Gene targeting in polymerase theta-deficient arabidopsis thaliana

Tol, N. van; Schendel, R. van; Bos. A; Kregten, M. van; Pater, B.S. de; Hooykaas, P.J.J.; Tijsterman, M.

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **109,** 112–125 doi: 10.1111/tpj.15557

Gene targeting in polymerase theta-deficient *Arabidopsis thaliana*

Niels van Tol[1], Robin van Schendel[2], Alex Bos[1], Maartje van Kregten[1], Sylvia de Pater[1], Paul J.J. Hooykaas[1] and Marcel Tijsterman[1,2]*

[1]*Institute of Biology Leiden, Leiden University, Sylviusweg 72, Leiden, 2333 BE, The Netherlands, and*

[2]*Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, Leiden, 2300 RC, The Netherlands*

Received 26 May 2021; revised 19 October 2021; accepted 25 October 2021; published online 28 October 2021.

*For correspondence (e-mail M.Tijsterman@lumc.nl)

SUMMARY

***Agrobacterium tumefaciens*-mediated transformation has been for decades the preferred tool to generate transgenic plants. During this process, a T-DNA carrying transgenes is transferred from the bacterium to plant cells, where it randomly integrates into the genome via polymerase theta (Polθ)-mediated end joining (TMEJ). Targeting of the T-DNA to a specific genomic locus via homologous recombination (HR) is also possible, but such gene targeting (GT) events occur at low frequency and are almost invariably accompanied by random integration events. An additional complexity is that the product of recombination between T-DNA and target locus may not only map to the target locus (true GT), but also to random positions in the genome (ectopic GT). In this study, we have investigated how TMEJ functionality affects the biology of GT in plants, by using *Arabidopsis thaliana* mutated for the *TEBICHI* gene, which encodes for Polθ. Whereas in TMEJ-proficient plants we predominantly found GT events accompanied by random T-DNA integrations, GT events obtained in the *teb* mutant background lacked additional T-DNA copies, corroborating the essential role of Polθ in T-DNA integration. Polθ deficiency also prevented ectopic GT events, suggesting that the sequence of events leading up to this outcome requires TMEJ. Our findings provide insights that can be used for the development of strategies to obtain high-quality GT events in crop plants.**

Keywords: *Agrobacterium tumefaciens*, *Arabidopsis thaliana*, T-DNA integration, polymerase theta, homologous recombination, true gene targeting, ectopic gene targeting.

INTRODUCTION

The transferred DNA (T-DNA) of *Agrobacterium tumefaciens* can be custom designed to transfer sequences of interest into plant cells. Such T-DNAs may already be expressed in the extrachromosomal state before integration and may then be lost, a process known as transient expression, or may stably integrate into the host genome. It has been observed that the integration of a single copy of the T-DNA provides the best guarantee for stable transgene expression (Forsbach et al., 2003). However, T-DNAs are known to predominantly integrate as repeats of multiple copies (Grevelding et al., 1993; Jorgensen et al., 1987; Jupe et al., 2019; van Kregten et al., 2016) that are often subject to gene silencing (Mlotshwa et al., 2010; Schubert et al., 2004; Stam et al., 2000). In addition, T-DNA integration is known to be accompanied by deletions, inversions, and translocations of chromosomal arms (Gang et al., 2019; Jupe et al., 2019; Nacry et al., 1998; Seagrist et al., 2018; Tax and Vernon, 2001). These factors complicate the search for high-quality transgenic plant lines with stable transgene expression, and cause noise in and after large-scale screens for targeted genome engineering events.

Random T-DNA integration is particularly undesired when highly precise gene editing via homology directed repair (HDR) is required. Physiological HDR at a site of DNA damage progresses through invasion of a homologous sequence, usually the sister chromatid, acting as a template for repair. This step can be exploited in genome engineering strategies by artificially providing a homologous sequence in which desired mutations or selectable markers have been incorporated, a process commonly known as gene targeting (GT) (reviewed in Huang and Puchta, 2019). Upon entering the plant cell nucleus such a sequence can act as a repair template for HDR via one-sided invasion of the genomic locus, leading to incorporation of the desired modifications via a mechanism that is generally thought to be synthesis-dependent strand

annealing (SDSA) (Huang and Puchta, 2019; Puchta, 1998). Such GT events, however, occur infrequently in plants, and require large-scale screens to be identified (Hanin et al., 2001; Offringa et al., 1990; Paszkowski et al., 1988; Risseeuw et al., 1995). Still then, most GT events are uncontrollably accompanied by random integrations of the T-DNA elsewhere in the genome, which need to be removed by backcrossing in subsequent generations (e.g., Lidder and Sonnino, 2012; Wolter et al., 2019). During GT experiments HDR can also lead to alteration of the T-DNA instead of the genomic locus, resulting in T-DNA copies that have acquired sequences coming from the targeted locus, but are located elsewhere in the genome. These events are commonly referred to as ectopic GT (EGT) events (Endo et al., 2006; Hanin et al., 2001; Jia et al., 2012; Offringa et al., 1993; Risseeuw et al., 1995). At present, little is known with respect to the molecular mechanism underlying EGT. SDSA-based models have been proposed (Huang and Puchta, 2019; Puchta, 1998), but insight into the biochemistry and the genetic requirements of EGT is *grosso modo* lacking.

For random integration of T-DNA in plant genomes, many different models have been proposed over the last few decades (e.g., Gheysen et al., 1991; Mayerhofer et al., 1991; Tinland, 1996; Tzfira et al., 2004). Recent work in the model species *Arabidopsis thaliana* has shed new light on to this matter by revealing that integration is accomplished by an alternative end joining mechanism, termed polymerase theta (Polθ)-mediated end joining (TMEJ). It was proposed that Polθ facilitates the capture of the T-DNA's single-stranded 3′ end by resected genomic double strand break (DSB) ends (van Kregten et al., 2016), directing stable integration via a mechanism that has not yet been elucidated. A hallmark of Polθ activity in many species, including human, is the occasional presence of templated insertions at sites where DNA ends are joined together (reviewed in Schimmel et al., 2019). Such templated insertions were recognized already 30 years ago at sites of T-DNA integration, where they have been termed "filler" sequences (De Buck et al., 1999; Mayerhofer et al., 1991; Salomon and Puchta, 1998; Windels et al., 2003). A recent systematic analysis of over 10 000 transgenic Arabidopsis lines generated via floral dip transformation revealed that approximately 50% of T-DNA-genome junctions showed such fillers. Of the remaining junctions without fillers, close to 90% displayed use of microhomology to the T-DNA (Kleinboelting et al., 2015). Consistent with both these features being indicators for Polθ action, no transformants containing stable T-DNA integration events have so far been reported in Polθ-deficient plants (van Kregten et al., 2016; Nishizawa-Yokoi et al., 2020).

In the present study, we have investigated GT in Arabidopsis plants that are mutated for the *TEBICHI* (*TEB*) gene (Inagaki et al., 2006, 2009; van Kregten et al., 2016), which encodes for Polθ. We show that *teb-5* plants are proficient for GT, although the frequency was found to be reduced as compared with wild-type plants. The products of GT in Polθ-deficient plants were, however, more clean: devoid of random integration events, which accompanied almost all GT events in Polθ-proficient plants, and not resulting from ectopic GT, which make up a relatively large fraction of the recombinants in wild-type plants. These findings indicate that Polθ is not needed for GT, corroborate the crucial role for Polθ in the process of T-DNA integration, and, in addition, reveal new insights into the process of EGT.

RESULTS

Experimental design of the study

To address GT experimentally we made use of a previously established assay for HDR at the endogenous *PPO* locus (At4g01690) of Arabidopsis (Hanin et al., 2001; de Pater et al., 2009, 2013, 2018). The PPO enzyme is essential for chlorophyll biosynthesis and disruptions of the *PPO* open reading frame (ORF) are lethal to photoautotrophically grown Arabidopsis plants (Duke et al., 1991). This HDR assay is based on introducing a DSB in the endogenous *PPO* locus, meanwhile also providing cells with a repair template that is homologous to the *PPO* locus at either side of the DNA break. This template, however, is not identical but contains mutations in codon 426 of the *PPO* coding sequence (TAC>ATG), which confer an amino acid change (Y426M) rendering the PPO enzyme resistant to the herbicide Butafenacil (Butafenacil^R) (Hanin et al., 2001). A single base pair (TCA>TTA) substitution at codon 305 is also included in the repair template (e.g., Hanin et al., 2001; de Pater et al., 2018), but is not required for Butafenacil^R (de Pater et al., 2018). This scheme allows for selection of GT events by screening progeny of transformed plants for Butafenacil^R (Duke et al., 1991).

In this study, GT events were generated via floral dip transformation (Clough and Bent, 1998) of wild-type Columbia-0 (Col-0) plants and Polθ-deficient *teb-5* mutant plants (Inagaki et al., 2006, 2009; van Kregten et al., 2016) with the *Agrobacterium tumefaciens* strain AGL1 containing binary vector pDE-Cas9-PPO+RT (Strunks, 2019). In addition, we transformed immunodeficient *efr-1* mutant (Wu et al., 2014) and Polθ-deficient *efr-1 teb-5* double mutant plants. The *efr-1* mutant is a T-DNA insertion knock-out of the *EFR* gene, encoding for the immunoreceptor binding to the elongation factor Tu MAMP (Zipfel et al., 2006), rendering it hypersusceptible to Agrobacterium infection (Wu et al., 2014). We hypothesized that this sensitized genetic background may benefit the experimental setup by increasing the number of cells in which GT can take place.

A schematic overview of the experimental design is shown in Figure 1. In brief, the T-DNA harbors a *CAS9* expression cassette under control of the *UBIQUITIN* (*UBI*)

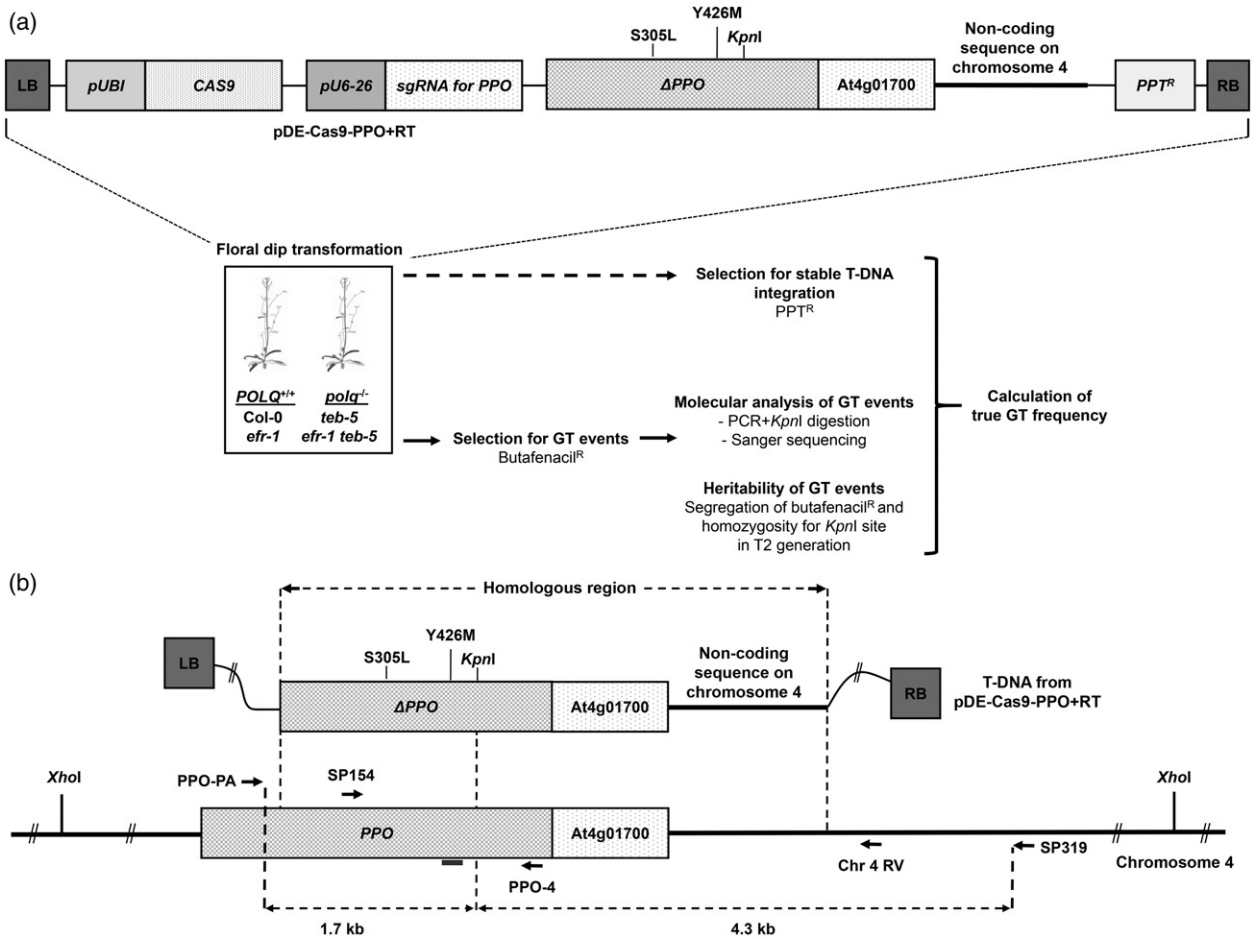

Figure 1. Overview of the experimental design of the study.
(a) Polymerase theta (Polθ)-proficient (Col-0 and *efr-1*) plants and Polθ-deficient (*teb-5* and *efr-1 teb-5*) plants were floral dipped with *Agrobacterium tumefaciens* strain AGL1 containing the binary vector pDE-Cas9-PPO+RT, encoding for a sgRNA cassette targeting the Arabidopsis *PPO* locus (At4g01690), a *CAS9* expression cassette under control of the ubiquitin promoter (*pUBI*), a PPT resistance marker (*PPT*^{R}), 2139 bp of sequence homology to the *PPO* locus carrying mutations in codons 305 (S305L) and 426 (Y426M), and silent mutations introducing a *Kpn*I restriction site. This stretch of homology (*ΔPPO*) was (i) truncated at the 5′ end and therefore lacks signals required for gene expression, including the ATG start codon, and the sequence encoding for the N-terminal chloroplast transit peptide, (ii) at the 3′ end extends into the directly adjacent At4g01700 locus due to (possible) overlap in regulatory sequences, and (iii) further downstream includes a stretch of non-coding sequence of chromosome 4. Approximately one-fourth of the obtained floral dip seed pools were selected for PPT^R to determine the frequency of stable T-DNA integration as a measure for transformation efficiency. Approximately three-fourths of the seeds were selected for Butafenacil^R to screen for gene targeting (GT) events. GT events were analyzed at the molecular level in the T1 and T2 generations.
(b) Molecular analysis of GT events at the *PPO* locus. *PPO* loci from Butafenacil^R T1 plants were amplified with primer combination PPO-PA and SP319. Primer sites were located outside of the homology provided on the T-DNA, so that only genomic *PPO* loci could be amplified. In addition, it should be noted that the protospacer (indicated with gray bar) for Cas9 overlapped with codon 426 and therefore only targeted the wild-type sequence, and not the *ΔPPO* repair template. The resulting polymerase chain reaction (PCR) products were digested with *Kpn*I to detect GT events. PCR products containing *Kpn*I sites were subsequently sequenced by Sanger sequencing with forward primer SP154. Inheritance of GT events was assessed via PCR and *Kpn*I digestion using the primer combination SP154 and Chr4 RV.

promoter, an sgRNA expression cassette under control of the PolIII *U6-26* promoter, a PPT resistance gene, and a segment of homology to the *PPO* locus in which were incorporated: (i) the single base substitution in codon 305 (TCA>TTA) leading to amino acid change S305L, (ii) the substitution of all three bases in codon 426 (TAC>ATG) leading to amino acid change Y426M, and (iii) silent mutations (TGTAAC>GGTACC; mutations underlined) leading to a novel *Kpn*I restriction site (Figure 1a). When transferred into wild-type target cells, the T-DNA may be transiently present, possibly resulting in transient Cas9 expression and DSB induction at the *PPO* locus, and/or stably integrate into the genome, resulting in stable Cas9 expression followed by DSB induction at the *PPO* locus. In both cases, the repair template on the T-DNA can be used for HDR of the DSB. Genomic incorporation of the substitution leading to the amino acid change Y426M in the PPO protein results in Butafenacil^R (de Pater et al., 2018).

Polθ-deficient *teb-5* plants do not have developmental or reproductive defects at standard growth conditions

In previous studies, Polθ-deficient *teb* mutants were described to have developmental defects along with mildly stunted growth when grown *in vitro* (Inagaki et al., 2006, 2009). In a recent study, however, *teb* mutants were found to develop comparatively normally, with severe phenotypic defects only occurring stochastically among a relatively small percentage of individuals, likely due to replicative stress triggered by environmental conditions (Nisa et al., 2021). Regardless, we anticipated that such phenotypic defects could introduce a bias in the outcome of Agrobacterium-mediated transformation of *teb* plants, as most protocols (in particular floral dip and root transformation) run over extended periods and might therefore be affected by differences in plant growth rate. To address this concern, a number of growth characteristics of soil grown *teb-5* plants were examined. We quantified rosette surface area of *teb-5* plants in different stages of vegetative growth; rosette surface area is considered an accurate and non-destructive proxy for overall growth and development (Leister et al., 1999; van Tol et al., 2017). Upon flowering (at approximately 4 weeks after germination) *teb-5* plants appeared on average slightly smaller than wild-type Col-0 plants, but this difference is not statistically significant (Figure 2a). In addition, the primary inflorescences of *teb-5* plants were not significantly shorter than those of Col-0 plants were, and did not display any developmental defects (Figure 2b). Importantly, *teb-5* plants on average produced the same amounts of seeds as the wild type, and these seeds were viable (Figure 2c). These results demonstrated that, in our greenhouse conditions, soil grown *teb-5* plants are not significantly affected in terms of growth and development, and that no experimental bias is likely to be introduced when using *teb-5* mutant plants for quantitative floral dip experiments. Apart from phenotypic characteristics, we also inspected the genomes of *teb-5* plants; as Polθ is involved in DNA repair (Inagaki et al., 2006; Mara et al., 2019; Nisa et al., 2021), the genomes of *teb-5* mutant plants might accumulate mutations, specifically large deletions (Koole et al., 2014; van Schendel et al., 2015). To address this notion, we generated so-called mutation accumulation (MA) lines by propagating randomly selected *teb-*

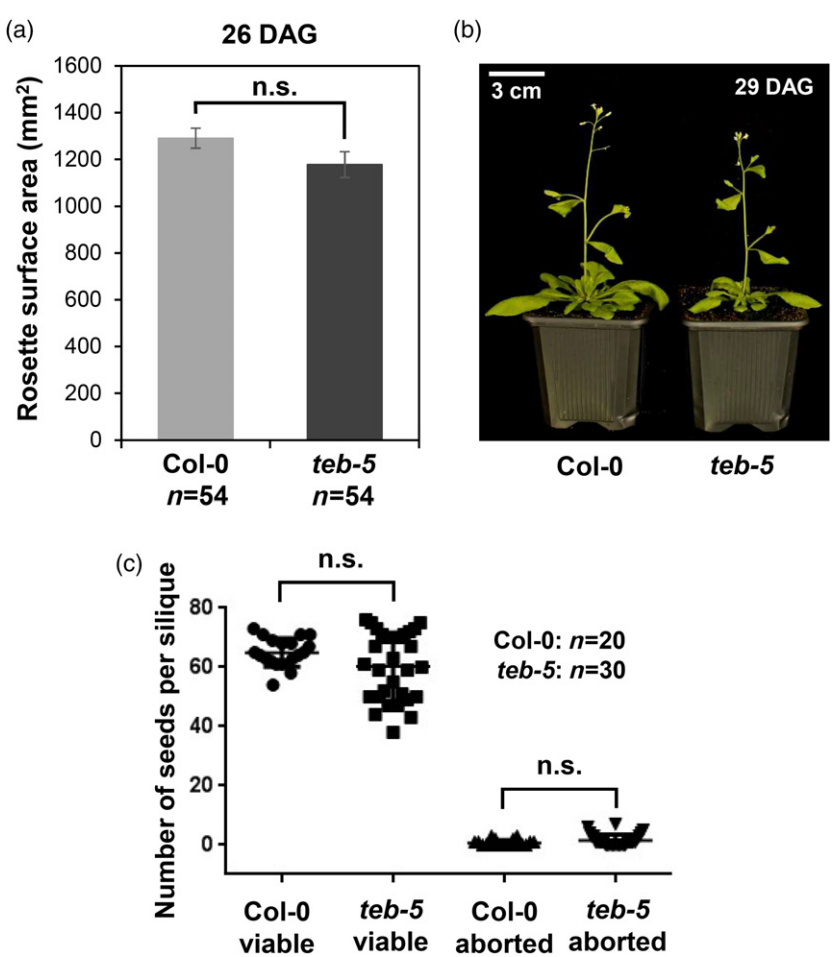

Figure 2. Growth phenotypes of wild-type Columbia-0 (Col-0) and polymerase theta-deficient *teb-5* plants.
(a) Rosette surface area of Col-0 and *teb-5* plants just before the start of flowering (26 days after germination [DAG]). Error bars represent SEM values ($n = 54$ per genotype). n.s., non-statistically significant difference at $P < 0.05$, as determined with a heteroscedastic T-test assuming unequal variance between samples.
(b) Representative photos of flowering Col-0 and *teb-5* plants at 29 DAG.
(c) Seed yield of Col-0 and *teb-5* plants. Numbers of viable and aborted (non-viable) seeds obtained from fully ripened siliques of Col-0 and *teb-5* plants ($n = 20$ and $n = 30$, respectively). n.s., non-statistically significant difference at $P < 0.05$, as determined with a heteroscedastic T-test assuming unequal variance between samples.

5 individuals in parallel over the course of five subsequent generations followed by whole genome sequencing. We found the single nucleotide variant (SNV) types and frequencies in the *teb-5* MA lines to be comparable with those previously reported for Col-0 (Exposito-Alonso et al., 2018; Ossowski et al., 2010), and identified only seven non-SNV variants, i.e., small deletions, in a total of 85 plant generations (17 lines; five generations) (Table S1). We conclude that Polθ deficiency does not trigger genomic instability and is thus unlikely to compromise GT analysis indirectly.

Polθ-deficient *teb-5* plants are devoid of stable T-DNA integration

In a previous study we showed that the frequency of stable T-DNA integration after floral dip transformation in Polθ-deficient *teb* mutants is at least 100–200 fold lower than in wild-type Col-0 plants (van Kregten et al., 2016). To (i) address the degree of resistance in *teb*-5 mutant plants against T-DNA integration, and (ii) set the T-DNA integration frequency in Col-0 to calculate the transformation efficiency, we selected pools of 100 000 seeds from each of five (Col-0 and *teb-5*) or three (*efr-1* and *efr-1 teb-5*) independent floral dip transformations for resistance to PPT, conferred by expression of the PPT resistance marker on the T-DNA. Polθ-proficient genotypes Col-0 and *efr-1* displayed average stable transformation frequencies of 1.4% and 1.2%, respectively (Table 1; data for individual replicates in Table S2), with stable transformation frequency being defined as the percentage of floral dip seeds germinating into PPT-resistant primary transformants. In contrast, the Polθ-deficient *teb-5* and *efr-1 teb-5* genetic backgrounds were completely devoid of stable T-DNA integration (Table 1; Table S2), not only corroborating previous findings, but realizing a 10-fold lower limit of stable T-DNA integration frequency for floral dip transformation in Polθ-deficient plants, to at least 1000–2000 fold lower than in the wild-type.

GT in the Polθ-deficient *teb-5* genetic background

To investigate GT in Polθ-deficient Arabidopsis, we screened pools ranging from 200 000 to 500 000 seeds derived from five (Col-0 and *teb-5*) or three (*efr-1* and *efr-1 teb-5*) independent floral dip transformations, respectively, with Agrobacterium strain AGL1 containing pDE-Cas9-PPO+RT for Butafenacil^R. From the Polθ-proficient Col-0 and *efr-1* genetic backgrounds we obtained, in total, 34 and 19 T1 plants, respectively, from which the Butafenacil^R was inherited to the T2 generation (Table 1). In further molecular analysis for GT events we also included a few sterile Butafenacil^R T1 plants for which stable transmission of the marker to the next generation obviously could not be verified beforehand.

To analyze plants for a potential true GT (TGT) outcome, we generated a 6-kb polymerase chain reaction (PCR) product containing the *PPO* locus with forward and reverse primer combination PPO-PA and SP319 (Figure 1b), which are both located outside of the repair template provided on the T-DNA. As incorporation of all the modifications from the repair template into the genomic *PPO* locus would introduce a novel, diagnostic *Kpn*I site, we digested this 6-kb PCR product with the restriction enzyme *Kpn*I, in the case of a targeted *PPO* locus yielding two bands of 4.3 and 1.7 kb, respectively (Figures 1b and 3a). We found that 33 of 53 potential GT outcomes that were obtained from Polθ-proficient genetic backgrounds contained a *PPO* locus with

Table 1 Output of gene targeting experiments at the *PPO* locus in Polθ-proficient (Col-0 and *efr-1*) and Polθ-deficient (*teb-5* and *efr-1 teb-5*) Arabidopsis plants

	Col-0	*teb-5*	*efr-1*	*efr-1 teb-5*
Stable transformation frequency (%PPT^R T1 plants among floral dip seed pools) ± SD	1.4 ± 0.4	0.0 ± 0.0	1.2 ± 0.6	0.0 ± 0.0
Floral dip seeds screened for GT events	1 875 000	1 770 000	1 000 000	815 000
T1 plants with heritable Butafenacil^R[a]	34	2	19	1
Sterile T1 plants (no T2 progeny)	1	1	0	0
Non-Mendelian segregation of Butafenacil^R in T2[a]	1	0	0	0
T1 plants containing TGT events[b]	23	3	10	1
T1 plants with heritable TGT events[c]	21	2	10	1
TGT frequency[d]	1.1×10^{-5}	1.1×10^{-6}	1.0×10^{-5}	1.2×10^{-6}

Presented data are the combined results of five (Col-0 and *teb-5*) or three (*efr-1* and *efr-1 teb-5*) independent experimental replicates. Data collected from the individual replicates are presented in Table S2. Data collected from individual Butafenacil^R T1 plants are presented in Tables S3 and S4. Col-0, Columbia-0; TGT, true gene targeting.

[a]Butafenacil^R:Butafenacil^S is 3:1 or 4:0 among T2 plants (for heterozygous and homozygous T1 parents, respectively).

[b]Determined by polymerase chain reaction amplification with primer combination PPO-PA and SP319, followed by *Kpn*I digestion and/or sequencing of the amplicon.

[c]Loss of wild-type *PPO* loci among approximately one-fourth of T2 plants and/or Mendelian segregation of Butafenacil^R.

[d]Number of T1 plants harboring true and heritable GT events divided by the total number of floral dip seeds screened, × 100%.

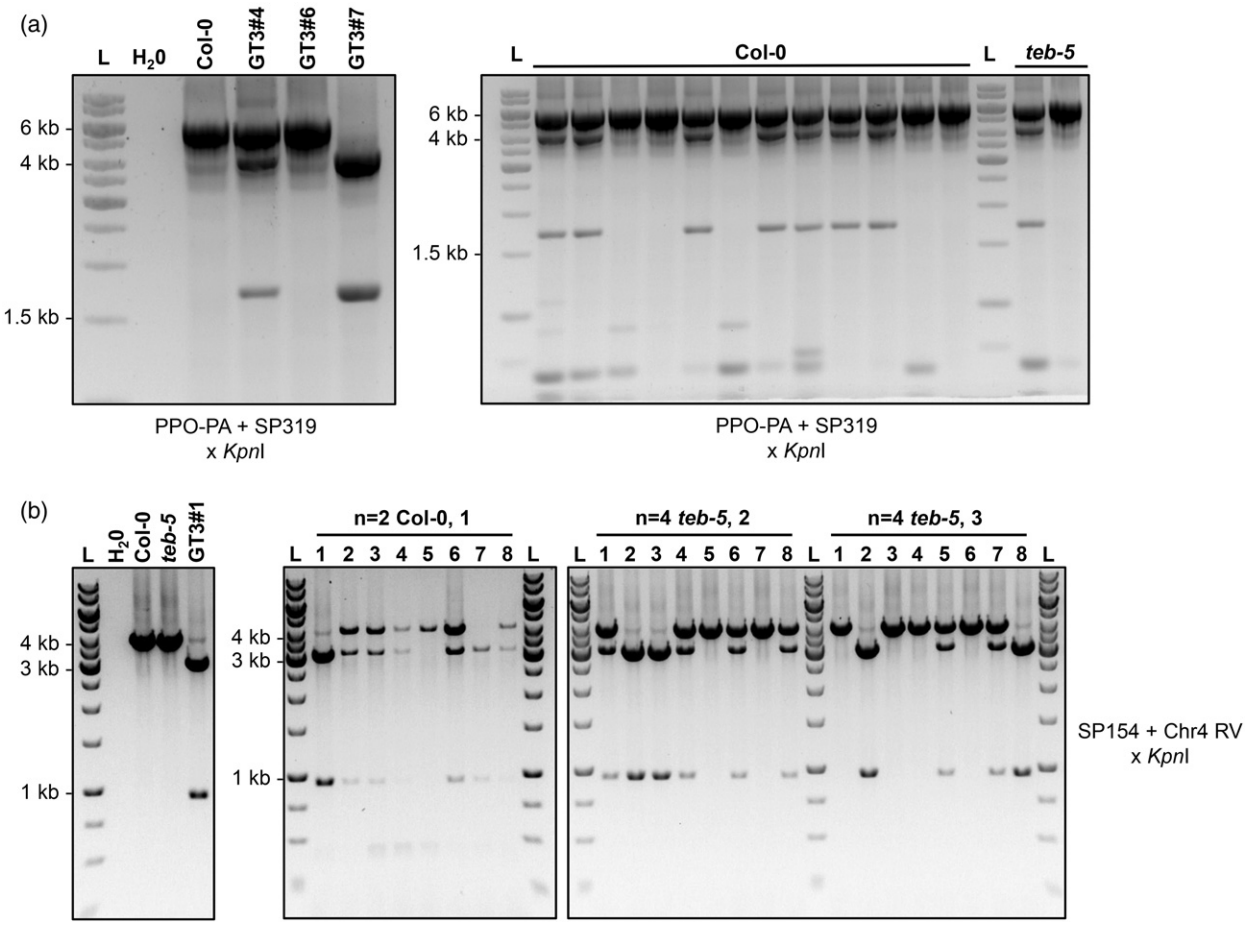

Figure 3. Molecular analysis of gene targeting (GT) events at the *PPO* locus.
(a) Gel electrophoresis of *Kpn*I digested polymerase chain reaction (PCR) products generated with primer combination PPO-PA and SP319 from genomic DNA samples isolated from heterozygous (GT3#4), wild-type (GT3#6), or homozygous (GT3#7) T2 segregants of previously described true GT line "GT#3" (de Pater et al., 2018; left panel) and from Butafenacil^{R} primary transformants generated in the Columbia-0 (Col-0) and *teb-5* genetic backgrounds in this study, isolated from experimental replica $n = 2$ (right panel). PCR products containing wild-type *PPO* loci yield 6 kb PCR fragments that are resistant to *Kpn*I digestion; PCR products containing a *PPO* locus with a GT event yield two fragments of 4.3 and 1.7 kb, respectively, after *Kpn*I digestion. Transformants showing all three bands (of 6, 4.3, and 1.7 kb) are therefore heterozygous for a GT event, containing one wild-type and one targeted *PPO* locus.
(b) Gel electrophoresis of *Kpn*I digested PCR products generated with primer combination SP154 and Chr4 RV from genomic DNA samples of wild-type Col-0 plants, *teb-5* plants, from previously described stable GT line GT3#1 (de Pater et al., 2018), which is homozygous for a targeted *PPO* locus (left panel), and from T2 segregants of representative Butafenacil^{R} primary transformant individual Col-0 1 from replicate $n = 2$, and *teb-5* individuals 2 and 3 from replicate $n = 4$ (right panels). PCR products from wild-type *PPO* loci yield 4 kb PCR fragments that are resistant to *Kpn*I digestion; PCR products from *PPO* loci with a GT event yield two fragments of 0.9 and 3.1 kb, respectively, after *Kpn*I digestion. Transformants showing all three bands (of 4, 3.1, and 0.9 kb) are therefore heterozygous for a GT event, containing one wild-type and one targeted *PPO* locus.

a *Kpn*I site diagnostic for TGT (Figure 3a), 32 of which were heterozygous (Table S3), while one may be homozygous for a modified *PPO* locus, as the PCR product was fully digested by *Kpn*I (Table S3). The remaining Butafenacil^{R} transformants (20) may harbor EGT events, which are analyzed in more detail below. For the 33 candidate TGT events we found Butafenacil^{R} to be inherited in a Mendelian manner in 31 cases (Table 1; Table S3). In one case, i.e., the plant suspected to have both alleles modified by GT, all T2 offspring were Butafenacil^{R}; in the remaining case, the T1 plant did not produce progeny (Table S3).

To provide further genetic evidence for TGT at the *PPO* locus, we genotyped unselected T2 segregants for loss of the wild-type *PPO* allele. To this end, we generated a 4 kb PCR product with primer combination SP154 and Chr4 RV (Figure 1b), which was subsequently digested with *Kpn*I, producing 3.1- and 0.9-kb fragments in case of a modified allele, while the non-targeted allele yields a 4-kb product. We found, on average, a quarter of the investigated T2 offspring of lines with suspected GT events to yield a fully digested PCR product, demonstrating homozygosity for the targeted endogenous allele (Figure 3b).

We next confirmed the presence of the Butafenacil^R conferring Y426M mutation in 15 of 15 GT-transformed plants that we analyzed by sequencing, next to a previously generated TGT line (GT3; de Pater et al., 2018) and untransformed Col-0 as controls (Figure 4). In two cases, the substitution of codon 426 coincided with the substitution of codon 305 (Figure 4), indicating that the vast majority of gene conversion tracts are smaller than 604 bp, and confirming that S305L is not needed for resistance. Based on the above criteria for TGT, we calculated the overall TGT frequencies in the Col-0 and *efr-1* genetic backgrounds to be 1.1×10^{-5} and 1.0×10^{-5}, respectively, when normalized for the total number of floral dip seeds screened (Table 1).

From the combined experimental replicates generated in the Polθ-deficient *teb-5* and *efr-1 teb-5* genetic backgrounds we isolated three heritable Butafenacil^R transformants, and one Butafenacil^R primary transformant that was sterile (Table 1; Table S3). All four T1 plants were heterozygous for the *PPO* allele carrying the Y426M edit, with three also containing the *Kpn*I site, as well as one plant that had the substitution at codon 305, in the heterozygous state (Figure 4). We next verified that these TGT events had been generated in Polθ-deficient conditions by confirming that the transformants were in the *teb-5* genetic background (to rule out cross-contamination in the screen). The overall TGT frequency in the *teb-5* and *efr-1 teb-5* genetic backgrounds was calculated to be 1.1×10^{-6} and 1.2×10^{-6}, respectively (Table 1), which is a 10-fold reduction compared with the wild-type background ($P = 0.016$ for *teb-5* vs. Col-0). Altogether, these data show that Polθ-deficient plants are proficient for GT, but that the frequency of TGT is reduced compared with Polθ-proficient genetic backgrounds.

Polθ deficiency eliminates experimental noise accompanying random T-DNA integration

The T2 progeny of primary transformants containing TGT events in the *teb-5* background proved fully sensitive to PPT (Table 2; Table S3), strongly suggesting that no random T-DNA integration had occurred in these plants. To verify the absence of T-DNA in these individuals, we also analyzed for the presence of the *CAS9* encoding ORF (adjacent to the LB of the T-DNA; Figure 1a) by PCR, to exclude that a truncated T-DNA lacking the PPT resistance marker had integrated. Indeed, none of the transformed *teb-5* plants contained the *CAS9* ORF (Table 2; Table S3). In contrast, 26 of 33 plants with TGT events in the Polθ-proficient backgrounds also contained the *CAS9* ORF, and 29 displayed segregation of PPT^R among the T2 progeny, indicative of random T-DNA integration elsewhere in the genome (Table 2; Table S3). Three of these 29 individuals apparently harbored truncated T-DNAs

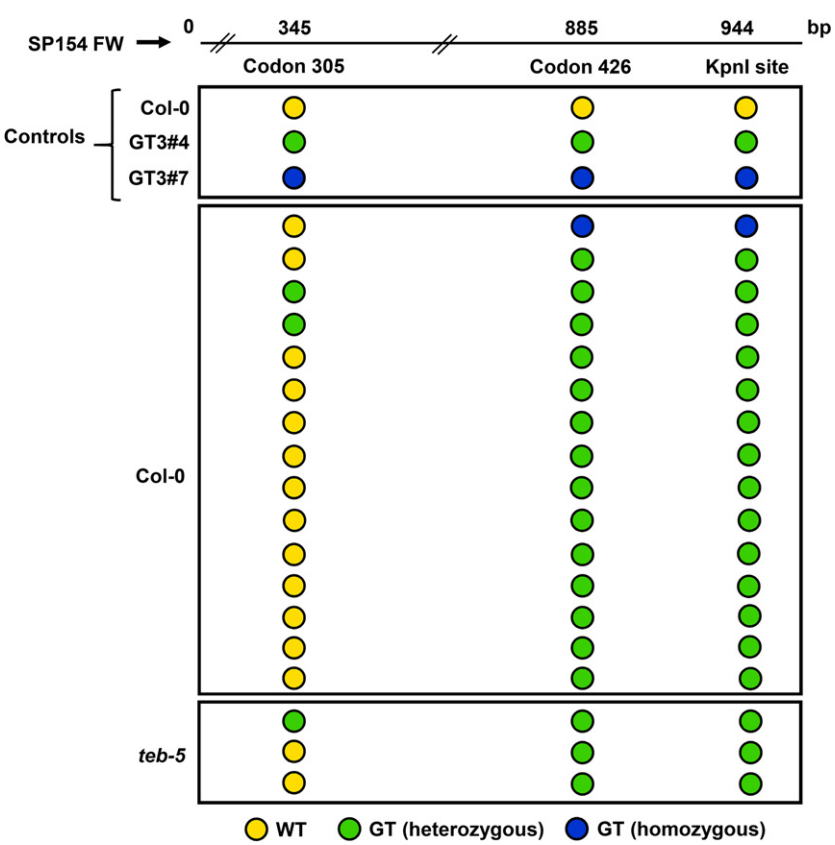

Figure 4. Overview of *PPO* genotypes of Butafenacil^R T1 plants containing true gene targeting (GT) events in wild-type (WT) Columbia-0 (Col-0) background and polymerase theta-deficient *teb-5* or *efr-1 teb-5* backgrounds.
Genotypes were determined via Sanger sequencing of *PPO* polymerase chain reaction products (PPO-PA+SP319) generated with sequencing primer SP154. Possible modifications incorporated through GT (at codon 305, codon 426, and/or the *Kpn*I site) are at the indicated distances in bp from the sequencing primer. Codons with WT sequences are indicated with yellow circles. Mutated codons incorporated through GT are indicated with either green circles (heterozygous state) or blue circles (homozygous state). *PPO* polymerase chain reaction products of WT Col-0 plants (WT codons) and lines containing true GT events in the heterozygous or homozygous state generated in a previous study (GT3#4 and GT3#7, respectively; de Pater et al., 2018) were sequenced as controls.

Table 2 Properties of Butafenacil^R T1 plants containing *PPO* loci sensitive or resistant to *Kpn*I digestion

	Sensitive to *Kpn*I digestion				Resistant to *Kpn*I digestion			
Butafenacil^R individuals	Col-0	*teb-5*	*efr-1*	*efr-1 teb-5*	Col-0	*teb-5*	*efr-1*	*efr-1 teb-5*
Total	23	3	10	0	25	7	13	2
Sterile (no T2 progeny)	1	1	0	0	3	2	0	0
Heritable Butafenacil^{Ra}	21	2	10	0	11	0	9	1
Heritable PPT^R	19	0	10	0	11	0	9	0
CAS9 ORF from the T-DNA detected[b]	16	0	10	0	11	0	7	0
Clean TGT events[c]	2	3	0	0	0	0	0	1
Clean heritable TGT events[d]	2	2	0	0	0	0	0	1

ORF, open reading frame; TGT, true gene targeting.
[a]Butafenacil^R:Butafenacil^S ratio in T2 generation of 3:1 or 4:0, respectively.
[b]Determined via polymerase chain reaction for *CAS9* with primer combination *CAS9* FW and RV.
[c]Determined by polymerase chain reaction amplification with primer combination PPO-PA and SP319, followed by *Kpn*I digestion and/or sequencing of the amplicon, and/or Mendelian segregation of Butafenacil^R in the T2 generation.
[d]Loss of wild-type *PPO* loci among approximately one-fourth of T2 plants.

containing the PPT resistance marker, but lacked the *CAS9* encoding ORF.

To substantiate further that TGT events derived from transforming *teb-5* plants, in contrast to Col-0 plants (P = 0.004 for *teb-5* vs. Col-0), are devoid of DNA remnants coming from the transformation reagents, we performed whole genome sequencing on the three heritable TGT lines derived from *teb-5* and one TGT line derived from Col-0, which served as a control. Figure 5 presents sequence coverage plots for the Col-0 TGT line and one of the *teb-5* TGT lines; homozygosity for the *teb*-5 mutated allele was verified by inspecting the genomic *TEB* locus (Figure 5a). Examination of the *PPO* locus (Figure 5b) clearly revealed that no additional *PPO* sequences were introduced in any of the *teb-5* TGT plants. Instead, and in perfect agreement with the expectation of TGT converting only one allele, we found for all three lines an approximately 50–50 coverage of the sequences encoding wild-type and altered *PPO*. In sharp contrast, we find for the TGT case in Col-0 that the copy number of the *PPO* sequence located on the T-DNA is dramatically increased (Figure 5b), suggesting the gain of approximately 30–40 copies of T-DNA via the transformation process. In agreement, only 2% of the reads covering codon 426 had the wild-type sequence. Finally, we mapped whole genome sequencing reads to the T-DNA as well as the bacterial vector sequence. While all three *teb*-5 TGT lines were completely devoid of T-DNA sequence and bacterial vector sequence, we found, in line with the *PPO* mapping data, that the Col-0 TGT line contained approximately 30–40 genome equivalents of almost the entire T-DNA (Figure 5c).

From our combined data we conclude that, whereas TGT events in the Polθ-proficient background are almost invariably accompanied by random T-DNA integration events, TGT events in the *teb-5* background are completely free of such events (Table 2).

From the output of the GT screens we noted that 20 of 53 of the Butafenacil^R individuals in the Polθ-proficient genetic backgrounds contained *PPO* loci that were resistant to *Kpn*I digestion (Table 2; Table S4) and did not contain the Y426M mutation (Figure S1), suggesting that these plants either contained EGT events (Endo et al., 2006; Hanin et al., 2001; Jia et al., 2012; Wolter et al., 2018), or for other yet unknown reasons escaped selection. The experimental noise caused by EGT events is a known phenomenon, and likely a consequence of recombination of the repair template on the T-DNA with the genomic *PPO* locus to acquire the 5′ gene expression and chloroplast translocation signals (Hanin et al., 2001), followed by random integration of the recombined T-DNA. Indeed, all these plants displayed segregation of Butafenacil^R along with PPT^R (Table 2; Table S4), suggesting cosegregation of Butafenacil^R with a T-DNA insertion. It is important to note that, as opposed to TGT events, Butafenacil^R segregates with the T-DNA in the case of EGT events, and may therefore show non-Mendelian segregation patterns in the T2 generation.

To address further the possibility of recombination between the T-DNA and the genome, we searched for such recombinant molecules: we hypothesized that a recombinant product conferring Butafenacil^R would manifest in PCR amplification of the transcriptional unit of the endogenous allele in plants for which we have shown that the endogenous locus is unmodified. To this end, an approximately 2.1-kb PCR amplicon was chosen that would only allow for amplification of a recombinant T-DNA, and not of non-recombinant molecules (see Figure 1a: forward primer PPO-PA and reverse primer PPO-4, the latter of which binds within the repair template on the T-DNA). We subsequently performed PCR on genomic DNA of all 20 Butafenacil^R non-TGT plants and digested the product with *Kpn*I. We found a significant *Kpn*I-sensitive subpopulation of molecules of

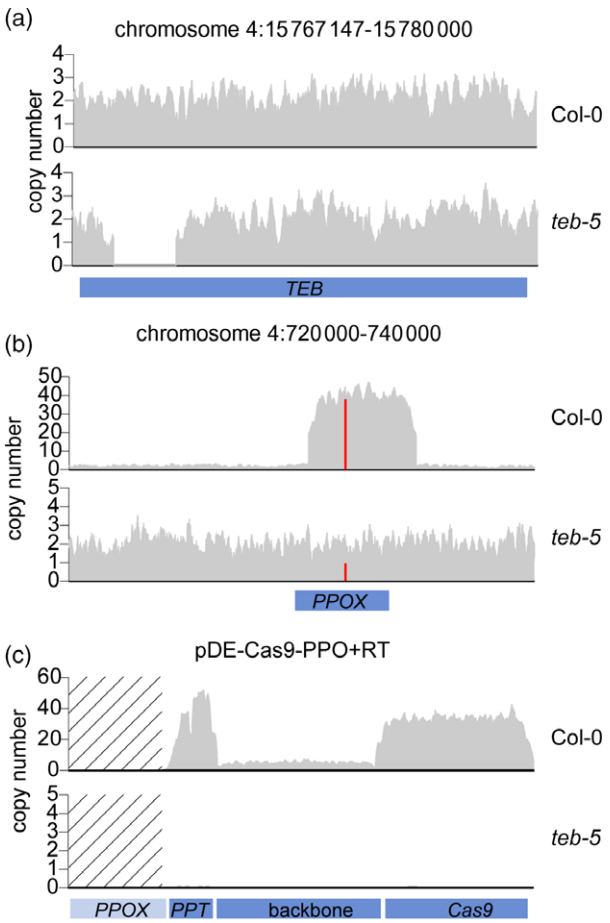

Figure 5. Whole genome sequencing analysis of true gene targeting (TGT) lines.
(a) Copy number representation of the genomic *TEB* locus in TGT lines of the indicated genotypes. Col-0, Columbia-0.
(b) Copy number representation of the genomic *PPO* locus for the TGT lines displayed in (a). Red vertical lines display the fraction of sequences containing the codon 426-encoding ATG triplet, either incorporated in the genomic locus, or present on the T-DNA. The *PPO* locus is here referred to as "*PPOX*" to distinguish it from the *PPO*/*TOPP2* locus in the reference genome.
(c) Copy number representation of sequences that uniquely map to the binary vector pDE-Cas9-PPO+RT that was used in this study; reads that mapped to the endogenous *PPOX* locus, or to the pBIN-ROK2 vector that was used to generate *teb-5*, were filtered out.

approximately 1.7 and 0.4 kb in 17 of 20 plants (Figure S2), arguing for *in vivo* recombination of the T-DNA with the endogenous *PPO* locus. This observation, together with the demonstration of unaltered *PPO* endogenes, supports the idea of EGT events in these individuals (Figure S2). Importantly, EGT events were not found in Polθ-deficient backgrounds (Table 2), suggesting that TMEJ plays an essential role in the mechanism underlying EGT.

DISCUSSION

Here, we have found that Polθ-deficient Arabidopsis plants are proficient for GT. Whereas GT events in Polθ-proficient genetic backgrounds are typically accompanied by random integration of T-DNA copies, GT events in *teb-5*, which in our study were all proven to be TGT events, are without such undesired noise.

Previous studies on GT have demonstrated that availability of homology to a sequence close to the DSB site (e.g., Baltes et al., 2014; Hahn et al., 2018) and design of the repair template (e.g., Hoshijima et al., 2016; Huang and Puchta, 2019; Puchta, 1998; Song and Stieger, 2017) can affect the frequency and nature of GT events. When investigating the nature of TGT events at codon 426 of the *PPO* locus we found that most events included the substitutions leading to a *Kpn*I site, which was approximately 60 bp downstream of the induced DSB, but lacked the substitution at codon 305 (Figure 4), approximately 540 bp upstream of it. Gene conversion tracts were thus rather small, resembling mammalian cells in which short gene conversion tracts, often less than 60 bp, have been observed in GT experiments (Elliott et al., 1998). In rare cases that had longer gene conversion tracts, all silent mutations present in the repair donor were incorporated. Such longer gene conversion tracts have recently also been reported in plants (Huang et al., 2021). It is still unclear for Agrobacterium-mediated GT whether the T-DNA substrate is captured by the genomic DSB after it is converted into a double-stranded configuration, as most models now envision, or whether it is assimilated as a single stranded molecule.

We found the frequency of GT in the Polθ-deficient genetic background to be reduced as compared with Col-0, which may argue for a yet unidentified role for Polθ in HR itself, as was recently suggested for the model alga species *Chlamydomonas reinhardtii* (Sizova et al., 2021). Such a role, however, contradicts numerous studies performed in other eukaryotic model systems (e.g., Mara et al., 2019; McVey et al., 2016; Zelensky et al., 2017). Other logical explanations for reduced GT in *teb-5* mutants that do not implicate Polθ directly are:

i A reduced induction of GT-stimulating DSBs in the endogenous *PPO* locus. Because Polθ deficiency eliminates random T-DNA integration, Cas9 expression in *teb-5* has become solely reliant on non-integrated transient expression. In Polθ-proficient backgrounds, however, DSBs can be induced by such transient expression as well as by stable expression resulting from genomic integration of the expression cassette on the T-DNA.
ii A reduced availability of repair donors. The notion of TGT events in *teb-5* indicates that non-integrated DNA can be used as a donor for GT. However, it is unknown to what extent potentially dozens of integrated T-DNA copies (in some cells perhaps in relatively close proximity to the DSB) may facilitate HDR of the introduced DSB. The latter may in fact explain the relatively low

GT frequency in plants, if GT is indeed dependent on previous nearby random integration of donor DNA.

Interestingly, GT can produce disadvantageous outcomes, i.e., DNA molecules resulting from recombination between donor and template present at seemingly random positions in the genome. This phenomenon, called EGT, creates experimental noise in screens and undesirable genome alterations in transformed plants (Endo et al., 2006; Hanin et al., 2001; Offringa et al., 1993; Risseeuw et al., 1995; Wolter et al., 2018). In the presented experiments, we show that the repair template on the T-DNA may acquire gene and protein expression signals from the chromosomal *PPO* locus, leading to Butafenacil^R that cosegregates with integrated T-DNA rather than with the *PPO* locus. The frequency with which such EGT events occur varies (de Pater et al., 2013; Hanin et al., 2001; this study), but can build up to comprise two-thirds of the output in TGT screens, likely depending on the reagents and methodologies used. Relatively little is known about the molecular mechanism underlying EGT, but the finding that *teb-5* mutant backgrounds lack EGT events points to an essential role for Polθ, although it should here be noted that also the TGT frequency is lowered. The recombination reaction between a transiently present repair template on the T-DNA and the genomic target locus leading to GT has previously been proposed to occur via SDSA (Huang and Puchta, 2019; Puchta, 1998). It is presently unknown which molecule (the T-DNA or the genome) triggers the recombinogenic process, and which serves as a donor. As already discussed earlier, it is also unclear whether T-DNA integrates before or after interacting with homologous sequences present in the genome. One can envisage a scenario wherein EGT occurs after initial integration of the T-DNA at an ectopic site followed by (ectopic) HR with the target locus. In that scenario, EGT would resemble so-called "*in planta* GT" between a locus and a previously integrated homologous T-DNA. In EGT, recombination between the T-DNA and the genome may also precede (Polθ-mediated) integration, but one would need to envision the existence of truncated T-DNA molecules to be able to acquire sequences from the genomic locus via an HR-resembling pairing and extension mechanism before integration elsewhere in the genome via TMEJ. Finally, recombination leading to EGT may take place during the process of T-DNA integration at a random genomic break, whereby the captured, but not-yet fully integrated T-DNA recombines with the donor locus before integration is completed. All three scenarios entail T-DNA capture by the genome, explaining the observed requirement for Polθ.

One concern when employing deficiency for a protein as a tool in genome editing strategies is how a plant tolerates loss of the respective protein. Arabidopsis Polθ-deficient plants were first described by Inagaki et al. (2006, 2009). In those studies, *in vitro* grown *teb* mutants containing different loss-of-function alleles were reported to have developmental defects (e.g., abnormal phyllotaxy, highly serrated and asymmetric leaves, severely disturbed organization of the shoot apical meristem). However, another study reports these defects to occur in only 10–15% of *teb* individuals, arising in a stochastic manner (Nisa et al., 2021). This percentage was found enhanced upon exposure to environmental stress, indicating that growth conditions affect the occurrence of these phenotypes. In our study, we found *teb-5* plants not significantly affected in terms of growth and development, particularly not in the flowering stages, and did not accumulate genomic scars. We thus concur with the suggestion made by Nisa et al. (2021), that the observed differences are likely due to differences in growth conditions between laboratories. In addition, *in vitro* growth on minimal media and exposure of the roots to light may exacerbate otherwise mild phenotypic differences between soil grown plants. It may very well be that species and context-dependent differences exist regarding the importance of Polθ proficiency for plant development, perhaps depending on the predominant DNA repair mechanisms in particular species. A recent study concerning *POLθ* function in the moss species *Physcomitrella patens* (Mara et al., 2019), showed that knockout plants were unaffected in terms of development and genome stability, similar to our findings for Arabidopsis. In addition, rice plants deficient for Polθ were found not to display phenotypic defects (Nishizawa-Yokoi et al., 2020). However, the latter report mentions potential problematic regeneration of transformed somatic tissue. These authors propose that a lack of obtaining transgenic progeny by transforming *teb* mutant somatic cells is due to fragility of the target tissue, and suggest the existence of tissue-specific alternative mechanisms of T-DNA capture (Nishizawa-Yokoi et al., 2020).

Despite the lack of gross developmental aberrations, a better and much more versatile strategy to use Polθ depletion in genome engineering approaches, particularly for crop developments, would be temporary inhibition. In that respect, it is noteworthy that this protein recently entered the limelight (e.g., Schrempf et al., 2021) as a target for anticancer treatment, as Polθ functionality promotes cell growth of tumor cells that are deficient in homologous recombination (Ceccaldi et al., 2015; Feng et al., 2019; Hwang et al., 2020). Numerous initiatives are currently ongoing to identify small molecule Polθ inhibitors (Cleary et al., 2020; Higgins and Boulton, 2018), of which two were recently published (Zatreanu et al., 2021; Zhou et al., 2021).

In conclusion, the findings described here provide important new insights into the process of GT via Agrobacterium-mediated transformation. In the future, these insights might be used to facilitate the development

of a one-step GT protocol, in which Polθ deficiency is employed possibly to allow for transient expression-based gene editing and GT strategies for targeted sequence modification.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

All *in vitro* cultured plants in this study were grown in a climate-controlled growth chamber at a temperature of 21°C, 50% relative humidity, approximately 50 μmol m^{-2} sec^{-1} of photosynthetically active radiation, and at a 16-h photoperiod. These conditions are referred to below as "tissue culture conditions." At 10–11 days after germination, herbicide-resistant (PPT^R or $Butafenacil^R$) T1 plants were transferred to soil and further cultivated in a climate-controlled growth chamber at a temperature of 20°C, 70% relative humidity, approximately 150 μmol m^{-2} sec^{-1} of photosynthetically active radiation, and at a 16 h photoperiod. These conditions are referred to below as "growth chamber conditions." T1 plants were allowed to complete their life cycle for the harvesting of T2 seeds. Soil-grown primary transformants and T2 plants were also grown at growth chamber conditions. All seeds were stratified for 3–6 days at 4°C before being placed in the growth chamber. Floral dip transformations (described further below) were performed in the wild-type Col-0 background, the Polθ-deficient *teb-5* background (Inagaki et al., 2006; van Kregten et al., 2016), the immunodeficient *efr-1* mutant (Wu et al., 2014), and the homozygous *efr-1 teb-5* double mutant (generated for this study).

Phenotypic characterization

Photos of plants and rosette surface area values were collected and processed as described previously (van Tol et al., 2017). The numbers of normally set and spontaneously aborted seeds were determined as described previously (Conger et al., 2011). Quantitative phenotypic data were inspected for normal distribution via Q–Q plots, and were statistically analyzed using heteroscedastic T-tests assuming unequal variance between samples.

MA lines and whole genome sequencing

For the generation of MA lines 17 *teb-5* plants were randomly selected in parallel. The seeds of these plants were collected and germinated into F1 plants. For each of the 17 lines a randomly selected F1 individual was propagated for the collection of F2 seeds. This process was repeated for five consecutive generations, finally yielding 17 independent populations of F5 plants. Genomic DNA was isolated from leaf material of randomly selected F5 plants, and from plants with the parental *teb-5* genotype using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Genomic DNA was sequenced using an Illumina Novaseq 6000 (Illumina, San Diego, CA, USA) machine according to protocol provided by the manufacturer. Image analysis, base calling, and error calibration were performed using the standard Illumina software. Raw reads were mapped to the *A. thaliana* reference genome (TAIR10) by BWA (Li and Durbin, 2009), and GATK's HaplotypeCaller (DePristo et al., 2011) was used for SNV calling. Pindel (Ye et al., 2009) was used to identify larger structural variations. Variants compared with the parental genotype were marked as true events if (i) they were *de novo* variants, thus not supported by other MA lines, if (ii) they were covered by both forward and reverse reads, and if (iii) they were covered at least five times. Mutation rates for different lines were calculated as the number of SNVs per generation divided by the number of generations grown.

Floral dip transformation

Floral dip transformations were performed according to a modified version of the originally published protocol (Clough and Bent, 1998). Plants of the respective genotypes were floral dipped at approximately 40 days after germination with Agrobacterium strain AGL1 (Lazo et al., 1991) containing binary vector pDE-Cas9-PPO+RT (Strunks, 2019), in a solution of 5% sucrose and 0.02% Silwet77 at a final $OD_{600\,nm}$ of approximately 0.8. Open flower buds, open flowers and siliques were removed from the plants before transformation. Floral dipped plants were incubated maximally for 24 h in growth chamber conditions in plastic bags, which were then opened. A suspension of AGL1 containing pDE-Cas9-PPO+RT in 5% sucrose and 0.03% Silwet77 was applied to newly grown out flower buds (that had therefore not been transformed via floral dip previously) by pencil brushing at 3–4 days after floral dip transformation, after which the plants were again incubated in closed plastic bags maximally for 24 h. Floral dip seed pools were harvested approximately 4 weeks after completion of the transformation protocol.

Selection for stable T-DNA integration and GT

Approximately 100 000 seeds from the respective floral dip seed pools were sterilized and plated on MA medium containing 0.5% sucrose, 100 μg ml^{-1} nystatin, 100 μg ml^{-1} Timentin, and 15 μg ml^{-1} PPT to select for primary transformants (T1 generation) containing stably integrated T-DNA constructs. Transformation frequency was subsequently calculated by dividing the total number of PPT^R transformants by the total number of floral dip seeds screened. Primary transformants containing GT events were selected for by plating the remainders of the floral dip seed pools (300 000–500 000 seeds) on MA medium containing 0.5% sucrose, 100 μg ml^{-1} nystatin, 100 μg ml^{-1} Timentin, and 50 nM Butafenacil. $Butafenacil^R$ T1 plants were counted, transferred to soil at 10–11 days after germination, and further cultivated until flowering.

Analysis of GT events

Genomic DNA was isolated from flower buds of $Butafenacil^R$ T1 plants using the CTAB protocol (Richards et al., 2001). A PCR amplicon containing the endogenous *PPO* locus was subsequently generated using the forward and reverse primer combination PPO-PA and SP319 (Table S5). The resulting 6-kb PCR fragment was digested with the restriction enzyme *Kpn*I (Thermo Scientific, Waltham, MA, USA), yielding 4.3- and 1.7-kb bands in the case of a GT event, and a resistant 6-kb product in the case of a wild-type *PPO* locus. Undigested PCR products were sequenced by Sanger sequencing (Macrogen Europe, Amsterdam, The Netherlands) using the primer SP154 (Table S5). The genotype of primary transformants containing TGT events in the *teb-5* background was verified using PCR reactions with primer combination SALK_015581 FW and SALK_015581 RV (to detect the wild-type *TEB* allele), and with combination SALK_015581 FW and pBIN LB RV (to detect the *teb-5* insertion allele) (Table S5).

Heritability of GT events

T2 seeds of $Butafenacil^R$ T1 plants were plated on MA medium containing 0.5% sucrose, 100 μg ml^{-1} nystatin, 100 μg ml^{-1} Timentin, and 50 nM Butafenacil to select for $Butafenacil^R$ T2 plants, and on half-strength Murashige and Skoog medium containing 100 μg ml^{-1} nystatin, 100 μg ml^{-1} Timentin, and 15 μg ml^{-1} PPT to select for PPT^R T2 plants. The numbers of $Butafenacil^R$,

Butafenacil^S, PPT^R, and PPT^S T2 seedlings were subsequently counted at 10–11 days after germination. Genomic DNA was extracted from single leaves of unselected T2 plants using a quick DNA isolation protocol (Edwards et al., 1991). A 4-kb PCR product was generated with the primer combination SP154 and Chr4 RV (Table S5) using this genomic DNA as a template, which was subsequently digested with *Kpn*I, yielding 4.0, 3.1, and 0.9 kb fragments in the case of heterozygous segregants, 3.1 and 0.9 kb fragments in the case of homozygous segregants, and a 4 kb undigested product in the case of wild-type segregants lacking *Kpn*I sites.

PCR analysis of random integration and EGT events

The presence of random T-DNA inserts was determined at the molecular level via detection of the full *CAS9* ORF (adjacent to the LB of the T-DNA) using the primer combination CAS9 FW and CAS9 RV (Table S5). Conclusions were supported using the inheritance of PPT^R in the T2 generation (PPT resistance marker is adjacent to the RB of the T-DNA). EGT events were detected in Butafenacil^R individuals by *Kpn*I digestion of the 2.1-kb PCR product generated with the primer combination PPO-PA and PPO-4 (Table S5). Presence of a recombinant T-DNA would yield a subpopulation of molecules of 1.74 and 0.36 kb.

Whole genome sequencing of TGT lines

Genomic DNA was isolated from pools of unselected T2 seedlings from the three *teb-5* lines harboring heritable TGT events, and from a Col-0 TGT line ($n = 4$ Col-0 individual 7; Table S3) as a control, using the CTAB protocol (Richards et al., 2001). Genomic DNA was sequenced using an Illumina Novaseq 6000 machine according to a protocol provided by the manufacturer. Image analysis, base calling, and error calibration were performed using the standard Illumina software. Raw reads were mapped to the *A. thaliana* reference genome (TAIR10) by BWA (Li and Durbin, 2009). Coverage on the respective regions in the genome was determined using samtools mpileup. For the data presented in Figure 5c, the reads were mapped to the sequence of binary vector pDE-Cas9-PPO+RT, and to pBIN-ROK2, which contains the T-DNA used for generating the SALK *teb-5* and *efr-1* mutants. Only reads with a mapping quality of 60 were used to determine coverage in Figure 5c, and the region harboring the genomic *PPO* target locus was excluded from the analysis.

Statistical analyses

TGT frequencies of individual experimental replicates (normalized for the total number of seeds screened within each replicate) were compared using the non-parametric Mann–Whitney test, with $P < 0.05$ as the threshold for statistical significance. The ratios between the numbers of clean versus non-clean heritable TGT events (with "clean" being defined as being devoid of random integration based on PCR analysis and segregation of PPT^R) obtained from the total number of seeds screened per plant genotype were compared via chi-squared tests, with $P < 0.05$ as the threshold for statistical significance.

ACKNOWLEDGEMENTS

We wish to thank the Genome Editing team of Trait Research, BASF Innovation Center Ghent (Belgium) for fruitful discussions and for funding parts of this research. This work was also supported by an ALW OPEN grant (OP.269) from the Netherlands Organization for Scientific Research for Earth and Life Sciences to MT. We would like to thank Dr. Bert van der Zaal (Institute of Biology Leiden) for providing seeds of the homozygous *efr-1 teb-5* double mutant.

AUTHOR CONTRIBUTIONS

NvT, SdP, PH, and MT designed the study. NvT, AB, MvK, and SdP performed the experiments. NvT, MvK, and RvS analyzed the data. NvT, PH, and MT wrote the manuscript. All authors have read and approved the manuscript.

CONFLICT OF INTERESTS

PH, MvK, and MT are inventors on patent application US 2019/0390208 A1 entitled "Methods for transfecting plants and for reducing random integration events."

DATA AVAILABILITY STATEMENT

All data generated for this study have been included in the article or in the Supporting Information. Raw whole genome sequencing data have been deposited in the NCBI Sequence Read Archive (project number PRJNA772250). Constructs, bacterial strains and seed stocks are available from the corresponding author upon request.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. *PPO* genotypes of Butafenacil^R T1 plants with PPO-PA+SP319 PCR products resistant to *Kpn*I digestion.

Figure S2. Detection of ectopic GT events in the 20 Butafenacil^R primary transformants with *Kpn*I-resistant PPO-PA+SP319 fragments.

Table S1. Whole genome sequencing analysis of mutation accumulation lines generated in the *teb-5* background

Table S2. Output of gene targeting experiments at the *PPO* locus in Polθ-proficient (Col-0 and *efr-1*) and Polθ-deficient (*teb-5* and *efr-1 teb-5*) Arabidopsis plants

Table S3. Butafenacil^R (GT events) and PPT^R (stable T-DNA inserts) among the T2 progeny of Butafenacil^R T1 plants containing GT events based on PCR analysis of the *PPO* locus and *Kpn*I digestion of the PCR product

Table S4. Butafenacil^R (GT events) and PPT^R (stable T-DNA inserts) among the T2 progeny of Butafenacil^R T1 plants with *Kpn*I resistant PCR products generated with primer combination PPO-PA and SP319

Table S5. Oligo DNA sequences used for PCR analysis

REFERENCES

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A. & Voytas, D.F. (2014) DNA replicons for plant genome engineering. *Plant Cell*, **26**, 151–163.

Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I. *et al.* (2015) Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature*, **518**, 258–262.

Cleary, J.M., Aguirre, A.J., Shapiro, G.I. & D'Andrea, A.D. (2020) Biomarker-guided development of DNA repair inhibitors. *Molecular Cell*, **78**, 1070–1085.

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

Conger, R., Chen, Y., Fornaciari, S., Faso, C., Held, M.A., Renna, L. *et al.* (2011) Evidence for the involvement of the Arabidopsis SEC24A in male transmission. *Journal of Experimental Botany*, **62**, 4917–4926.

De Buck, S., Jacobs, A., Van Montagu, M. & Depicker, A. (1999) The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. *The Plant Journal*, **20**, 295–304.

de Pater, S., Klemann, B. & Hooykaas, P.J.J. (2018) True gene-targeting events by CRISPR/Cas-induced DSB repair of the PPO locus with an ectopically integrated repair template. *Scientific Reports*, **8**, 3338.

de Pater, S., Neuteboom, L.W., Pinas, J.E., Hooykaas, P.J. & van der Zaal, B.J. (2009) ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. *Plant Biotechnology Journal*, **7**, 821–835.

de Pater, S., Pinas, J.E., Hooykaas, P.J. & van der Zaal, B.J. (2013) ZFN-mediated gene targeting of the Arabidopsis protoporphyrinogen oxidase gene through Agrobacterium-mediated floral dip transformation. *Plant Biotechnology Journal*, **11**, 510–515.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Duke, S.O., Lydon, J., Becerril, J.M., Sherman, T.D., Lehnen, L.P. & Matsumoto, H. (1991) Protoporphyrinogen oxidase-inhibiting herbicides. *Weed Science*, **39**, 465–473.

Edwards, K., Johnstone, C. & Thompson, C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, **19**, 1349.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A. & Jasin, M. (1998) Gene conversion tracts from double-strand break repair in mammalian cells. *Molecular and Cellular Biology*, **18**, 93–101.

Endo, M., Osakabe, K., Ichikawa, H. & Toki, S. (2006) Molecular characterization of true and ectopic gene targeting events at the acetolactate synthase gene in Arabidopsis. *Plant and Cell Physiology*, **47**, 372–379.

Exposito-Alonso, M., Becker, C., Schuenemann, V.J., Reiter, E., Setzer, C., Slovak, R. *et al.* (2018) The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS Genetics*, **14**, e1007155.

Feng, W., Simpson, D.A., Carvajal-Garcia, J., Price, B.A., Kumar, R.J., Mose, L.E. *et al.* (2019) Genetic determinants of cellular addiction to DNA polymerase theta. *Nature Communications*, **10**, 4286.

Forsbach, A., Schubert, D., Lechtenberg, B., Gils, M. & Schmidt, R. (2003) A comprehensive characterization of single-copy T-DNA insertions in the Arabidopsis thaliana genome. *Plant Molecular Biology*, **52**, 161–176.

Gang, H., Liu, G., Zhang, M., Zhao, Y., Jiang, J. & Chen, S. (2019) Comprehensive characterization of T-DNA integration induced chromosomal rearrangement in a birch T-DNA mutant. *BMC Genomics*, **20**, 311.

Gheysen, G., Villarroel, R. & Van Montagu, M. (1991) Illegitimate recombination in plants: a model for T-DNA integration. *Genes & Development*, **5**(2), 287–297.

Grevelding, C., Fantes, V., Kemper, E., Schell, J. & Masterson, R. (1993) Single-copy T-DNA insertions in arabidopsis are the predominant form of integration in root-derived transgenics, whereas multiple insertions are found in leaf-disks. *Plant Molecular Biology*, **23**(4), 847–860.

Hahn, F., Eisenhut, M., Mantegazza, O. & Weber, A.P.M. (2018) Homology-directed repair of a defective glabrous gene in Arabidopsis With Cas9-based gene targeting. *Frontiers in Plant Science*, **9**, 424.

Hanin, M., Volrath, S., Bogucki, A., Briker, M., Ward, E. & Paszkowski, J. (2001) Gene targeting in Arabidopsis. *The Plant Journal*, **28**, 671–677.

Higgins, G.S. & Boulton, S.J. (2018) Beyond PARP-POLθ as an anticancer target. *Science*, **359**, 1217–1218.

Hoshijima, K., Jurynec, M.J. & Grunwald, D.J. (2016) Precise genome editing by homologous recombination. *Methods in Cell Biology*, **135**, 121–147.

Huang, T.K., Armstrong, B., Schindele, P. & Puchta, H. (2021) Efficient gene targeting in *Nicotiana tabacum* using CRISPR/SaCas9 and temperature tolerant LbCas12a. *Plant Biotechology Journal*, **19**, 1314–1324. https://doi.org/10.1111/pbi.13546

Huang, T.K. & Puchta, H. (2019) CRISPR/Cas-mediated gene targeting in plants: finally a turn for the better for homologous recombination. *Plant Cell Reports*, **38**, 443–453.

Hwang, T., Reh, S., Dunbayev, Y., Zhong, Y., Takata, Y., Shen, J. *et al.* (2020) Defining the mutation signatures of DNA polymerase theta in cancer genomes. *NAR Cancer*, **2**, zcaa017.

Inagaki, S., Nakamura, K. & Morikami, A. (2009) A link among DNA replication, recombination and gene expression revealed by genetic and genomic analysis of TEBICHI gene of *Arabidopsis thaliana*. *PloS Genetics*, **5** (8), e1000613. https://doi.org/10.1371/journal.pgen.1000613

Inagaki, S., Suzuki, T., Ohto, M.A., Urawa, H., Horiuchi, T., Nakamura, K. *et al.* (2006) Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *The Plant Cell*, **18**, 879–892.

Jia, Q., Bundock, P., Hooykaas, P.J.J. & de Pater, S. (2012) *Agrobacterium tumefaciens* T-DNA integration and gene targeting in *Arabidopsis thaliana* non-homologous end-joining mutants. *Journal of Botany*, **2012**, 989272. https://doi.org/10.1155/2012/989272

Jorgensen, R., Snyder, C. & Jones, J.D.G. (1987) T-DNA is organized predominantly in inverted repeat structures in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. *Molecular and General Genetics*, **207**, 471–477.

Jupe, F., Rivkin, A.C., Michael, T.P., Zander, M., Motley, S.T., Sandoval, J.P. *et al.* (2019) The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS Genetics*, **15**, e1007819.

Kleinboelting, N., Huep, G., Appelhagen, I., Viehoever, P., Li, Y. & Weisshaar, B. (2015) The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. *Molecular Plant*, **8**(11), 1651–1664.

Koole, W., van Schendel, R., Karambelas, A.E., van Heteren, J.T., Okihara, K.L. & Tijsterman, M. (2014) A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nature Communications*, **5**, 3216.

Lazo, G.R., Stein, P.A. & Ludwig, R.A. (1991) A DNA transformation–competent *Arabidopsis* genomic library in *Agrobacterium*. *Nature Biotechnology*, **9**, 963–967.

Leister, D., Varotto, C., Pesaresi, P., Niwergall, A. & Salamini, F. (1999) Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiology and Biochemistry*, **37**, 671–678.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Lidder, P. & Sonnino, A. (2012) Biotechnologies for the management of genetic resources for food and agriculture. *Advances in Genetics*, **78**, 1–167.

Mara, K., Charlot, F., Guyon-Debast, A., Schaefer, D.G., Collonnier, C., Grelon, M. *et al.* (2019) POLQ plays a key role in the repair of CRISPR/Cas9-induced double-stranded breaks in the moss *Physcomitrella* patens. *New Phytologist*, **222**, 1380–1391.

Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K. *et al.* (1991) T-DNA integration: a mode of illegitimate recombination in plants. *EMBO Journal*, **10**(3), 697–704.

McVey, M., Khodaverdian, V.Y., Meyer, D., Cerqueira, P.G. & Heyer, W.D. (2016) Eukaryotic DNA polymerases in homologous recombination. *Annual Review of Genetics*, **50**, 393–421.

Mlotshwa, S., Pruss, G.J., Gao, Z., Mgutshini, N.L., Li, J., Chen, X. *et al.* (2010) Transcriptional silencing induced by *Arabidopsis* T-DNA mutants is associated with 35S promoter siRNAs and requires genes involved in siRNA-mediated chromatin silencing. *The Plant Journal*, **64**, 699–704.

Nacry, P., Camilleri, C., Courtial, B., Caboche, M. & Bouchez, D. (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics*, **149**, 641–650.

Nisa, M., Bergis, C., Pedroza-Garcia, J., Drouin-Wahbi, J., Mazubert, C., Bergounioux, C. *et al.* (2021) The plant DNA polymerase theta is essential for the repair of replication-associated DNA damage. *The Plant Journal*, **106**, 1197–1207.

Nishizawa-Yokoi, A., Saika, H., Hara, N., Lee, L., Toki, S. & Gelvin, S.B. (2020) *Agrobacterium* T-DNA integration in somatic cells does not require the activity of DNA polymerase θ. *New Phytologist*, **229**, 2859–2872.

Offringa, R., de Groot, M.J., Haagsman, H.J., Does, M.P., van den Elzen, J.P. & Hooykaas, P.J. (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium* mediated transformation. *EMBO Journal*, **9**(10), 3077–3084.

Offringa, R., Frankevandijk, M.E.I., Degroot, M.J.A., Vandenelzen, P.J.M. & Hooykaas, P.J.J. (1993) Nonreciprocal homologous recombination between *Agrobacterium* transferred DNA and a plant chromosomal locus. *Proceedings of the National Academy of Sciences of the United State of America*, **90**, 7346–7350.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J.I., Warthmann, N., Clark, R.M., Shaw, R.G. *et al.* (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92–94.

Paszkowski, J., Baur, M., Bogucki, A. & Potrykus, I. (1988) Gene targeting in plants. *EMBO Journal*, **7**(13), 4021–4026.

Puchta, H. (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. *The Plant Journal*, **13**(3), 331–339.

Richards, E., Reichardt, M. & Rogers, S. (2001) Preparation of genomic DNA from plant tissue. *Current Protocols in Molecular Biology*, Chapter, 2, Unit 23. https://doi.org/10.1002/0471142727.mb0203s27

Risseeuw, E., Offringa, R., Frankevandijk, M.E.I. & Hooykaas, P.J.J. (1995) Targeted recombination in plants using *Agrobacterium* coincides with additional rearrangements at the target locus. *The Plant Journal*, **7**, 109–119.

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

Schimmel, J., van Schendel, R., den Dunnen, J.T. & Tijsterman, M. (2019) Templated insertions: a smoking gun for Polymerase Theta-mediated end joining. *Trends in Genetics*, **35**(9), 632–644.

Schrempf, A., Slyskova, J. & Loizou, J.I. (2021) Targeting the DNA repair enzyme polymerase theta in cancer therapy. *Trends Cancer*, **7**, 98–111.

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S. & Schmidt, R. (2004) Silencing in *Arabidopsis* T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *The Plant Cell*, **16**, 2561–2572.

Seagrist, J.F., Su, S.H. & Krysan, P.J. (2018) Recombination between T-DNA insertions to cause chromosomal deletions in *Arabidopsis* is a rare phenomenon. *PeerJ*, **6**, e5076.

Sizova, I., Kelterborn, S., Verbenko, V., Kateriya, S. & Hegemann, P. (2021) *Chlamydomonas* POLQ is necessary for CRISPR Cas9-mediated GT. *G3 (Bethesda)*, **11**(7), jkab114.

Song, F. & Stieger, K. (2017) Optimizing the DNA donor template for homology-directed repair of double-strand breaks. *Molecular Therapy – Nucleic Acids*, **7**, 53–60.

Stam, M., de Bruin, R., van Blokland, R., van der Hoorn, R.A., Mol, J.N. & Kooter, J.M. (2000) Distinct features of post-transcriptional gene silencing by antisense transgenes in single copy and inverted T-DNA repeat loci. *The Plant Journal*, **21**, 27–42.

Strunks, G.D. (2019) *Strategies for the improvement of genome editing in* Arabidopsis thaliana. PhD thesis. https://openaccess.leidenuniv.nl/bitstream/handle/1887/77743/fulltext.pdf?sequence=1

Tax, F.E. & Vernon, D.M. (2001) T-DNA-associated duplication/translocations in *Arabidopsis*. Implications for mutant analysis and functional genomics. *Plant Physiology*, **126**, 1527–1538.

Tinland, B. (1996) The integration of T-DNA into plant genomes. *Trends in Plant Science*, **1**(6), 178–184.

Tzfira, T., Li, J., Lacroix, B. & Citovsky, V. (2004) Agrobacterium T-DNA integration: molecules and models. *Trends in Genetics*, **20**(8), 375–383.

van Kregten, M., de Pater, S., Romeijn, R., van Schendel, R., Hooykaas, P.J. & Tijsterman, M. (2016) T-DNA integration in plants results from polymerase-theta-mediated DNA repair. *Nature Plants*, **2**, 16164.

van Schendel, R., Roerink, S.F., Portegijs, V., van den Heuvel, S. & Tijsterman, M. (2015) Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. *Nature Communications*, **6**, 7394.

van Tol, N., Rolloos, M., Pinas, J.E., Henkel, C.V., Augustijn, D., Hooykaas, P.J. *et al.* (2017) Enhancement of *Arabidopsis* growth characteristics using genome interrogation with artificial transcription factors. *PLoS One*, **12**, e0174236.

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

Wolter, F., Klemm, J. & Puchta, H. (2018) Efficient in planta gene targeting in *Arabidopsis* using egg cell-specific expression of the Cas9 nuclease of *Staphylococcus aureus*. *The Plant Journal*, **94**, 735–746.

Wolter, F., Schindele, P. & Puchta, H. (2019) Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biology*, **19**, 176.

Wu, H.Y., Liu, K.H., Wang, Y.C., Wu, J.F., Chiu, W.L., Chen, C.Y. *et al.* (2014) AGROBEST: an efficient *Agrobacterium*-mediated transient expression method for versatile gene function analyses in *Arabidopsis* seedlings. *Plant Methods*, **10**, 19.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, **25**, 2865–2871.

Zatreanu, D., Robinson, H.M.R., Alkhatib, O., Boursier, M., Finch, H., Geo, L. *et al.* (2021) Polθ inhibitors elicit BRCA-gene synthetic lethality and target PARP inhibitor resistance. *Nature Communications*, **12**(1), 3636. https://doi.org/10.1038/s41467-021-23463-8

Zelensky, A.N., Schimmel, J., Kool, H., Kanaar, R. & Tijsterman, M. (2017) Inactivation of Pol theta and C-NHEJ eliminates off-target integration of exogenous DNA. *Nature Communications*, **8**, 66.

Zhou, J., Gelot, C., Pantelidou, C., Li, A., Yücel, H., Davis, R.E. *et al.* (2021) A first-in-class Polymerase Theta Inhibitor selectively targets Homologous-Recombination-Deficient Tumors. *Nature Cancer*, **2**(6), 598–610.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T. *et al.* (2006) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell*, **125**, 749–760.