Molecular analysis of the role of polymerase theta in gene targeting in Arabidopsis thaliana

Kralemann, L.E.M.; Tol, N. van; Hooykaas, P.J.J.; Tijsterman, M.

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) **118,** 255–262 doi: 10.1111/tpj.16689

Molecular analysis of the role of polymerase theta in gene targeting in *Arabidopsis thaliana*

Lejon E. M. Kralemann[1], Niels van Tol[1], Paul J. J. Hooykaas[1] and Marcel Tijsterman[1,2,*]

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

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

Received 16 October 2023; revised 5 January 2024; accepted 7 February 2024; published online 25 February 2024.

*For correspondence (e-mail m.tijsterman@lumc.nl).

SUMMARY

Precise genetic modification can be achieved via a sequence homology-mediated process known as gene targeting (GT). Whilst established for genome engineering purposes, the application of GT in plants still suffers from a low efficiency for which an explanation is currently lacking. Recently reported reduced rates of GT in *A. thaliana* deficient in polymerase theta (Polθ), a core component of theta-mediated end joining (TMEJ) of DNA breaks, have led to the suggestion of a direct involvement of this enzyme in the homology-directed process. Here, by monitoring homology-driven gene conversion in plants with CRISPR reagent and donor sequences pre-integrated at random sites in the genome (*in planta* GT), we demonstrate that Polθ action is not required for GT, but instead suppresses the process, likely by promoting the repair of the DNA break by end-joining. This finding indicates that lack of donor integration explains the previously established reduced GT rates seen upon transformation of Polθ-deficient plants. Our study additionally provides insight into ectopic gene targeting (EGT), recombination events between donor and target that do not map to the target locus. EGT, which occurs at similar frequencies as "true" GT during transformation, was rare in our *in planta* GT experiments arguing that EGT predominantly results from target locus recombination with nonintegrated T-DNA molecules. By describing mechanistic features of GT our study provides directions for the improvement of precise genetic modification of plants.

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

INTRODUCTION

An efficient procedure for precise, targeted modification of genes and exchange of alleles in plant genomes would enormously accelerate plant breeding to meet the challenges posed by climate change and a growing world population. Desired genetic changes can be accomplished by the process of gene targeting (GT), which is based on homologous recombination (HR) between an introduced donor DNA molecule and a target locus, but this procedure is inefficient in plants and thus requires strong selection and screening of large populations of transformants. The frequency of such events is not only low, but positive events are in most cases (>90%) accompanied by the co-occurrence of random integrations of the donor molecules elsewhere in the genome (Puchta, 1998; van Tol et al., 2022). Additionally, many positive events turn out to be no true GT events upon closer inspection (Huang & Puchta, 2019; van Tol et al., 2022). These include "one-sided" events, in which one side of the donor molecule recombines with the target locus, but the other side attaches to the genome via end-joining, as well as ectopic GT (EGT) events, in which the donor molecule acquires sequences from the target locus and integrates elsewhere in the genome.

Over the last decade, the application of new gene editing tools and especially the enzymes of the various CRISPR/Cas systems has led to a considerable increase in the efficiency of targeted mutagenesis, and has also led to an increased efficiency of GT (Huang & Puchta, 2019). Another way to stimulate GT is by having the donor sequence pre-integrated elsewhere into the genome via highly efficient end-joining. Initially, it was thought that DNA breaks within the target locus as well as flanking the GT donor were required to induce recombination between the homologous sequences (Fauser et al., 2012); however, it became later apparent that a double-strand break (DSB) at the target locus was sufficient and that "liberation" of

the donor sequence was not essential (de Pater et al., 2018). While this two-step "*in planta* gene targeting" approach is more efficient than the direct approach, it still requires great effort and the randomly integrated donor copies need to eventually be removed by crossing. Therefore, to increase the efficiency of GT in plants, it is of paramount importance to increase our knowledge concerning the molecular mechanisms underlying this process.

In many current protocols for GT in plants, the vehicle for introducing donor sequences as well as DSB-inducing CRISPR reagents is T-DNA, an initially ssDNA molecule that is delivered into plant cells through Agrobacterium-mediated transformation. T-DNA integration into the genome of *Arabidopsis thaliana* is absent in mutants in which the *TEBICHI (TEB)* gene, encoding polymerase theta (Polθ), has been corrupted (van Kregten et al., 2016). Polθ facilitates end joining repair of DSBs by capturing 3′ single strand ends, aligning them at complimentary bases (also called microhomologies), and subsequently extending one 3′ end using the annealed other strand as a template (Chan et al., 2010; Koole et al., 2014). In this process, termed theta-mediated end-joining (TMEJ), also extrachromosomal DNA, such as the free 3′ end of the T-strand, can be captured at a DSB (Kralemann et al., 2022; van Kregten et al., 2016). *teb* mutant plants, or conditions in which the protein's function is otherwise (e.g., pharmacologically) inhibited, may therefore be considered a useful background to perform GT, requiring no labor-intensive removal of randomly integrated T-DNAs. Also the search for infrequent GT events amongst successfully transformed plants would be greatly facilitated by having lost all random integrations. Indeed, we previously found GT events in *teb* mutant plants to be devoid of any background integration (van Tol et al., 2022). Unfortunately, however, the frequency of GT events in *teb* mutant plants was found to be 10-fold reduced as compared to GT frequencies measured in wild-type plants (van Tol et al., 2022). An even greater reduction in the GT frequency has been described in the accompanying article (Merker et al., 2024) for plants deficient for both canonical nonhomologous end joining (cNHEJ) as well as for Polθ. These findings may indicate that plant Polθ is involved in HR in plants, which would, however, contrast with homologs in animals (McVey et al., 2016). Another potential explanation for a reduced GT frequency in *teb* mutant plants could be the lack of T-DNA integration: an inability to integrate T-DNA in Polθ deficient plants may compromise the efficient induction of GT-stimulating DSBs in the target (by T-DNA encoded CRISPR enzymes), or may lead to a reduced availability of repair donors (part of the T-DNA), or both. To provide insight into these outstanding questions, we have here analyzed GT in plants that contain CRISPR cassettes and donor templates pre-integrated in the host genome.

RESULTS

Polθ decreases GT frequency

In order to determine whether Polθ is required for HR we carried out an *in planta* GT experiment with the same set of materials we used previously in direct GT experiments (van Tol et al., 2022) with the endogenous *PPO* gene of *Arabidopsis thaliana* as the target (de Pater et al., 2009, 2013, 2018; Hanin et al., 2001). Disruptions of this gene are lethal to photoautotrophically grown plants (Duke et al., 1991). The GT assay uses a T-DNA molecule (pDE-Cas9-PPO-RT) that contains both the donor template for homology-directed repair (HDR) and expression cassettes encoding CRISPR reagents directed at the *PPO* gene (Strunks, 2019). Upon successful GT, the donor template introduces a TAC > ATG change at codon 426 (conferring a Y to M amino acid substitution) rendering the PPO enzyme resistant to the herbicide butafenacil ($Buta^R$) (Hanin et al., 2001). Additionally, the repair template contains five other mutations that allow us to assess the length of the conversion tract, and it contains a phosphinothricin resistance (PPT^R) gene for the selection of transformants (Figure 1). By means of floral dip, we first generated plants in which the T-DNA was randomly integrated. To acquire plants that are competent to perform *in planta* GT, 50 PPT^R primary transformants were selfed and their offspring tested for $Buta^R$. Exactly 7 out of 50 plants showed evidence of *in planta* GT (already within a few hundred progenies), 1 of which (#22) was selected as a "GT inducer line" (Figure 2a). We then crossed this line with a Polθ mutant ($TEB^{-/-}$) plant to obtain Polθ proficient and deficient F2 plants capable of *in planta* GT. We collected 13 $TEB^{+/+}$ and 15 $TEB^{-/-}$ F2 individuals containing the GT reagents within their genomes. Siblings of these plants were tested for $Buta^R$ to exclude that GT events had already taken place in the germline of F1 plants. To derive GT frequencies the selected 28 F2 individuals were selfed, and their offspring (F3) tested for $Buta^R$ (Figure 2b). From the estimated 170 000 offspring of $TEB^{+/+}$ parents, 31 $Buta^R$ seedlings were obtained, while from 247 000 offspring of $TEB^{-/-}$ parents a total of 114 $Buta^R$ seedlings were recovered (Table 1). To inspect whether $Buta^R$ indeed resulted from *in planta* GT at the endogenous *PPO* locus, we performed PCR and sequence analysis (Figure 3). As controls, we sequenced several F2s from all F1 lines as well as untransformed wild type plants (Figure S1). A custom R script was used to analyze the sequencing data (metadata input and raw output of the program can be found in Data S1 and S2). Out of 145 analyzed $Buta^R$ individual plants 139 showed the incorporation of the expected nucleotide substitutions at codon 426 (TAC>ATG) of the *PPO* locus ("Site 4" in the figures). Most plants also showed the co-incorporation of some or all of the other substitutions that were present in the donor DNA (Figure 1). None of the

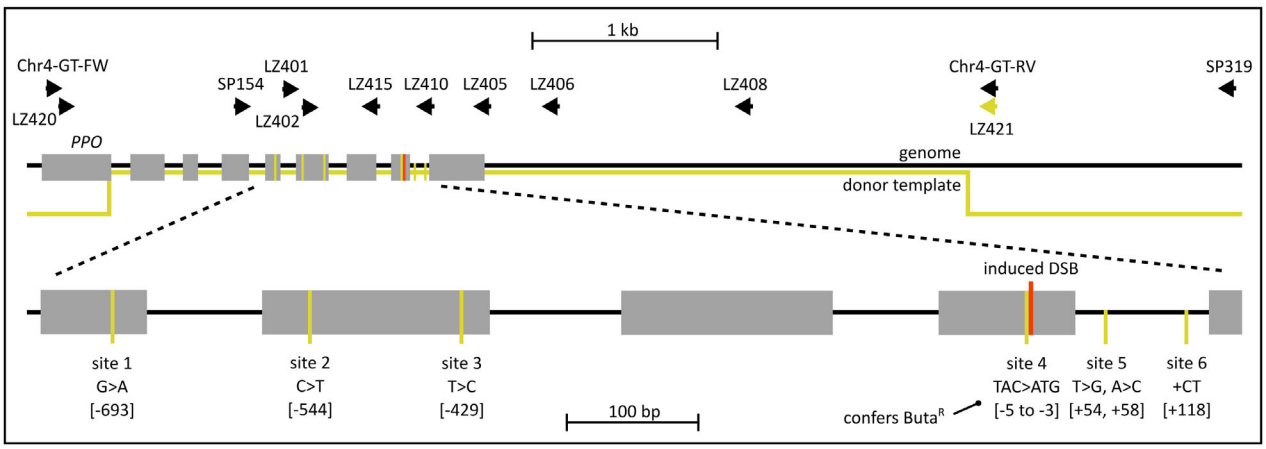

Figure 1. Map of the *PPO* locus. Gray bars indicate the exons of the *PPO* locus (At4g01690) on the genome (black horizontal line). The donor template is indicated in yellow, and the region where the donor template matches the genome is where the lines are close together. The bottom line and bars represent a detailed view around the induced DSB (red), with all 6 sites for which the donor template is different from the reference genome indicated as vertical yellow lines. The arrows indicate the primers used to amplify the endogenous *PPO* locus (Chr4-GT-FW + Chr4-GT-RV (or SP319)), to amplify ectopic gene targeting events (LZ420 + LZ421), and for sanger sequencing of the amplified segments (the other arrows).

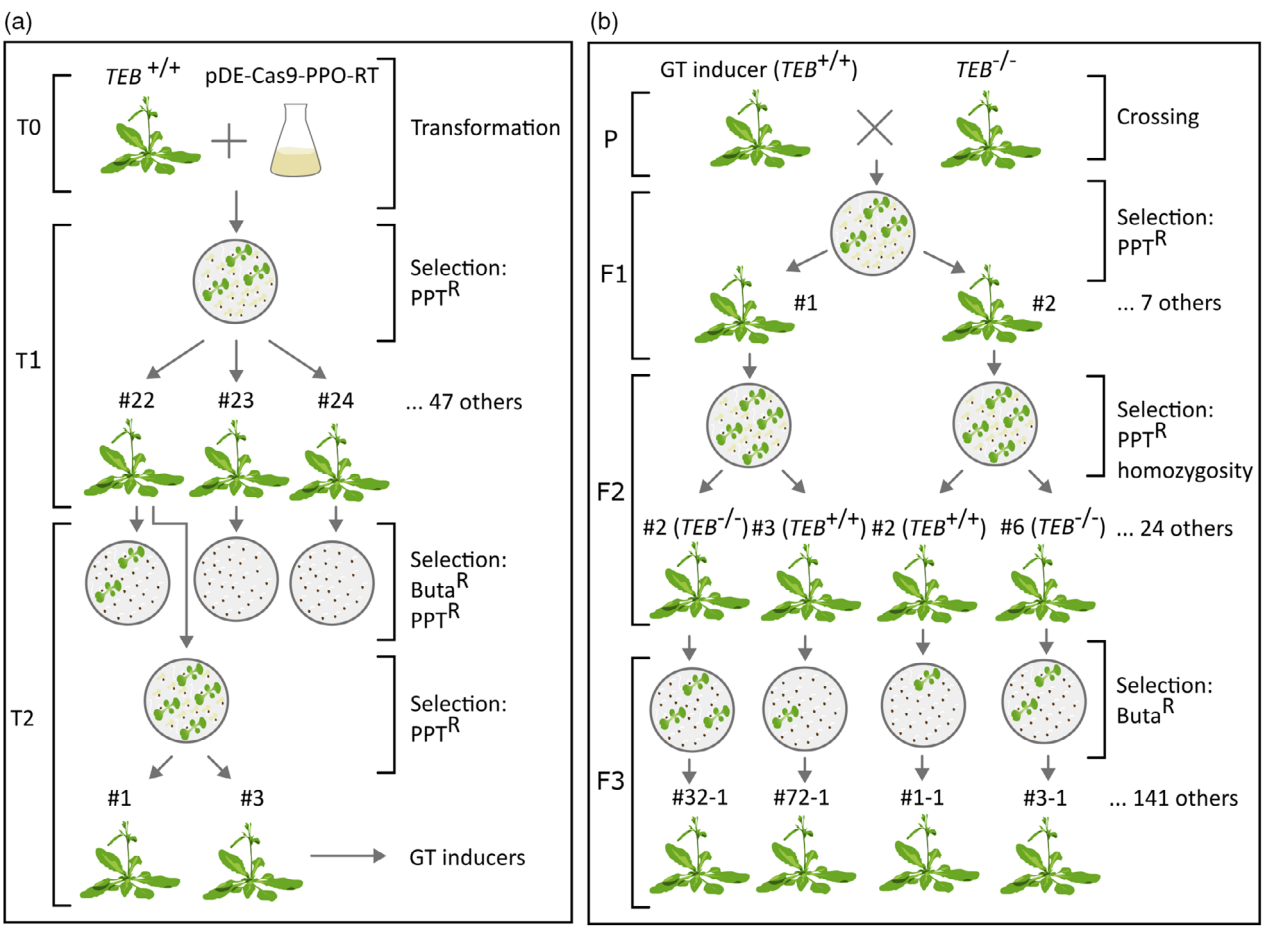

Figure 2. Schematic overview of the *in planta* GT experiment. The generation of GT inducer plant lines by floral dip transformation, PPT selection of the resulting seeds, selection of a suitable line by analyzing the F2 for presence of GT, pattern of inheritance of ppt^R, and normal appearance (a). These GT inducers were crossed with *teb* mutant plants, the resulting F1 selected for PPT^R, and the F2 selected for PPT^R and genotyped to select homozygous plants (*TEB*$^{+/+}$ and *TEB*$^{-/-}$). The F3 individuals of the selected F2s were selected for Buta^R and analyzed by sanger sequencing (b).

Table 1 GT does not require Polθ

TEB$^{+/+}$							*TEB*$^{-/-}$						
T2	F1	F2	F3				T2	F1	F2	F3			
ID	ID	ID	Tested	Buta^{R}	ATG at codon 426	TGT %	ID	ID	ID	Tested	Buta^{R}	ATG at codon 426	TGT %
1	1	1	6000	0	0	0.00	1	1	2	21 000	7	7	0.03
		3	13 000	1	1	0.01		2	6	13 000	4	4	0.03
		12	12 000	0	0	0.00			12	13 000	7	7	0.05
	2	2	16 000	3	3	0.02		3	5	20 000	7	6	0.03
	4	1	12 000	1	1	0.01			11	17 000	5	5	0.03
		2	15 000	1	1	0.01			13	9000	1	1	0.01
		5	20 000	1	1	0.01		4	7	22 000	6	6	0.03
		6	13 000	2	2	0.02	3	1	2	14 000	2	0	0.00
		12	17 000	1	1	0.01			5	6000	2	2	0.03
3	1	3	11 000	2	2	0.02		2	1	23 000	14	13	0.06
		4	11 000	13	13	0.12			9	21 000	29	29	0.14
	4	8	12 000	4	4	0.03		3	3	15 000	10	10	0.07
	5	4	12 000	2	2	0.02			4	19 000	2	2	0.01
								4	5	19 000	8	7	0.04
								5	6	15 000	10	9	0.06
Total			170 000	31	31	0.02	Total			247 000	114	108	0.04

Total number of F3 individuals tested for Buta^{R}, number of Buta^{R} positive F3 individuals, the number of Buta^{R} F3 individuals that also showed GT of codon 426, and the percentage of TGT positive individuals (based on codon 426 sequencing) to the total number tested. The left part of the table shows data for the offspring of Polθ proficient F2 plants, whereas the right displays the data for the offspring of Polθ deficient F2 plants. IDs are given for the F2s and their T2 and F1 progenitors.

control plants had any of these substitutions. This analysis reveals that in almost all Buta^{R} seedlings, whether Polθ proficient or deficient, a "true GT" (TGT) event had taken place. The calculated median TGT frequency across T2 individuals is 8.3×10^{-5} for *TEB*$^{+/+}$ and 3.3×10^{-4} for *TEB*$^{-/-}$, revealing that a lack of Polθ functionality enhances the frequency of *in planta* GT ~ 3-fold (Mann–Whitney *U*-test, $P = 0.0073$) (Figure 4). We conclude that Polθ is not required for HR in *Arabidopsis thaliana*, in sharp contrast, it inhibits this process.

Two pathways for ectopic GT

We next wished to understand what mechanisms other than TGT produced the residual Buta^{R} plants. For 2 plants (out of 145) we were unable to amplify the *PPO* locus to determine the nucleotide composition at codon 426 ("undetermined" in Figure 3). These plants (#12-1 and #12-2) produced hardly any viable offspring, suggesting a possible truncation of the *PPO* locus. For 4 other plants we verified the absence of a gene conversion allele. Remarkably, one of these plants (#66-4) was heterozygous for another substitution at codon 426 (TAC > GTA), conferring a Y to V amino acid substitution that likely provides Buta^{R} (Li et al., 2003); codon 425 was also altered in this plant (AAC > AAT), but encodes the same amino acid. We have tested the offspring of this plant and confirmed the presence of Buta^{R} segregating with the mutation (Figure S2). The three remaining plants, which were homozygously wild type at codon 426 (plants #23-1, #47-1, and #71-2) also had offspring with Buta^{R} segregating at a 3:1 ratio. These must therefore either contain an undiscovered Buta^{R}-conferring mutation elsewhere in the genome, or contain a product of ectopic GT (EGT), whereby the incomplete *PPO* donor sequence in the T-DNA obtained the missing 5′ end from the endogenous *PPO* locus by recombination.

To test the latter hypothesis, we designed primers (LZ420 and LZ421, Figure 1) that produce a 5 kb PCR product only if a complete *PPO* gene connected to T-DNA has been formed by recombination. Such a product can conceivably be formed by a one-sided recombination reaction in which a 3′ break end from the donor invades the target, or vice versa, and extends beyond the homologous segment. For 2 of the 3 plants (plants #47-1 and #71-2), such a 5 kb PCR product was produced and sequencing established that the product contained the T-DNA derived *PPO* sequences (at all 6 positions where the T-DNA is different from the endogenous locus, the T-DNA variant was found). From this analysis we conclude that 2 plants manifested the Buta^{R} phenotype due to EGT, while the remaining plant (#23-1) possibly contains a rare and currently unknown Buta^{R}-conferring genome alteration. We thus calculate the EGT frequency for *in planta* GT in our experimental setup to be 1.4% of Buta^{R} plants, which is 10–50 × less than what is measured for EGT in experiments using similar reagents assaying GT directly upon transformation (de Pater et al., 2013; Hanin et al., 2001; van Tol et al., 2022).

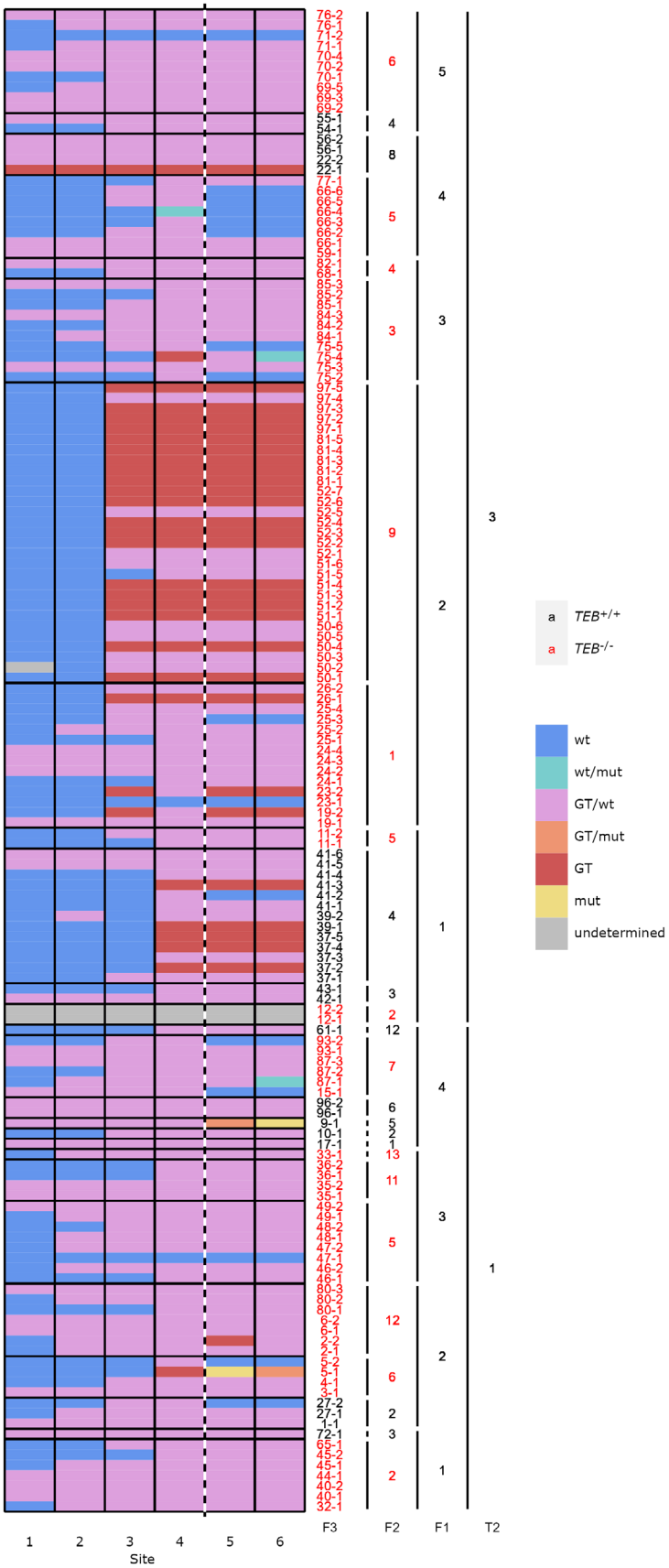

Figure 3. Sanger sequencing across the *PPO* locus shows true gene conversion of most Buta^{R} F3 plants. Each Buta^{R} F3 individual (rows) has been sanger sequenced with forward and reverse primers to investigate the extent of gene conversion with respect to the sites (*x*-axis) for which the donor template differs from the reference genome. The color of each of the 6 rectangles in each row indicates whether the sequence at these sites is identical to the reference ("wt"), identical to the donor template ("GT"), or is mutated ("mut"). If multiple alleles are detected (double-peaks) the status of both alleles is reported. "undetermined" indicates no (clean) sequence could be obtained. On the right are the names of the F3 plants, and their F2, F1, and T2 progenitors. The names of the F2s and F3s are colored to indicate the genetic background, either Polθ proficient (black) or deficient (red).

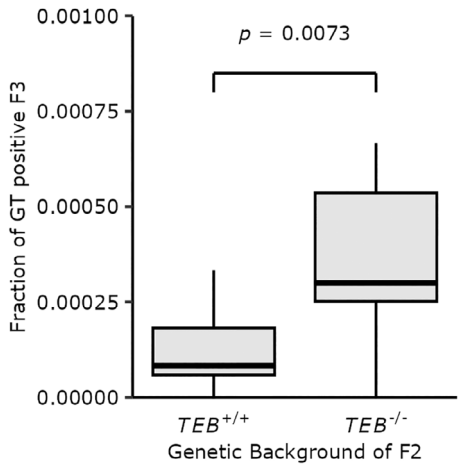

Figure 4. Absence of Polθ enhances GT. Boxplots summarizing the fraction of GT positive F3 individuals (of the total pool of offspring) per F2 parent, based on Buta^R and sanger sequencing of codon 426. A Mann–Whitney *U*-test was performed to test for a significant distribution shift.

We therefore suggest that the major recombination pathway producing EGT involves nonintegrated T-DNA, whereby T-DNA recombines with the genomic target locus, picking up neighboring sequences, prior to integrating at another genomic location. Our finding of *in planta* EGT argues also for a minor source, in which recombination occurs between genomically dispersed homologous sequences, potentially being a source of reciprocal translocations.

DISCUSSION

Here, by performing *in planta* GT in plants that are either proficient or deficient for Polθ, we conclusively establish that this enzyme is not required for HR in *Arabidopsis thaliana*. In contrast, Polθ acts as a suppressor of GT; the frequency being 3-fold higher in a Polθ mutant background. In addition, by molecular analysis of the GT outcomes, we discovered that EGT is rare in the situation where only integrated donor molecules are present (1.4%); however, EGT may comprise up to two-thirds of events in transformation-linked ("direct") GT protocols (de Pater et al., 2013, Hanin et al., 2001, van Tol et al., 2022). The notion that the frequency of EGT is greatly reduced in our *in planta* GT experiments as compared to protocols where GT is sought for via transformation of the required reagents, led us to suggest two distinct routes to EGT events: (i) a pathway in which nonintegrated T-DNA molecules engage with the (cleaved) target locus, such that DNA is being copied, followed by (completion of) integration of the modified T-DNA elsewhere in the genome. We cannot formally exclude the possibility that one end of the T-DNA is already attached to the genome prior to engaging with the target locus. (ii) a less efficient pathway, in which donor and target located at different genomic positions undergo one-sided recombination that results in either T-DNA sequence acquisition of the target locus, or target sequence acquisition of the T-DNA locus.

Three recent studies have reported a lower GT efficiency in plants deficient for Polθ in transformation experiments (Merker et al., 2024; Sizova et al., 2021; van Tol et al., 2022). These results are in contrast with findings obtained in other biological systems, which demonstrated that Polθ antagonizes HR and inhibits GT (Ceccaldi et al., 2015; Mara et al., 2019; Zelensky et al., 2017). Our current study, combined with our earlier work, resolves this apparent contradiction by demonstrating that Polθ mutant plants are not impaired in homology-driven gene conversion, and reveal that instead the defect in the integration of the GT reagents, the donor template and CRISPR cassette, likely underlies the lower GT efficiency in transformation experiments. Without T-DNA integration the transient availability of CRISPR and donor reagents restricts the window in which GT can occur. The apparent inhibitory effect of Polθ on (*in planta*) GT we have here revealed is likely caused by end-joining of the DSB introduced in the target locus, thus blocking an HR route. A similar situation was reported for cNHEJ deficiency (Endo et al., 2016; Qi et al., 2013). Simultaneous mutation of TMEJ and cNHEJ is therefore expected to further enhance the efficiency of GT, as was recently described in mammalian cells (Saito et al., 2017; Zelensky et al., 2017).

It has been well documented that HR between endogenous target loci and exogenously supplied donor molecules can lead to TGT events, but also to EGT events (Endo et al., 2006; Hanin et al., 2001; Iida & Terada, 2005; Lieberman-Lazarovich et al., 2013; Offringa et al., 1990). In the case of TGT, the endogenous locus acquires sequences from the donor, while this is the reverse for EGT. The frequency with which EGT events occur may depend on numerous factors (such as the configuration/design of the donor molecule and the selection that is applied), but can comprise up to two-thirds of the total number of GT events (de Pater et al., 2013; Hanin et al., 2001; van Tol et al., 2022). One scenario for EGT is T-DNA reacting with the target locus, picking up sequences prior to integration elsewhere in the genome (Huang & Puchta, 2019). Our data suggests the existence of a second scenario, in which T-DNA first integrates and then participates in HR with the target locus. Because we find only two EGT events amongst more than a hundred Buta^R plants in our current *in planta* GT experiment, we believe that most EGT events in direct GT experiments are formed prior to or during integration. It is also more logical: upon entering a cell, a T-DNA (either full length or abrogated) has a recombinogenic 3′ end, which can be captured by a spontaneously occurring DSB, leading to T-DNA integration, but which can also be viewed as an origin for a HR reaction. However, once integrated, the T-

DNA is devoid of such recombinogenic end and EGT may then rely on a spontaneous DSB within the genomically present T-DNA sequence (e.g., during DNA replication). This explanation is supported by the observation that EGT can reach up to 50% of all GT events for *in planta* GT experiments where the integrated donor locus is liberated by CRISPR/Cas, hence again creating recombinogenic ends at the donor site (Wolter & Puchta, 2019).

Decades of work in a variety of model systems have revealed that GT efficiency is affected by multiple factors, and that a detailed understanding of the underlying principles will benefit the design of more efficient methods. Previous elegant work in plants has shown that high levels of nuclease expression stimulates GT, which leads to a highly efficient two-step GT procedure, in which plant lines with high Cas9 nuclease expression are first selected, which are subsequently transformed with donor T-DNA (Miki et al., 2018). Also a higher abundance of donor copies has been seen to improve GT efficiency (Baltes et al., 2014).

We here present data supporting a scenario where integration of the donor T-DNA into the genome is an important first step in GT approaches employing transformation. Future efforts should therefore be directed towards identifying why nonintegrated T-DNA (or configurations of T-DNAs currently used in transformation protocols) makes such a poor TGT donor. If solved, inhibition of end joining could then serve to prevent unwanted, yet highly efficient T-DNA integrations which pollute the genome with undesired GT reagents.

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. ~10 days (8 days for F1s, 12 days for F2s, and 10 days for the rest) after sowing imbibed, stratified seeds on selective *in vitro* culture medium (half-strength MS macronutrients, B5 micronutrients, 3.67 g/l FeNaEDTA, 0.5% sucrose, 0.1% MES, 0.7% agar, pH 5.8, 100 μg/ml nystatin, 100 μg/ml Timentin, and either 15 μg/ml phosphinothricin or 50 nM butafenacil), surviving seedlings were counted and transferred to soil and further cultivated in a climate-controlled growth chamber at a temperature of 20°C, 70% relative humidity, ~150 μmol m^{-2} sec^{-1} of photosynthetically active radiation, and at a 16 h photoperiod. Plants were all of the Col-0 ecotype; Polθ mutants contained the *teb-5* insertional mutation (Inagaki et al., 2006; van Kregten et al., 2016).

Transformation

Transformation was performed by floral dip as described in van Tol et al. (2022) using Agrobacterium strain AGL1 (Lazo et al., 1991) containing binary vector pDE-Cas9-PPO-RT (Strunks, 2019). Primary transformants were selected on phosphinothricin.

Analysis of GT events

Leaf samples from the F3 plants and control plants were processed using the dilution protocol of the Phire Plant Direct PCR kit (Thermo Fisher Scientific, Waltham, MA, USA), upon which that kit was used for PCR following the manufacturer's instructions. For the analysis of TGT the endogenous *PPO* locus was amplified with the primers Chr4-GT-FW and Chr4-GT-RV for most plants. #5-1 and #23-1 did not yield a product with that primer combination, but a product was obtained with Chr4-GT-FW and SP319 (the latter is further downstream than Chr4-GT-RV). The ectopic gene targeting product was amplified with primers LZ420 and LZ421. Primers were removed from all PCR products using the Genejet PCR Purification kit (Thermo Fisher Scientific, Waltham, MA USA) upon which they were Sanger sequenced using various primers, listed in Table S1, at Macrogen Europe (Amsterdam, The Netherlands). Sanger sequencing results were analyzed using a custom made R program, which utilizes the sangerseqR package for signal decomposition (Hill et al., 2014). The program was run on strict mode (calling is restricted to noise-free data) and with GT_peak_ratio = 0.25 (secondary peaks must have at least 0.25× the intensity of the primary peaks, otherwise they are ignored). The metadata input can be found in Data S1 and S2 contains the raw output.

Buta^R inheritance analysis

Seeds were sown on a grid with 100 spots evenly spaced over a petri dish (so up to ~100 seeds per plate). After ~10 days, the plates were photographed, and each seed/seedling was put in one of 3 categories: (i) nongerminated, (ii) seeds with white/yellow tissue emerging, and (iii) green seedlings. The final category mostly contained healthy green plants, but also some stunted plants.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We wish to thank the Genome Editing team of Trait Research, BASF Innovation Center Ghent (Belgium) for fruitful discussions. 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.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

SUPPORTING INFORMATION

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

Figure S1. No pre-existing gene targeting in the F2 generation.

Figure S2. Buta^R inheritance by the F4 generation.

Table S1. Oligonucleotides used in this study.

Data S1. Metadata for GT analysis program.

Data S2. Output of GT analysis program.

OPEN RESEARCH BADGES

This article has earned Open Data and Open Materials badges. Data and materials are available at https://doi.org/10.17026/LS/AEAQST and https://github.com/LejonKralemann/SangerGTAnalysis.

DATA AVAILABILITY STATEMENT

Raw Sanger sequencing data (AB1 format) has been deposited on the DANS Data Station Life Sciences (https://doi.org/10.17026/LS/AEAQST). The custom R program for the analysis of the sanger sequencing data is available on GitHub (https://github.com/LejonKralemann/SangerGTAnalysis).

REFERENCES

Baltes, N.J., **Gil-Humanes, J.**, **Cermak, T.**, **Atkins, P.A.** & **Voytas, D.F.** (2014) DNA replicons for plant genome engineering. *The 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.

Chan, S.H., **Yu, A.M.** & **McVey, M.** (2010) Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in drosophila. *PLoS Genetics*, **6**, e1001005.

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.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.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.

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.

Endo, M., **Mikami, M.** & **Toki, S.** (2016) Biallelic gene targeting in rice. *Plant Physiology*, **170**, 667–677.

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.

Fauser, F., **Roth, N.**, **Pacher, M.**, **Ilg, G.**, **Sánchez-Fernández, R.**, **Biesgen, C.** ***et al.*** (2012) In planta gene targeting. *Proceedings of the National Academy of Sciences*, **109**, 7535–7540.

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

Hill, J.T., **Demarest, B.L.**, **Bisgrove, B.W.**, **Su, Y.C.**, **Smith, M.** & **Yost, H.J.** (2014) Poly peak parser: method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Developmental Dynamics*, **243**, 1632–1636.

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.

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

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. *Plant Cell*, **18**, 879–892.

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.

Kralemann, L.E.M., **de Pater, S.**, **Shen, H.**, **Kloet, S.L.**, **van Schendel, R.**, **Hooykaas, P.J.J.** ***et al.*** (2022) Distinct mechanisms for genomic attachment of the 5′ and 3′ ends of agrobacterium T-DNA in plants. *Nature Plants*, **8**, 526–534.

Lazo, G.R., **Stein, P.A.** & **Ludwig, R.A.** (1991) A DNA transformation–competent Arabidopsis genomic library in agrobacterium. *Bio/Technology*, **9**, 963–967.

Li, X., **Volrath, S.L.**, **Nicholl, D.B.**, **Chilcott, C.E.**, **Johnson, M.A.**, **Ward, E.R.** ***et al.*** (2003) Development of protoporphyrinogen oxidase as an efficient selection marker for agrobacterium tumefaciens-mediated transformation of maize. *Plant Physiology*, **133**, 736–747.

Lieberman-Lazarovich, M., **Melamed-Bessudo, C.**, **de Pater, S.** & **Levy, A.A.** (2013) Epigenetic alterations at genomic loci modified by gene targeting in Arabidopsis thaliana. *PLoS One*, **8**, e85383.

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. *The New Phytologist*, **222**, 1380–1391.

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.

Merker, L., **Feller, L.**, **Dorn, A.** & **Puchta, H.** (2024) Deficiency of both classical and alternative end joining pathways leads to a synergistic defect in DSB repair but not to an increase in homology-dependent gene targeting in Arabidopsis. *The Plant Journal*.

Miki, D., **Zhang, W.**, **Zeng, W.**, **Feng, Z.** & **Zhu, J.-K.** (2018) CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. *Nature Communications*, **9**, 1967.

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

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

Qi, Y., **Zhang, Y.**, **Zhang, F.**, **Baller, J.A.**, **Cleland, S.C.**, **Ryu, Y.** ***et al.*** (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. *Genome Research*, **23**, 547–554.

Saito, S., **Maeda, R.** & **Adachi, N.** (2017) Dual loss of human POLQ and LIG4 abolishes random integration. *Nature Communications*, **8**, 16112.

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

Strunks, G. (2019) Strategies for the improvement of genome editing in *Arabidopsis thaliana*: PhD Thesis. Leiden University Scholarly Publications, Leiden, the Netherlands.

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 Tol, N., **van Schendel, R.**, **Bos, A.**, **van Kregten, M.**, **de Pater, S.**, **Hooykaas, P.J.J.** ***et al.*** (2022) Gene targeting in polymerase theta-deficient *Arabidopsis thaliana*. *The Plant Journal*, **109**, 112–125.

Wolter, F. & **Puchta, H.** (2019) In planta gene targeting can be enhanced by the use of CRISPR/Cas12a. *The Plant Journal*, **100**, 1083–1094.

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.