Strategies for the improvement of genome editing in Arabidopsis thaliana

Strunks, G.D.

Chapter 5

Developing a strategy for the identification of gene targeting mutants in *Arabidopsis thaliana*

Gary D. Strunks[1,2], Bart J. P. M. Klemann[1,3], Paul J. J. Hooykaas[1] and Sylvia de Pater[1]

[1]Molecular and Developmental Genetics, Institute of Biology, Leiden University, 2333 BE, the Netherlands
[2]Current affiliation: Naktuinbouw, 2371 GD Roelofarendsveen, the Netherlands
[3]Current affiliation: Dümmen Orange, 2678 PS De Lier, the Netherlands

Abstract

Gene targeting (GT) is a molecular technique that exploits homologous recombination to change an endogenous gene. Gene targeting in plants is of interest for both fundamental research and plant biotechnology. In plants, gene targeting is a very rare event due to the low efficiency of homologous recombination in somatic cells and low transformation efficiency of the repair template DNA. Although the advent of sequence-specific nucleases such as those based on the CRISPR/Cas system have enhanced both the efficiencies of targeted mutagenesis and gene targeting, there is still room for improvement. In this study, we used zinc-finger artificial transcription factor (ZF-ATF)-mediated genome interrogation in an attempt to identify plants with enhanced gene targeting frequencies. We used a collection of ZF-ATF Arabidopsis lines that expressed ATFs consisting of three finger ZFs (3F-ATFs) fused to the EAR transcriptional repressor. Each of these ZF-ATF in theory could cause changes in the expression of numerous genes. Plants harboring these ATF expression constructs were screened for GT at the *PPO* gene. A total of eleven true GT events and one ectopic GT event were found and 3F-EAR sequences in these plants were determined. To find out whether the identified ZF-ATFs enhanced GT, wild type plants were retransformed with the identified 3F-EAR expression constructs and screened again for GT at *PPO*. One retransformed plant line harboring 3F-EAR4, showed an enhanced GT frequency. To get more insight into the effect of 3F-EAR4, GT screening in 3F-EAR4 retransformants was performed at the *PPO*, *CRU3* and *ADH1* genes. However, in these experiments, GT frequencies did not differ substantially from control plants. The initially observed seemingly enhanced GT frequency at the *PPO* locus was therefore most likely due to experimental variation and not due to an effect of the ZF-ATF construct.

Introduction

A valuable tool for genome editing in plants is gene targeting (GT), a technique that exploits the endogenous homologous recombination (HR) machinery to change a gene of interest. GT events are very rare in plants, mainly due to the low efficiency of HR in somatic cells (1). Several approaches to increase GT efficiency in plants, including manipulation of the DSB repair machinery, positive-negative selection and the targeted induction of DSBs with sequence-specific nucleases, have been successful. However, GT efficiency needs to be increased further in order to make GT a feasible tool for genome editing in crops (2).

The current understanding is that GT utilizes the synthesis-dependent strand annealing (SDSA) pathway of HR in which both the 5' and 3' homology arms present on the GT repair template integrate via HR independently (3–5). Although many factors that are involved in HR have been discovered, true regulators of GT have thus far remained elusive. Discovering and manipulating such regulators might increase the efficiency and feasibility of GT.

In this study, we investigated whether artificially induced differential gene expression in Arabidopsis can lead to phenotypes with intrinsically higher levels of GT. The method of genome interrogation makes it possible to uncover novel phenotypes by expressing the genome differentially using zinc-finger artificial transcription factors (ZF-ATFs) (6). A typical ZF-ATF with a DNA binding domain of three ZFs can bind to hundreds or even thousands of loci in eukaryotic genomes. When these ATFs bind in the vicinity of gene control regions, they can affect gene expression patterns on a genome-wide scale (6). This can be brought about effectively by attaching an effector domain to the ZF-ATF, for instance the VP16 activator domain (7) or the EAR repressor domain (8, 9).

The first genome interrogation experiments using ZF-ATFs were performed in single-celled organisms and mammalian cell cultures, and novel phenotypes where discovered by applying selection for those phenotypes (10–14). The first proof of principle study in multicellular organisms was performed in plants (15). In this study, conducted in our lab, Arabidopsis plants were transformed with a library of 4200 3F-ATFs in which the 3F DNA binding domains were fused to the VP16 transcription activation domain from the herpes simplex virus. Furthermore, 3F-ATFs were expressed under the *ribosomal protein 5A* (*RPS5a*) promoter, which is mainly active in zygotes, early embryo's and meristematic tissue (16). This ensures that ZF-ATF-mediated genome-wide differential gene expression is confined to dividing, undifferentiated cells which undergo transformation and where gene targeting can take place. An ATF was identified that induced a substantial increase in somatic homologous

recombination (HR) (15). In a follow-up study, it was shown that the ATF was indeed responsible for this phenotype by acting as an ectopic master regulator of a set of endogenous genes, and the resulting increase in somatic HR was much greater than when each of the individual genes was overexpressed (17). In more recent studies conducted in our lab, genome interrogation in plants has successfully uncovered novel phenotypes including high salinity tolerance and enhanced growth characteristics (18, 19). In the latter study, beside the VP16 effector, a library of ~700 3F-ATFs was used where the 3F domain was fused to the ERF-associated Amphiphilic Repression (EAR) motif, a transcriptional repressor (19). Here, we used two 3F-EAR sublibraries to screen for an enhanced GT phenotype.

Materials and Methods

Plant material and growth conditions

Arabidopsis thaliana plants of the Columbia-0 ecotype were used for all transformations and as wild-type control. Seeds were germinated on MA or ½ MS medium in a climate-controlled growth chamber at 20˚C with a light intensity of 75 µmol m^{-2} s^{-1} during 16h/day photoperiod and 50% humidity. Seedlings that were transferred to soil were grown in a climate-controlled growth chamber with a light intensity of 200 µmol m^{-2} s^{-1} during 16h/day photoperiod at 20˚C and 70% humidity.

Screening the 3F-EAR library for GT at PPO

Plants harboring 3F-EAR constructs from ZF-library pools 1 and 5 (19) were transformed with the *PPO* GT repair construct (PSDM3900) as described (20). A total of 101 lines from 3F-EAR library 1 and 43 lines from 3F-EAR library 5 were transformed, with about 16 – 20 plants per line. To select for GT events at *PPO*, T1 seeds were plated on MA solid medium containing 0.5% sucrose supplemented with timentin (100 µg/mL), nystatin (100 µg/mL) and 50 nM butafenacil. To determine the transformation frequency of the *PPO* GT repair construct, a portion of T1 seeds was selected on timentin (100 µg/mL), nystatin (100 µg/mL) and phosphinothricin (PPT) (15 µg/mL). Genomic DNA was isolated from butafenacil-resistant plants using the CTAB method (21) and plants were analysed for GT by PCR. Primers PPO-PA and PPO-4 were used to detect 5’ GT, while primers PPO-PA and PPO-319 were used to detect true GT at both the 5’ and 3’ ends. Subsequently, PCR products were digested with *Kpn*I,

which should result in 1.7 and 0.3 kb fragments in case of a 5' GT event, and 4.3 and 1.7 kb fragment in case of a true GT event.

Transformation with reconstituted 3F-EAR T-DNA constructs

DNA fragments encoding the 3F-EAR were PCR amplified from genomic DNA from plants with GT events, using primers in the *RPS5a* promoter and the *nos* terminator sequence in the 3F-EAR T-DNA (Table S1). PCR-amplified 3F sequences were first cloned into pJET1.2 (Thermo Fisher Scientific), and sequenced using the pJET1.2 forward and reverse primers to determine the 3F combination. The 3F sequences were then cloned into pRF-pRPS5a-EAR-Kana (19) with *Sfi*I to generate reconstituted binary vectors, which were introduced into the *Agrobacterium* strain AGL1 by electroporation (den Dulk-Ras and Hooykaas 1995). Col-0 plants were transformed and T1 retransformants were selected on MA solid medium without sucrose supplemented with timentin (100 µg/mL), nystatin (100 µg/mL) and kanamycin (35 µg/mL). T1 retransformants were transferred to soil after 2 – 3 weeks, and allowed to set seed. T2 seeds were harvested and selected for the 3F T-DNAs on ½ MS medium supplemented with kanamycin (35 µg/mL). T2 plants were used in further GT experiments at the *PPO*, *CRU3* and *ADH1* loci. For convenience, the plant lines retransformed with the 7 identified 3F sequences were designated 3F-EAR1 – 3F-EAR7.

Cloning of Cas9-GT constructs

T-DNA constructs containing a CRISPR/Cas9 expression cassette and a GT repair template were constructed for the *PPO*, *CRU3* and *ADH1* genes. The Cas9-PPO-RT construct was cloned by digestion of the *PPO* repair template from pSDM3900 (20) with *Sma*I and *Sal*I. The resulting fragment with the PPO-RT was cloned into pEntry4, digested with *Eco*RV and *Sal*I. pEntry4-PPO-RT was digested with *Nco*I and blunted with Klenow fragment, followed by *Bam*HI digestion. The resulting PPO-RT fragment was cloned into *Sma*I/*Bam*HI-digested pDE-Cas9-PPO (pSDM3905) to create pDE-Cas9-PPO-RT.

To create the Cas9-CRU2-CRU-GFP-PPT construct, the 35S-PPT-tNOS sequence was PCR amplified from pCambia3300 with primers SP585/SP586, both containing a *Pst*I restriction site, cloned into pJET1.2 (ThermoFisher Scientific) and sequenced. The 35S-PPT-tNOS fragment was digested from pJET with *Pst*I and cloned into *Pst*I-digested vector pCambia2300-CRU-GFP (Chapter 4, pSDM3913). pUbi-Cas9-CRU2 was digested from vector pDE-Cas9-CRU2 (Chapter 2 and 4, pSDM3908) with *Nhe*I and blunted with Klenow fragment

(ThermoFisher Scientific), followed by *Rsr*II digestion. The released pUbi-Cas9-CRU2 fragment was isolated and cloned into pCambia2300-CRU-GFP-PPT, that was digested with *EcoR*I and blunted with Klenow fragement, followed by *Rsr*II digestion. The resulting vector was designated Cas9-CRU2-CRU-GFP-PPT.

To create the Cas9-ADH2-ADH-PPT-RT construct the ADH-PPT repair template was first constructed by PCR amplifying the 5' and 3' *ADH1* homology arms from genomic DNA with primers GS33/GS34 (*ADH1* 5' homology, 823 bp) and GS35/GS36 (*ADH1* 3' homology, 804 bp). Fragments were cloned into pJET1.2 and the 5' *ADH1* homology arm was cloned into linearized pJET-35S-PPT-tNOS with *Xba*I, which was also present in the 5' *ADH1* primer tails. To check for the correct orientation of the 5' ADH1 fragment, pJET-5'ADH-35S-PPT-tNOS was digested with *Xma*I and *EcoR*V. The 3' *ADH1* homology arm was cloned into linearized pJET-5'ADH-35S-PPT-tNOS with *BamH*I. Correct orientation of 3' *ADH1* was confirmed with *EcoR*I digestion. 5'ADH-35S-PPT-tNOS-3'ADH was digested from pJET with *Bcu*I and *Pme*I and cloned into *Bcu*I/*Pme*I-digested vectors Cas9-ADH-1 (pSDM3916) and Cas9-ADH-2 (pSDM3917) (Chapter 2), resulting in Cas9-ADH1-ADH1-PPT-RT (pSDM3918) and Cas9-ADH2-ADH-PPT-RT (pSDM3919).

Screening for GT at PPO in retransformant plant lines harboring reconstituted 3F-EAR constructs

T2 plants retransformed with 3F-EAR1 to 3F-EAR7 were transformed with Agrobacterium strain AGL1 containing the pDE-Cas9-PPO-RT construct using the floral dip method (22). Primary transformant seeds were plated on MA solid medium containing 0.5% sucrose supplemented with timentin (100 µg/mL), nystatin (100 µg/mL), and 50 nM butafenacil to select for GT events at *PPO*. A portion of T1 seeds was selected on timentin (100 µg/mL), nystatin (100 µg/mL) and PPT (15 µg/mL) to determine the transformation frequency of the pDE-Cas9-PPO-RT construct. Genomic DNA was isolated from butafenacil-resistant primary transformants using the CTAB method (21), and plants were analysed for GT at *PPO* using the same method as with the initial GT screening using primers PPO-PA and PPO-4 to detect 5' GT, and primers PPO-PA and PPO-319 to detect true GT at both the 5' and 3' ends.

Screening for GT at CRU3 and ADH1 in retransformant plant lines

For GT analysis at the *CRU3* and *ADH1*, T2 plants from plant lines 3F-EAR4-2, 4-4 and 4-6 were transformed with Cas9-CRU2-CRU-GFP-PPT and Cas9-ADH1/ADH2-ADH-PPT constructs by floral dipping. T1 seeds were plated on MA solid medium without sucrose, supplemented with timentin (100 µg/mL), nystatin (100 µg/mL) and 15 µg/mL PPT. Genomic DNA was isolated from pools containing on average 10 PPT-resistant primary transformant seedlings using the fast extraction method (23), and plants were analysed for GT with PCR. When GT events were detected in pools, individual seedlings from that pool were analysed for GT again.

For GT detection at *CRU3*, the complete *CRU3* integration site was first PCR amplified with primers SP261/SP608 for 15 cycles, resulting in a PCR product of 7678 bp after a GT event or 4810 bp for the unmodified CRU3 locus. Using this PCR product as template, two PCRs were performed to detect 5' GT and 3' GT, using primer combinations SP261/SP262 and primers SP604/SP608, respectively. For GT detection at *ADH1*, two PCRs were performed for 5' GT and 3' GT using primer combinations GS41/GS44 (1132 bp) and GS45/GS46 (1130 bp), respectively.

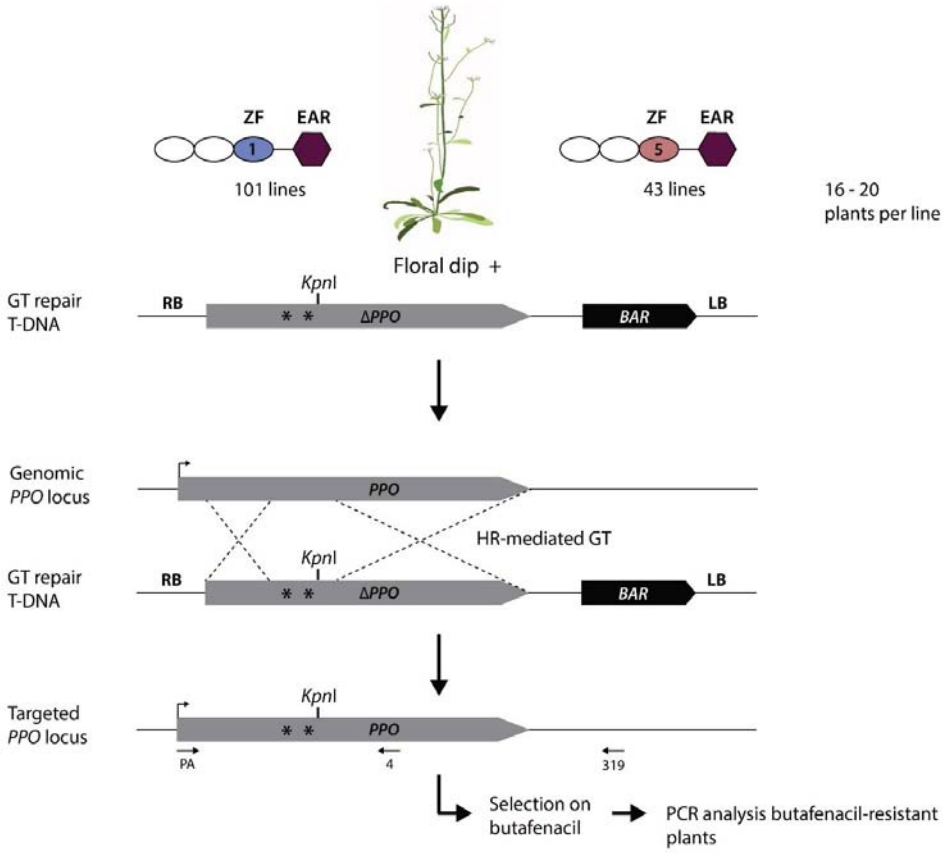


Figure 1. Overview of screening strategy for GT at *PPO* in ZF-EAR plant lines. Plant lines harboring ZF-EAR constructs from ZF-EAR pool 1 (blue oval) and pool 5 (red oval) were transformed with a GT repair template T-DNA for GT at *PPO* using floral dip. A total of 101 plant lines from pool 1 and 43 plant lines of pool 5 were transformed. About 16 – 20 plants per line were used for transformation. Primary transformant seeds were selected on medium containing butafenacil to select for GT events at *PPO*. Butafenacil-resistant plants were PCR-analysed with primer pairs PPO-PA and PPO-4, and PPO-PA and PPO-319 for GT, followed by *Kpn*I digestion of the PCR product, to confirm GT events at the molecular level. Transformation efficiency of the GT repair T-DNA was assessed by selection of primary transformant seeds on PPT using the *BAR* selection marker.

Results

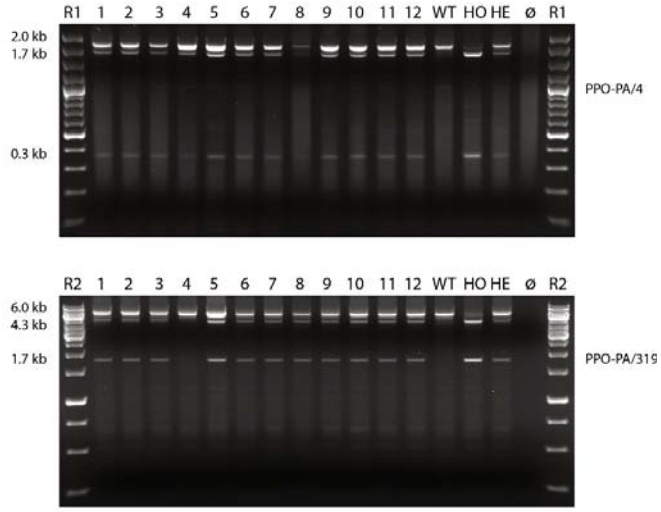
Plants of a 3F-EAR library used for gene targeting at the PPO gene

We undertook a genome interrogation approach to find plants with enhanced GT. In order to find such plants, a collection of Arabidopsis plant lines expressing different 3F-EAR effectors were screened for gene targeting at the *protoporphyrinogen oxidase* (*PPO*) gene, which is involved in chlorophyll and heme synthesis. The herbicide butafenacil can be used to inhibit the enzyme encoded by *PPO*, which causes plant death due to the formation of reactive oxygen species. The PPO enzyme can be made insensitive to butafenacil by two specific amino acid changes, for which the mutations can be introduced via homology direct repair (HDR) using an artificial GT repair template (24). The collection of plant lines used was generated previously, and is composed of 7 subpools, in which each plant contains an 3F-EAR construct consisting of an array of three 5'-GNN-3' binding ZFs that were translationally fused to the EAR transcription repression domain (19).

To identify plants with an enhanced GT phenotype, 101 plant lines from 3F-EAR subpool 1 and 43 plant lines from 3F-EAR subpool 5 were transformed by floral dip (16 – 20 plants per line) with the GT repair construct for the *PPO* gene, which consists of a 5' truncated *PPO* gene containing the two mutations that confer butafenacil resistance (S305L and Y426M) and a *Kpn*I site at position E445/A446 to assist in molecular analysis of GT events (pSDM3900) (20, 25) (Figure 1).

Seeds collected after floral dip transformation were selected on butafenacil. Plants that survived butafenacil selection were transferred to soil and analyzed by PCR for GT events. Due to the presence of the *Kpn*I site in the *PPO* repair template, plants rendered butafenacil-resistant through a GT event should have incorporated this new *Kpn*I-site at the *PPO* locus. In order to obtain a first indication that the butafenacil-resistance was due to GT, the 5' end of the *PPO* locus was recovered with PCR using primers PPO-PA and PPO-4 resulting in a 2 kb fragment that can be cleaved into a 1.7 kb and a 0.3 kb fragment after digestion with *Kpn*I after GT. Similarly, the entire *PPO* locus was amplified by PCR using primers PPO-PA and PPO-319 rendering a 6 kb fragment that can be cleaved in 4.3 kb and 1.7 kb fragments after GT (20, 25). From the butafenacil-resistant seedlings screened, heterozygous GT events in plants harboring 3F-EAR effectors from both pool 1 and pool 5 were detected, meaning that recombination took place on one of the two *PPO* alleles, as indicated by both the presence of *Kpn*I-sensitive and *Kpn*I-resistant PCR product. Eleven true GT events were detected, indicated by *Kpn*I-sensitivty

A



B

Sample	PA / 4	KpnI	PA / 319	KpnI	ZF-sequence	#
1. Pool-1 30 1	+	+	+	+	GCG-GGC-GGG	1
2. Pool-1 32 1	+	+	+	+	GAC-GCC-GGG	2
3. Pool-1 36 1	+	+	+	+	GCG-GGC-GGG	1
4. Pool-1 44 1	+	+	+	-	GCG-GGC-GGG	1
5. Pool-1 67 3	+	+	+	+	GAT-GGT-GGG	3
6. Pool-1 90 5	+	+	+	+	GGG-GTG-GGG	4
7. Pool-1 95 4	+	+	+	+	GAA-GCC-GGG	5
8. Pool-1 102 1	+	+	+	+	-	
9. Pool-1 136 1	+	+	+	+	-	
10. Pool-5 3 1	+	+	+	+	-	
11. Pool-5 9 2	+	+	+	+	GTT-GCG-GCC	6
12. Pool-5 27 1	+	+	+	+	GTA-GAT-GCC	7

Figure 2. GT events detected with initial screening of ZF-EAR lines. **A**. PCR was performed on 12 butafenacil-resistant plant lines transformed with ZF-EAR constructs and as controls on wild-type (WT), homozygous GT (HO) and heterozygous (HO) plant lines. Primers PPO-PA and PPO-4 were used for 5' GT (A) and primers PPO-PA and SP319 for true GT (B), followed by *Kpn*I digestion and analysed on 1.5% (top gel) or 0.7% (bottom gel) agarose gels. Ø is PCR control without template. R1, 100 bp+ ruler (Thermo Fisher Scientific). R2, 1 kb ruler (Thermo Fisher Scientific). **B**. Overview of GT events and 3F-sequence DNA binding sequence determined for each plant. Each 3F sequence was given a number and sequences that were found multiple times are shown in red.

of both the PPO-PA/PPO-4 and PPO-PA/PPO-319 PCR products (Figure 2A). One plant (#4) only showed *Kpn*I-sensitivity of the PPO-PA/PPO-4 PCR product, and can be considered an ectopic GT event.

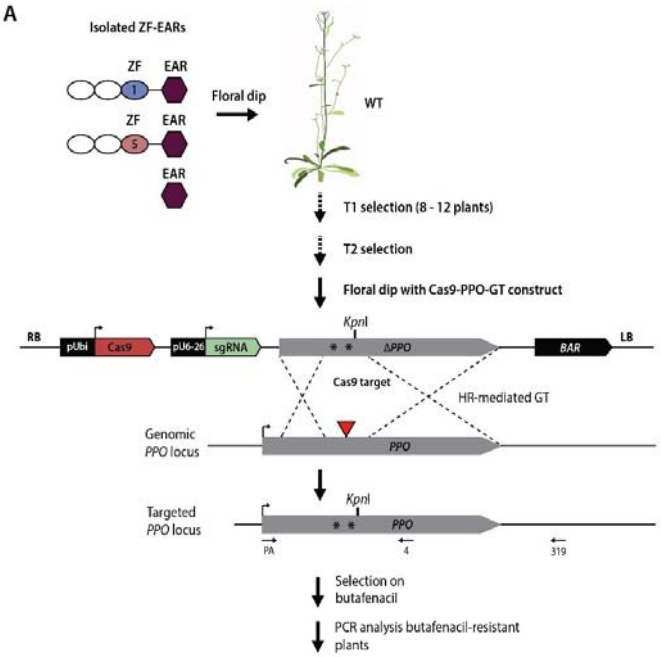
From the plants that contained a GT event the sequences encoding the 3F-EAR effector were PCR-amplified using a forward primer binding in the *RPS5a* promoter and a reverse primer binding in the *nos* terminator of the pRF-pRPS5a-3F-EAR-Kana T-DNA (Table 1), and the 3F-EAR DNA targets were determined by sequencing and named after the triplet it binds to (Figure 2B). PCR products could be generated in seven pool 1 plants and two pool 5 plants. Remarkably, 3F-EAR1, binding to triplets GCG-GGC-GGG, was present in three of the plants with GT events from pool 1. The 3F sequence of three plants (8, 9 and 10) could not be PCR-amplified, suggesting that the 3F-EAR T-DNA was absent in these plants (Figure 2B).

Screening for GT at PPO in reconstituted plant lines with identified ZF-ATFs

We continued to investigate if any of the identified 3F-EAR effectors could cause enhanced GT. Reconstituted T-DNA pRF-pRPS5a-EAR-Kana constructs were made harboring the isolated 3F sequences, and wild type plants were subsequently transformed with these constructs and several plant lines were selected. T2 lines harboring the reconstituted 3F-EAR T-DNAs were then transformed with a T-DNA containing a CRISPR/Cas9 expression cassette for DSB induction at the *PPO* gene as well as the *PPO* repair template. In a previous genome interrogation study using 3F-EAR effectors, it has been shown that expression of the EAR effector alone can already have an phenotypic effect (19). To correct for this effect, plants transformed with T-DNA constructs containing only the EAR domain were used as control, beside WT plants.

Reanalysis of the sequences of each of the isolated 3F-EAR effector encoding genes showed that several contained mutations in the 3F coding sequences, that would result in different proteins (Figure S1). The sequences of 3F-EAR3 and 4 have a T deletion in the first 3F sequence, resulting in a frameshift and a premature stop codon. The sequence of 3F-EAR5 has a substitution from G to A in the third 3F sequence (missense mutation), resulting in one amino acid change and leaving most of the 3F-EAR intact. The sequence of 3F-EAR7 has a G deletion in the first 3F sequence, also resulting in a frame shift and a premature stop codon, producing truncated proteins upon translation. The sequence of 3F-EAR1, 2 and 6 were correct and would result in functional 3F-EAR effectors. The mutations were both discovered in 3F-EAR sequences from reconstituted plant lines and in 3F-EAR sequences present in the original

A



B

# (independent lines)	TGT PCR+KpnI	Plated seeds on Buta (10^4)	Transform. Freq. on PPT (%)	Total transformants tested on Buta	TGT freq. (%)
ZF-EAR1 (8)	1	40	0.16	640	0.15
ZF-EAR2 (10)	1	45	0.08	413	0.24
ZF-EAR3 (10)	1	50	0.35	1750	0.06
ZF-EAR4 (11)	10	85	0.03	255	4.4*
ZF-EAR5 (8)	4	40	0.44	1760	0.23
ZF-EAR6 (11)	2	45	0.34	1530	0.13
ZF-EAR7 (12)	0	55	0.16	880	<0.1
EAR (5)	4	37.5	0.63	2362	0.17
WT	0	12.5	0.14	175	<0.6

Figure 3. Screening for GT at *PPO* in reconstituted ZF-EAR plant lines. **A.** The 3F-EAR encoding fragments, designated 3F-EAR1-7, were isolated from plants with GT events from the initial screening, and reconstituted ZF-EAR T-DNA constructs were generated. Col-0 plants were transformed with these constructs, together with empty-EAR constructs as a negative control, and selected up to the T2 generation. T2 plants were transformed with the Cas9-PPO-GT construct, containing the PPO-GT repair template and Cas9 and sgRNA expression casettes. Primary transformant seeds were plated on medium containing butafenacil to select for GT events at *PPO*. PCR analysis and *Kpn*I digestion of the PCR product was performed on butafenacil-resistant plants to detect GT events at the molecular level. **B.** Overview of TGT events at *PPO* in ZF-EAR1-7 lines, EAR lines containing an EAR construct not fused to a ZF, and WT plants retransformed with the Cas9-PPO-GT construct. * is statistically significant ($P < 0.05$, Fisher's exact test).

3F-EAR library plant lines. Because the truncated 3F-EARs might still have an effect on GT, we decided to continue with the GT screening at *PPO*.

To screen for GT at *PPO*, T1 seeds were selected on butafenacil and positive plants were analyzed by PCR using primers PPO-PA/PPO4 for 5' GT and PPO-PA/SP319 for true GT, and subsequent *Kpn*I digestion (Figure 3). Because GT frequency (number of GT events per (random) integration event) is dependent on transformation frequency of the repair template (random integration events), a portion of T1 seeds was selected on PPT to calculate the corresponding number of transformants that was selected on butafenacil. Heterozygous TGT events were found in plants harboring 3F-EAR1 to 3F-EAR6 as well as in plants harboring an empty EAR construct, indicated by both *Kpn*I-sensitivity and *Kpn*I-resistance of the PPO-PA/PPO-4 and PPO-PA/SP319 PCR products. Transformation frequencies for the *PPO* repair template differed substantially between plant lines, making a fair comparison of GT frequencies between 3F-EAR, empty EAR and WT plants difficult (Figure 3). All in all, 3F-EAR4 plants showed the highest increase in GT frequency, with 10 GT events corresponding to a significantly higher GT frequency of 4.4% ($P = 0.0067$, Fisher's exact test) (Figure 3). The average transformation frequency of the 3F-EAR4 construct was measured by selecting a portion of seeds from plant lines 3F-EAR4-1, 4-4, 4-7 and 4-10, resulting in a transformation frequency of 0.03% for this pool of plant lines. This was suspiciously low, and therefore decreased the number of corresponding plants tested on butafencil drastically compared to the other plant lines. However, GT is still increased substantially at 3F-EAR4 plants compared to the other plant lines when looking at the number of GT events relative to the total number of seeds selected on butafenacil (10 GT events in 850.000 seeds versus 1 or 2 GT events in 400 – 450.000 seeds) (Figure 3B). Plants retransformed with the 3F-EAR1 construct, which was found three times in the initial GT screening, did not show an increase in GT.

A

Plant line	GT PCR+KpnI	Plated seeds on Buta (10^6)	Transform. Freq. on PPT (%)	Total transformants tested on Buta	GT freq. (%)
ZF-EAR4-1	0	5	0.28	142	<0.7
ZF-EAR4-2	1	5	0.53	266	0.38
ZF-EAR4-3	1	5	0.51	254	0.39
ZF-EAR4-4	0	5	0.74	368	<0.27
ZF-EAR4-5	3	5	0.44	220	1.36
ZF-EAR4-6	1	5	0.48	240	0.42
WT1	1	10	0.32	316	0.32
WT2	1	10	0.44	440	0.23
WT3	1	10	0.31	308	0.32

B

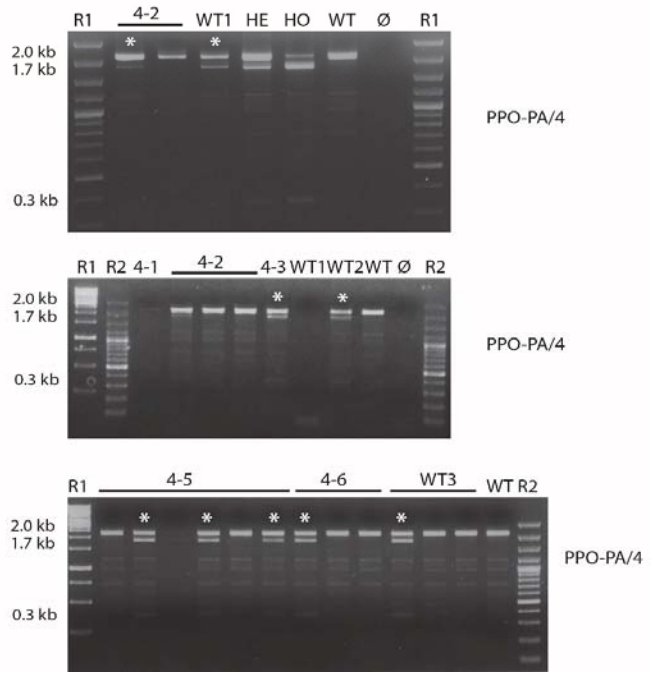


Figure 4. Overview of GT events at *PPO* in reconstituted ZF-EAR4 lines. **A.** Number of GT events and GT frequencies at *PPO* in T2 3F-EAR4-1 to 4-6 reconstituted plant lines**,** and wild type plants**. B.** GT events in butafenacil resistant 3F-EAR4-2, 4-3, 4-5, 4-6 and wild type plants found with PCR using primers PPO-PA/4 and *Kpn*I digestion of the PCR product. Heterozygous GT events are indicated by a white star. HE is a plant line with a validated heterozygous GT event. HO is a plant line with a validated homozygous GT event. WT is wild type control. Ø is PCR control without template. R1 is the 1 kb ruler (ThermoFisher Scientific), R2 is the 100 bp+ ruler (ThermoFisher Scientific).

Screening for GT at *PPO* in reconstituted 3F-EAR4 plant lines

Plants harboring 3F-EAR4 showed the highest GT frequency at *PPO*. However, because of the very low *PPO* repair template transformation frequency resulting in a low number of butafencil-tested transformants, we decided to repeat this GT experiment with each of the ten plant lines containing reconstituted 3F-EAR4 individually. After transformation with the GT construct a total of $5*10^4$ seeds of T2 3F-EAR4 lines 3F-EAR4-1 to 4-6 and $1*10^5$ seeds of wild type were selected on butafenacil. Subsequently, butafenacil-resistant seedlings were screened for GT with PCR using primers PPO-PA/PPO-4 followed by digested with *Kpn*I. A portion of seeds was used to determine the transformation frequency of the *PPO* repair template for each transformed plant line. Transformation frequencies ranged from 0.28 to 0.74 % (Figure 4A), which was better compared to the very low transformation frequency for the reconstituted 3F-EAR4 plants in the first experiment.

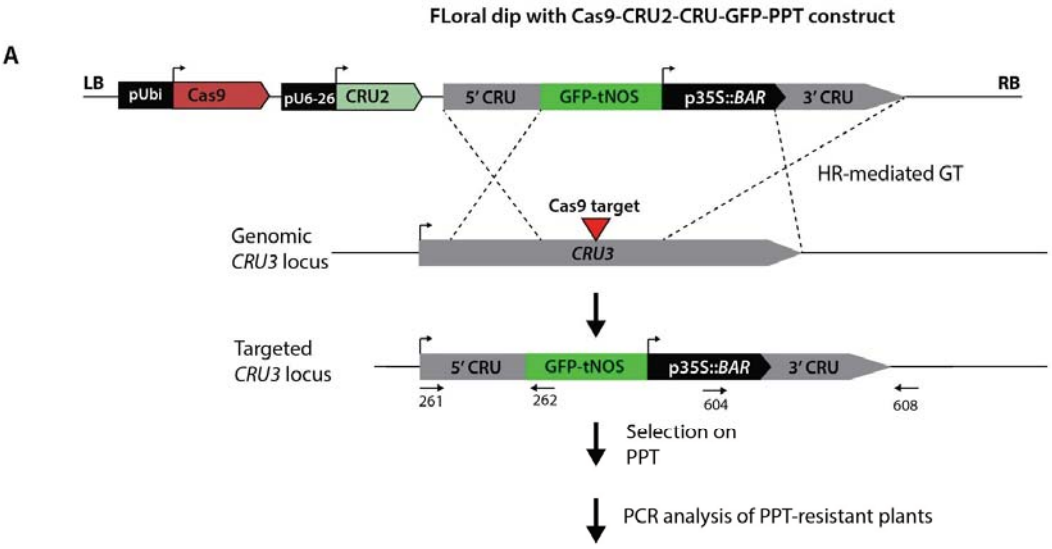
Heterozygous GT events were found in butafenacil-resistant seedlings of lines 3F-EAR4-2, 4-3, 4-5, 4-6, and in wild type plants, indicated by both *Kpn*I-sensitivity and *Kpn*I-resistance of the PPO-PA/PPO-4 PCR product (Figure 4A, B). The highest GT frequency was found for line 3F-EAR4-5, with 3 GT events from a total of 220 transformants, corresponding to a GT frequency of 1.36%, which was, although still higher than in wild-type plants, not statistically significant (P = 0.31, Fisher's exact test).

Screening for GT at *CRU3* in reconstituted 3F-EAR4 plant lines

Plants harboring 3F-EAR4 showed the highest GT frequency at *PPO*. Therefore, 3F-EAR4 plants were also analyzed for GT at the *CRU*3 gene using the GT repair template described in Chapter 4, but with an added *BAR* gene under control of the 35S promoter, 3' of the GFP-tNOS sequence, conferring resistance to PPT. In this way, it is not necessary to determine the transformation frequency of the GT repair template separately in order to calculate the number of transformants to determine GT frequency. GT frequency could directly be determined with PCR analysis as the proportion of total transformants that show PPT resistance.

3F-EAR4-2, 4-4 and 4-6 T2 plant lines were transformed with a GT binary vector containing both the Cas9-CRU2 expression cassette (Chapter 2, 4) and the CRU-GFP repair template (Chapter 4) with an added 35S-PPT expression cassette 5'of the GFP-tNOS sequence (Chapter 4) (Figure 5A). Plants were selected on PPT and PPT-resistant plants were analyzed for GT with PCR.

A



B

Plant line	TGT	5' GT	3' GT	PPTr plants	TGT freq. (%)	TGT + EGT freq. (%)
ZF-EAR 4-2	2	8	0	716	0.28	1.3
ZF-EAR 4-4	3	3	0	554	0.54	1.1
ZF-EAR 4-6	4	5	1	767	0.52	1.3
WT	1	4	0	341	0.29	1.5

Figure 5. Screening for GT at *CRU3* in reconstituted ZF-EAR4 plant lines **A.** T2 plant lines 4-2, 4-4 and 4-6 harboring ZF-EAR4 transformed with the Cas9-CRU2-CRU-GFP-PPT construct, containing the CRU-GT repair template with BAR as a selection marker, conferring PPT resistance, and Cas9-CRU2 expression casettes. Primary transformant seeds were plated on medium containing PPT. To select for GT events at *CRU3*, PCR analysis was performed on PPT-resistant plants to detect GT events at the molecular level. **B.** Overview of GT events at *CRU3* in ZF-EAR4-2, 4-4 and 4-6 lines and wild-type plants retransformed with the Cas9-CRU2-CRU-GFP-PPT construct. TGT is true gene targeting. EGT is ectopic gene targeting.

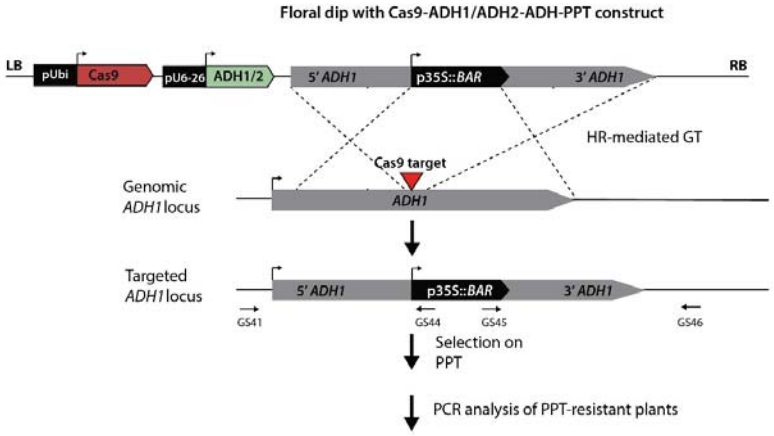
To detect GT at CRU3, the complete *CRU3* GT integration site was first PCR-amplified using primers SP261 and SP608, resulting in a 4810 bp fragment for the unmodified *CRU3* locus and 7678 bp fragment after a GT event. This PCR product was subsequently used as a template for PCR at the 5' and 3' GT integration sites using primers SP261/SP262 (5' GT), which should result in a 1133 bp fragment, and SP604/SP608 (3' GT) which should result in a 3500 bp fragment. T1 seeds were selected on PPT. A total of 2387 PPT-resistant plants from lines 3F-EAR4-2 (716), 4-4 (554), 4-6 (767) and WT Col-0 (341) were screened. True GT (TGT) and ectopic GT (EGT) events were found in all 3F-EAR4 lines and wild type plants (Figure 5B, S2). The two TGT events found in 3F-EAR4-2 plants corresponded to a TGT frequency of 0.28%, comparable to wild type plants. TGT frequency in 3F-EAR4-4 and 4-6 plants were almost doubled with frequencies of 0.54 % and 0.52 %, although not significantly higher compared to wild type plants (P = 1, Fisher's exact test). Furthermore, GT frequencies in general were much lower than observed at the *PPO* gene.

Screening for GT at *ADH1* in reconstituted 3F-EAR4 plant lines

In order to investigate if an increase in GT frequency in 3F-EAR4 plants could be detected at yet another target locus, plants were also screened for GT at the alcohol dehydrogenase 1 (*ADH1*) gene. GT at this gene has been shown to be successful by others (26, 27). Two GT constructs were developed, harboring a CRISPR/Cas9 expression construct with sgRNAs for *ADH1* as described earlier (sgRNA ADH1 and ADH2, Chapter 2), and an *ADH1* repair template consisting of a p35S-*BAR* sequence flanked by *ADH1* homology arms of approximately 800 bp each. T2 plant lines 3F-EAR4-2, 4-4 and 4-6 and T2 empty EAR lines 15, 16 and 17, together with wild type plants, were transformed with these constructs and screened for GT at *ADH1* by selection on PPT. PCR for both 5' GT and 3' GT was performed with primer pairs GS41/GS44 (5' GT) and GS45/GS46 (3' GT), which should render PCR products of 1132 bp and 1130 bp, respectively, when true GT has occurred (Figure 6A). First, leaves of approximately 10 PPT-resistant seedlings were pooled and PCR for GT detection was performed on genomic DNA isolated from these pools. When GT events were detected, individual plants from these pools were screened again. One true GT event was detected in a 3F-EAR4-4 plant transformed with the Cas9-ADH1-ADH-PPT construct out of 98 transformants, resulting in a ~1% GT frequency, which is not statistically significant (P = 0.47, Fisher's exact test) (Figure 6B). Several ectopic GT events for either 5' or 3' were detected in 3F-EAR4-4 and 4-6 plants for both Cas9-ADH constructs (Figure 6B, Figure S3). No GT events were found in 3F-EAR4-2, empty EAR and

wild type plants. Since only one true GT event and several ectopic GT events were detected, the numbers were too low to draw any conclusions.

A



B

Plant line	TGT	5' GT	3' GT	PPTr plants	TGT freq. (%)	TGT and EGT freq. (%)
Cas9-ADH1						
Nr 4-2	0	0	0	71	-	-
Nr 4-4	1	2	1	98	1 %	4.1%
Nr 4-6	0	0	0	29	-	-
EAR15	0	0	0	57	-	-
EAR16	0	0	0	57	-	-
EAR17	0	0	0	32	-	-
WT	0	0	0	111	-	-
Cas9-ADH2						
Nr 4-2	0	0	0	104	-	-
Nr 4-4	0	3	1	152	-	2.6%
Nr 4-6	0	0	0	30	-	-
EAR15	0	0	0	56	-	-
EAR16	0	0	0	56	-	-
EAR17	0	0	0	32	-	-
WT	0	0	0	47	-	-

Figure 6. Screening for GT at *ADH1* in reconstituted ZF-EAR4 plant lines **A.** T2 plant lines 4-2, 4-4 and 4-6 harboring ZF-EAR4 transformed with the Cas9-ADH1-ADH-PPT or Cas9-ADH2-ADH-PPT constructs, containing the ADH1 GT repair template with *BAR* as a selection marker, conferring PPT resistance, or Cas9-ADH1 and Cas9-ADH2 expression casettes, respectively. Primary transformant seeds were plated on medium containing PPT. To select for GT events at *ADH1*, PCR analysis was performed on PPT-resistant plants to detect GT events at the molecular level. **B.** Overview of GT events at *ADH1* in ZF-EAR4-2, 4-4 and 4-6 lines, wild-type plants retransformed with the Cas9-ADH1-ADH-PPT or Cas9-ADH2-ADH-PPT constructs. TGT is true gene targeting. EGT is ectopic gene targeting.

Discussion

In this study, we attempted to identify plants with higher GT frequency using the genome interrogation method in Arabidopsis plant lines harboring 3F-EAR artificial transcription factor encoding genes. In previous studies, the *PPO* GT system was successfully used to screen for GT events (20, 24, 25). Initial GT screening at the *PPO* gene yielded GT events in plants from 3F-EAR libraries 1 and 5. One of the 3F-EAR effectors, 3F-EAR1, yielded 3 GT events. The 3F-EAR library is constructed in such a way that the same 3F-EAR construct is independently transformed multiple times. In this way, 3F-EAR1 could have been overrepresented in the library. Indeed, upon retesting we found that this 3F-EAR fusion did not have any effect on GT frequency in reconstituted plants. However, in the initial screening only a *PPO* GT repair template was presented, while CRISPR/Cas9 and the *PPO* GT repair template was used for GT screening in reconstituted plants. It is therefore possible that the effect of the 3F-EAR1 fusion on GT in the initial screening was masked by an overall GT enhancement by CRISPR/Cas9 in the reconstituted plants.

Unfortunately, mutations were discovered in four of the 3F sequences, which we retrieved. The single nucleotide deletions in 3F-EAR3, 4 and 7 resulted in frameshifts and premature stop codons, while a substitution in 3F-EAR5 resulted in one amino acid change, but left the 3F-EAR protein largely intact. Nevertheless, we decided to continue screening for enhanced GT with these constructs, as starting over with intact 3F-EAR constructs was not possible due to time constraints and, more importantly, it could still be that the retrieved 3F-EAR3, 4 and 7 constructs might still have been able to trigger enhanced GT.

Screening for GT at *PPO* in reconstituted 3F-EAR lines in the first experiment showed that GT frequency was significantly increased in plant lines harboring the 3F-EAR4 construct. However, because transformation frequency of the *PPO* repair template is determined separately from GT frequency, a low transformation frequency results in a low number of total transformants in the calculation of GT frequency. As the transformation frequency of the *PPO* repair template was much lower in 3F-EAR4 plant lines compared to the other 3F-EAR plant lines, this skewed the results and artificially increased the apparent GT frequency. Transformation frequency of the *PPO* repair template in reconstituted 3F-EAR4 plants in a second experiment was much higher, which subsequently resulted in a lower GT frequency than determined in the first screening of reconstituted 3F-EAR4 plants. Thus, the low transformation frequency observed for 3F-EAR4 plants in the first experiment was likely due to inefficient PPT selection. However, GT was still substantially increased in 3F-EAR4 plants compared to

the other plant lines when looking at the number of GT events relative to the total number of butafenacil-selected seeds. However, it must be noted that the amount of 3F-EAR4 seeds screened was more than double the amount of seeds harboring the EAR construct without 3F fusion, and almost seven times more than wild type seeds. In the second experiment, the same amount of 3F-EAR4 seeds and wild type seeds were screened, and the amount of GT events detected in all 3F-EAR lines combined was doubled compared to wild type. This indicates that the 3F-EAR4 construct might still have had an effect as trigger of an enhanced GT phenotype.

In order to have only one selection system for both transformation frequency and GT frequency, we developed a GT system for *CRU3* and *ADH1* based on PPT selection to directly screen for GT events in PPT-resistant plants. This approach showed a slight increase in GT frequency at *CRU3,* but this increase was not statistically significant. Almost no GT events were found for *ADH1*. However, especially in the *ADH1* GT experiment, the number of plants analyzed for GT was low and larger sample sizes are needed for statistical significance.

When comparing the *PPO* GT system with the GT systems developed for *CRU3* and *ADH1*, the *PPO* GT system was the most efficient of the three, and GT events yielded by this system were the most convincing. This might be explained by the fact that during a GT event at *PPO* the endogenous *PPO* gene is replaced by the same sequence that only differs in a few point mutations. HR-mediated repair might work more effectively with this donor DNA, compared to the integration of a foreign sequence flanked by homology arms during GT at *CRU3* and *ADH1*. However, overall the total number of GT events found with all three systems was low, and more plants need to be analyzed to give a conclusive answer.

The main reason we chose to work with the 3F-EAR library instead of the 3F-VP16 was based on results from previous genome interrogation experiments in Arabidopsis that showed that the presence of 3F-EAR fusions in primary transformants correlated with larger phenotypic variation compared to 3F-VP16 fusions (19). However, in reconstituted 3F-EAR lines this correlation was lost, and it was shown that the observed changes in gene expression were most likely caused by a global effect of the EAR domain itself, rather than the presence of a specific 3F domain (19). In reconstituted lines harboring 3F-ATFs fused to the VP16 transcriptional activator the observed transcriptional changes were truly 3F-specific and causative for the phenotype. Although we corrected for the unspecific effect of the EAR domain by using plant lines only expressing the EAR domain as a control library for GT, in the end the 3F-EAR library seems less suitable for discovering genes that are causative for a unique novel phenotype compared to the 3F-VP16 library. With 15 3F subpools transformed and 4278 primary

transformants, the 3F-VP16 library is more complex than the 3F-EAR library, which should increase the chance of finding novel unique phenotypes. It would therefore be better to use the 3F-VP16 library to screen for enhanced GT phenotypes in future experiments. However, a fairly large number of primary 3F transformants need to be screened to sufficiently cover the genome. In our experiment, the 144 primary 3F-EAR transformants screened comprised $1/5^{th}$ of the total library of 700 transformants. As each 3F effector has a chance to bind to both strands of the Arabidopsis genome of approximately 130 Mb, one 3F sequence can bind to 991 potential primary targets ($(1/4)^9 * 2 * 130.000.000 = 991$). With a total of 4096 possible 3F (3 x GNN) combinations, at least 32 3Fs are needed to cover the Arabidopsis genome once (130.000.000 / 991 / 4096 = 32). With 144 3Fs screened, a genome coverage of ~4.5 times was ensured. However, in this calculation it is assumed that each 3F can bind to a unique target, which in reality most likely does not happen due to highly similar 3Fs. Thus, a 4.5 times coverage is likely an overestimate of the real coverage. Also, because GT frequencies are generally low, it is still necessary to select amongst hundreds of thousands of seeds in order to find statistically significant numbers of GT events. This makes such experiments time consuming and logistically challenging.

In conclusion, using the genome interrogation method combined with an efficient GT selection system can potentially yield novel phenotypes with enhanced GT frequency. Even when plants with strongly enhanced GT can be found, the next step of discovering novel genes involved in this pathway is also challenging. In order to identify multiple ZF-ATFs that are really causative for transcriptional changes that lead to the observed phenotype, one ideally needs to identify several genome interrogation mutants with similar phenotypes to find shared transcriptional changes that are causative for the phenotype. However, at the plant level it is difficult to find similar phenotypes in small population of genome interrogation mutants. A solution that proved to be helpful is the so-called nearest active neighbor/nearest inactive neighbor (NIN/NAN) approach (6, 17). In this approach, a set of ZF-ATFs that are highly similar to the one that triggers the newly observed phenotype of interest is used. When these ZF-ATFs also trigger the same phenotype, one can perform comparative transcriptome analyses between plants expressing the ZF-ATFs that trigger the phenotype (NANs) and plants expressing the ZF-ATFs that do not trigger the phenotype (NINs). In this way, a set of genes was identified that contributed to an enhanced somatic homologous recombination phenotype (17).

Acknowledgements

We would like to thank the Puchta lab for supplying the pEn-Chimera and the pDe-Cas9 vectors. We also would like to thank Dr. Bert van der Zaal and Dr. Niels van Tol for supplying the seeds for the 3F-EAR libraries and for advice during this study. This work was financially supported by the Partnership Program STW-Rijk Zwaan of the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs, Agriculture and Innovation (STW12428).

References

1. Offringa R, et al. (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation Remko. *EMBO J* 9(10):3077–3084.
2. Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. *Int J Dev Biol* 57:629–637.
3. Puchta H (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. *Plant J* 13(3):331–339.
4. Puchta H, Fauser F (2014) Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J* 78(5):727–741.
5. Fauser F, et al. (2012) In planta gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
6. van Tol N, van der Zaal BJ (2014) Artificial transcription factor-mediated regulation of gene expression. *Plant Sci* 225:58–67.
7. Sadowski I, Ma J, Triezenbergt S, Ptashne M (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335:563–564.
8. Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-takagi M (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* 13:1959–1968.
9. Hiratsu K, Ohta M, Matsui K, Ohme-Takagi, M (2002) The SUPERMAN protein is an active repressor whose carboxy-terminal repression domain is required for the development of normal flowers. *FEBS Lett* 514:351–354.
10. Park KS, et al. (2003) Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol* 21(10):1208–1214.
11. Beltran A, Liu Y, Parikh S, Temple B, Blancafort P (2006) Interrogating genomes with combinatorial artificial transcription factor libraries: asking zinc finger questions. *Assay Drug Dev Technol* 4(3):317–331.
12. Lee J, et al. (2011) Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library. *PLoS One* 6(7):1–10.
13. Lee JY, Yang KS, Jang SA, Sung BH, Kim SC (2011) Engineering butanol-tolerance in Escherichia coli with artificial transcription factor libraries. *Biotechnol Bioeng* 108(4):742–749.
14. Lee JY, et al. (2008) Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in Escherichia coli. *Nucleic Acids Res* 36(16):e102.

15. Lindhout BI, Pinas JE, Hooykaas PJJ, van der Zaal BJ (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J* 48(3):475–483.
16. Weijers D, et al. (2001) An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development* 128(21):4289–4299.
17. Jia Q, et al. (2013) Zinc finger artificial transcription factor-based nearest inactive analogue/nearest active analogue strategy used for the identification of plant genes controlling homologous recombination. *Plant Biotechnol J* 11(9):1069–1079.
18. van Tol N, Pinas J, Schat H, Hooykaas PJJ, van der Zaal BJ (2016) Genome interrogation for novel salinity tolerant Arabidopsis mutants. *Plant Cell Environ* 39(12):2650–2662.
19. van Tol N, et al. (2017) Enhancement of Arabidopsis growth characteristics using genome interrogation with artificial transcription factors. *PLoS One* 12(3):e0174236.
20. de Pater S, Pinas JE, Hooykaas PJJ, van der Zaal BJ (2013) ZFN-mediated gene targeting of the Arabidopsis protoporphyrinogen oxidase gene through Agrobacterium-mediated floral dip transformation. *Plant Biotechnol J* 11(4):510–5.
21. de Pater S, Neuteboom LW, Pinas JE, Hooykaas PJJ, van der Zaal BJ (2009) ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. *Plant Biotechnol J* 7(8):821–35.
22. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735–43.
23. Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* 19(6):1349.
24. Hanin M, et al. (2001) Gene targeting in Arabidopsis thaliana. *Plant J* 28(6):671–677.
25. de Pater S, Klemann BJPM, Hooykaas PJJ (2018) True gene-targeting events by CRISPR/Cas-induced DSB repair of the PPO locus with an ectopically integrated repair template. *Sci Rep* 8(1):3338.
26. Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *Plant J*:348–359.
27. Schiml S, Fauser F, Puchta H (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *Plant J* 80:1139–1150.

Supplementary information

Table S1. Primers used for cloning and PCR

Primer	**Sequence (5' – 3')**	**Used for**
RPS5a	GCCCAAACCCTAAATTTCTCATC	3F-EAR sequence, forward
NOSt	CAAGACCGGCAACAGGAT	3F-EAR sequence, reverse
PPO-PA	GTGACCGAGGCTAAGGATCGT	*PPO* GT forward
PPO-4	CATGAAGTTGTTGACCTCAATC	*PPO* GT reverse (5' GT)
SP319	CTATCAAAGAGCACAGACAGC	*PPO* GT reverse (true GT)
SP261	CTCAGCAATCTCCTCGTTG	5' GT *CRU3* forward
SP262	TCGCCCTTGCTCACCAT	5' GT *CRU3* reverse
SP604	ATGAGCCCAGAACGACGCCC	3' GT *CRU3* forward
SP608	CAGAAACAGAGCACCAAATGGG	3' GT *CRU3* reverse
SP585	GGATCCTGCAGTAATTCGGGGGATCTGGATTTTAG	35S-PPT-Tnos forward
SP586	CACTGCAGGCTAGAGCAGCTTGCAAC	35S-PPT-Tnos reverse
GS41	ATTCCAACTTGATGACCAAG	5' GT *ADH1* forward
GS44	GCAATGATGGCATTTGTAGG	5' GT *ADH1* reverse
GS45	GTGTGAGTAGTTCCCAGATAAGG	3' GT *ADH1* forward
GS46	TGAGTCTTGAAGGCATCGTC	3' GT *ADH1* reverse

3F-EAR3: GAT-GGT-GGG

GAT sequence: T-deletion

```
correct GGCGAGAAGCCTTATGCCTGTCCTGAGTGCGGGAAGTCTTTTAGCACCTCGGGTAATCTCGTCCGTCACCAACGTACGCATACC
-T      GGCGAGAAGCCTTATGCCTGTCCTGAGTGCGGGAAGTCTT-TAGCACCTCGGGTAATCTCGTCCGTCACCAACGTACGCATACC
```

Resulting amino acid sequence 3F-EAR3:

Correct sequence:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPECGKSFSTSGNLVRHQRTHTGEKPYACPECGKSFSTSGHLVRHQRTHTGE
KPYACPECGKSFSRSDKLVRHQRTHTGELGGGSGEKPGKKTSGQAGQLDLDLELRLGFA*

-T:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPECGKSLAPRVISSVTNVRIPARSPMRALSVASPLAPRVIWCVTNVHIPAK
NRMLVQNVVSPFLVAISSFATKGLIPASWVAVPEKSPGRKPRARPANWIWIWN*

3F-EAR4: GGG-GTG-GGG

1st GGG sequence: T-deletion

```
correct GGCGAAAAACCGTATGCTTGTCCAGAATGTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAAGGACTCATACC
-T      GGCGAAAAACCGTATGCTTGTCCAGAA-GTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAAGGACTCATACC
```

Resulting amino acid sequence 3F-EAR4:

Correct sequence:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPECGKSFSRSDKLVRHQRTHTGEKPYACPECGKSFSRSDELVRHQRTHTGE
KPYACPECGKSFSRSDKLVRHQRTHTGELGGGSGEKPGKKTSGQAGQLDLDLELRLGFA*

-T:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPEVVSPFLVAISSFATKGLIPAKSRMRAPNVGSPFLVATSLFATKGPIPAK
NRMLVQNVVSPFLVAISSFATKGLIPASWVAVPEKSPGRKPRARPANWIWIWN*

3F-EAR5: GAA-GCC-GGG

GGG sequence: G/A substitution

```
correct  GGCGAAAAACCGTATGCTTGTCCAGAATGTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAAGGACTCATACC
G/A      GGCGAAAAACCGTATGCTTGTCCAGAATGTGGTAAGTCCTTTTCTCGTAGCAATAAGCTCGTTCGCCACCAAAGGACTCATACC
```

Resulting amino acid sequence 3F-EAR5:

Correct sequence:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPECGKSFSQSSNLVRHQRTHTGEKPYACPECGKSFSDCRDLARHQRTHTGE
KPYACPECGKSFSRSDKLVRHQRTHTGELGGGSGEKPGKKTSGQAGQLDLDLELRLGFA*

G/A:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPECGKSFSQSSNLVRHQRTHTGEKPYACPECGKSFSDCRDLARHQRTHTGE
KPYACPECGKSFSRSNKLVRHQRTHTGELGGGSGEKPGKKTSGQAGQLDLDLELRLGFA*

3F-EAR7: GTA-GAT-GCC

GTA sequence: G-deletion

```
correct GGCGAGAAGCCTTACGCTTGCCCCGAGTGTGGCAAATCGTTCTCCCAGTCCTCCTCGTTGGTCCGTCATCAACGGACGCATACC
-T      GGCGAGAAGCCTTACGCTTGCCCCGAGT-TGGCAAATCGTTCTCCCAGTCCTCCTCGTTGGTCCGTCATCAACGGACGCATACC
```

Resulting amino acid sequence 3F-EAR7:

Correct sequence:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPECGKSFSQSSSLVRHQRTHTGEKPYACPECGKSFSTSGNLVRHQRTHTGE
KPYACPECGKSFSDCRDLARHQRTHTGELGGGSGEKPGKKTSGQAGQLDLDLELRLGFA*

-G:
MADYKDDDDKRPLEPPKKKRKVELAGTEAQAALEPTGEKPYACPELANRSPSPPRWSVINGRIPARSLMPVLSAGSLLAPRVISSVTNVRIPAR
SLTLARSVASHSATAVISQGINAPIPASWVAVPEKSPGRKPRARPANWIWIWN*

Figure S1. Overview of mutations in 3F-EAR3, 4, 5 and 7 sequences. Shown are the mutations in the 3F sequences and the resulting change in amino acid sequence of the encoded 3F-EAR protein highlighted in grey. *indicates a stop codon.

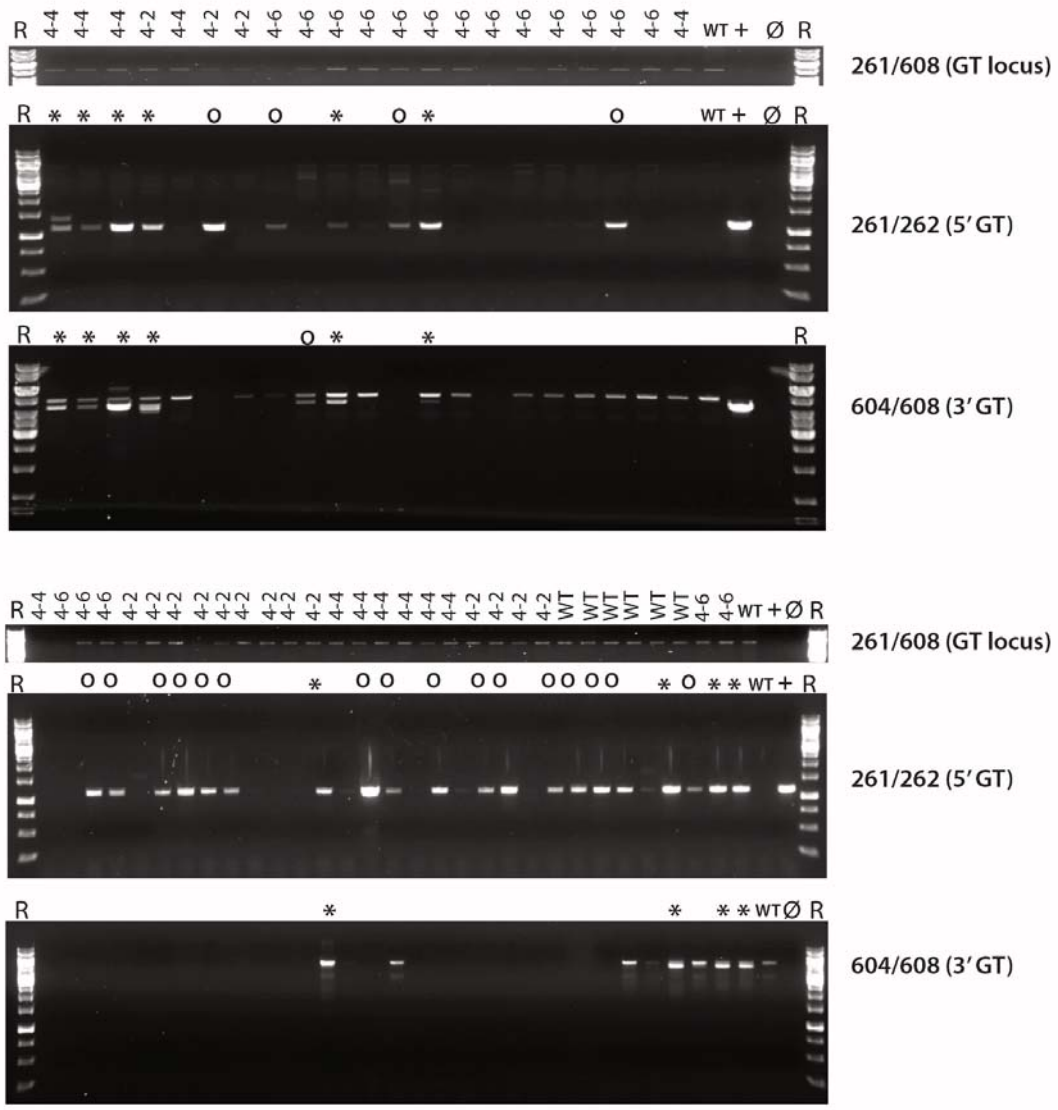


Figure S2. *CRU3* GT events detected with PCR in ZF-EAR4 lines 4-2, 4-4 and 4-6 and WT Col-0. The *CRU3* target was first amplified using primers SP261/SP608. This PCR product was used as a template for 5' GT and 3' GT detection using primers SP261/SP262 and SP604/SP608, respectively. True GT events are indicated with a star. 5' or 3' ectopic GT events are indicated with a circle. WT is wild-type Col-0. + is putative positive true GT control. Ø is negative control (PCR reaction without template). R is the 1 kb ruler (ThermoFisher Scientific).

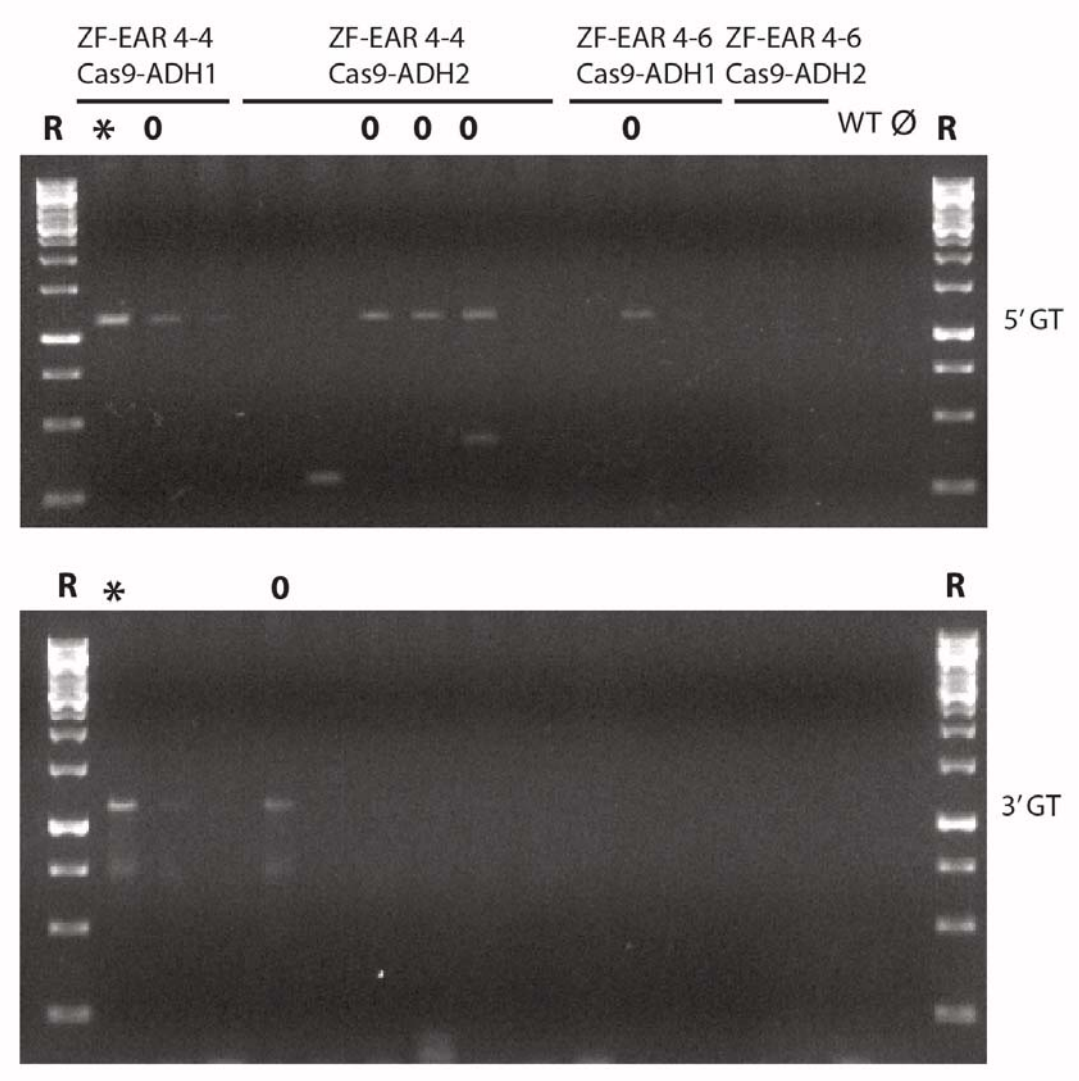


Figure S3. *ADH1* GT events detected with PCR in ZF-EAR4 lines 4-4 and 4-6. 5' GT and 3' GT was detected using primers GS41/44 and GS45/46, respectively. True GT events are indicated with a star. 5' or 3' ectopic GT events are indicated with a circle. WT is wild-type Col-0, Ø is PCR control without template. R is the 1 kb ruler (ThermoFisher Scientific).