allele encodes a protein still capable of this activity. One prediction that follows from our genetic analyses is that while single cNHEJ and *mre11* mutant plants are proficient for AMT, double mutants may not be. This is indeed what we observe: whereas 30 - 60 % of calli derived from AMT-treated *ku70*, *ku80* and *mre11* mutant plants form shoots on selective medium (which we use as a proxy for stable T-DNA integration), we find none in *ku70 mre11* and *ku80 mre11* double mutant plants (Fig. 2h + 2i, Supplementary Fig. 5). Corroborating the absence of shoots, we also found a dramatic reduction in the number of junctions in *mre11 ku70* , and (to a somewhat lesser extent) in *mre11 lig4* calli using TRANSGUIDE competition experiments (Supplementary Fig. 6). Expression of a T-DNA encoded β-glucuronidase (GUS) marker demonstrates that the absence of T-DNA integration in the double mutants is not caused by impaired T-DNA transfer (Supplementary Fig. 7).

The notion of cNHEJ being proficient in attaching the 5' end of the T-DNA to the genome when MRE11 is impaired argues for another activity able to remove VirD2. The fact that most RB-genome junctions are without loss of the T-DNA's outermost 5' nucleotides suggests the action of an enzyme able to cleave the phosphotyrosyl bond between VirD2 and the 5' phosphate of the DNA, as such generating a ligatable end that can be used by cNHEJ. Previous work in a variety of biological systems has identified the tyrosyl-DNA phosphodiesterase 2 (TDP2) to possess such biochemical activity[30], hence we next assayed *Arabidopsis* plants deficient for the orthologous protein. Root tissue from such *tdp2* mutant plants was efficiently transformed by Agrobacterium as visualized by shoot formation from selected calli, demonstrating that TDP2 is not essential for T-DNA integration (Fig. 2i + 2j). However, similar to mutations in cNHEJ, also TDP2 deficiency alters the junctional spectrum, specifically of RB-genome junctions, which shifts towards a typical TMEJ profile (Fig. 2k, Supplementary Fig. 8). This outcome is consistent with a model where TDP2 acts to facilitate cNHEJ and in line with this interpretation, we find that AMT is severely impaired in *mre11 tdp2* double mutant plants (Fig 2i + 2j, Supplementary Fig. 5).

We next reasoned that mutant backgrounds that have impaired T-DNA integration because of an inability to capture the 5' end would be proficient for AMT in situations where 3' attachment of a T-DNA is sufficient to produce cells that stably transmit T-DNA. Such T-DNAs have been previously created: T-DNAs that at their 5' side contain so-called telomere repeat arrays (TRAs), being long stretches of sequence exclusively consisting of $(TTTAGGG)_n$, are able to trigger the formation of new telomeres following genomic capture at their 3' end[31] (see Fig. 3a for a schematic representation). Two types of outcomes are found upon AMT of TRA-containing T-DNAs *i.e.* type I: canonical T-DNA integration at a random position in the genome, and type II: telomere formation-dependent integration, which goes together with loss of DNA positioned between the new and former telomere[31]. Likely because of provoking haplo-insufficiency

(providing counter-selection for viability) type II integrations are preferentially found proximal to chromosomal ends (within ~2.5 mb) in full grown plants. We next performed AMT experiments using TRA-containing T-DNA (in parallel to control T-DNAs) in the aforementioned genetic backgrounds. A *lig4* mutant background was used to assay cNHEJ deficiency as Ku is involved in maintaining telomere homeostasis and also strongly affects *de novo* telomere formation[31-33]. In agreement with cNHEJ being required for AMT in plants with disturbed MRE11 function we found profoundly reduced shoot formation in *lig-4 mre11* mutant plants transformed with control T-DNA, although not to the same extent as observed for *ku70 mre11* and *ku80 mre11*, which failed to produce shoots altogether (Fig. 3b + 3c). However, successful AMT with a telomere-forming T-DNA construct did not require functional cNHEJ in the *mre11-2* mutant background (Fig. 3b + 3d), supporting the conclusion that cNHEJ action is specific to genomic attachment of the 5' end of T-DNAs. In agreement with the prediction that these integrations are predominantly of type II, we found upon inspection by TRANSGUIDE a profound overrepresentation of LB junctions mapping near the ends of chromosomes (Fig. 3e, Supplementary Fig. 9). The finding that AMT was reduced for *mre11 tdp2* mutant roots even with TRA-containing T-DNA, yet not in the respective single mutants (Fig. 3b +3d), argues that 5' covalently bound VirD2 is also a blocking entity to *de novo* telomere formation.

Following our previous elucidation of how, during AMT, the 3' end of a T-DNA molecule is attached to the plant genome, we have here identified the mechanisms by which the 5' end can be attached. In contrast to T-DNA's 3' end, which because of its chemical composition (*i.e.* a 3' hydroxyl at the terminus of a ssDNA molecule) is an ideal substrate for TMEJ, the structure of the 5' end needs additional processing to create a ligatable end. Our data suggests that MRE11 acts to liberate the 5' end to facilitate TMEJ, whereas TDP2 acts to allow genomic attachment via cNHEJ.

Given the biochemical properties of both MRE11 and TDP2, *i.e.* acting on dsDNA, we consider it likely that single-stranded T-DNA molecules are first converted to a double-stranded configuration prior to 5' attachment. One potential mechanism for such conversion is genomic capture of the T-DNA 3' end followed by DNA synthesis using the genomic end as a primer. In this way a new "extended" DSB end is created (see Fig. 3f) in which the VirD2 protein blocks 5' to 3' resection. Such a structure is conceptually similar to a meiotic SPO11-bound DSB-end or to a stalled TOP2 cleavage complex; substrates that for protein removal to facilitate repair depend either on MRE11 or on TDP2. However, the observation of relatively proficient "transient" expression of T-DNA-encoded genes in plants deficient for Pol θ argues for dsDNA formation also in the absence of genomic capture. It is conceivable that free-floating T-DNA molecules can also react with each other via the identified end-joining mechanisms prior to genomic capture, a process that may underlie two yet unexplained AMT phenomena: i) extrachromosomal T-circles[34,35], and ii) T-DNA conglomerates that were recently found to make up a large proportion of AMT outcomes[36,37].

The observation of cNHEJ-mediated attachment of T-DNA 5' ends also in Pol θ proficient cells reveals that a proportion of the integrations have used both pathways, *i.e.* cNHEJ for 5' and TMEJ for 3' attachment, as was previously hypothesized[38]. This finding may explain many seemingly contradictory

observations in mutant analysis that has confounded AMT research for several decades: the usage of cNHEJ over MRE11-stimulated TMEJ to capture the 5' end may be context dependent with respect to the AMT protocol, the reagents used, and the tissue that is targeted. cNHEJ repairs DSBs in G1 and in pre-replicative DNA in S phase[39], whereas recent work in mammalian cells argues for TMEJ in late-S/G2/M phases of the cell cycle[40], and it is thus tempting to speculate that the cell-cycle stage of the host cell when infected may dictate pathway choice and AMT outcome. Indeed, comparing the genome-T-DNA junction signature of AMT events derived from somatic transformation with those from germline transformation reveals that TMEJ is more prominently used to attach the 5' end of T-DNA in germ cells[9,12] (Supplementary Fig. 2).

Apart from providing a mechanistic understanding, we aim to unravel the biology of (T-)DNA integration to allow for improved biotechnological strategies to develop transgenic crops. Recent work demonstrated that homology-directed gene targeting in Pol θ-deficient plants goes without undesired integration of AMT reagents[41], which otherwise contaminates gene targeting in wild-type conditions. Here, we find that a combinatorial inhibition of *MRE11* and cNHEJ activities, for which inhibitors are available, also precludes random integration. We envisage that an increased understanding on how exogenously provided DNA molecules interact with the genome of a host plant can help in developing precise genome-engineering approaches to benefit crop development.

Methods

Plant lines and growth conditions. Insertional mutants used in this study: *ku70* [19,42] (Col-0, SALK_123144), *ku70* [19,32] (Ws), *ku80* [19] (SALK_016627), *lig4* [42] (SALK_044027), *tdp2-1* (SALK _043413), *teb-5* [9,43] (SALK_018851), *mre11-2* [19,29]. Double mutants were created by crossing single mutants. Plants were grown on soil at 20°C in a 16 h light/8 h dark cycle.

Root transformation for TRANSGUIDE and shoot formation assay. Root transformations were performed as described previously[44], using disarmed *Agrobacterium tumefaciens* strain AGL1[45] harbouring either pUBC (pUBC-YFP-Dest[46] with *ccdB* cassette removed), pUBC-2 (same as pUBC-YFP, but the sequence between secondary TRANSGUIDE primer and LB or RB nick was replaced by a semi-random 56 bp sequence), pWY82[47], or pCAS9 (pDE-CAS9[48,49] with gRNA against *PPO1;* AT4G01690), or pCAMBIA3301 (Cambia). After co-culture root explants were transferred to shoot induction medium with vancomycin and timentin to kill off remaining bacteria, and phosphinotricin to select for transformed plant cells. After 3 weeks of selection either calli were harvested for TRANSGUIDE analysis (20 per sample), or were transferred to fresh selection medium for assaying shoot formation. After a total of 6 weeks of selection, plates were photographed and calli were scored for shoot formation (without prior knowledge of the genotype); any leaf-like protrusions from callus tissue was considered shoot tissue.

Junction enrichment and sequencing. Enrichment of T-DNA-genome junctions was similar to the GUIDEseq procedure[50]. DNA extraction was performed with the Wizard genomic DNA isolation kit

(Promega; Madison, WI, USA). Sonication was performed with a Bioruptor (Diagenode; Liège, Belgium) for 6 cycles (30 seconds on, 30 seconds off) on 'high' intensity. End repair, A-tailing, and Y-adapter ligation was performed with the NEBNext ultra II kit (New England Biolabs; Ipswich, MA, USA), and the library amplification was performed with Phusion polymerase (Thermo Scientific; Waltham, MA, USA). See Supplementary Table 1 for the primers that were used. Sequencing was performed on the Illumina MiSeq (300 bp paired end, v3 chemistry, at LGTC; Leiden, The Netherlands) and on the Illumina NovaSeq 6000 (150 bp paired end, v1.5 chemistry, at GenomeScan BV; Leiden, The Netherlands). Samples were demultiplexed using bcl2fastq2 conversion software v2.2 (Illumina; San Diego, CA, USA).

Junction calling. Reads were clipped to 150 bp and adapters removed (Trimmomatic[51]). Reads with identical molecular identifier (adapter UMI + 6 bp from forward read + 6 bp from reverse read) were combined into consensus sequences using custom software. Mapping was done with BWA-mem[52], using the default settings. Reads with identical unique molecular identifier were combined into consensus sequences, and any remaining optical duplicates were excluded from the analysis. Read pairs were grouped into junctions based on their genomic positions. Second-in-pair reads were required to start with the secondary T-DNA primer and end with a genomic sequence. These reads were used to determine the exact genomic position, as well as filler and homology sequences and deletion length. First-in-pair reads (anchors) were counted for each junction, and indicated the number of fragments present in the sample that support the junction. For each junction we generated a consensus sequence and calculated the percentage of reads exactly matching the consensus (consensus match). The junctions were then filtered i) for duplicate positions between samples (barcode hopping was accounted for), ii) for number of anchors (at least 3), and iii) for the consensus match (at least 75%). For most analyses (except for junction number comparison) we applied an additional filter for fair comparison, because distances between primer and border were not constant: homology ≤ 57 bp, filler ≤ 22 bp, end deletion ≤ 26 bp.

Competition assay. Roots were transformed with either pUBC (barcode 1) or pUBC-2 (barcode 2). 10 calli were collected per sample, and equal DNA amounts of 2 samples with different barcodes were combined prior to junction enrichment. During junction calling the reads were assigned to the sample of origin using the barcode. Junctions with duplicate positions within a sample pair were removed.

Junction validation. Using the same DNA samples as used for TRANSGUIDE, we performed up to two PCRs (nested) followed by Sanger sequencing (Macrogen Europe BV; Amsterdam, The Netherlands) to determine the correctness of the called junctions. Junctions were selected semi-randomly, making sure different types of junctions (filler/ non-filler, intact/non-intact, etc) were included. See Supplementary Table 1 for the primers that were used.

GUS staining. After co-cultivation, root explants were stained overnight in phosphate buffer (pH 7.3) containing 1 mM $K_3Fe(CN)_6$,1 mM $K_4Fe(CN)_6$, 10 mM Na_2EDTA, 0.1% SDS, 0.1% Triton X-100 and 2 mM X-gluc, and destained using 70% ethanol.

References

1. Bevan, M. W. & Chilton, M.-D. T-DNA of the Agrobacterium Ti and Ri plasmids. *Annual review of genetics* **16**, 357–384 (1982).
2. Stachel, S. E., Timmerman, B. & Zambryski, P. Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from Agrobacterium tumefaciens to plant cells. *Nature* **322**, 706–712 (1986).
3. Ward, E. R. & Barnes, W. M. VirD2 protein of Agrobacterium tumefaciens very tightly linked to the 5'end of T-strand DNA. *Science* **242**, 927 (1988).
4. Scheiffele, P., Pansegrau, W. & Lanka, E. Initiation of Agrobacterium tumefaciens T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of superhelical T-border DNA in vitro. *J Biol Chem* **270**, 1269–1276, doi:10.1074/jbc.270.3.1269 (1995).
5. van Kregten, M., Lindhout, B. I., Hooykaas, P. J. & van der Zaal, B. J. Agrobacterium-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a type IV secretion system translocation signal. *Molecular plant-microbe interactions* **22**, 1356–1365 (2009).
6. Winans, S. C. Two-way chemical signaling in Agrobacterium-plant interactions. *Microbiological reviews* **56**, 12–31 (1992).
7. Citovsky, V. & Zambryski, P. Transport of nucleic acids through membrane channels: snaking through small holes. *Annual review of microbiology* **47**, 167–197 (1993).
8. Kim, S. I., Veena & Gelvin, S. B. Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant J* **51**, 779–791, doi:10.1111/j.1365-313X.2007.03183.x (2007).
9. van Kregten, M. *et al.* T-DNA integration in plants results from polymerase-theta-mediated DNA repair. *Nat Plants* **2**, 16164, doi:10.1038/nplants.2016.164 (2016).
10. Schimmel, J., van Schendel, R., den Dunnen, J. T. & Tijsterman, M. Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining. *Trends Genet* **35**, 632–644, doi:10.1016/j.tig.2019.06.001 (2019).
11. Ramsden, D. A., Carvajal-Garcia, J. & Gupta, G. P. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nature Reviews Molecular Cell Biology*, 1–16 (2021).
12. Kleinboelting, N. *et al.* The Structural Features of Thousands of T-DNA Insertion Sites Are Consistent with a Double-Strand Break Repair-Based Insertion Mechanism. *Mol Plant* **8**, 1651–1664, doi:10.1016/j.molp.2015.08.011 (2015).
13. Tinland, B. The integration of T-DNA into plant genomes. *Trends in plant science* **1**, 178–184 (1996).
14. Tzfira, T., Li, J., Lacroix, B. & Citovsky, V. Agrobacterium T-DNA integration: molecules and models. *TRENDS in Genetics* **20**, 375–383 (2004).
15. Shilo, S. *et al.* T-DNA-genome junctions form early after infection and are influenced by the chromatin state of the host genome. *PLoS Genet* **13**, e1006875, doi:10.1371/journal.pgen.1006875 (2017).
16. Nishizawa-Yokoi, A. *et al.* Agrobacterium T-DNA integration in somatic cells does not require the activity of DNA polymerase theta. *New Phytol* **229**, 2859–2872, doi:10.1111/nph.17032 (2021).

17. Friesner, J. & Britt, A. B. Ku80-and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. *The Plant Journal* **34**, 427–440 (2003).
18. Li, J. *et al.* Involvement of KU80 in T-DNA integration in plant cells. *Proceedings of the National Academy of Sciences* **102**, 19231-19236 (2005).
19. Jia, Q., Bundock, P., Hooykaas, P. J. J. & de Pater, S. Agrobacterium tumefaciens T-DNA Integration and Gene Targeting in Arabidopsis thaliana Non-Homologous End-Joining Mutants. *Journal of Botany* 2012, 1-13, doi:10.1155/2012/989272 (2012).
20. Mestiri, I., Norre, F., Gallego, M. E. & White, C. I. Multiple host-cell recombination pathways act in Agrobacterium-mediated transformation of plant cells. *Plant J* **77**, 511–520, doi:10.1111/tpj.12398 (2014).
21. Gallego, M. E., Bleuyard, J. Y., Daoudal-Cotterell, S., Jallut, N. & White, C. I. Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in Arabidopsis. *Plant J* **35**, 557–565, doi:10.1046/j.1365-313x.2003.01827.x (2003).
22. van Attikum, H. *et al.* 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, doi:10.1093/nar/gkg458 (2003).
23. Park, S. Y. *et al.* 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, doi:10.1111/tpj.12779 (2015).
24. Vaghchhipawala, Z. E., Vasudevan, B., Lee, S., Morsy, M. R. & Mysore, K. S. Agrobacterium may delay plant nonhomologous end-joining DNA repair via XRCC4 to favor T-DNA integration. *Plant Cell* **24**, 4110–4123, doi:10.1105/tpc.112.100495 (2012).
25. Hartsuiker, E., Neale, M. J. & Carr, A. M. Distinct requirements for the Rad32Mre11 nuclease and Ctp1CtIP in the removal of covalently bound topoisomerase I and II from DNA. *Molecular cell* **33**, 117–123 (2009).
26. Hartung, F. *et al.* The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. *The Plant Cell* **19**, 3090–3099 (2007).
27. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
28. Puizina, J., Siroky, J., Mokros, P., Schweizer, D. & Riha, K. Mre11 deficiency in Arabidopsis is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. *The Plant Cell* **16**, 1968–1978 (2004).
29. Bundock, P. & Hooykaas, P. Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in Arabidopsis MRE11 mutants. *Plant Cell* **14**, 2451–2462, doi:10.1105/tpc.005959 (2002).
30. Zeng, Z., Cortés-Ledesma, F., El Khamisy, S. F. & Caldecott, K. W. TDP2/TTRAP is the major 5'-tyrosyl DNA phosphodiesterase activity in vertebrate cells and is critical for cellular resistance to topoisomerase II-induced DNA damage. *Journal of Biological Chemistry* **286**, 403–409 (2011).

31. Nelson, A. D., Lamb, J. C., Kobrossly, P. S. & Shippen, D. E. Parameters affecting telomere-mediated chromosomal truncation in Arabidopsis. *Plant Cell* **23**, 2263–2272, doi:10.1105/tpc.111.086017 (2011).
32. Bundock, P., van Attikum, H. & Hooykaas, P. Increased telomere length and hypersensitivity to DNA damaging agents in an Arabidopsis KU70 mutant. *Nucleic acids research* **30**, 3395–3400 (2002).
33. Riha, K., Watson, J. M., Parkey, J. & Shippen, D. E. Telomere length deregulation and enhanced sensitivity to genotoxic stress in Arabidopsis mutants deficient in Ku70. *The EMBO Journal* **21**, 2819–2826 (2002).
34. Bakkeren, G., Koukolikova-Nicola, Z., Grimsley, N. & Hohn, B. Recovery of Agrobacterium tumefaciens T-DNA molecules from whole plants early after transfer. *Cell* **57**, 847–857 (1989).
35. Singer, K., Shiboleth, Y. M., Li, J. & Tzfira, T. Formation of complex extrachromosomal T-DNA structures in Agrobacterium tumefaciens-infected plants. *Plant Physiol* **160**, 511–522, doi:10.1104/pp.112.200212 (2012).
36. Jupe, F. *et al.* The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS Genet* **15**, e1007819, doi:10.1371/journal.pgen.1007819 (2019).
37. Pucker, B., Kleinbolting, N. & Weisshaar, B. Large scale genomic rearrangements in selected Arabidopsis thaliana T-DNA lines are caused by T-DNA insertion mutagenesis. *BMC Genomics* **22**, 599, doi:10.1186/s12864-021-07877-8 (2021).
38. Levy, A. A. T-DNA integration: Pol theta controls T-DNA integration. *Nat Plants* **2**, 16170, doi:10.1038/nplants.2016.170 (2016).
39. Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nature cell biology* **19**, 1–9, doi:https://doi.org/10.1038/ncb3452 (2016).
40. Llorens-Agost, M. *et al.* POLθ-mediated end joining is restricted by RAD52 and BRCA2 until the onset of mitosis. *Nature cell biology* **23**, 1095–1104 (2021).
41. van Tol, N. *et al.* Gene targeting in Polymerase Theta-deficient Arabidopsis thaliana. *The Plant Journal* (2021).
42. Du, Y., Hase, Y., Satoh, K. & Shikazono, N. Characterization of gamma irradiation-induced mutations in Arabidopsis mutants deficient in non-homologous end joining. *J Radiat Res* **61**, 639–647, doi:10.1093/jrr/rraa059 (2020).
43. Inagaki, S. *et al.* Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *Plant Cell* **18**, 879–892, doi:10.1105/tpc.105.036798 (2006).
44. Vergunst, A. C., de Waal, E. C. & Hooykaas, P. J. in *Arabidopsis Protocols* 227-244 (Springer, 1998).
45. Lazo, G. R., Stein, P. A. & Ludwig, R. A. A DNA transformation–competent Arabidopsis genomic library in Agrobacterium. *Bio/technology* **9**, 963-967 (1991).
46. Grefen, C. *et al.* A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *The Plant*

Journal **64**, 355–365 (2010).

47. Yu, W., Lamb, J. C., Han, F. & Birchler, J. A. Telomere-mediated chromosomal truncation in maize. *Proc Natl Acad Sci U S A* **103**, 17331–17336, doi:10.1073/pnas.0605750103 (2006).
48. Fauser, F., Schiml, S. & Puchta, H. Both CRISPR/C as-based nucleases and nickases can be used efficiently for genome engineering in A rabidopsis thaliana. *The Plant Journal* **79**, 348–359 (2014).
49. Shen, H., Strunks, G. D., Klemann, B. J., Hooykaas, P. J. & de Pater, S. CRISPR/Cas9-Induced Double-Strand Break Repair in Arabidopsis Nonhomologous End-Joining Mutants. *G3 (Bethesda)* **7**, 193-202, doi:10.1534/g3.116.035204 (2017).
50. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187–197 (2015).
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).

Figures

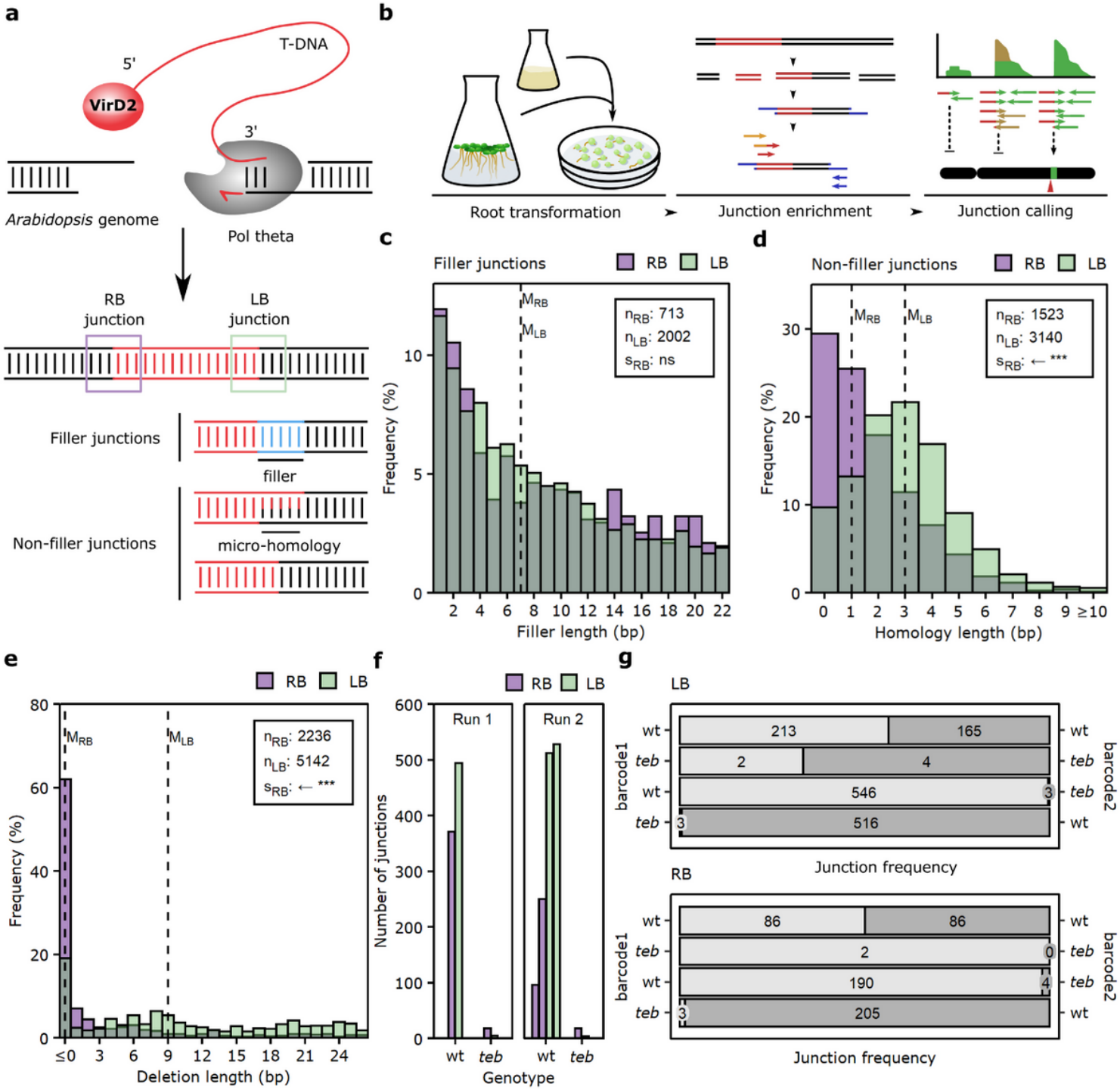

Figure 1

TRANSGUIDE reveals different characteristics for LB and RB T-DNA junctions. a, Tentative model for genomic capture of T-DNA via polymerase theta action. b, Schematic overview of TRANSGUIDE, see Methods section for details. c, d, e, Overlapping histograms showing the frequency of different lengths of filler (c) presence of micro-homology at the junctions (d) and loss of T-DNA (e) for LB junctions (light green) and RB junctions (purple). The medians (dashed lines), the number of observations (n), and shifts in the RB distribution relative to LB (s) are indicated. One-sided Wilcoxon rank-sum tests were performed to find the direction and the significance of the shifts; ns: p≥0.05, *: p<0.05, **: p<0.01, ***: p<0.001. The

arrow indicates the direction of the shift. f, Comparison of T-DNA-genome junction number between wt and teb callus obtained in two different sequencing runs (Run 1 and 2) from different callus samples (bars). g, Number of junctions in competitive TRANSGUIDE, in which equimolar amounts of genomic DNA of two samples with differently barcoded T-DNA (barcode 1 in light grey, and barcode 2 in dark grey) were combined.

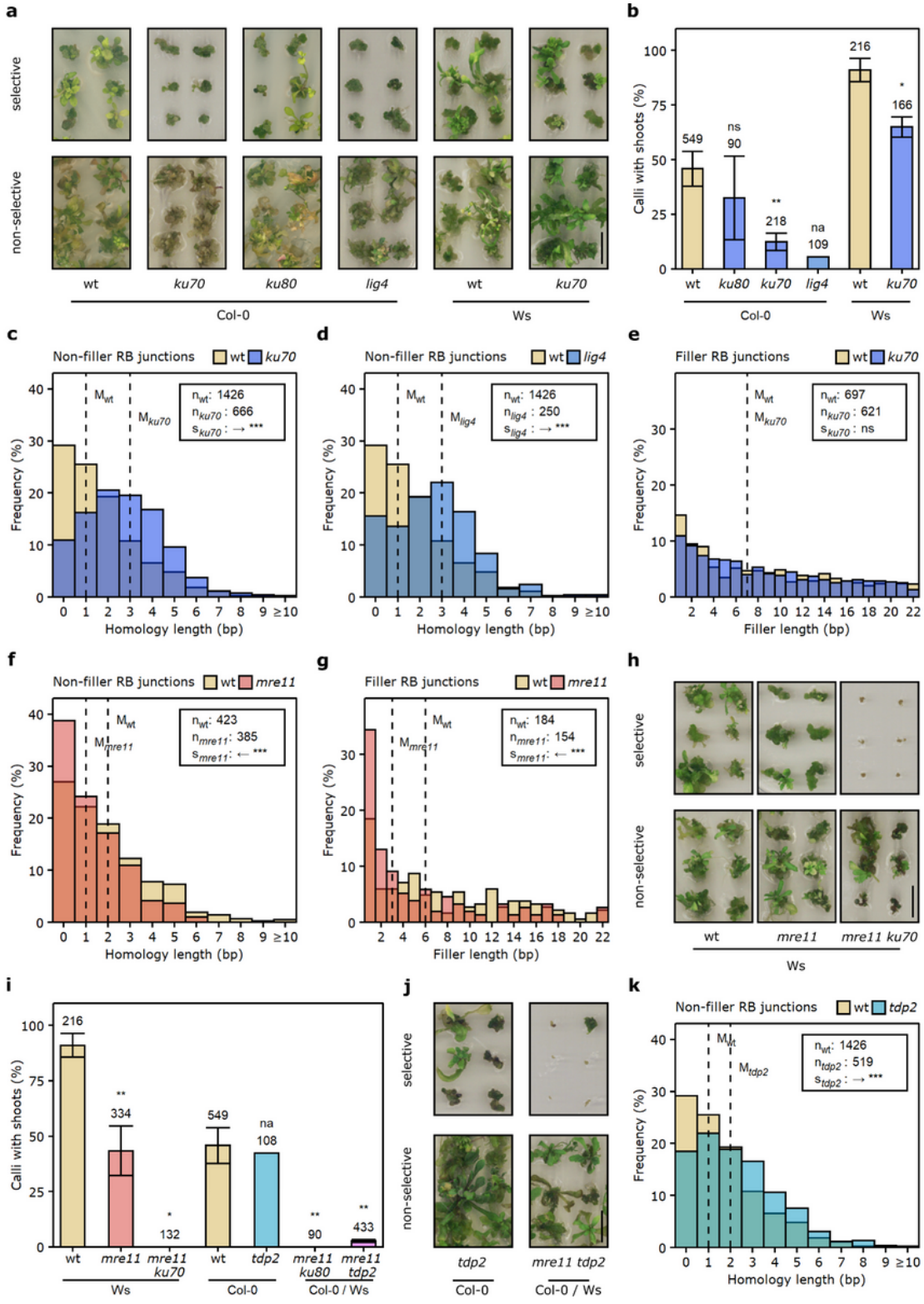

Figure 2

TMEJ and cNHEJ function redundantly in genomic capture of T-DNA's 5' end. a, Photographs of representative calli of various genotypes (wt, ku70, ku80, lig4) with two different genetic backgrounds (Col-0 and Ws) after transformation with pCAMBIA3301, grown on either selective (+PPT) or non-selective medium. b, Bar charts showing the average percentage of calli with shoot tissue (a measure of T-DNA integration success) after transformation with pCAMBIA3301. Error bars indicate the standard error of the mean. c - g, Overlapping histograms showing the frequency of junctions with the indicated degree of micro-homology (c, d, f), or filler presence (e, g) for wt (yellow) or mutant (shades of blue for cNHEJ mutants ku70 and lig4, and light red for mre11) junctions. The medians (dashed lines), the number of observations (n), and shifts in the mutant distribution relative to wt (s) are indicated. h, Photographs of representative calli of various genotypes (wt, mre11, mre11 ku70) after transformation with pCAS9, grown on either selective or non-selective medium. i, Bar charts showing the average percentage of calli with shoot tissue after transformation with pCAMBIA3301. Error bars indicate the standard error of the mean. j, Photographs of representative calli of tdp2 and mre11 tdp2 mutants with two different genetic backgrounds (Ws and a Col-0 / Ws mix) after transformation with pCAMBIA3301, grown on either selective or non-selective medium. k, Overlapping histograms showing the frequency of junctions with the indicated degree of micro-homology for wt (yellow) or tdp2 mutant (cyan). One-sided Student's t-tests were performed to test for significant reductions in T-DNA integration efficiency (b, i); ns: $p>=0.05$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, na: not enough observations. Mutants were compared to the wt of the same genetic background, except for mutants with mixed background, which were compared to the Col-0 wt. One-sided Wilcoxon rank-sum tests were performed to find the direction and significance level of the shifts in homology and filler distributions (c-g, k); ns: $p\geq0.05$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$. Scale bars, 1 cm in a, h, j.

Figure 3

MRE11 and TDP2 are required for 5' attachment of T-DNA. a, Schematic to illustrate that subsequent to genomic capture of its 3' end (1.), 5’ telomere repeat array (TRA)-containing T-DNAs can be resolved either via de novo telomere formation (2a.) or via regular RB capture (2b.). In case of telomere formation, loss of a part of the broken chromosome may ensue. b, Photographs of representative calli of various genotypes (wt, lig4, tdp2, mre11, mre11 lig4, mre11 tdp2) with three different genetic backgrounds (Col-0, Ws, or a

mix of Col-0 and Ws) after transformation without TRA (pUBC, - TRA) or with TRA (pWY82, + TRA) grown on selective medium. Scale bar, 1 cm. c, d, Bar charts showing the average percentage of calli with shoot tissue (a measure of T-DNA integration success) after transformation with pUBC (c) or pWY82 (d). Error bars indicate the standard error of the mean. One-sided Student's t-tests were performed to test whether mutants had decreased integration compared to the corresponding wt (in the case of Col-0/Ws the Col-0 wt was used); ns: $p \geq 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. Mutants were compared with wt of the same genetic background, with the exception of mre11 lig4 and mre11 tdp2, which were compared to the wt of the Col-0 background. Averages are based on different independent experiments; numbers above the bars indicate the total numbers of calli that were examined. e, Density plots showing the relative frequency of LB junctions after transformation with pWY82 (+ TRA) or pUBC (- TRA) along all chromosome arms, comparing wt (yellow) and mutants (other colors). Mutants were compared to wt of the same genetic background, with the exception of mre11 lig4 and mre11 tdp2, which were compared to the wt of the Col-0 background. 0 % indicates centromeric position and 100 % telomeric; n indicates the number of mutant junctions. One-sided Wilcoxon rank-sum tests were performed to find the direction and significance level of the shifts (s) in relative position; ns: $p \geq 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. Only junctions that are represented by more than 20 different DNA molecules (thus representing events that are compatible with multiple cell divisions) were included in this analysis. f, Proposed model of T-DNA integration: capture of the T-DNAs 3’ end essentially requires polymerase theta; genomic capture of the 5' end can be accomplished by cNHEJ after removal of VirD2 by TDP2, or via TMEJ after this end is processed by MRN (of which MRE11 is the enzymatic core) to generate a 3’ overhang, suitable for TMEJ action.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigures0211206.docx
- Supplementarydata120211206.xlsx
- Supplementarydata220211206.xlsx
- Supplementarydata320211206.xlsx
- SupplementaryTable120211206.xlsx