



Universiteit
Leiden
The Netherlands

Strategies for the improvement of genome editing in *Arabidopsis thaliana*

Strunks, G.D.

Citation

Strunks, G. D. (2019, September 11). *Strategies for the improvement of genome editing in Arabidopsis thaliana*. Retrieved from <https://hdl.handle.net/1887/77743>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/77743>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/77743> holds various files of this Leiden University dissertation.

Author: Strunks, G.D.

Title: Strategies for the improvement of genome editing in *Arabidopsis thaliana*

Issue Date: 2019-09-11

**Strategies for
the improvement
of genome editing in
*Arabidopsis thaliana***

Gary Strunks

Gary Strunks

Strategies for the improvement of genome editing in *Arabidopsis thaliana*

PhD thesis, Leiden University, 2019

Copyright © Gary Strunks, 2019. All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the copyright holder.

ISBN | 978-94-6375-495-8

Layout | Gary Strunks and Ika Putranto

Cover illustration and design | Ika Putranto

Printing | Ridderprint BV, the Netherlands

Strategies for the improvement of genome editing in
Arabidopsis thaliana

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 11 september 2019
klokke 11.15 uur

door

Gary Daniël Strunks

geboren te Amsterdam, Nederland

18 maart 1986

Promotor: prof.dr. P.J.J. Hooykaas

Copromotor: dr. B.S. de Pater

Promotiecommissie: prof.dr. G.P. van Wezel (voorzitter)

prof.dr. R. Offringa (secretaris)

Overige commissieleden: prof.dr. Y. Bai, Wageningen UR

dr. F.A. Krens, Wageningen UR

To Ika and my parents

Contents

Chapter 1	General introduction	9
Chapter 2	Sequence-specific nucleases for genome editing in plants: TALENs versus CRISPR/Cas9	47
Chapter 3	CRISPR/Cas9-induced double strand break repair in <i>Arabidopsis</i> nonhomologous end-joining mutants	63
Chapter 4	<i>In planta</i> gene targeting during meiosis	89
Chapter 5	Developing a strategy for the identification of gene targeting mutants in <i>Arabidopsis thaliana</i>	121
Chapter 6	Summarizing discussion	149
Chapter 7	Samenvatting	157
Curriculum Vitae		165

Chapter 1

General introduction

Gary Strunks

Introduction

The genome is subject to various exogenous and endogenous assaults that can cause DNA damage. DNA damage that only affects a single strand can be repaired relatively easily, as one intact strand is still present. However, damage that affects both strands such as double-strand breaks (DSBs) can be particularly harmful because no intact complementary strand is available as a repair template. DSBs can arise in various ways. They can be induced in a programmed manner, such as in V(D)J recombination and at the initiation of meiotic recombination, but also following replication fork stalling and collapse or by exogenous damaging agents like ionizing irradiation and reactive oxygen species (1). The inability to repair DSBs can cause genomic instability and mutations, which in turn can lead to cancer predisposition and cell death (2).

In order to maintain genome integrity, cells have evolved complex and highly conserved mechanisms to detect and repair DNA lesions. Single-strand damage can be repaired by mechanisms such as base excision repair (BER), mismatch repair (MMR) and nucleotide excision repair (NER). Generally, DSBs can be repaired by two main pathways: homologous recombination (HR) and nonhomologous end-joining (NHEJ). HR involves DSB repair by using the homologous information of an intact DNA sequence provided by a sister chromatid or homologous chromosome. The use of homologous sequences to guide DSB repair ensures accurate repair. In contrast, DSB repair through NHEJ involves direct ligation of two double strand ends, without the use of homologous sequence information to guide repair. Although generally considered a conservative repair mechanism, NHEJ-mediated DSB repair can occasionally lead to insertions and deletions at the break site (3). In general, HR is a minor pathway in somatic cells of higher eukaryotes, because the efficiency of it is dependent on the presence of homologous DNA sequences at specific points during the cell cycle, whereas NHEJ is the predominant DSB repair pathway which can act throughout the cell cycle (4).

From a plant biotechnological point of view, the HR and NHEJ pathways have been important subjects of study for their role in gene targeting and targeted mutagenesis, respectively. Gene targeting (GT) is a molecular technique that employs HR to change an endogenous gene or add novel genetic sequences, while targeted mutagenesis is the generation of mutations at a specific genomic site as a result of DSB induction and imperfect repair through NHEJ. However, in plants and most other higher eukaryotes, the efficiency of GT has remained very low due to the low frequency of somatic HR. A breakthrough came when it was found that GT frequency can be significantly increased with the use of artificial sequence-specific nucleases, that can induce DSBs at a DNA sequence of interest.

In this chapter, the DNA damage response and the two main pathways of DSB repair, NHEJ and HR, will first be reviewed. This will be followed by a discussion on GT and the different approaches aimed at increasing GT efficiency in plants. Finally, an outline of this thesis will be given.

The DNA damage response

When DNA damage occurs a complex signal transduction pathway known as the DNA damage response (DDR) is activated. The DDR senses DNA damage which leads to the activation of downstream effectors that in turn affect various processes, including transcriptional responses, chromatin remodeling, cell cycle arrest, DNA repair and apoptosis or senescence. Cell cycle arrests provide the cell with enough time to repair the damage. However, if the damage is beyond repair, cells can go into apoptosis or senescence (5). Master regulators of the DDR are members of the phosphatidylinositol 3-OH kinase-like kinases (PIKKs) family, which connect the initial DSB signal to downstream effectors in the DDR. The two most studied master regulators of the DDR are ataxia telangiectasia mutated (ATM) and ATM- and RAD3-related (ATR) (Figure 1) (5–8).

The DDR has been well-described in mammals. It starts with the recognition of the break by the MRE11/RAD50/NBS1 (MRN) complex in mammals and plants or the MRE11/RAD50/Xrs2 (MRX) complex in yeast (9, 10). The MRN complex activates ATM through auto phosphorylation, which in turn activates a multitude of DSB repair factors through phosphorylation, most importantly the histone variant H2AX, producing γ H2AX foci (Figure 1) (11, 12). These foci promote amplification of the DDR signal by acting as recruitment platforms for chromatin modifying factors, DDR signaling proteins and DNA repair proteins (13). ATM is recruited to DSBs, while ATR mainly responds to the presence of ssDNA, bound by the ssDNA-binding replication protein A (RPA), which is a characteristic of replication errors (14). However, during DSB repair ATM and ATR are usually both activated, as DSB processing may result in ssDNA formation after resection and DSB formation through replication fork stalling also involves ssDNA (8). In mammals, ATR and ATM propagate the DDR signal further by phosphorylating downstream protein kinases CHK1 and CHK2, respectively, as well as effector protein p53 (15–17). CHK1 and 2 also phosphorylate p53 to further stabilize it (18, 19). Subsequently, active p53 can regulate the transcription of various downstream proteins, including cyclin-dependent kinases (CDKs) and DNA repair factors, leading to cell cycle arrest and DSB repair (Figure 1) (5). The budding yeast orthologs of ATM,

ATR, CHK1 and CHK2 are Tel1, Mec1, ScChk1 and ScRad53, respectively (20, 21). As p53 is absent in yeast, it is thought that downstream processes such as DSB repair and cell-cycle regulation are directly activated by Chk1 and Rad53 in yeast (22).

Plant orthologs of the factors in the initial stages of the DDR, such as the MRN complex subunits and the master regulators ATM and ATR, have been identified (10, 23, 24). Surprisingly, orthologs of CHK1, CHK2 and p53 have not been found. Instead, in plants this role is thought to be played by the transcription factor suppressor of gamma response 1 (SOG1). Active ATM and ATR can activate SOG1 by phosphorylation, which in turn can induce the transcription of a plethora of genes involved in the DDR, including CDK inhibitors, DNA repair proteins and factors involved in programmed cell death (Figure 1) (5, 8, 25–27).

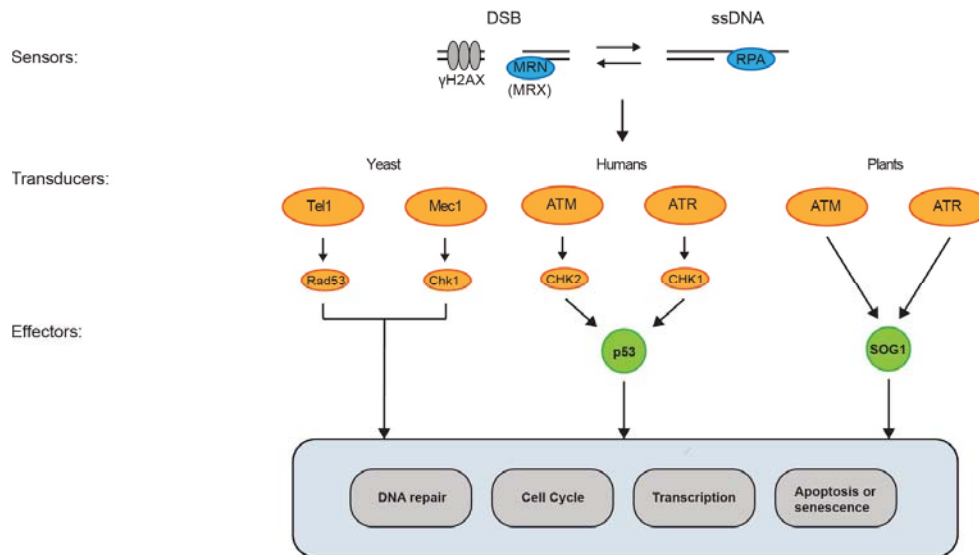


Figure 1. The DNA damage response. The pathway is activated when DSBs with blunt or resected ends are bound by the MRN complex and RPA, respectively. This leads to activation of the transducers ATM and ATR, which propagate the signal via phosphorylation of CHK1 and CHK2. ATM, ATR and CHK1 and CHK2 all play a part in the activation of downstream effector p53, which in turn regulates the transcription of various genes involved in DNA repair, cell cycle arrest and apoptosis or senescence. In plants, the role of CHK1/2 and p53 is replaced by the transcription factor SOG1, while in yeast orthologs of CHK1/2, Chk1/Rad53, directly activate downstream processes.

Non-homologous end joining

Non-homologous end joining (NHEJ) repairs DSBs through direct rejoining of chromosome ends. It is the major DSB repair pathway in most higher eukaryotes. In contrast to HR it is not dependent on long stretches of homologous sequences and extensive DNA synthesis, and is available throughout the cell cycle (28). End joining can be distinguished into classical NHEJ (C-NHEJ) and alternative end joining (A-EJ). C-NHEJ repairs DSB ends precisely if ligatable, but when DSB ends have non-ligatable termini end processing must be performed. This end processing can result in small deletions or insertions at the break site. A-EJ involves DSB end resection to expose small microhomologies surrounding the DSB. These microhomologies can be utilized for repair, which results in deletions. NHEJ involves three stages: recognition of the DSB-ends, end-processing of the DSB-ends, and ligation of the DSB-ends. The important factors that are involved in these steps for both C-NHEJ and A-EJ will be discussed in more detail.

Classical NHEJ (C-NHEJ)

The factors involved in C-NHEJ act as a multi-protein machine. A core complex is capable of recognizing and aligning broken DNA ends and recruiting processing factors that make broken ends ligatable. Processing factors include end-cleaning enzymes, nucleases and polymerases (29). Processed ends are ligated by a ligase complex. The C-NHEJ core complex consists of the KU70/80 heterodimer and the DNA ligase complex DNA ligase IV (Lig4)-X-ray repair cross-complementing protein 4 (XRCC4)-XRCC4-like factor (XLF, also called Cernunnos) (30). Most C-NHEJ core complex proteins were first identified in yeast and mammals, but plant homologs have been identified for most of these factors.

End binding

C-NHEJ is initiated by the binding of the KU heterodimeric complex KU70/KU80 to the broken DNA ends at the DSB site. The KU heterodimer is a highly abundant complex and has a high affinity for broken DNA ends, and loads onto them in a sequence independent manner by threading them through a central channel formed by the heterodimer (31). Once bound, the KU heterodimeric complex protects the broken DNA ends from resection and aligns the broken ends so that processing factors can act on them. Furthermore, the KU-heterodimeric complex also acts as a scaffold for the recruitment of downstream NHEJ factors, including DNA-dependent protein kinase, catalytic subunit (DNA-PKcs) (32), XRCC4-Lig4 (32, 33), XLF (34),

and DNA polymerases (35). Except DNA-PKcs, which is only present in animals, all these factors are conserved in animals, fungi and plants. The KU heterodimer is conserved in Arabidopsis and KU-deficient plants are hypersensitive to DSB-inducing agents (36–39).

End processing

In an ideal situation, DNA termini with blunt ends or compatible overhangs and that possess a 5' phosphate and a 3' hydroxyl can be directly ligated by the C-NHEJ-specific DNA ligase Lig4. However, naturally occurring DSBs such as those caused by irradiation or reactive oxygen species, or DSBs with hairpin ends, cannot be ligated directly and need end processing. DNA end processing enzymes are specifically targeted to various kinds of non-ligatable ends. Processing activities can be removal of non-ligatable residues, DNA end resection by nucleases, DNA gap filling by polymerases, and addition of 5' phosphate and removal of 3' phosphate groups by polynucleotide kinase/phosphatase (PNKP).

A well-characterized end processing enzyme in humans is Artemis. Artemis is a nuclease with intrinsic 5' exonuclease activity on ssDNA. However, upon activation by and in complex with DNA-PKcs, Artemis gains endonuclease activity on 5' and 3' DNA overhangs (40, 41). This makes Artemis a key factor in the removal of incompatible 5' and 3' overhangs. Furthermore, Artemis is also able to process hairpin ends produced during V(D)J recombination (42). The closest yeast and plant homolog of Artemis is Sensitive to Nitrogen Mustard 1 (ScSNM1/AtSNM1). Studies in yeast have shown that SNM1 has 5'-exonuclease activity (43). In Arabidopsis SNM1 plays a role in NHEJ-mediated repair of oxidatively-induced DSBs and SNM1 over-expression positively enhanced NHEJ, presumably through end processing by its exonuclease activity (44, 45).

Although Artemis is responsible for most DSB end processing, other nucleases may also be involved. One candidate is Aprataxin and PNKP-like factor (APLF), which possesses both endonuclease and exonuclease activity *in vitro* and therefore may act as a DSB end processing factor (46–48). Furthermore, besides its function as a DSB sensor, the MRN-complex has been shown to be involved in end processing by bridging the DNA ends and acting as an endonuclease followed by 3' – 5' exonuclease activity in collaboration with CtBP-interacting protein (CtIP, Sae2 in yeast) (49, 50). The protein encoded by the Arabidopsis gene *COM1* is closely related to CtIP/Sae2 (51, 52). More recently it has been shown that maize *ZmCOM1* also plays a role in mitotic DSB repair (53). Another protein that might function as end

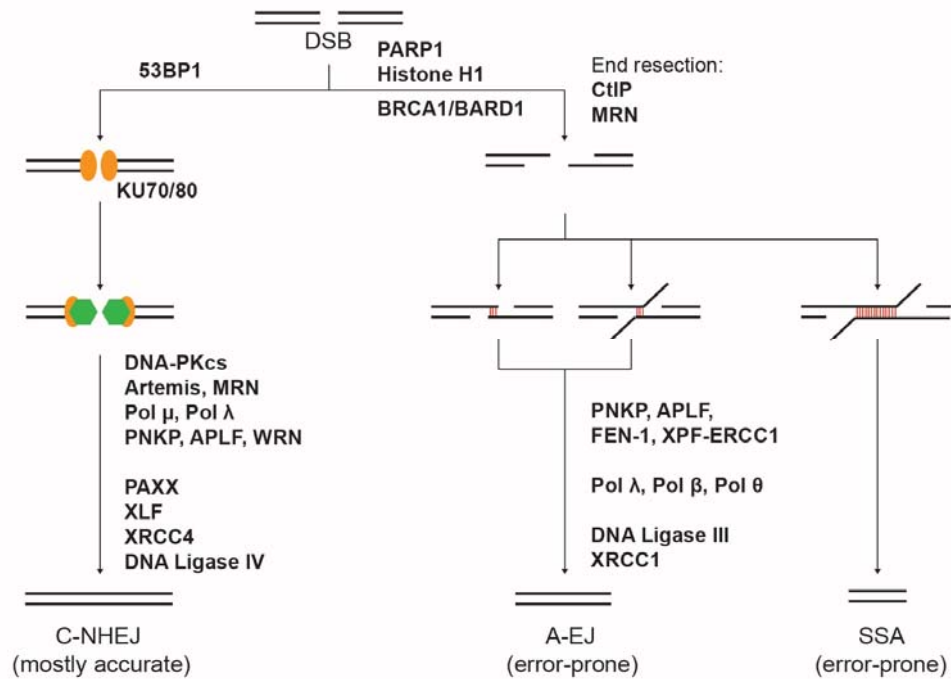


Figure 2. DSB repair by end joining. The C-NHEJ pathway is favored when 53BP1 is counteracting end resection by CtIP/MRN-complex. C-NHEJ is initiated by binding of KU70/80 to the break ends, followed by recruitment and alignment of break ends by DNA-PKcs and recruitment of various end processing factors that make the break suitable for ligation by the DNA Ligase IV complex. C-NHEJ repairs mostly accurately, as end resection is prevented. DSBs can also be repaired by A-EJ. A-EJ can be initiated by binding of PARP proteins to the break ends, stimulated by histone H1. Recruitment of the MRN complex to PARP leads to limited end resection. As a consequence, small microhomologies surrounding the break are uncovered and used for DSB repair, resulting in small or larger deletions depending on the location of the microhomologies. Similar to A-EJ, repair by SSA involves the use of homologies surrounding the DSB, although these homology stretches are usually longer.

processing factor is Werner syndrome ATP-dependent helicase (WRN) (54–56). In Arabidopsis, AtRecQ2 has been identified as a functional ortholog of the helicase part of human WRN, while AtWRNexo is orthologous to the exonuclease domain of human WRN (57, 58). A plant homolog of APLF has not yet been discovered.

Before ligation of broken ends is possible, addition of correct chemical end groups to non-ligatable termini is necessary. In mammals, the enzyme polynucleotide kinase 3'-phosphatase (PNKP) is responsible for this activity. PNKP can both phosphorylate 5'-OH termini and dephosphorylate 3'-phosphate termini. PNKP activity has been shown to be dependent on the presence of XRCC4 and DNA-PKcs, indicating its role in C-NHEJ (59). Furthermore, PNKP has been shown to interact with XRCC4, thereby directly linking end processing to ligation (60). In Arabidopsis the modular protein Zinc finger DNA 3' Phosphoesterase (ZDP) dephosphorylates 3'-phosphate ends at SSBs and DSBs to generate 3'-OH termini. However, this enzyme is devoid of kinase activity at 5'-OH termini (61–63).

End-processing can result in small DNA gaps that must be filled prior to ligation. Four members of the mammalian X-family polymerases that can fulfill this activity during NHEJ are Pol β , Pol λ , Pol μ and terminal deoxynucleotidyl transferase (TdT). TdT is exclusively expressed in early B- and T lymphocytes during V(D)J recombination (64). Pol λ , Pol μ and TdT can interact with Ku, XRCC4 and Ligase IV via their BRCA1 C Terminus (BRCT) domain (35, 65). While TdT activity is template-independent (66), the other three are template dependent but differ in the templates they can act on. In plants, Pol λ seems to be the only representative of the X-family polymerases and has been implicated in the repair of DNA damage induced by reactive oxygen species, UV-B irradiation, high salinity and DNA cross-linking agents, as well as in meiotic recombination (67–71).

Ligation

The final step of NHEJ is the ligation of compatible DNA ends. This step is carried out by Lig4, which interacts via its BRCT domain to XRCC4 (72). This stimulates Lig4 activity and stabilizes the complex (73). In mammals the protein XRCC4-Like Factor (XLF, also called Cernunnos), which shows structural similarities to XRCC4, interacts with the Lig4-XRCC4 complex. XLF stimulates ligation of DNA ends by the complex and also may play an important role in aligning the two DNA ends prior to ligation (74–76). The recently discovered protein paralog of XRCC4 and XLF (PAXX) has been shown to interact with Lig4/XRCC4/XLF and with KU and plays a role in NHEJ, presumably by stabilizing the KU heterodimer at DSB sites

and promoting the assembly of other NHEJ factors (77–79). Arabidopsis orthologs AtLIG4 and AtXRCC4 have been shown to interact via their BRCT domain and AtLIG4 expression is increased in response to γ -irradiation (80). Furthermore, the *Atlig4* mutant is hypersensitive to DNA damaging agents MMS and X-rays (81). All in all these observations indicate a role for the AtLIG4-XRCC4 complex in DSB repair. Plant orthologs of XLF and PAXX have not been discovered.

Alternative end joining (A-EJ)

When C-NHEJ is compromised due to the absence of one or more key factors, end joining is still possible through other end joining pathways. These pathways collectively termed alternative end joining (A-EJ) are independent of the Ku70/80 and XRCC4/Lig4 complexes and typically involve more extensive end resection resulting in larger deletions. Mechanisms of A-EJ are still poorly understood but key factors have been identified, including the breast cancer 1 (BRCA1), MRN complex, CtIP, poly (ADP-ribose) polymerase 1 (PARP1), Pol θ , XRCC1 and DNA Ligase III (Lig3) (2). Similar to C-NHEJ, A-EJ can be subdivided in distinct steps of DSB sensing and binding, end processing and ligation.

A-EJ is initiated by sensing of DSBs by PARP1, which is also involved in SSB repair through recruitment of the XRCC1/Lig3 complex to promote ligation. PARP1 is thought to initiate A-EJ by binding to DSBs and catalyzes the poly(ADP-ribosyl)ation of other factors at DSBs (82). PARP-1 activity is enhanced by histone H1, and together the PARP-1/H1 complex may contribute to the alignment of DSB ends, similar to Ku/DNA-PKcs in C-NHEJ (83). Of the PARP superfamily, PARP2 and PARP3 have also been implicated in DNA repair (84, 85). The Arabidopsis orthologs of PARP1 and PARP2 have been shown to be involved in A-EJ (86).

Poly(ADP-ribosyl)ation at DSBs by PARP1 recruits many components of the A-EJ machinery to the DSB, including BRCA1, BRCA1-associated RING domain protein 1 (BARD1) and the MRN complex, as indicated by reduced MRN retention at DSBs in PARP1 deletion mutants or after treatment with PARP1 inhibitors (87, 88). Binding of poly(ADP-ribose) to BARD1 in turn recruits BRCA1. BRCA1 can stimulate DSB end resection by interacting with the MRN/CtIP complex and by displacing 53BP1 (89). In turn, 53BP1 can suppress MRN/CtIP-mediated end resection to initiate C-NHEJ. Thus, 53BP1 and BRCA1 are important decision factors in the initiation of either C-NHEJ or A-EJ (8). A consequence of MRN/CtIP-mediated end resection is that small microhomologies surrounding the DSB can be uncovered and utilized for DSB repair, which results in smaller or larger deletions, depending

on the distance between the microhomologies (Figure 2). Hence, A-EJ is often associated with microhomology-mediated end joining (MMEJ). MMEJ shares similarities to another DSB repair pathway called single-strand annealing (SSA). However, SSA utilizes longer stretches of homology (Figure 2) (90).

Several end processing enzymes have been linked to A-EJ. X-family polymerases Pol λ and Pol β are likely candidates for end processing in A-EJ. *In vitro* experiments showed that Pol λ and Pol β are involved in gap filling when 3'-ssDNA had terminal microhomology, although with different specificities (66, 91). More recently, the A-family polymerase Pol θ has been shown to play an important role in A-EJ in *Drosophila*, *C. elegans* and mammalian cells (92–95). Pol θ mediates end joining by capturing 3'-ssDNA ends with short microhomologies of 1 – 3 nt and extending one 3' DNA end by using the captured sequence as a template (96, 97). Pol θ -mediated EJ (TMEJ) is also possible when microhomologies lie further away in long 3'-ssDNA overhangs as a result of end resection, thereby creating non-homologous 3' ssDNA tails that need to be removed prior to extension (92, 95–97). A candidate for this removal is flap endonuclease-1 (FEN-1), which was shown to be involved in A-EJ in *Xenopus* and mammalian cells (98, 99). FEN-1 is thought to remove the 3' ssDNA tails through its endonuclease activity. The yeast homolog of FEN-1, Rad27, is involved in NHEJ and interacts with Pol4, the only X-family polymerase in yeast (100, 101). Furthermore, FEN-1 interacts with Pol β and coordinates gap-filling synthesis and 3'-ssDNA tail removal (102). It might be that FEN-1 functions in a similar way in A-EJ in higher eukaryotes through interaction with X-family polymerases. A rice ortholog of FEN-1 has been discovered (103). Other nucleases that might function in this step are the SSA factor xeroderma pigmentosum group F (XPF)-excision repair cross-complementation group 1 (ERCC1) nuclease complex (104), APLF and Artemis-DNA-PKcs (30). The functional homolog of XPF-ERCC1 in yeast is Rad1-10 (105, 106). Homologs of XPF-ERCC1 have also been identified in *Arabidopsis* (106, 131, 132). Yeast and plant homologs of APLF have not yet been discovered.

Addition of a phosphate group to 5'-OH termini by PNKP is an important end processing step. Beside interaction with XRCC4, PNKP also interacts with XRCC1, suggesting it plays a role as 5'-OH kinase in both C-NHEJ and A-EJ (107).

The ligation step of A-EJ is facilitated by the XRCC1/LIG3 complex in mammals. Studies have shown increased XRCC1/Lig3 complex formation in C-NHEJ deficient cells, and association of XRCC1/LIG3 with the MRN complex at DSBs (108). Another mammalian ligase that has been implicated in A-EJ is Ligase I (LIG1), which is also involved in SSB repair and

the BER and NER pathways (109). In the absence of LIG3 and LIG4, cells were still able to perform A-EJ through LIG1 activity, suggesting redundant roles of LIG1 and LIG3 (109, 110). Plants are lacking LIG3 but not LIG1, and a role for LIG1 in A-EJ in plants has been implicated (111).

Table 1. Proteins involved in C-NHEJ and A-EJ in human, budding yeast and Arabidopsis.

<i>Homo sapiens</i>	<i>Saccharomyces cerevisiae</i>	<i>Arabidopsis thaliana</i>	Arabidopsis gene number	Function
53BP1	Rad9	-		DNA damage checkpoint, suppression of MRN/CtIP-mediated end resection
BRCA1	-	BRCA1	At4g21070	Stimulation of MRN/CtIP-mediated end resection
BARD1	-	BARD1	At1g04020	Recruitment of BRCA1 to DSBs
KU70	Ku70	KU70	At1g16970	DSB end binding, alignment
KU80	Ku80	KU80	At1g48050	DSB end binding, alignment
DNA-PKcs	-	-		Protein kinase, alignment
Artemis	Smn1/Pso2	SNM1	At3g26680	DSB end processing
CtIP	Sae2	COM1	At3g52115	DSB end resection
PNKP	Tpp1	ZDP	At3g14890	DSB end processing
APLF	-	-		DSB end processing
WRN	-	WRNexo, RECQ2	At4g13870, At1g31360	DSB end processing
Pol β	-	-		DNA Polymerase, DSB end processing
Pol λ	Pol4	Pol λ	At1g10520	DNA Polymerase, DSB end processing
Pol μ	-	-		DNA Polymerase, DSB end processing
TdT	-	-		DNA Polymerase, DSB end processing
XRCC4	Lif1	XRCC4	At3g23100	Ligation, complex with LIG4
XLF/Cernunnos	Nej1	-		Ligation, complex with XRCC4/LIG4
PAXX	-	-		Ligation, complex with XRCC4/XLF/LIG4, KU-binding
DNA ligase IV (LIG4)	Dnl4	LIG4	At5g57160	ATP-dependent DNA ligase
PARP 1	-	PARP 1	At2g31320	DNA end binding, NAD+ADP-ribosyltransferase
PARP 2	-	PARP 2	At4g02390	DNA end binding, NAD+ADP-ribosyltransferase
PARP 3	-	PARP 3	At5g22470	DNA end binding, NAD+ADP-ribosyltransferase
Histone H1	Histone H1	Histone H1.1, 1.2, 1.3	At1g06760, At2g30620, At2g18050	DSB end binding, alignment
MRE11-RAD50-NBS1 (MRN complex)	Mre11-Rad50-Xrs2 (MRX complex)	MRE11-RAD50-NBS1 (MRN complex)	At5g54260-At2g31970-At3g02680	DSB end processing, DNA damage checkpoints
XPF-ERCC1	Rad1-Rad10	XPF-ERCC1	At5g41150-At3g05210	DSB end processing
FEN-1	Rad27	FEN1 or SAV6	At5g26680	DSB end processing
Pol Q	-	Pol θ (Tebichi)	At4g32700	DNA polymerase, DSB end processing
XRCC1	-	XRCC1	At1g80420	Ligation, complex with LIG1/LIG3
DNA ligase I (LIG1)		LIG1	At1g08130	ATP-dependent DNA ligase
DNA ligase III (LIG3)	-	-		ATP-dependent DNA ligase

Homologous Recombination

Homologous recombination (HR) is a highly conserved mechanism that utilizes DNA homology to direct DSB repair. This homology can be provided in the form of a sister chromatid available during the S/G2 phase of the cell cycle, or a homologous chromosome during meiosis (112, 113). Because of the use of homology to guide DSB repair HR is considered a more precise repair mechanism than NHEJ and indeed usually results in accurate repair. HR can be subdivided into four steps: 1) DNA end resection and nucleoprotein filament formation, 2) DNA homolog search and strand invasion, 3) DNA heteroduplex extension, and 4) resolution/dissolution of the D-loop (63). For each factor described below, the budding yeast homolog, together with known human and plant homologs will be mentioned.

HR events are initiated by 5' – 3' resection at the DSB ends to produce 3' single-strand overhangs (Figure 3). This is initially performed by the MRX-complex (MRN-complex in human and plants) in cooperation with Sae2 (CtIP in humans and COM1 plants) and BARD1/BRCA1, similar to end resection during A-EJ (89). This initial end resection removes 50 – 100 nucleotides from the 5' end and is then taken over by two parallel pathways, one involving exonuclease 1 (Exo1) and the other involving the Sgs1-Top3-Rmi1 (STR) helicase complex in collaboration with Dna2, for more extensive resection spanning a several kbp-long region (114). In humans and plants, the functional homologs for the STR complex are BLM-TOPOIII α -RMI1-RMI2 and RECQ4A-TOP3A-RMI1-RMI2, respectively (115–117). A plant homolog of Dna2, named JHS1, was recently identified by Jia *et al.* (118).

The 3'-ssDNA tails are coated with RPA, which consists of three subunits (RPA1-3) that are encoded by multiple paralogs in plants (*RPA1A-E*, *RPA2A/B*, *RPA3A/B*) (119). In yeast RPA is replaced by recombination factor Rad51 by mediation of Rad52, to form a right-handed helical polymer consisting of a nucleoprotein filament known as the presynaptic filament (Figure 3). Although Rad52 is conserved in mammals and plants, the BRCA2 protein functions as the mediator in these organisms (120). Each Rad51 filament covers 18-19 DNA bases and ~6 protein monomers per helical turn of the polymer (121). Rad51 paralogs Rad55 and Rad57 are thought to stabilize the filament (121). In humans and plants, any of the Rad51 paralogs RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3 may perform this function (115). Complexes formed by these paralogs can act in somatic recombination (RAD51B-RAD51C-RAD51D-XRCC2) and meiotic recombination (RAD51C-XRCC3) (122). After assembly of the presynaptic filament strand, it may invade duplex DNA molecules to search for homology, thereby creating a displacement loop (D-loop) (Figure 3). Once homology is found,

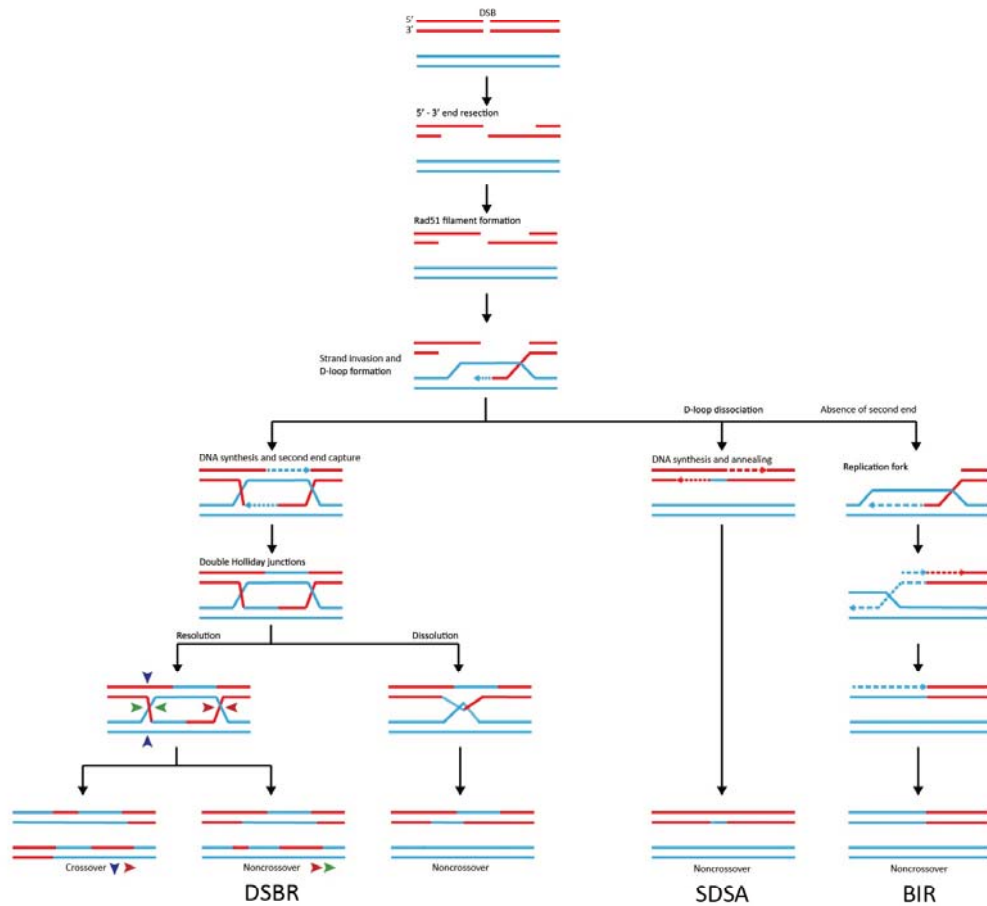


Figure 3. HR-mediated DSB repair. When a DSB occurs, it can be repaired by several HR subpathways, including canonical double-strand break repair (DSBR), synthesis-dependent strand annealing (SDSA) and break-induced replication (BIR). A DSB is first processed by 5' – 3' end resection to form ssDNA overhangs. Generation of these overhangs are crucial for Rad51 filament formation, which in turn aid in homology search and strand invasion, resulting in a D-loop. In the DSBR subpathway, the second end is captured and gaps are filled through DNA synthesis. This results in double Holliday junctions (dHJs), which can either be dissolved into a noncrossover product, or resolved into either crossover or noncrossover products. A DSB can also be repaired through the SDSA pathway when the D-loop dissociates. Gaps are filled through DNA synthesis and annealed, resulting in a noncrossover product. When a second end is absent, the BIR pathway can be used by forming a replication fork and synthesizing the strand using the sister chromatid as a template.

homologous pairing by Rad51 is stimulated by Rad54, a member of the Swi2/Snf2 chromatin remodeling/ATPase protein enzyme family. Rad54 can displace Rad51 *in vitro* in an ATP-dependent manner, and is thought to facilitate the recycling of Rad51 or expose the 3' end of paired DNA to initiate DNA synthesis (121). This DNA synthesis is thought to be performed by polymerase δ together with the required proliferative cell nuclear antigen (PCNA) using the paired strand as a template strand (115, 123, 124). In addition to polymerase δ , polymerases η , κ and ζ have been shown to be involved in DNA synthesis during D-loop extension, although they generate shorter extension products than polymerase δ (124).

From the moment of D-loop formation and –extension, progression into one of three HR sub-pathways is possible. The pathway known as double-strand break repair (DSBR) is the canonical HR pathway. In DSBR DNA synthesis eventually leads to the capture of the second end. Ligation of the 5' end of the break with the 3' of the extended displaced strand, then leads to formation of a double Holliday junction (dHJ) intermediate (Figure 3) (115, 125, 126). The dHJ intermediates can either be dissolved or resolved into two separate DNA duplexes. Dissolution of the dHJ requires the helicase activity of the Sgs1-Top3-Rmi1 complex to move the Holliday junctions close together and its topoisomerase activity to separate the interlinked strands, resulting in non-crossover (NCO) products (115, 127). This is the predominant pathway in somatic cells. The dHJ intermediate can also be resolved by nucleolytic cleavage of the Holliday junctions, leading to crossover (CO) products when the inner strands of one junction and the outer strands of the other junction are cleaved, or NCO products when the inner strands of both junctions are cleaved (115). Several resolvases have been identified that facilitate Holliday junction resolution in yeast, including Mus81-Mms4, Slx1-Slx4, Yen1, and Rad1-Rad10 (MUS81-EME1, SLX1-SLX4, GEN1, and XPF-ERCC1 in human) (115, 128). Two EME1 homologs have been identified in Arabidopsis, EME1A and EME1B. Both are able to form complexes with Arabidopsis AtMUS81 (129). Two paralogs of GEN1, named AtGEN1 and SEND1, have been identified in Arabidopsis (130), as well as homologs of XPF-ERCC1 (106, 131, 132).

COs generated by the DSBR pathway are formed during meiotic recombination. However, COs are not favorable in mitotic cells, as they can generate large genomic rearrangements (133). Therefore, the resolvases involved in DSBR become active only late in the cell cycle (134). To prevent the generation of COs during mitotic and meiotic recombination, the DSBR pathway can be suppressed by D-loop dissociation by the Mph1 helicase (FANCM in human and Arabidopsis). Annealing the displaced extended strand to the

Table 2. Proteins involved in mitotic HR in human, budding yeast and Arabidopsis.

<i>Homo sapiens</i>	<i>Saccharomyces cerevisiae</i>	<i>Arabidopsis thaliana</i>	Arabidopsis gene number	Function
MRE11-RAD50-NBS1 (MRN-complex)	Mre11-Rad50-Xrs2 (MRX-complex)	MRE11-RAD50-NBS1 (MRN-Complex)	At5g54260-At2g31970-At3g02680	DSB end resection, DNA damage checkpoints
CtIP	Sae2	COM1	At3g52115	DSB end resection
BRCA1	-	BRCA1	At4g21070	Interaction with MRN/CtIP, stimulation of DSB end resection
BARD1	-	BARD1	At1g04020	Recruitment of BRCA1 to DSBs
EXO1	Exo1	EXO1A, EXO1B	At1g29630, At1g18090	DSB end processing, end resection
BLM-TOPOIII α -RMI1-RMI2	Sgs1-Top3-Rmi1 (STR complex)	RECQ4A, TOP3A, RMI1-RMI2 (RTR complex)	At1g10930, At5g63920, At5g63540, At1g08390	DSB end processing, end resection, HJ dissolution
DNA2	Dna2	JHS1	At1g08840	DSB end processing, end resection
RPA1, RPA2, RPA3	RPA1, RPA2, RPA3	RPA1A-E RPA2A, B RPA3A, B	At2g06510 (<i>RPA1A</i>), At5g08020 (<i>RPA1B</i>), At5g45400 (<i>RPA1C</i>), At5g61000 (<i>RPA1D</i>) At4g19130 (<i>RPA1E</i>); At2g24490 (<i>RPA2A</i>), At3g02920 (<i>RPA2B</i>); At3g52630 (<i>RPA3A</i>), At4g18590 (<i>RPA3B</i>)	ssDNA binding, initiation of presynaptic filament formation, recruitment of RAD51
RAD51, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3	Rad51, Rad55, Rad57	RAD51, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3	At5g20850, At2g28560, At2g45280, At1g07745, At5g64520, At5g57450	ssDNA binding, presynaptic filament formation/stabilization, recombination mediator
RAD52	Rad52	RAD52-1, RAD52-2	At1g71310, At5g47870	RAD51/RPA interactor, ssDNA binding, presynaptic filament, recombination mediator, second-end capture
BRCA2	-	BRCA2A, BRCA2B	At4g00020, At5g01630	RAD51 interactor, ssDNA binding, presynaptic filament, recombination mediator
RAD54, RAD54B	Rad54	RAD54	At3g19210	Homologous pairing stimulation, D-loop formation
Pol δ	Pol δ	POLD2, POLD3, POLD4	At2g42120, At1g78650, At1g09815	DNA synthesis, D-loop extension
Pol η	Pol η	POL2	At1g08260, At2g27120	DNA synthesis, D-loop extension
PCNA	PCNA	PCNA1, PCNA2	At1g07370, At2g29570	DNA synthesis, D-loop extension
MUS81-EME1	Mus81-Mms4	MUS81-EME1A-EME1B	At4g30870, At2g21800, At2g22140	Endonuclease, part of resolvase complex, HJ resolution
SLX1-SLX4	Slx1-Slx4	-	-	Endonuclease, part of resolvase complex, HJ resolution
GEN1	Yen1	GEN1, SEND1	At1g01880, At3g48900	Endonuclease, part of resolvase complex, HJ resolution
XPF-ERCC1	Rad1-Rad10	XPF-ERCC1	At5g41150, At3g05210	Endonuclease, part of resolvase complex, HJ resolution
FANCM	Mph1	FANCM	At1g35530	Helicase activity, D-loop dissociation

complementary strand at the other side of the break is followed by DNA synthesis by the synthesis-dependent strand annealing (SDSA) pathway. This results exclusively in NCO products (115) (Figure 3). Sometimes, a second end is not present and only one end of the DSB shares homology with other genomic sequences. In this case, cells rely on the break-induced replication (BIR) sub-pathway (Figure 3). During BIR, the broken chromosome is restored by strand invasion of the single broken end and copying of the homologous chromosome template, resulting in a NCO product. This mechanism may rescue stalled or broken replication forks (135).

Homologous recombination and gene targeting

Gene targeting (GT) is a molecular genetic technique to change or replace an endogenous gene by HR. A gene of interest can be changed or replaced by providing the cell with a repair template, consisting of the DNA sequence of interest flanked by sequences homologous to the target gene, which are used for HR-mediated integration. A true GT event is characterized by HR-mediated recombination reactions with both homology arms independently, according to the SDSA model of HR. However, homologous recombination may take place only at one of the homology arms, while the other end integrates by an end-joining process (134). Also so called ectopic GT events can also occur, whereby the repair template is initially extended at one side via a gene conversion mechanism, after which the whole repair template is integrated elsewhere in the genome (136).

GT has the potential to induce subtle changes, completely knock out target genes or insert foreign DNA sequences at preselected sites. The significant advantage of specifically targeting DNA to change a sequence of interest is that accidental change of endogenous genes by random integration events, or variation in expression patterns of the transgene itself due to position effects, are avoided. Besides its applicability for fundamental research, GT is a promising tool for plant biotechnology. With GT, desired crop traits can be produced much more rapidly and precisely than with conventional breeding methods. GT is highly efficient in budding yeast, and more than 30 years ago pioneering research showed that it was possible to efficiently generate loss-of-function mutations with GT in mouse embryonic stem cells (137, 138). However, GT in other animal cells and plants remained challenging for a long time.

In plants, it was shown that a non-functional kanamycin resistance gene could be restored via HR at low frequency by direct DNA transformation of tobacco protoplasts with plasmids harbouring the missing part of the kanamycin gene (139). Two years later, molecular

evidence showing true GT was provided in tobacco plant cells in which a defective kanamycin resistance gene had been repaired through *Agrobacterium*-mediated T-DNA transformation (140). Although these and subsequent studies showed that GT in plants was possible, the low GT frequencies ranging from one GT event per $10^4 - 10^6$ transformation events made GT not yet feasible as a tool for agricultural biotechnology (136, 141–145). There are two main reasons why high frequency GT is still challenging in plants. Firstly, most DSBs in somatic cells of higher eukaryotes are repaired by NHEJ, while GT relies on HR-mediated repair. Thus, the GT repair template will mostly integrate randomly via NHEJ, instead of targeted via HR. Secondly, GT frequency relies on the transformation frequency of the GT repair template. When the transformation frequency is low, GT frequency will be low.

Over the years, several approaches have been used to improve the GT frequency in plants, including manipulation of the NHEJ and HR machinery, positive-negative selection of GT events, the use of artificial sequence-specific nucleases (SSNs) for targeted DSB induction and release of an integrated GT repair template from the genome, and amplification of the GT repair template (146, 147). Manipulation of the DSB repair machinery in favor of GT was attempted in different ways. In the first attempts, enzymes of the HR machinery from organisms in which GT is efficient were expressed in plants. Tobacco plants expressing the RecA strand exchange protein from *E. coli* showed increased rates of intrachromosomal HR. However, this did not significantly improve GT frequency (148). In another study, the RAD54 gene from budding yeast was overexpressed in *Arabidopsis*, which was reported to result in an enhancement of GT frequency in ScRAD54 overexpression plant lines (149). In a follow-up study, egg cell-specific ScRAD54 expression resulted in a 10-fold enhanced GT frequency compared to wild-type (150). In a more recent study it was reported that a combination of *Arabidopsis* RAD52 (AtRAD52) overexpression and RNA interference (RNAi) had a moderate positive effect on GT frequency (151). In other approaches, the intrinsic DSB repair machinery was altered to increase GT by suppressing NHEJ. This approach was successful in yeasts and fungi, where blocking NHEJ by deletion of Ku or Lig4 enhanced GT frequency substantially (152–154). However, deletion of *KU70*, *KU80*, or *LIG4* did not result in an enhancement of GT (39). An enhancement in GT frequency was however reported for rice *LIG4* mutants (155). The presence of multiple independent end-joining pathways in plants is the likely reason that full suppression of end-joining in favor of efficient HR-mediated GT is still challenging (156). A promising recent discovery that might open the door to improved GT in plants is that the disruption of *TEBICHI*, the *Arabidopsis* homolog of Pol θ , abolishes random end-joining-

mediated T-DNA integrations below detection levels (157). This discovery may lead to the development of more efficient GT methods in which random integration events can be avoided.

Positive-negative selection (PNS) is another approach aiming to improve GT efficiency in plants. In this approach, the selection of GT events relies on a positive selectable marker placed on the GT repair template, and negative selectable markers placed outside the homology arms. In this way, GT events can be selected for using the positive selectable marker, and random integrations of the GT repair template T-DNA can be selected against using the negative selectable markers, thereby enriching GT events in the total number of transformations. Because PNS does not rely on the selection of specific phenotypes, it can be applied for any gene of interest. PNS has been successfully applied for GT at various endogenous genes in rice (for review, see Endo and Toki 2014). However, in *Arabidopsis* PNS has not led to improved GT efficiency, possibly because GT events are often accompanied by random integration events.

Targeted DSB-induction

Another approach to improve GT frequency is the induction of a DSB at the gene of interest. It was already known that DSBs induced by naturally occurring nucleases can induce recombination events, such as those that occur during meiotic recombination or mating type switching (159–162). In pioneering experiments in plants, it was shown that HR and GT were enhanced by inducing a targeted DSB with the meganuclease *I-SceI* at artificial targets (163–166) (Figure 4). Although these experiments showed the proof of principle, meganucleases were very difficult to adjust for targeting of natural genes, as the DNA recognition site and nuclease catalytic centre overlap. However, researchers nowadays have several artificial sequence-specific nucleases (SSNs) at their disposal to induce DSBs at almost any genomic locus of interest. These nucleases can be designed in such a way that their DNA recognition site is of sufficient length that it occurs only at the target of interest and nowhere else in the genome. Three classes of SSNs have been used for both NHEJ-mediated targeted mutagenesis and GT: zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and those based on the CRISPR/Cas9 (“clustered regularly interspaced short palindromic repeats/CRISPR-associated”) system (Figure 4).

ZFNs consist of arrays of zinc-finger (ZF) DNA binding domains fused to the nuclease domain of the type II restriction endonuclease *FokI*, which cleaves DNA as a dimer producing a break with a 5' overhang (167, 168). Obligate *FokI* heterodimers have been developed to

prevent homodimerization at the target and increase specificity and efficiency (169). Each ZF domain can recognize a specific DNA-triplet, and ZF domains can be combined in such a way that a specific sequence of interest can be recognized. For example, a 3-finger ZFN can recognize a sequence of 9 bp and functioning as a heterodimer it can recognize a total of 18 bp, a sequence length which is usually unique in the genomes of higher eukaryotes. Several studies, including from our lab, have shown that targeted DSB induction by ZFNs can enhance GT frequency substantially in several plant species (170–175). However, nowadays ZFNs are used less often, mainly due to the laborious design process. Furthermore, it can be difficult to predict efficient ZFN targets, due to the fact that individual ZFN domains can influence each other (176).

Similar to ZFNs, TALENs consist of DNA binding domains fused to the *FokI* nuclease domain (177). The DNA binding domains are derived from plant promoter-binding proteins produced by the plant pathogen *Xanthomonas sp.* and consist of an array of 13 – 28 repeats, each consisting of 34 highly conserved amino acids. The amino acid residues at position 12 and 13, also known as repeat variable diresidues (RVDs), determine the binding to one of the four DNA bases in a one to one correspondence (177–179). In this way, TALENs can be designed to target almost any genomic sequence, provided that the binding sequence of each TALE is preceded by a 5' T (178, 179). With publicly available methods such as the Golden Gate cloning, TALENs can be designed and assembled efficiently (180, 181). Several studies showed that TALENs can be used successfully for GT in plants (182, 183).

The most recent SSN class is the CRISPR/Cas system (184). It is derived from an adaptive immune system present in bacteria and archaea, where it targets the degradation of foreign viral or plasmid DNA. Site-specific binding of CRISPR/Cas9 requires the binding of two short RNA molecules: the CRISPR RNA (crRNA) and the trans-activating RNA (tracrRNA). The tracrRNA ensures recruitment of the Cas9 nuclease, while the crRNA ensures binding of the Cas9-RNA complex to a defined DNA sequence of 20 nt called the protospacer. A protospacer must be flanked on the 3' by a protospacer adjacent motif (PAM), which interacts with the Cas9 PAM-interacting (PI) domain (185, 186). The Cas9-tracrRNA-crRNA complex finds the DNA target by first probing for a proper PAM, after which the flanking sequence is investigated for potential crRNA-DNA complementarity. When the correct protospacer and PAM is found, Cas9 triggers melting of the local dsDNA, followed by invasion of the crRNA strand and RNA-DNA base pairing (187). The CRISPR/Cas9 system was made

simpler by fusing the crRNA and tracrRNA into a chimeric single-guide RNA (sgRNA), while retaining the functions of both RNAs (184).

The first Cas9 orthologue that was used as a nuclease for genome engineering was Cas9 isolated from *Streptococcus pyogenes* (SpCas9). Cas9 operates via its HNH and RuvC-like endonuclease domains, which cleave both DNA strands to induce a DSB 3–4 bp from the NGG PAM sequence. It was assumed that Cas9 induces a blunt end DSB. However, recent evidence suggests that Cas9 induces a staggered DSB with a 1 bp overhang (188). An advantage of the CRISPR/Cas9 system is that, because of the direct RNA-DNA recognition, new sgRNAs can easily be designed to target new sequences. So far, CRISPR/Cas9 has mostly been used for targeted mutagenesis in plants (for a review, see 180). Beside targeted mutagenesis, several studies also showed improved HR-mediated repair and GT using CRISPR/Cas9 (190–194).

Beside the Cas9 nuclease, nickases are also available in which either the HNH or RuvC-like endonuclease domain has been inactivated. DSBs with long or short 5' or 3' overhangs can be generated by using nickases in pairs, each one inducing a single strand break on either the same or on the complementary strand (195, 196). Experiments in plants showed that the Cas9 nickase was able to induce HR-mediated repair at a higher rate than the Cas9 nuclease (193, 197). Furthermore, a study in human cells showed that 5' overhang DSBs produced by paired Cas9 nickases are more often repaired by HR than blunt end DSBs induced by the Cas9 nuclease, suggesting that paired nickases are a more attractive option for GT (198).

More recently, Cas9 orthologues and other Cas variants were isolated that expand the possibilities for Cas-mediated genome engineering. The Cas9 orthologues SaCas9 and StCas9, isolated from *Staphylococcus aureus* and *Streptococcus thermophilus*, respectively, use different PAMs than SpCas9, which enables targeting of a wider variety of sequences (197, 199–201). The Cas variant Cas12a, formerly known as Cpf1, induces a DSB with a staggered 5' 5-bp overhang, instead of the blunt cut of Cas9 (202). Several Cas12a orthologues have shown robust mutagenesis frequencies in human cells and plants (for a review, see 200). Beside efficient targeted mutagenesis, the staggered DSB with complementary overhangs that Cas12a induces might work more efficiently for targeted integrations than the blunt DSB of Cas9 by shifting the repair mode from NHEJ to HR, similar to repair of a 5' overhang DSB produced by paired nickases. The first studies using Cas12a in rice and algae showed that rates of HR-mediated DSB repair and GT are indeed enhanced compared to Cas9 (204, 205).

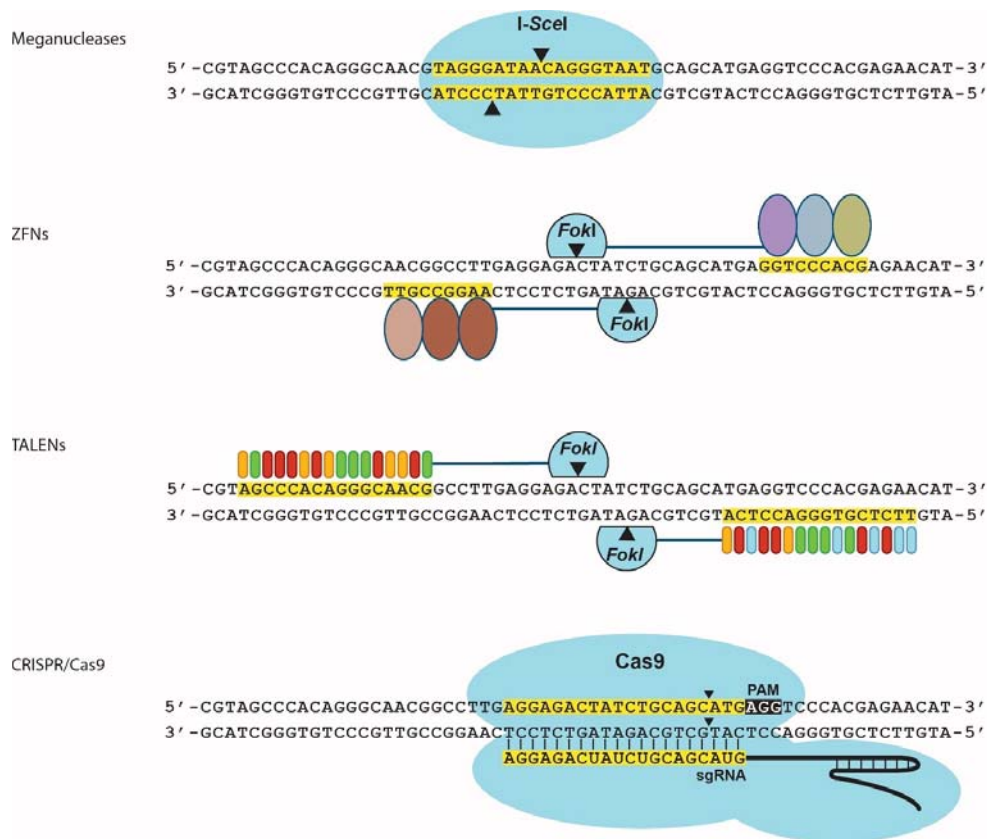


Figure 4. SSNs used for targeted mutagenesis and GT: meganucleases, ZFNs, TALENs and CRISPR/Cas9. The recognition sequences of the DNA binding domains for the meganuclease I-SceI, ZFNs and TALENs, and the CRISPR/Cas9 protospacer are shown in yellow. The PAM is shown in black. The nuclease cutting sites are indicated by black triangles.

***In planta* GT**

In classical GT experiments in plants, recombination events mostly take place during the transformation process when the repair template T-DNA becomes integrated in the genome. This makes GT dependent on T-DNA transformation efficiency, and very challenging for plant species that are difficult to transform. A GT method called *in planta* GT has been developed to overcome this issue (193, 194, 206). *In planta* GT is based on stable ectopic integration of the repair template T-DNA, on which the repair template is flanked by SSN sequences. SSN expression will simultaneously induce a break at the target gene and release a linearized repair template from the genome, which can be used as a substrate for GT. Because this can potentially happen in every transformed cell independently, GT events may take place during the plant's lifetime and events that happen in the germline will be transmitted to the next generation. Recently, our group showed that *in planta* GT is possible at low frequencies without excision of the repair template, suggesting that repair template excision is not a prerequisite for GT (191).

A limiting factor for efficient *in planta* GT may be that only one copy of the ectopically integrated repair template is present in each cell. To overcome this issue, researchers have equipped the repair template T-DNA with sequences derived from gemini viruses, which will lead to rolling circle replication upon expression of replication protein Rep (175). This system was successfully applied to enhance GT in tobacco, wheat and tomato (207–209).

Outline of the thesis

Although GT frequencies in plants have improved since the first experiments, there is still room for improvement. Especially for use as a genome engineering tool in various crops, GT frequencies need to be enhanced to feasible levels. The two main roadblocks that need to be overcome are the low rate of HR in somatic cells compared to the rate of NHEJ, and the low transformation efficiencies in plants that impact GT frequency. The first objective of this project was to elucidate whether GT efficiency would increase by performing *in planta* GT in meiocytes, cells that already have a higher rate of HR. The second objective was to find new *Arabidopsis* mutants with increased GT frequencies and to identify genes involved in this phenotype. In the end this may lead to a better understanding of the mechanisms underlying GT and these may be used to realize higher GT frequencies in plants.

The first step in the development of our *in planta* GT system was constructing and selecting SSNs that could efficiently induce DSBs. In **chapter 2**, the targeted mutagenesis

efficiencies of TALENs and CRISPR/Cas9 were analyzed at the *Arabidopsis cruciferin 3* (*CRU3*) and *alcohol dehydrogenase 1* (*ADH1*) genes. Although both classes of nucleases were able to induce DSBs, overall the CRISPR/Cas9 system performed better. CRISPR/Cas9-mediated targeted mutagenesis was especially efficient when a protospacer containing GG at the 3' end was used.

In **chapter 3** DSB repair in *Arabidopsis* NHEJ mutants was studied in order to get more insight into the DSB repair mechanisms in plants and how they possibly could be further exploited for genome engineering. Targeted mutagenesis using CRISPR/Cas9 was performed in a set of *Arabidopsis* single, double and triple EJ mutants.

Chapter 4 describes the development of an *in planta* GT system for the *CRU3* gene. The CRISPR/Cas9 system described in chapter 2 was used to induce DSBs at both the *CRU3* target gene and at sequences flanking a pre-inserted GT repair template, which will lead to an excised repair template that can be freely used in each transformed cell. In order to direct GT to more HR-efficient cells such as meiocytes, Cas9 was expressed under various promoters that are active in germline and meiotic tissues.

In **chapter 5** attempts to generate and select new *Arabidopsis* phenotypes with higher GT frequencies are described. New GT phenotypes were generated and selected for by a method called “genome interrogation”, which employs zinc-finger artificial transcription factors (ZF-ATFs) to generate genome-wide alterations in gene expression levels.

References

1. Goodarzi AA, Jeggo PA (2013) The repair and signaling responses to DNA double-strand breaks. *Advances in Genetics* 82:1–45.
2. Frit P, Barboule N, Yuan Y, Gomez D, Calsou P (2014) Alternative end-joining pathway(s): Bricolage at DNA breaks. *DNA Repair* 17:81–97.
3. Bétermier M, Bertrand P, Lopez BS (2014) Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* 10(1):e1004086.
4. Mao Z, et al. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* 7(18):2902–2906.
5. Hu Z, Cools T, De Veylder L (2016) Mechanisms used by plants to cope with DNA damage. *Annu Rev Plant Biol* 67(1):439–462.
6. Falck J, Coates J, Jackson SP (2005) Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* 434:605–611.
7. Matsuoka S, et al. (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316:1160–1167.
8. Yoshiyama K, Sakaguchi K, Kimura S (2013) DNA damage response in plants: conserved and variable response compared to animals. *Biology* 2(4):1338–1356.
9. Rupnik A, Lowndes NF, Grenon M (2010) MRN and the race to the break. *Chromosoma* 119(2):115–135.
10. Amiard S, et al. (2010) Distinct roles of the ATR kinase and the Mre11-Rad50-Nbs1 complex in the maintenance of chromosomal stability in *Arabidopsis*. *Plant Cell* 22(9):3020–3033.
11. Fernandez-Capetillo O, Lee A, Nussenzweig M, Nussenzweig A (2004) H2AX: The histone guardian of the genome. *DNA Repair* 3(8–9):959–967.
12. Stucki M, Jackson SP (2006) γ H2AX and MDC1: Anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair* 5(5):534–543.
13. Sirbu BM, Cortez D (2013) DNA damage response: three levels of DNA repair regulation. *Cold Spring Harb Perspect Biol* 5(8):a012724–a012724.
14. Zou L, Elledge SJ (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300(5625):1542–1548.
15. Banin S, et al. (1998) Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* 281(5383):1674–1677.
16. Liu Q, et al. (2000) Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint. *Genes Dev* 14(12):1448–1459.
17. Matsuoka S, et al. (2000) Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc Natl Acad Sci U S A* 97(19):10389–94.
18. Chehab NH, Malikzay A, Appel M, Halazonetis TD (2000) Chk2/hCds1 functions as a DNA damage checkpoint in G₂ by stabilizing p53. *Genes Dev* 14(3):278–288.

19. Shieh SY, Ahn J, Tamai K, Taya Y, Prives C (2000) The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate, p53 at multiple DNA damage-inducible sites. *Genes Dev* 14(3):289–300.
20. Stracker TH, Usui T, Petrini JHJ (2009) Taking the time to make important decisions: The checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. *DNA Repair* 8(9):1047–1054.
21. Craven RJ, Greenwell PW, Dominska M, Petes TD (2002) Regulation of genome stability by TEL1 and MEC1, yeast homologs of the mammalian ATM and ATR genes. *Genetics* 161(2):493–507.
22. Wahl GM, Carr AM (2001) The evolution of diverse biological responses to DNA damage: Insights from yeast and p53. *Nat Cell Biol* 3(12):277–286.
23. Garcia V, et al. (2003) AtATM is essential for meiosis and the somatic response to DNA damage in plants. *Plant Cell Online* 15(1):119–132.
24. Friesner JD, Liu B, Culligan K, Britt AB (2005) Ionizing radiation-dependent γ -H2AX focus formation requires Ataxia Telangiectasia Mutated and Ataxia Telangiectasia Mutated and Rad3-related. *Mol Biol Cell* 16(1):2566–2576.
25. Yoshiyama KO, et al. (2013) ATM-mediated phosphorylation of SOG1 is essential for the DNA damage response in Arabidopsis. *EMBO Rep* 14(9):817–22.
26. Johnson RA, et al. (2018) SUPPRESSOR OF GAMMA RESPONSE 1 links DNA damage response to organ regeneration. *Plant Physiol* 176:1665–1675.
27. Ogita N, et al. (2018) Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in Arabidopsis. *Plant J* 94(3):439–453.
28. Chapman JR, Taylor MRG, Boulton SJ (2012) Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* 47(4):497–510.
29. Waters C a., Strande NT, Wyatt DW, Pryor JM, Ramsden D a. (2014) Nonhomologous end joining: A good solution for bad ends. *DNA Repair* 17:39–51.
30. Chang HHY, Pannunzio NR, Adachi N, Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 18(8):495–506.
31. Walker JR, Corpina RA, Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412:607–614.
32. Calsou P, Delteil C, Frit P, Drouet J, Salles B (2003) Coordinated assembly of Ku and p460 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment. *J Mol Biol* 326(1):93–103.
33. Nick McElhinny SA, Snowden CM, McCarville J, Ramsden DA (2000) Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* 20(9):2996–3003.
34. Yano K, et al. (2008) Ku recruits XLF to DNA double-strand breaks. *EMBO Rep* 9(1):91–6.
35. Mahajan KN, Nick McElhinny S a, Mitchell BS, Ramsden D a (2002) Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* 22(14):5194–202.
36. Bundock P, van Attikum H, Hooykaas P (2002) Increased telomere length and hypersensitivity to DNA damaging agents in an Arabidopsis KU70 mutant. *Nucleic Acids Res* 30(15):3395–3400.

37. Tamura K, Adachi Y, Chiba K, Oguchi K, Takahashi H (2002) Identification of Ku70 and Ku80 homologues in *Arabidopsis thaliana*: Evidence for a role in the repair of DNA double-strand breaks. *Plant J* 29(6):771–781.
38. Gallego ME, Bleuyard JY, Daoudal-Cotterell S, Jallut N, White CI (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in *Arabidopsis*. *Plant J* 35(5):557–565.
39. Jia Q, Bundock P, Hooykaas PJJ, de Pater S (2012) *Agrobacterium tumefaciens* T-DNA integration and gene targeting in *Arabidopsis thaliana* non-homologous end-joining mutants. *J Bot* 2012:1–13.
40. Goodarzi AA, et al. (2006) DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J* 25(16):3880–9.
41. Gu J, et al. (2010) DNA-PKcs regulates a single-stranded DNA endonuclease activity of Artemis. *DNA Repair* 9(4):429–437.
42. Ma Y, Pannicke U, Schwarz K, Lieber MR (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108(6):781–794.
43. Li X, Hejna J, Moses RE (2005) The yeast Snm1 protein is a DNA 5'-exonuclease. *DNA Repair* 4(2):163–170.
44. Molinier J, Stamm M-E, Hohn B (2004) SNM-dependent recombinational repair of oxidatively induced DNA damage in *Arabidopsis thaliana*. *EMBO Rep* 5(10):994–9.
45. Johnson RA, Hellens RP, Love DR (2011) A transient assay for recombination demonstrates that *Arabidopsis* SNM1 and XRCC3 enhance non-homologous recombination. *Genet Mol Res* 10(3):2104–2132.
46. Kanno S, et al. (2007) A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses. *EMBO J* 26(8):2094–2103.
47. Li S, et al. (2011) Polynucleotide kinase and aprataxin-like forkhead-associated protein (PALF) acts as both a single-stranded DNA endonuclease and a single-stranded DNA 3' exonuclease and can participate in DNA end joining in a biochemical system. *J Biol Chem* 286(42):36368–36377.
48. Grundy GJ, et al. (2013) APLF promotes the assembly and activity of non-homologous end joining protein complexes. *EMBO J* 32(1):112–25.
49. Lengsfeld BM, Rattray AJ, Bhaskara V, Ghirlando R, Paull TT (2007) Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 Complex. *Mol Cell* 28(4):638–651.
50. Williams RS, et al. (2008) Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* 135(1):97–109.
51. Uanschou C, et al. (2007) A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. *EMBO J* 26(24):5061–5070.
52. Ji J, et al. (2012) The role of OsCOM1 in homologous chromosome synapsis and recombination in rice meiosis. *Plant J* 72(1):18–30.
53. Wang Y, Jiang L, Zhang T, Jing J, He Y (2018) ZmCom1 is required for both mitotic and meiotic recombination in maize. *Front Plant Sci* 9:1–14.

54. Li B, Comai L (2000) Functional interaction between Ku and the Werner syndrome protein in DNA end processing. *J Biol Chem* 275(37):28349–28352.
55. Perry JJP, et al. (2006) WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat Struct Mol Biol* 13(5):414–422.
56. Shamanna RA, et al. (2016) WRN regulates pathway choice between classical and alternative non-homologous end joining. *Nat Commun* 7:13785.
57. Kobbe D, Blanck S, Demand K, Focke M, Puchta H (2008) AtRECQ2, a RecQ helicase homologue from *Arabidopsis thaliana*, is able to disrupt various recombinogenic DNA structures in vitro. *Plant J* 55(3):397–405.
58. Plchova H, Hartung F, Puchta H (2003) Biochemical characterization of an exonuclease from *Arabidopsis thaliana* reveals similarities to the DNA exonuclease of the human Werner syndrome protein. *J Biol Chem* 278(45):44128–44138.
59. Chappell C, Hanakahi LA, Karimi-Busheri F, Weinfeld M, West SC (2002) Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *Embo J* 21(11):2827–2832.
60. Koch CA, et al. (2004) Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J* 23(19):3874–85.
61. Petrucco S, Volpi G, Bolchi A, Rivetti C, Ottonello S (2002) A nick-sensing DNA 3'-repair enzyme from *Arabidopsis*. *J Biol Chem* 277(26):23675–23683.
62. Martínez-Macías MI, et al. (2012) A DNA 3' phosphatase functions in active DNA demethylation in *Arabidopsis*. *Mol Cell* 45(3):357–370.
63. Spampinato CP (2017) Protecting DNA from errors and damage: an overview of DNA repair mechanisms in plants compared to mammals. *Cell Mol Life Sci* 74(9):1693–1709.
64. Bertocci B, De Smet A, Weill JC, Reynaud CA (2006) Nonoverlapping functions of DNA polymerases Mu, Lambda, and Terminal Deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. *Immunity* 25(1):31–41.
65. Ma Y, et al. (2004) A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol Cell* 16(5):701–713.
66. Nick McElhinny S a, et al. (2005) A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* 19(D):357–366.
67. Uchiyama Y, Kimura S, Yamamoto T, Ishibashi T, Sakaguchi K (2004) Plant DNA polymerase λ , a DNA repair enzyme that functions in plant meristematic and meiotic tissues. *Eur J Biochem* 271(13):2799–2807.
68. Amoroso A, et al. (2011) Oxidative DNA damage bypass in *Arabidopsis thaliana* requires DNA polymerase λ and proliferating cell nuclear antigen 2. *Plant Cell* 23(2):806–822.
69. Roy S, Choudhury SR, Singh SK, Das KP (2011) AtPol λ , a homolog of mammalian DNA polymerase λ in *arabidopsis thaliana*, is involved in the repair of UV-B induced DNA damage through the dark repair pathway. *Plant Cell Physiol* 52(2):448–467.
70. Roy S, Choudhury SR, Sengupta DN, Das KP (2013) Involvement of AtPol in the repair of high salt- and DNA cross-linking agent-induced double strand breaks in *Arabidopsis*. *Plant Physiol* 162(2):1195–1210.

71. Bakshi S, Sihi S, Sengupta DN (2016) Activity of DNA polymerase λ in spikelets of rice and maize. *Biol Plant* 60(4):788–792.
72. Deshpande RA, Wilson TE (2007) Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA Repair* 6(10):1507–1516.
73. Grawunder U, et al. (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388(6641):492–5.
74. Ahnesorg P, Smith P, Jackson SP (2006) XLF interacts with the XRCC4-DNA Ligase IV complex to promote DNA nonhomologous end-joining. *Cell* 124(2):301–313.
75. Buck D, et al. (2006) Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* 124(2):287–299.
76. Akopiants K, et al. (2009) Requirement for XLF/Cernunnos in alignment-based gap filling by DNA polymerases λ and μ for nonhomologous end joining in human whole-cell extracts. *Nucleic Acids Res* 37(12):4055–4062.
77. Ochi T, et al. (2015) PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* 347(6218):185–188.
78. Xing M, et al. (2015) Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat Commun* 6:6233.
79. Tadi SK, et al. (2016) PAXX Is an accessory c-NHEJ factor that associates with Ku70 and has overlapping functions with XLF. *Cell Rep* 17(2):541–555.
80. West CE, Waterworth WM, Jiang Q, Bray CM (2000) Arabidopsis DNA ligase IV is induced by γ -irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. *Plant J* 24(1):67–78.
81. van Attikum H, et al. (2003) The Arabidopsis AtLIG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA. *Nucleic Acids Res* 31(14):4247–4255.
82. Robert I, Dantzer F, Reina-san-martin B (2009) Parp1 facilitates alternative NHEJ , whereas Parp2 suppresses IgH / c-myc translocations during immunoglobulin class switch recombination. *J Exp Med* 206(5):1047–1056.
83. Rosidi B, et al. (2008) Histone H1 functions as a stimulatory factor in backup pathways of NHEJ. *Nucleic Acids Res* 36(5):1610–1623.
84. Amé JC, Spenlehauer C, De Murcia G (2004) The PARP superfamily. *BioEssays* 26(8):882–893.
85. Boehler C, et al. (2011) Poly(ADP-ribose) polymerase 3 (PARP3), a newcomer in cellular response to DNA damage and mitotic progression. *Proc Natl Acad Sci* 108(7):2783–2788.
86. Jia Q, den Dulk-Ras A, Shen H, Hooykaas PJJ, de Pater S (2013) Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in Arabidopsis thaliana. *Plant Mol Biol* 82(4–5):339–51.
87. Haince JF, et al. (2008) PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. *J Biol Chem* 283(2):1197–1208.
88. Li M, Yu X (2015) The role of poly(ADP-ribosyl)ation in DNA damage response and cancer chemotherapy. *Oncogene* 34(26):3349–3356.
89. Takaoka M, Miki Y (2018) BRCA1 gene: function and deficiency. *Int J Clin Oncol* 23(1):36–44.

90. McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24(11):529–38.
91. Crespan E, Czabany T, Maga G, Hübscher U (2012) Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases λ and β on normal and repetitive DNA sequences. *Nucleic Acids Res* 40(12):5577–5590.
92. Chan SH, Yu AM, McVey M (2010) Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet* 6(7):1–16.
93. Roerink SF, Schendel R, Tijsterman M (2014) Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res* 24(6):954–962.
94. Koole W, et al. (2014) A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun* 5:3216.
95. Mateos-Gomez PA, et al. (2015) Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* 518(7538):254–257.
96. Kent T, Chandramouly G, McDevitt SM, Ozdemir AY, Pomerantz RT (2015) Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol* 22(3):230–237.
97. Wyatt DW, et al. (2016) Essential roles for polymerase θ -mediated end joining in the repair of chromosome breaks. *Mol Cell* 63:1–12.
98. Göttlich B, Reichenberger S, Feldmann E, Pfeiffer P (1998) Rejoining of DNA double-strand breaks in vitro by single-strand annealing. *Eur J Biochem* 258(2):387–395.
99. Liang L, et al. (2005) Modulation of DNA end joining by nuclear proteins. *J Biol Chem* 280(36):31442–31449.
100. Tseng HM, Tomkinson AE (2004) Processing and joining of DNA ends coordinated by interactions among Dn14/Lif1, Pol4, and FEN-1. *J Biol Chem* 279(46):47580–47588.
101. Budd ME, Campbell JL (1997) A yeast replicative helicase, Dna2 helicase, interacts with yeast FEN-1 nuclease in carrying out its essential function. *Mol Cell Biol* 17(4):2136–2142.
102. Prasad R, Dianov GL, Bohr VA, Wilson SH (2000) FEN1 stimulation of DNA polymerase β mediates an excision step in mammalian long patch base excision repair. *J Biol Chem* 275(6):4460–4466.
103. Kimura S, et al. (2000) Plant homologue of flap endonuclease-1: Molecular cloning, characterization, and evidence of expression in meristematic tissues. *Plant Mol Biol* 42(3):415–427.
104. Ahmad A, et al. (2008) ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Mol Cell Biol* 28(16):5082–5092.
105. Bardwell AJ, Bardwell L, Tomkinson AE, Friedberg EC (1994) Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. *Science* 265(5181):2082–2085.
106. Macaisne N, Vignard J, Mercier R (2011) SHOC1 and PTD form an XPF-ERCC1-like complex that is required for formation of class I crossovers. *J Cell Sci* 124(16):2687–2691.
107. Weinfeld M, Mani RS, Abdou I, Aceytuno RD, Glover JNM (2011) Tidying up loose ends: The role of polynucleotide kinase/phosphatase in DNA strand break repair. *Trends Biochem Sci* 36(5):262–271.

108. Della-Maria J, et al. (2011) Human Mre11/human Rad50/Nbs1 and DNA ligase IIIa/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. *J Biol Chem* 286(39):33845–33853.
109. Arakawa H, et al. (2012) Functional redundancy between DNA ligases I and III in DNA replication in vertebrate cells. *Nucleic Acids Res* 40(6):2599–2610.
110. Paul K, et al. (2013) DNA Ligases I and III cooperate in alternative non-homologous end-joining in vertebrates. *PLoS One* 8(3):e59505.
111. Waterworth WM, et al. (2009) DNA ligase I deficient plants display severe growth defects and delayed repair of both DNA single and double strand breaks. *BMC Plant Biol* 9(1):79.
112. Karanam K, Kafri R, Loewer A, Lahav G (2012) Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol Cell* 47(2):320–329.
113. Székvölgyi L, Kunihiro O, Nicolas A (2015) Initiation of meiotic homologous recombination: Flexibility , impact of histone modifications, and chromatin remodeling. *Cold Spring Harb Perspect Biol* 7:a016527.
114. Mimitou EP, Symington LS (2011) DNA end resection-Unraveling the tail. *DNA Repair* 10(3):344–348.
115. Mazón G, Mimitou EP, Symington LS (2010) SnapShot: Homologous recombination in DNA double-strand break repair. *Cell* 142(4):646, 646.e1.
116. Knoll A, Schröpfer S, Puchta H (2014) The RTR complex as caretaker of genome stability and its unique meiotic function in plants. *Front Plant Sci* 5:1–6.
117. Röhrig S, Schröpfer S, Knoll A, Puchta H (2016) The RTR complex partner RMI2 and the DNA helicase RTEL1 are both independently involved in preserving the stability of 45S rDNA repeats in *Arabidopsis thaliana*. *PLoS Genet* 12(10):1–23.
118. Jia N, Liu X, Gao H (2016) A DNA2 homolog is required for DNA damage repair, cell cycle regulation, and meristem maintenance in plants. *Plant Physiol* 171(1):318–333.
119. Aklilu BB, Culligan KM (2016) Molecular evolution and functional diversification of replication protein A1 in plants. *Front Plant Sci* 7(33):1–15.
120. Seeliger K, Dukowic-Schulze S, Wurz-Wildersinn R, Pacher M, Puchta H (2012) BRCA2 is a mediator of RAD51- and DMC1-facilitated homologous recombination in *Arabidopsis thaliana*. *New Phytol* 193(2):364–375.
121. San Filippo J, Sung P, Klein H (2008) Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 77:229–257.
122. Liu Y, Tarsounas M, O'Regan P, West SC (2007) Role of RAD51C and XRCC3 in genetic recombination and DNA repair. *J Biol Chem* 282(3):1973–1979.
123. Li X, Stith CM, Burgers PM, Heyer WD (2009) PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase δ . *Mol Cell* 36(4):704–713.
124. Sebesta M, et al. (2013) Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. *DNA Repair* 12(9):691–698.
125. Holliday R (1964) A mechanism for gene conversion in fungi. *Genet Res* 5:282–304.
126. McIlwraith MJ, West SC (2008) DNA repair synthesis facilitates RAD52-mediated second-end capture during DSB repair. *Mol Cell* 29(4):510–516.

127. Wu L, Hickson IO (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426(6968):870–874.
128. Svendsen JM, Harper JW (2010) GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. *Genes Dev* 24(6):521–536.
129. Geuting V, et al. (2009) Two distinct MUS81-EME1 complexes from Arabidopsis process Holliday junctions. *Plant Physiol* 150(2):1062–1071.
130. Bauknecht M, Kobbe D (2014) AtGEN1 and AtSEND1, Two paralogs in Arabidopsis, possess Holliday junction resolvase activity. *Plant Physiol* 166(1):202–216.
131. Dubest S, Gallego ME, White CI (2002) Role of the AtRad1p endonuclease in homologous recombination in plants. *EMBO Reports* 3(11):1049–1054.
132. Dubest S, Gallego ME, White CI, Pascal B (2004) Roles of the AtErccl protein in recombination. *Plant J* 39:334–342.
133. Moynahan ME, Jasin M (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* 11(3):196–207.
134. Matos J, West SC (2014) Holliday junction resolution: Regulation in space and time. *DNA Repair* 19:176–181.
135. Lydeard JR, Jain S, Yamaguchi M, Haber JE (2007) Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448(7155):820–823.
136. Hanin M, et al. (2001) Gene targeting in Arabidopsis thaliana. *Plant J* 28(6):671–677.
137. Doetschman T, et al. (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* 330(6148):576–578.
138. Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51(3):503–512.
139. Paszkowski J, Baur M, Bogucki A, Potrykus I (1988) Gene targeting in plants. *EMBO J* 7(13):4021–6.
140. 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.
141. Lee KY, Lund P, Lowe K, Dunsmuir P (1990) Homologous recombination in plant cells after Agrobacterium-mediated transformation. *Plant Cell* 2(5):415–425.
142. Halfter U, Morris PC, Willmitzer L (1992) Gene targeting in Arabidopsis thaliana. *Mol Gen Genet* 231(2):186–193.
143. Hrouda M, Paszkowski J (1994) High fidelity extrachromosomal recombination and gene targeting in plants. *MGG Mol Gen Genet* 243(1):106–111.
144. Miao Z -H, Lam E (1995) Targeted disruption of the TGA3 locus in Arabidopsis thaliana. *Plant J* 7(2):359–365.
145. Risseuw E, Offringa R, Franke van Dijk MEI, Hooykaas PJJ (1995) Targeted recombination in plants using Agrobacterium coincides with additional rearrangements at the target locus. *Plant J* 7(1):109–119.
146. Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. *Int J Dev Biol* 57(6–7–8):629–637.
147. Puchta H (2017) ScienceDirect Applying CRISPR / Cas for genome engineering in plants : the best is yet to come. *Curr Opin Plant Biol* 36:1–8.

148. Reiss B, et al. (2000) RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by *Agrobacterium*. *Proc Natl Acad Sci* 97(7):3358–3363.
149. Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in *Arabidopsis* plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102(34):12265–12269.
150. Even-Faitelson L, Samach A, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA (2011) Localized egg-cell expression of effector proteins for targeted modification of the *Arabidopsis* genome. *Plant J* 68(5):929–937.
151. Samach A, Gurevich V, Avivi-Ragolsky N, Levy AA (2018) The effects of AtRad52 overexpression on homologous recombination in *Arabidopsis*. *Plant J*. 95: 30–40.
152. Attikum H Van, Bundock P, Hooykaas PJJ (2001) Non-homologous end-joining proteins are required for *Agrobacterium* T-DNA integration. *EMBO J* 20(22):6550–6558.
153. Kooistra R, Hooykaas PJJ, Steensma HY (2004) Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* 21(9):781–792.
154. Ninomiya Y, Suzuki K, Ishii C, Inoue H (2004) Highly efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining. *Proc Natl Acad Sci U S A* 101(33):12248–12253.
155. Endo M, Mikami M, Toki S (2016) Biallelic gene targeting in rice. *Plant Physiol* 170(2):667–677.
156. Charbonnel C, Allain E, Gallego ME, White CI (2011) Kinetic analysis of DNA double-strand break repair pathways in *Arabidopsis*. *DNA Repair* 10(6):611–9.
157. van Kregten M, et al. (2016) T-DNA integration in plants results from Polymerase Theta-mediated DNA repair. *Nat Plants* 2:1–6.
158. Endo M, Toki S (2014) Toward establishing an efficient and versatile gene targeting system in higher plants. *Biocatal Agric Biotechnol* 3(1):2–6.
159. Kolodkin AL, Klar AJS, Stahl FW (1986) Double-strand breaks can initiate meiotic recombination in *S. cerevisiae*. *Cell* 46(5):733–740.
160. Sun H, Treco D, Schultes NP, Szostak JW (1989) Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* 338(6210):87–90.
161. Fishman-Lobell J, Rudin N, Haber JE (1992) Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol Cell Biol* 12(3):1292–303.
162. de Massy B, Nicolas a (1993) The control in cis of the position and the amount of the ARG4 meiotic double-strand break of *Saccharomyces cerevisiae*. *EMBO J* 12(4):1459–66.
163. Puchta H, Dujon B, Hohn B (1993) Homologous recombination in plant cells is enhanced by *in vivo* induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res* 21(22):5034–5040.
164. Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc Natl Acad Sci U S A* 93(10):5055–5060.
165. 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.

166. Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* 17(20):6086–6095.
167. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93(3):1156–1160.
168. Smith J (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* 28(17):3361–3369.
169. Doyon Y, et al. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* 8(1):74–79.
170. Shukla VK, et al. (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459(7245):437–441.
171. Townsend J a, et al. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459(7245):442–5.
172. 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.
173. 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.
174. Qi Y, et al. (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. *Genome Res* 23(3):547–54.
175. Baltes NJ, Gil-Humanes J, Cermak T, Atkins P a, Voytas DF (2014) DNA replicons for plant genome engineering. *Plant Cell* 26(1):151–163.
176. Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. *Annu Rev Plant Biol* 64(1):327–350.
177. Christian M, et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186(2):756–761.
178. Boch J, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326(5959):1509–12.
179. Moscou M, Bogdanove A (2009) Recognition by TAL Effectors. *Science* 326:1501.
180. Cermak T, et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39(12):e82.
181. Sakuma T, et al. (2013) Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci Rep* 3:1–8.
182. Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* 30(5):390–392.
183. Zhang Y, et al. (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol* 161(1):20–27.
184. Jinek M, et al. (2012) A programmable dual-RNA – guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–822.

185. Jinek M, et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1215–1226.
186. Nishimasu H, et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156(5):935–949.
187. Jiang F, Doudna JA (2017) CRISPR – Cas9 structures and mechanisms. *AnnuRevBiophys* 46:505–529.
188. Zuo Z, Liu J (2016) Cas9-catalyzed DNA cleavage generates staggered ends: Evidence from molecular dynamics simulations. *Sci Rep* 5:1–9.
189. Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V (2015) Editing plant genomes with CRISPR/Cas9. *Curr Opin Biotechnol* 32:76–84.
190. Svitashv S, et al. (2015) Targeted mutagenesis, precise gene editing, and site-specific gene insertion in Maize using Cas9 and guide RNA. *Plant Physiol* 169(2):931–945.
191. 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.
192. 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.
193. 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.
194. Zhao Y, et al. (2016) An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci Rep* 6:1–11.
195. Ran FA, et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380–1389.
196. Mali P, et al. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31(9):815–838.
197. 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. *Plant J* 94(4):735–746.
198. Bothmer A, et al. (2017) Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* 8:1–12.
199. Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. *Plant Cell Rep* 35(7):1–10.
200. Kaya H, Mikami M, Endo A, Endo M, Toki S (2016) Highly specific targeted mutagenesis in plants using Staphylococcus aureus Cas9. *Sci Rep* 6:1–9.
201. Karvelis T, et al. (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol* 16(1):1–13.
202. Zetsche B, et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:1–13.
203. Schindele P, Wolter F, Puchta H (2018) Transforming plant biology and breeding with CRISPR/Cas9, Cas12, Cas13. *FEBS Lett* 592(12):1954–1967.
204. Begemann MB, et al. (2017) Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Sci Rep* 7(1):1–6.

- 205. Ferenczi A, Pyott DE, Xipnitou A, Molnar A (2017) Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. *Proc Natl Acad Sci* 114(51):201710597.
- 206. Fauser F, et al. (2012) In planta gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
- 207. Čermák T, Baltes NJ, Čegan R, Zhang Y, Voytas DF (2015) High-frequency, precise modification of the tomato genome. *Genome Biol* 16(1):232.
- 208. Gil-Humanes J, et al. (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J* 89(6):1251–1262.
- 209. Dahan-Meir T, et al. (2018) Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. *Plant J*:1–12.

Chapter 2

Sequence-specific nucleases for genome editing in plants: TALENs versus CRISPR/Cas9

**Gary D. Strunks^{1,2}, Bart J. P. M. Klemann^{1,3},
Paul J. J. Hooykaas¹ and Sylvia de Pater¹**

¹Molecular and Developmental Genetics, Institute of Biology,
Leiden University, 2333 BE, the Netherlands

²Current affiliation: Naktuinbouw, 2371 GD Roelofarendsveen,
the Netherlands

³Current affiliation: Dümmer Orange, 2678 PS De Lier,
the Netherlands

Abstract

Artificial sequence-specific nucleases such as TALENs and those based on the CRISPR/Cas9 system can be used as tools for plant genome editing. DNA double-strand breaks (DSBs) induced by these nucleases can be repaired by either nonhomologous end-joining (NHEJ) or homologous recombination (HR). Mutation occurs when DSBs are repaired imprecisely by NHEJ. Alternatively, DSBs form entry points for repair by HR and are thus a prerequisite for gene targeting via HR when an artificial repair template is introduced. Here, we compare two TALEN constructs (TALEN-CRU-1 and TALEN-CRU-2) and two CRISPR/Cas9 (Cas9-CRU-1 and Cas9-CRU-2) constructs for targeted mutagenesis efficiencies at the *Arabidopsis* cruciferin 3 (*CRU3*) gene. Wild-type plants were transformed with TALEN and CRISPR/Cas9 expression vectors and targeted mutagenesis efficiencies were determined by footprint analysis. Mutations at the repair junctions as a result of imperfect DSB repair were obtained in both TALEN and CRISPR/Cas9 plants, indicating both nucleases were expressed and induced DSBs at the *CRU3* target. However, both TALEN constructs performed poorly in terms of mutagenesis frequency. One of the CRISPR/Cas9 constructs, however, gave a significant increase in mutagenesis frequency. This Cas9-CRU-2 construct uniquely had a GG at the 3' end of the protospacer, which may be responsible for the enhanced efficiency. A similar conclusion could be drawn when comparing two CRISPR/Cas9 constructs targeting the *Arabidopsis* alcohol dehydrogenase 1 (*ADH1*) locus.

Introduction

Genome editing in plants can be achieved by introducing DNA double strand breaks (DSBs) in the genome by using artificial nucleases (1). DSBs can be repaired by one of the two DSB repair pathways: nonhomologous end-joining (NHEJ) and homologous recombination (HR). NHEJ and HR mostly leads to precise repair (2). However, repeated cycles of DSB induction and repair can eventually lead to indels at the repair junction by imprecise NHEJ. In this way, sequence-specific nuclease-induced DSB repair by NHEJ can be utilized for targeted mutagenesis of a desired genomic locus. Additionally, HR can be utilized for gene targeting when an artificial repair template lacking the DSB site is supplied (3).

Currently there are four main classes of artificial nucleases: modified meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and those based on the clustered regularly interspaced short palindromic repeat/CRISPR-associated 9 (CRISPR/Cas9) system (3, 4). Meganucleases such as I-*SceI* and I-*CreI* were utilized to induce targeted DSBs in the first break-through experiments to elucidate the basic DSB-repair mechanisms, and to demonstrate the proof of principle of targeted mutagenesis and gene targeting in plants (5–11). However, engineering meganucleases to target novel sequences is challenging because the nuclease domain and the DNA binding domain overlap.

DSB induction in novel targets became more feasible with ZFNs. These nucleases consist of a zinc-finger array as DNA binding domain and the *FokI* nuclease domain that can induce a DSB as a dimer (12). Each zinc-finger can recognize 3 bp, so with a pair of 3-finger ZFNs a sequence of 18 bp can be recognized. Being able to target novel sequences with ZFNs was a major breakthrough, although nowadays their use has largely been replaced by the more convenient TALENs and CRISPR/Cas9 system (11).

TALENs share similarities in design with ZFNs. Their DNA binding domains are derived from proteins produced by plant pathogens of the genus *Xanthomonas*. Like ZFNs, the DNA binding domain is fused to the *FokI* nuclease domain (13). The TALEN DNA binding domain consists of an array of 13 – 28 repeats, each consisting of 34 highly conserved amino acids. The amino acid residues at position 13 and 14 are called the repeat variable diresidues (RVDs) and determine the binding to one of each four DNA bases, so that there is a one to one correspondence with the DNA sequence (13–15). A prerequisite for proper TALEN activity is that the binding sequence of each TALE is preceded by a 5' T (14, 15). Because a TALEN pair can recognize 26 – 56 bp, unique genomic targets can be easily selected. Nowadays, TALENs

can be designed and assembled easily with publicly available kits, such as the Golden Gate kit (16). However, designing TALENs for multiple targets can still be relatively time-consuming.

More recently, the CRISPR/Cas9 system was developed (17). Originally, it functions as an adaptive immune system present in bacteria and archaea, where it targets the degradation of foreign viral or plasmid DNA. A short RNA molecule called CRISPR RNA (crRNA) binds to a DNA target sequence and together with another short RNA molecule, the trans-activating RNA (tracrRNA) recruits the Cas9 endonuclease that induces a DSB. To make the system simpler for application in eukaryotes, both short RNAs were fused into a chimeric single-guide RNA (sgRNA) without losing the function of both individual RNAs (17). A prerequisite for sgRNA design is the presence of a protospacer adjacent motif (PAM) of NGG flanking the 3' end of the target sequence, the so-called protospacer. The PAM interacts with the Cas9 PAM interacting domain (PI domain) (18, 19). One of the advantages of the CRISPR/Cas9 system is that the direct RNA-DNA recognition allows for rapid and convenient design of new sgRNAs for almost any target of interest, compared to the more laborious assembly of the new TAL effector arrays for TALENs.

TALENs and the CRISPR/Cas9 system have been used for plant genome editing (20, 21). Their activity greatly determines the efficiency of either targeted mutagenesis or gene targeting. Here, we used two different TALEN constructs and two different CRISPR/Cas9 constructs for targeted mutagenesis at the *Arabidopsis* cruciferin 3 (*CRU3*) gene, which encodes a seed storage protein. A loss of restriction site assay was used to determine the effectiveness of DSB induction at the *CRU3* target; footprint analyses as evidence for imprecise NHEJ-mediated DSB-repair was performed (4). Footprints could be readily obtained for both TALENs and the CRISPR/Cas9 system. However, mutagenesis efficiencies for most constructs remained low. Interestingly, CRISPR/Cas9 mutagenesis efficiency substantially increased when using a slightly different protospacer sequence.

Materials and Methods

Plant material and growth conditions

Plants of the *Arabidopsis* Columbia-0 ecotype were used as wild-type control and for all transformations. T1 or T2 seeds were germinated on medium in a climate-controlled growth chamber at 20°C and 50% humidity, with a light intensity of 75 $\mu\text{mol m}^{-2} \text{s}^{-1}$ during 16h/day photoperiod. Seedlings that were transferred to soil were grown in a climate-controlled growth chamber at 20°C and 70% humidity, with a light intensity of 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ during 16h/day photoperiod.

TALENs and CRISPR/Cas9 vector construction and plant transformation

The Golden Gate Kit (AddGene) was used for TALEN design and assembly as described (16). TALE repeat arrays with corresponding RVDs (Table 1) of the *CRU3* target DNA binding domain were assembled in vector pZHY500 (TALEN-CRU-1/2-left) or pZHY501 (TALEN-CRU-1/2-right). TALEs were cloned into vector pZHY013 using *Xba*I and *Bam*HI (TALEN-CRU-1/2-left) or *Nhe*I and *Bgl*II (TALEN-CRU-1/2-right). Subsequently, the TALEN pairs were cloned in the binary 35S T-DNA expression vector pMDC32 (22) via a Gateway LR reaction to create TALEN-CRU-1 (pSDM3906) and TALEN-CRU-2 (pSDM3907).

To assemble the CRISPR/Cas9 constructs, oligo's SP509/SP510 (CRU-1 sgRNA expression) (23), GS21/GS22 (CRU-2 sgRNA expression), GS29/GS30 (ADH-1 sgRNA expression) and GS31/GS32 (ADH-2 sgRNA expression) (Table 2) were annealed and cloned in *Bbs*I-digested pEn-Chimera (24). Subsequently, sgRNAs encoding genes were cloned in expression vector pDE-Cas9 (24) by a Gateway LR reaction, resulting in Cas9-CRU-1 (pSDM3903) (23), Cas9-CRU-2 (pSDM3908), Cas9-ADH-1 (pSDM3916) and Cas9-ADH-2 (pSDM3917).

Plant binary vectors were introduced into *Agrobacterium tumefaciens* strain AGL1 by electroporation (Den Dulk-Ras and Hooykaas, 1995). *Arabidopsis thaliana* wild type plants were transformed with T-DNAs containing nuclease expression cassettes via the floral dip method (25). T1 plants were selected on MA solid medium without sucrose supplemented with timentin (100 $\mu\text{g/mL}$), nystatin (100 $\mu\text{g/mL}$) and 15 $\mu\text{g/mL}$ hygromycin for TALEN T-DNA selection or 15 $\mu\text{g/mL}$ phosphinothricin for CRISPR/Cas9 T-DNA selection.

DNA isolation and footprint analysis

T2 seeds derived from independently selected T1 plants were germinated on ½ MS supplemented with 10 µg/ml hygromycin for TALEN T-DNA selection or 10 µg/ml phosphinothricin for CRISPR/Cas9 T-DNA selection, and after 10 days of growth, pools of 10 seedlings per plant line were disrupted to a powder under liquid N₂ in a tissue lyser (Retch, Haan, Germany). Genomic DNA was extracted using the CTAB method (26), and was subjected to predigestion with the appropriate restriction enzymes. Predigested genomic DNA was used for PCR with Phusion polymerase (Thermo Scientific) to amplify the nucleases target site, followed by digestion of the PCR products with *Pst*I, *Dde*I or *Eae*I and separated in agarose gels. Primers SP491 and SP492 were used for the *CRU3* target region, and primers GS48 and GS49 were used for the *ADH1* target region (Table 2). Restriction enzyme digestion-resistant fragments were visualized and then extracted from an agarose gel, cloned into pJet1.2 (Thermo Scientific) and sequenced by Macrogen Europe (Amsterdam, The Netherlands). Identical sequences in the same line were counted as one mutagenesis event as they could have been arisen by PCR amplification.

Table 1. TALEN RVDs

TALE-domain	RVDs
TALE-CRU-1 left	NG HD HD NI NN NN NG HD NN NG NN NI NN NN HD
TALE-CRU-1 right	HD HD NI HD NG HD HD NG HD NN HD NG HD NG HD NN NG NI
TALE-CRU-2 left	NI NN HD HD HD NI HD NI NN NN NN HD NI NI HD NN NN HD HD NG
TALE-CRU-2 right	NG HD NG HD NN NG NN NN NN NI HD HD NG HD NI

Table 2. Primers used for cloning and PCR reactions.

Primer	Sequence (5' – 3')	Used for
SP509	ATTGAGGAGACTATCTGCAGCATG	sgRNA cloning <i>CRU3</i> (Cas9-CRU-1 construct)
SP510	AAACCATGCTGCAGATAGTCTCCT	sgRNA cloning <i>CRU3</i> (Cas9-CRU-1 construct)
GS21	ATTGTCGTAGGGCTGTCTTAGAGG	sgRNA cloning <i>CRU3</i> (Cas9-CRU-2 construct)
GS22	AAACCCTCTAAGACAGCCCTACGA	sgRNA cloning <i>CRU3</i> (Cas9-CRU-2 construct)
GS29	ATTGCGTATCTTCGGCCATGAAGC	sgRNA cloning <i>ADH1</i> (Cas9-ADH-1 construct)
GS30	AAACGCTTCATGGCCGAAGATACG	sgRNA cloning <i>ADH1</i> (Cas9-ADH-1 construct)
GS31	ATTGATCTTCGGCCATGAAGCTGG	sgRNA cloning <i>ADH1</i> (Cas9-ADH-2 construct)
GS32	AAACCCAGCTTCATGGCCGAAGAT	sgRNA cloning <i>ADH1</i> (Cas9-ADH-2 construct)
SP491	GCTTCAGAACCAACAAGACAGC	<i>CRU3</i> target sense
SP492	TGAGCCTGACATACTCCAAG	<i>CRU3</i> target antisense
GS48	ACCACCGACAGATTATTCG	<i>ADH1</i> target sense
GS49	GGAGAATCTTGATTCACCATCG	<i>ADH1</i> target antisense

Estimation of mutation efficiency and statistics

To estimate TALEN and CRISPR/Cas9 mutagenesis efficiency, the *CRU3* and *ADH1* target sites were PCR-amplified using undigested genomic DNA. PCR products were digested with the appropriate restriction enzymes and analysed on agarose gels. A semi-quantitative analysis for mutation efficiency was performed with ImageJ (27) by dividing the intensity of the digestion-resistant band by the total intensity of all bands in a given lane (Qi *et al.* 2013; Shen *et al.* 2017). To see if the fraction of digestion-resistant fragment was significantly different between the TALEN-CRU-1, TALEN-CRU-2, Cas9-CRU-1 and Cas9-CRU-2 transformed lines, and between Cas9-ADH-1 and Cas9-ADH-2 transformed lines, one-tailed Mann-Whitney *U* tests were performed.

Results and Discussion

Mutagenesis at the CRU3 locus with TALENs and CRISPR/Cas9

With the aim of future gene targeting experiments, we compared TALEN- and CRISPR/Cas9-mediated mutagenesis of the *Arabidopsis CRU3* gene. Two TALEN expression constructs (TALEN-CRU-1 and TALEN-CRU-2) and two CRISPR/Cas9 expression constructs (Cas9-CRU-1 and Cas9-CRU-2) were designed that targeted a region on exon 3 of the *CRU3* gene (Figure 1). Nucleases were constitutively expressed under the CaMV (cauliflower mosaic virus) 35S promoter for TALENs, the Ubiquitin promoter for Cas9 and the U6-26 promoter for the CRISPR RNA (16, 23, 24). Wild-type plants were transformed with these constructs via *Agrobacterium*-mediated floral dip (25). After selection of several primary transformants on either hygromycin (TALEN constructs) or phosphinotricin (CRISPR/Cas9 constructs), the T2 progeny was used for further analysis. The *CRU3* target sequences were conveniently selected in the vicinity of a restriction enzyme recognition site in order to be able to assess nuclease activity using the loss of the restriction site as a proxy. The repeated cycles of DSB-induction by either the TALEN or CRISPR/Cas9 endonucleases near the restriction enzyme recognition site, followed by imprecise NHEJ-mediated repair, lead to footprints with mutated restriction sites. By performing PCR amplification of a region flanking the nuclease target site, followed by restriction digestion and gel electrophoresis, the presence of restriction digestion-resistant PCR product served as evidence for footprints at the repair junction due to imprecise NHEJ-mediated DSB repair (4, 23, 28). In this study, *Pst*I was used for the TALEN-CRU-1 and Cas9-CRU-1 assays and *Dde*I was used for the TALEN-CRU-2 and Cas9-CRU-2 assays.

To evaluate the presence of footprints at *CRU3* in plants transformed with TALEN and CRISPR/Cas9 constructs, genomic DNA was isolated from pools of 10 T2 seedlings that were grown for 10 days. To enrich for mutations that partly or completely remove the restriction sites, genomic DNA was subjected to an overnight pre-digestion with the appropriate restriction enzymes (4, 23). Predigested genomic DNA was used as a template for PCR-amplification of a 272 bp fragment containing the TALEN and Cas9 targets and the PCR products were subjected to *Pst*I or *Dde*I digestion and gel electrophoresis. Restriction digestion-resistant PCR fragments were cloned and sequenced. Indeed, footprints were detected in several plant lines transformed with each nuclease (Figure 2). Footprints mainly consisted of 1 bp insertions or small to larger deletions ranging from 1 to 50 bp that were sometimes accompanied by insertions. It was not possible to detect very large deletions with this experimental setup due to constraints by the PCR product length. Furthermore, the resulting sequences suggested that microhomology-mediated end-joining (MMEJ) was utilized for DSB repair. Plants with *CRU3* footprints did not show a distinct phenotype, presumably because of redundant functions with the other *Arabidopsis* cruciferin genes (29). Taken together, the results indicated both TALEN and CRISPR/Cas9 were active and could effectively induce DSBs at the *CRU3* locus.

Quantification of nuclease activity of TALENs and CRISPR/Cas9

To compare the efficiencies of our TALEN and CRISPR/Cas9 constructs, nuclease activity was semi-quantified (Figure 3). Non-predigested genomic DNA of wild-type, and of the TALEN-CRU-1, TALEN-CRU-2, Cas9-CRU-1 and Cas9-CRU-2 transformed lines was PCR-amplified, resulting in a fragment of 272 bp. PCR products were digested with *Pst*I (TALEN-CRU-1 and Cas9-CRU-1) resulting in a 147 bp and a 125 bp fragment or *Dde*I (TALEN-CRU-2 and Cas9-CRU-2), resulting in a 199 bp and a 73 bp fragment. The ratio of restriction digestion resistant- and non-resistant PCR product was determined using ImageJ (27). It must be noted that the observed resistant PCR product as a measure for efficiency is an underestimate of the real efficiency because the loss of restriction site assay is only able to detect mutations that affect the restriction site. Thus, small mutations that occur outside of the restriction site were not detected. Next generation sequencing is necessary for more precise quantification of nuclease efficiencies. However, larger mutations that affected the restriction sites appeared to be frequent enough to give estimates of nuclease efficiencies.

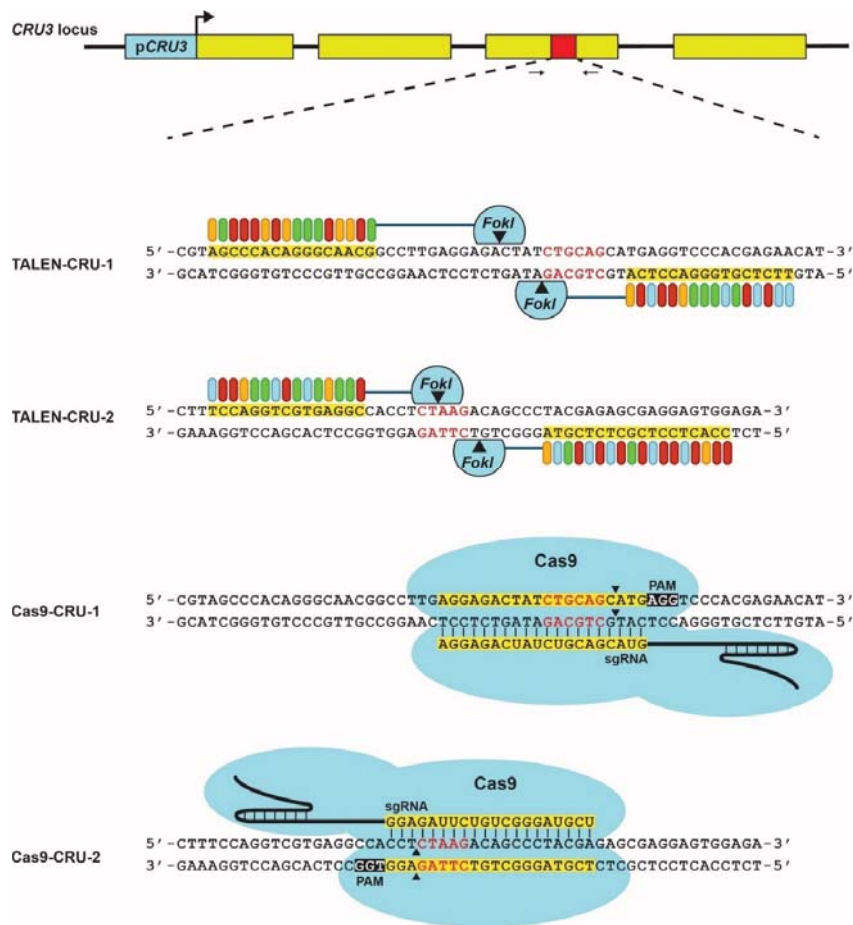


Figure 1. TALEN and CRISPR/Cas9 nucleases targeting the *CRU3* locus. A fragment of 272 bp on exon 3 (yellow), containing the nuclease target sites (red) was PCR-amplified with primers SP491 and SP492 (arrows) and digested with either *Pst*I (red, CTGCAG) for TALEN-CRU-1 and Cas9-CRU-1, resulting in a 147 and a 125 bp fragment, or *Dde*I (red, CTAAG) for TALEN-CRU-2 and Cas9-CRU-2, resulting in a 199 bp and a 73 bp fragment. TALE binding sequences and the CRISPR/Cas9 protospacers and DNA-binding part of the sgRNAs are shown in yellow. PAM sequences are shown in black. Nuclease cutting sites are indicated by black triangles.

WT AGCCACAGGGCAACGGCCTTGAGGAGACTATCTGCAGCA TGAGGTCCCACGAGA

TALEN-CRU-1

plant line 5 -6 AGCCACAGGGCAACGGCCTTGAGGAG-----TGCAGCATGAGGTCCCACGAGA
-11 AGCCACAGGGCAACGGCCTTGAGGAG-----CATGAGGTCCCACGAGA (3)
-11 AGCCACAGGGCAACGGCCTTGAGGAG-----GAGGTCCCACGAGA
plant line 6 -8 AGCCACAGGGCAACGGCCTTGAGGAG-----CAGCATGAGGTCCCACGAGA
plant line 7 -8 AGCCACAGGGCAACGGCCTTGAGGAG-----CAGCATGAGGTCCCACGAGA
-11 AGCCACAGGGCAACGGCCTTGAGGAG-----CATGAGGTCCCACGAGA (2)
plant line 10 -5 AGCCACAGGGCAACGGCCTTGAGGAG-----GCAGCATGAGGTCCCACGAGA
plant line 13 -8 AGCCACAGGGCAACGGCCTTGAGGAG-----CAGCATGAGGTCCCACGAGA (2)
plant line 15 -3 AGCCACAGGGCAACGGCCTTGAGGAGACTA---GCAGCATGAGGTCCCACGAGA
-3 AGCCACAGGGCAACGGCCTTGAGGAGACT---TGCAGCATGAGGTCCCACGAGA (2)
-5 AGCCACAGGGCAACGGCCTTGAGGAG-----TGCAGCATGAGGTCCCACGAGA (2)
-6 AGCCACAGGGCAACGGCCTTGAGGAG-----TGCAGCATGAGGTCCCACGAGA (3)
-7 AGCCACAGGGCAACGGCCTTGAGGAGACT-----GCATGAGGTCCCACGAGA (3)
-8 AGCCACAGGGCAACGGCCTTGAGGAG-----CAGCATGAGGTCCCACGAGA (6)
-11 AGCCACAGGGCAACGGCCTTGAGGAG-----CATGAGGTCCCACGAGA (3)
-15 AGCCACAGGGCAACGGCCTTGAGGAG-----GGTCCCACGAGA (4)

WT TCCAGGTCGTGAGGCCACCTCTAAGACAGCCC TACGAGCGAGGAGTGG

TALEN-CRU-2

plant line 3 -2 TCCAGGTCGTGAGGCCACCTCTA---CAGCCCTACGAGAGCGAGGAGTGG
-11 TCCAGGTCGTGAGGCCACCTCTA-----CGAGAGCGAGGAGTGG (2)
plant line 5 +1 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG
-1 TCCAGGTCGTGAGGCCACCTCTA---ACAGCCCTACGAGAGCGAGGAGTGG
-1 TCCAGGTCGTGAGGCCACCTCTA---GACAGCCCTACGAGAGCGAGGAGTGG
-6+2 TCCAGGTCGTGAGGCCACCT---A---CAGCCCTACGAGAGCGAGGAGTGG (2)
-2 TCCAGGTCGTGAGGCCACCTCTA---CAGCCCTACGAGAGCGAGGAGTGG (7)
-3 TCCAGGTCGTGAGGCCACCTCTA---CAGCCCTACGAGAGCGAGGAGTGG (3)
-7 TCCAGGTCGTGAGGCCACCT-----AGCCCTACGAGAGCGAGGAGTGG
-9 TCCAGGTCGTGAGGCCACCT-----CCTACGAGAGCGAGGAGTGG
-11 TCCAGGTCGTGAGGCCACCT-----CCTACGAGAGCGAGGAGTGG (3)
-13 TCCAGGTCGTGAGGCCACCT-----ACGAGAGCGAGGAGTGG
plant line 7 -2 TCCAGGTCGTGAGGCCACCTCTA---CAGCCCTACGAGAGCGAGGAGTGG (4)
-3 TCCAGGTCGTGAGGCCACCT---ACAGCCCTACGAGAGCGAGGAGTGG
-4 TCCAGGTCGTGAGGCCACCTC---ACAGCCCTACGAGAGCGAGGAGTGG (2)
-11 TCCAGGTCGTGAGGCCACCTCTA-----CGAGAGCGAGGAGTGG

WT AGCCACAGGGCAACGGCCTTGAGGAGACTATCTGCAGCATG AGGTCCCACGAGA

Cas9-CRU-1

plant line 1 -3 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT**BCA**---TGAGGTCCCACGAGA (2)
-6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT---TGAGGTCCCACGAGA
-13 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT**CA**-----CGAGA (2)
-20 AGCCACAGGGCAACGGCCTTGAGG-----TCCCACGAGA (2)
-50 AGCCACAGGG-----//-----ACGACCCTGCTC (2)
plant line 2 -10 AGCCACAGGGCAACGGCCTTGAGGAG-----ATGAGGTCCCACGAGA
-14 AGCCACAGGGCAACGGCCTTGAGGAG-----GGTCCCACGAGA
-14+6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---TTTTAT---GAGA (2)
-18+6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTG-----ACGTG-----GA
plant line 6 -5 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTG-----GAGGTCCCACGAGA
-6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT-----ATGAGGTCCCACGAGA
-9 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT-----GTCCCACGAGA

WT TCCAGGTCGTGAGGCCA CCTCTAAGACAGCCCTACGA GAGCGAGGAGTGG

Cas9-CRU-2

plant line 1 +1 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG
-1 TCCAGGTCGTGAGGCCACCTC---AAGACAGCCCTACGAGAGCGAGGAGTGG
-3 TCCAGGTCGTGAGGCCACCT---AAGACAGCCCTACGAGAGCGAGGAGTGG
-5 TCCAGGTCGTGAGGCCACCT---GACAGCCCTACGAGAGCGAGGAGTGG
-8 TCCAGGTCGTGAGGCC-----GACAGCCCTACGAGAGCGAGGAGTGG
plant line 3 +1 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG
-2 TCCAGGTCGTGAGGCCACCT---AAGACAGCCCTACGAGAGCGAGGAGTGG
-8 TCCAGGTCGTGAGGCCACCT---AGCCCTACGAGAGCGAGGAGTGG
-13 TCCAGGTCGTGAGGCCACCT-----ACGAGAGCGAGGAGTGG
plant line 6 -4 TCCAGGTCGTGAGGCCACCT---GACAGCCCTACGAGAGCGAGGAGTGG
-5 TCCAGGTCGTGAGGCCACCTC---CAGCCCTACGAGAGCGAGGAGTGG

Figure 2. TALEN and CRISPR/Cas9-mediated targeted mutagenesis of the *CRU3* target. Predigested genomic DNA of pools of T2 seedlings derived from individual primary transformants was PCR-amplified and digested with either *Pst*I or *Dde*I. Resistant PCR product was cloned and sequenced. The TALEN binding sequences and the sgRNA protospacers are shown in yellow. The PAM sequence is shown in black. The restriction sites for *Pst*I and *Dde*I are underlined in red. Indels were obtained from different independently transformed plant lines. Deletions are shown in dashes and insertions are in green. Possible microhomologies used for repair are shown in purple. The numbers of multiple clones that had the same footprint are shown at the right of the sequence. The numbers at the left of the sequence are deletion lengths (-) or insertions lengths (+).

To correct for background levels in all nuclease-transformed plant lines as a result of incomplete digestion, the average fraction of undigested bands in the wild type control was subtracted from those seen in nuclease-transformed plant lines (Figure 3A). After this correction, the fraction of digestion-resistant PCR product in all nine TALEN-CRU-1 lines and eight TALEN-CRU-2 lines was with an average of 0.01% and 0.7% very low (Figure 3A). The average fraction of *Pst*I-resistant PCR product in the ten Cas9-CRU-1 lines was with 2.0% still low, but significantly higher compared to TALEN-CRU-1 ($p = 5.9 \times 10^{-4}$, Mann-Whitney *U* test) and TALEN-CRU-2 ($p = 0.016$, Mann-Whitney *U* test) (Figure 3A, B). The highest levels of resistant PCR product were found in the nine Cas9-CRU-2 lines, with an average fraction of *Dde*I resistant PCR product of 19.9%, a 28-fold increase compared to TALEN-CRU-2 ($p = 2.0 \times 10^{-5}$, Mann-Whitney *U* test) and an almost 10-fold increase compared to Cas9-CRU-1 ($p = 1.1 \times 10^{-5}$, Mann-Whitney *U* test) (Figure 3A, B). The low fractions of digestion-resistant PCR products in the TALEN-transformed plants compared to the CRISPR/Cas9-transformed plants suggests a lower intrinsic activity of TALENs compared to CRISPR/Cas9 and therefore less mutations are generated in each cell. Thus, mutations are not easily detected without predigesting genomic DNA as a mutation enrichment step. Additionally, TALEN-induced 5' overhang DSBs may be faithfully repaired more easily than the primarily blunt-end DSBs of Cas9, and more rounds of DSB induction and repair are needed for mutations to occur.

The substantial increase in the fraction of resistant PCR product in Cas9-CRU-2 plants might be explained by the slight difference in protospacer sequence between Cas9-CRU-1 and Cas9-CRU-2. In a study in *C. elegans* it was shown that sgRNA design ending with a GG sequence dramatically increased targeted mutagenesis frequencies (30). Furthermore, experiments in human cell lines also showed a slight increase in mutagenesis efficiency when the protospacer ended with GG (31). Our results corroborate with these findings when comparing Cas9-CRU-1 with Cas9-CRU-2, and high activity was also observed in our laboratory for CRISPR/Cas9-mediated targeted mutagenesis at the *Arabidopsis* protoporphyrinogen oxidase (*PPO*) gene (23, Chapter 3).

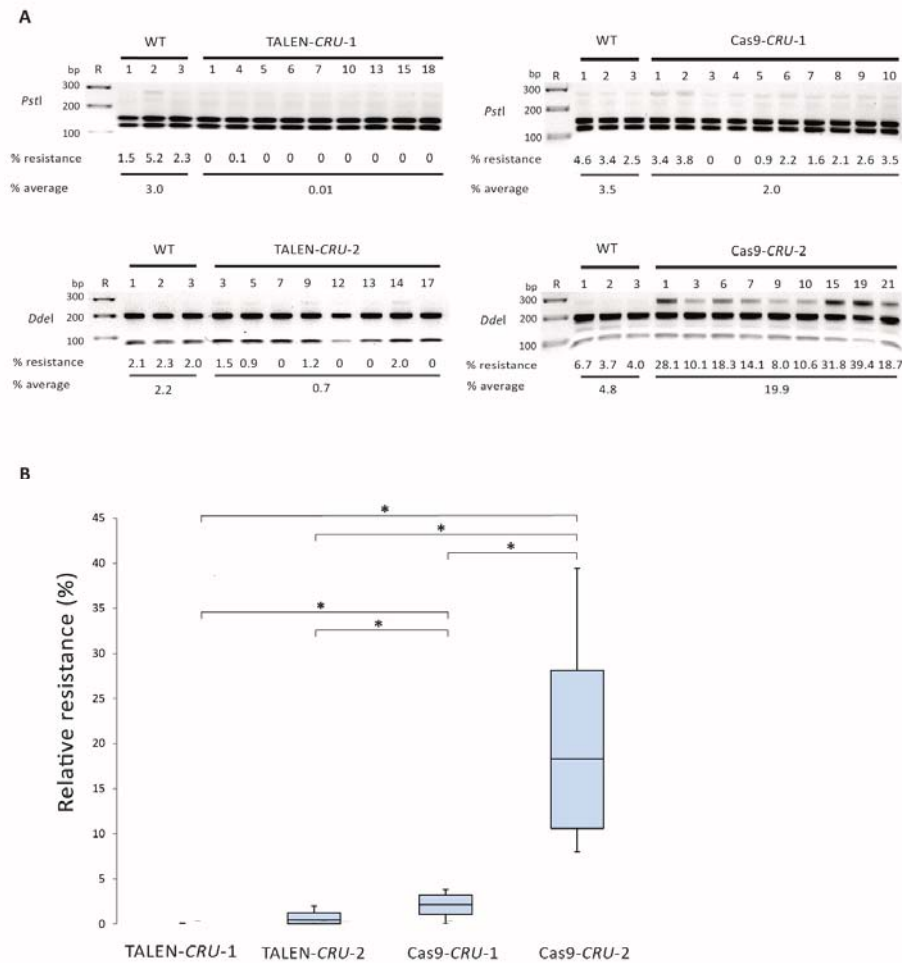


Figure 3. Semi-quantification of TALEN and CRISPR/Cas9 endonuclease-induced mutagenesis. **A.** The *CRU3* target was PCR-amplified from undigested genomic DNA from untransformed wild-type seedlings and T2 seedlings transformed with TALEN-CRU-1, TALEN-CRU-2, Cas9-CRU-1 and Cas9-CRU-2 constructs, resulting in a 272 bp fragment. PCR products were digested with *Pst*I (TALEN-CRU-1 and Cas9-CRU-1) resulting in a 147 bp and a 125 bp fragment or *Dde*I (TALEN-CRU-2 and Cas9-CRU-2), resulting in a 199 bp and a 73 bp fragment. The percentage of digestion resistant PCR fragment is shown below each lane. The fraction of *Pst*I- and *Dde*I-resistant PCR product is shown below each lane. This fraction is the normalized fraction obtained by subtracting the fraction seen in the wild-type samples from the original fraction measured in plant lines with each of the TALEN-CRU and Cas9-CRU constructs. R is the 100 bp+ ladder (Thermo Scientific). **B.** Box plots showing the average fraction of *Pst*I and *Dde*I-resistant PCR product in plant lines with the TALEN-CRU and Cas9-CRU constructs. The average fraction in plants with the Cas9-CRU-2 construct is significantly higher than the average fraction in plants with the Cas9-CRU-1 construct ($p = 1.1 \times 10^{-5}$), TALEN-CRU-1 construct ($p = 4.1 \times 10^{-5}$) and TALEN-CRU-2 construct ($p = 2.0 \times 10^{-5}$). The average fraction in plants with the Cas9-CRU-1 construct is also significantly higher than the average fraction in plants with the TALEN-CRU-2 construct ($p = 0.016$) and TALEN-CRU-1 ($p = 5.9 \times 10^{-4}$). *, p value ≤ 0.05 as determined by the one-tailed Mann-Whitney *U*-test.

Mutagenesis at the ADH1 locus

To investigate if a protospacer ending with GG also enhances CRISPR/Cas9 efficiency at another locus, we compared CRISPR/Cas9 efficiency at the *Arabidopsis* alcohol dehydrogenase 1 (*ADH1*) gene with two constructs, Cas9-ADH-1 and Cas9-ADH-2, using two slightly different protospacers. The protospacer of Cas9-ADH-1 was already successfully used in targeted mutagenesis and gene targeting experiments (24, 32). For Cas9-ADH-2 the protospacer was shifted 3 bp upstream so that it would end with GG (Figure 4A). PCR was performed on genomic DNA of pools of T2 seedlings from 12 plants transformed with the Cas9-ADH-1 construct and 12 plants transformed with the Cas9-ADH-2 construct, and PCR products were digested with *EaeI*. As done for the *CRU3* locus, the fractions of *EaeI*-resistant PCR product were corrected for background levels by subtracting the average fraction of *EaeI*-resistant PCR product seen in the wild type from the resistant fraction observed in each of the Cas9-ADH-transformed plant lines. The average fraction of *EaeI*-resistant PCR product in Cas9-ADH-2-transformed plants was with 18.3% significantly higher than that in Cas9-ADH-1-transformed plants with 4.7% on average ($p = 1,7 \cdot 10^{-5}$, Mann-Whitney *U* test), indicating that CRISPR/Cas9 efficiency was also enhanced at this locus when the protospacer ended with GG (Figure 4B, C). It should also be noted that because the Cas9-ADH-2 protospacer is shifted 3 bp upstream the *EaeI* site lies 3 bp further away from the Cas9 DNA cleavage site as compared to the Cas9-ADH-1 protospacer. Thus, larger deletions are necessary for Cas9-ADH-2 to detect *EaeI* resistance. Nevertheless, Cas9-ADH-2 still performed significantly better than Cas9-ADH-1.

It remains to be determined which sgRNA design is optimal in plants. The fact that increased Cas9 efficiency is observed in multiple organisms when using protospacers ending with GG may point towards a general protospacer design rule, although more in-depth analyses need to be performed in order to determine if this design is indeed more efficient in general.

Taken together, our results show that TALENs and CRISPR/Cas9 can be designed that are capable of inducing targeted DSBs at the *CRU3* and *ADH1* loci. However, the efficiency of the used TALEN constructs was very low. The CRISPR/Cas9 nucleases designed worked satisfactorily adequately at both the *CRU3* and *ADH1* loci, with a high increase in efficiency when using the Cas9-CRU-2 and Cas9-ADH-2 constructs with protospacers ending with GG. Given the fact that efficient targeted DSB induction is an important prerequisite for gene targeting, these results, together with the ease of use of the CRISPR/Cas9 system, made Cas9-CRU-2 and Cas9-ADH-2 the preferred nuclease constructs for further gene targeting experiments at the *CRU3* and *ADH1* genes

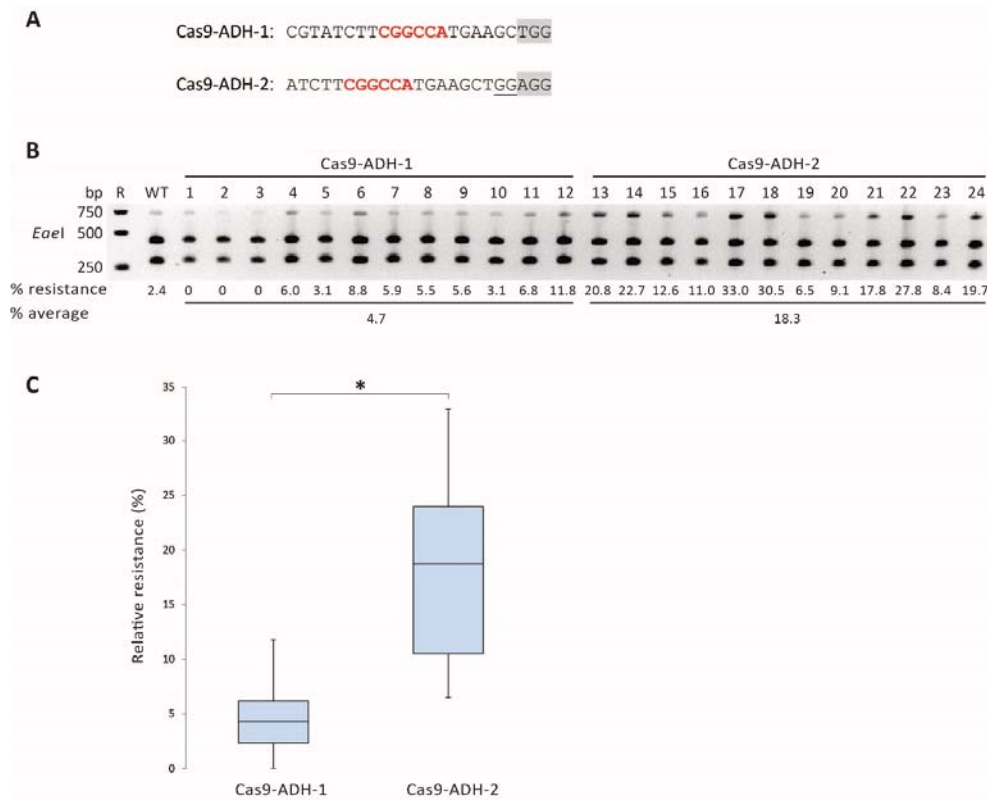


Figure 4. Semi-quantification of CRISPR/Cas9-mediated targeted mutagenesis at the *Arabidopsis ADH1* gene.

A. The protospacer of Cas9-ADH-2 ends with an GG motif (underlined), while the protospacer of Cas9-ADH-1 does not. PAM is highlighted in grey. The *EaeI* restriction enzyme recognition site is in red. **B.** The *ADH1* target was PCR-amplified from undigested genomic DNA from untransformed wild-type seedlings and pools of T2 seedlings of 12 plant lines transformed with Cas9-ADH-1 and 12 plant lines transformed with the Cas9-ADH-2 construct, resulting in a 717 bp fragment. PCR products were digested with *EaeI* resulting in a 429 bp and a 288 bp fragment. The fraction of *EaeI*-resistant PCR product is shown below each lane. This fraction is the normalized fraction obtained by subtracting the fraction seen in the wild-type samples from the original fraction measured in each Cas9-ADH line. R is the 1 kb ladder (Thermo Scientific). **C.** Box plots showing the average fraction of *EaeI*-resistance in plant lines with each of the Cas9-ADH constructs. Cas9-ADH-2 performs significantly better than Cas9-ADH-1 ($p = 1.7 \times 10^{-5}$). *, p value ≤ 0.05 as determined by the one-tailed Mann-Whitney *U*-test.

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 BSc internship students Manoah van der Velde, Kübra Kontbay, Jiyang Chan and Jun Lerou for their technical assistance. 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 (12428).

References

1. Weeks DP, Spalding MH, Yang B (2016) Use of designer nucleases for targeted gene and genome editing in plants. *Plant Biotechnol J* 14:483–495.
2. Bétermier M, Bertrand P, Lopez BS (2014) Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* 10(1):e1004086.
3. Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. *Int J Dev Biol* 57(6-7-8):629–637.
4. Voytas DF (2013) Plant Genome Engineering with Sequence-Specific Nucleases. *Annu Rev Plant Biol* 64(1):327–350.
5. Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc Natl Acad Sci U S A* 93(10):5055–5060.
6. Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J Exp Bot* 56(409):1–14.
7. Fauser F, et al. (2012) In planta gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
8. Roth N, et al. (2012) The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. *Plant J* 72(5):781–790.
9. Wei W, et al. (2012) A Role for Small RNAs in DNA Double-Strand Break Repair. *Cell* 149(1):101–112.
10. Ayar A, et al. (2013) Gene targeting in maize by somatic ectopic recombination. *Plant Biotechnol J* 11(3):305–14.
11. Puchta H, Fauser F (2014) Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J* 78(5):727–741.
12. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93(3):1156–1160.
13. Christian M, et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186(2):756–761.
14. Boch J, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326(5959):1509–12.
15. Moscou M, Bogdanove A (2009) Recognition by TAL Effectors. *Science* 326:1501.

16. Cermak T, et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39(12):e82.
17. Jinek M, et al. (2012) A programmable dual-RNA – guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–822.
18. Jinek M, et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997.
19. Nishimasu H, et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156(5):935–949.
20. Baltes NJ, Voytas DF (2015) Enabling plant synthetic biology through genome engineering. *Trends Biotechnol* 33(2):120–131.
21. Luo M, Gilbert B, Ayliffe M (2016) Applications of CRISPR / Cas9 technology for targeted mutagenesis , gene replacement and stacking of genes in higher plants. *Plant Cell Rep* 35(7):1439–1450.
22. Curtis MD, Grossniklaus U (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in plants. *Biotechnol Bioinform* 133:462–469.
23. Shen H, Strunks GD, Klemann BJPM, Hooykaas PJJ, de Pater S (2017) CRISPR/Cas9-induced double-strand break repair in *Arabidopsis* nonhomologous end-joining mutants. *G3:Genes|Genomes|Genetics* 7(1):193–202.
24. 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.
25. Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–743.
26. 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.
27. Schneider C a, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
28. Qi Y, et al. (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. *Genome Res* 23(3):547–54.
29. Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in *Arabidopsis* plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102(34):12265–12269.
30. Farboud B, Meyer BJ (2015) Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* 199(4):959–971.
31. Malina A, et al. (2015) PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. *Nat Commun* 6:10124.
32. 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.

Chapter 3

CRISPR/Cas9-induced double strand break repair in *Arabidopsis* nonhomologous end-joining mutants

Hexi Shen^{*1}, Gary D. Strunks^{*1,2}, Bart J. P. M.
Klemann^{1,3}, Paul J. J. Hooykaas¹, Sylvia de Pater¹

***These authors contributed equally to this work**

¹Molecular and Developmental Genetics, Institute of Biology,
Leiden University, 2333 BE, the Netherlands

²Current affiliation: Naktuinbouw, 2371 GD Roelofarendsveen,
the Netherlands

³Current affiliation: Dümme Orange, 2678 PS De Lier,
the Netherlands

This work is published as:

Shen H, Strunks GD, Klemann BJPM, Hooykaas PJJ, de Pater S (2017)
CRISPR/Cas9- induced double-strand break repair in *Arabidopsis*
nonhomologous end-joining mutants. *G3:Genes/Genomes/Genetics*
7(1):193–202.

Abstract

Double-strand breaks (DSBs) are one of the most harmful DNA lesions. Cells utilize two main pathways for DSB repair: homologous recombination (HR) and non-homologous end-joining (NHEJ). NHEJ can be subdivided into the KU-dependent classical NHEJ (c-NHEJ) and the more error-prone KU-independent backup-NHEJ (b-NHEJ) pathways, involving the poly (ADP-ribose) polymerases (PARPs). However, in absence of these factors, cells still seem able to adequately maintain genome integrity, suggesting the presence of other b-NHEJ repair factors or pathways independent from KU and PARPs. The outcome of DSB repair by NHEJ pathways can be investigated by using artificial sequence-specific nucleases such as CRISPR/Cas9 to induce DSBs at a target of interest. Here, we used CRISPR/Cas9 for DSB induction at the *Arabidopsis cruciferin 3* (*CRU3*) and *protoporphyrinogen oxidase* (*PPO*) genes. DSB repair outcomes via NHEJ were analysed using footprint analysis in wild-type plants and plants deficient in key factors of c-NHEJ (*ku80*), b-NHEJ (*parp1 parp2*) or both (*ku80 parp1 parp2*). We found that larger deletions of more than 20 bp predominated after DSB repair in *ku80* and *ku80 parp1 parp2* mutants, corroborating with a role of KU in preventing DSB end resection. Deletion lengths did not significantly differ between *ku80* and *ku80 parp1 parp2* mutants, suggesting that a KU- and PARP-independent b-NHEJ mechanism becomes active in these mutants. Furthermore, microhomologies and templated insertions were observed at the repair junctions in the wild type and all mutants. Since these characteristics are hallmarks of Polymerase θ -mediated DSB repair, we suggest a possible role for this recently discovered polymerase in DSB repair in plants.

Introduction

Double-strand breaks (DSBs) are one of the most lethal forms of DNA damage. DSBs can occur during normal cellular metabolism or can be induced by external factors, and highly threaten genomic integrity and cell survival (1). To repair DSBs, cells have two main pathways: Homologous Recombination (HR) and Non-Homologous End-Joining (NHEJ). Both of them function together to maintain genome integrity. NHEJ is the predominant pathway in higher eukaryotes and repair may lead to mutations at break sites, such as deletions, insertions and substitutions. At least two NHEJ pathways have been identified: the classic NHEJ pathway (c-NHEJ) and the backup-NHEJ pathway (b-NHEJ) also called alternative-NHEJ (a-NHEJ) or microhomology-mediated end-joining (MMEJ). The c-NHEJ is initiated by the recognition and binding of the KU heterodimer, consisting of KU70 and KU80 subunits, to DSBs (2). Once bound to a DSB, the KU heterodimer serves as a scaffold to recruit other c-NHEJ factors to the broken ends and promotes end-joining. Because KU is the key component of the c-NHEJ pathway, this pathway is also called KU-dependent NHEJ. In the absence of KU, other factors gain entry to the DSB site for repair by backup pathways. Although the b-NHEJ pathway was defined by involving multiple components, including poly (ADP-ribose) polymerase 1 (PARP1) the precise mechanism is still not clear (3). Furthermore, recently PARP1 was shown to be involved in repair of DSBs also in the presence of KU (4).

Nowadays, DSBs can be induced artificially at specific sites in the genome by sequence-specific artificial nucleases, which can be used to study DSB repair. These induced DSBs will be mainly repaired via NHEJ, which may lead to targeted mutagenesis. When repair restores the target site for the nuclease, the sequence will be cut again in the continuous presence of the nuclease. This cycle of cutting and repair will continue until incorrect repair destroys the target site. When a homologous sequence such as a sister chromatid, is present, DSB repair may also occur via HR, but this will inevitably also lead to restoration of the target site. A repair template without the target site may be provided by transformation or pre-inserted in the genome, and, when used for repair, lead to gene targeting (GT) (5, 6). The current genome editing tool kit comprises four classes of engineered nucleases: modified meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat/CRISPR-associated 9) system (5–7), of which the CRISPR/Cas9 system is the most easy and straightforward to use.

The CRISPR/Cas9 system is derived from an adaptive immune system present in bacteria and archaea, where it serves in degrading invading foreign plasmid or viral DNA. The type II CRISPR genomic locus encodes the Cas9 ('CRISPR-associated 9') endonuclease, which

can form a complex with two short RNA molecules: CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), which can be fused into a chimeric single-guide RNA (sgRNA) comprising the functions of both precursor RNAs (8). A sgRNA can be assembled to target any DNA sequence, with the prerequisite that a protospacer adjacent motif (PAM) sequence of NGG flanking the 3' end of the sgRNA target sequence is present, which interacts with the Cas9 PAM interacting domain (PI domain) (9, 10).

An in vitro study showed that the plant orthologs of PARP1 and PARP2 play a role in back-up end-joining, similar to its function in animals (11). However, the exact role of the PARP proteins in end-joining in plants is still unclear. Previous studies already demonstrated the feasibility of DSB-mediated targeted mutagenesis at artificial and endogenous loci in plants using ZFNs, TALENs and the CRISPR/Cas9 system (7, 12). Here, we investigate the role of KU80, PARP1 and PARP2 in NHEJ by using CRISPR/Cas9 for DSB-mediated targeted mutagenesis at the *Arabidopsis cruciferin 3* (*CRU3*) and *protoporphyrinogen oxidase* (*PPO*) genes. CRISPR/Cas9 nucleases were expressed in wild type and in *ku80*, *parp1 parp2* and *ku80 parp1 parp2* mutants. Footprint analysis in whole seedlings in the wild type and each of the three mutant genotype backgrounds demonstrated that key factors of NHEJ can affect the outcomes of targeted mutagenesis.

Materials and Methods

Plant material

The *ku80* (SALK_016627), *parp1* (GABI-Kat Line 692A05) and *parp2* (SALK_140400) T-DNA insertion lines (ecotype Col-0), the *parp1 parp2* double mutant and *ku80 parp1 parp2* triple mutant were described previously (11). More information about these lines can be found at <http://signal.salk.edu/cgi-bin/tdnaexpress> (13).

CRISPR/Cas9 vector construction and plant transformation

For the CRISPR/Cas9 constructs, oligo's SP509 and SP510 (*CRU3* target) and SP512 and SP513 (*PPO* target) (Table S2) were annealed and cloned in *Bbs*I digested pEn-Chimera (14). Subsequently, genes encoding sgRNAs were cloned in pDE-pUbi-Cas9 (14) by a Gateway LR reaction, resulting in Cas9-PPO (pSDM3905) and Cas9-CRU (pSDM3903), respectively.

Plant binary vectors were introduced into *Agrobacterium tumefaciens* strain AGL1 by electroporation. *Arabidopsis thaliana* plants of the Col-0 ecotype (wild type, *ku80*, *parp1 parp2*, *ku80 parp1 parp2*) were transformed with T-DNAs containing nuclease expression cassettes, using the floral dip method (15). T1 seeds were grown on MA solid medium without sucrose, supplemented with timentin (100 µg/mL), nystatin (100 µg/mL) and 15 µg/mL phosphinothricin for CRISPR/Cas9 T-DNA selection.

DNA isolation and footprint analysis

T2 seeds derived from independently selected T1 plants were germinated on ½ MS supplemented with 10 µg/mL phosphinothricin for T-DNA selection and pools of 10 seedlings per plant line of 10 days old were disrupted to a powder under liquid N₂ in a TissueLyser (Retch, Haan, Germany). Genomic DNA was extracted by the CTAB method (16). For predigestion, one µg of genomic DNA was digested with *Pst*I (Cas9-CRU analysis) or *Fau*I (for Cas9-PPO analysis) overnight and precipitated. Undigested or predigested DNA was used for PCR with Phusion polymerase (Thermo Scientific) to amplify the nucleases target sites, followed by digestion of the PCR products with *Pst*I or *Fau*I and separated in agarose gels. PCR primers are shown in Table S2. Primers SP245 and SP248 were used for the Cas9-CRU target region and primers SP392 and SP538 were used for the Cas9-PPO target region. The resistant fragments were isolated from gel, cloned into pJet1.2 (Thermo Scientific) and sequenced by MacroGen Europe (Amsterdam, The Netherlands). Identical sequences in the same line were considered

as one mutagenesis event since they might have resulted from PCR amplification. Two-tailed Mann-Whitney tests were performed for statistical analysis of deletion- and insertion lengths.

Estimation of relative number of mutations

To estimate the relative number of Cas9-induced mutations the target sites were amplified using undigested genomic DNA. PCR products were digested with the appropriate restriction enzymes and analysed on agarose gels. The intensity of bands was quantified using ImageJ software. The relative number of mutation was calculated by dividing the intensity of the digest-resistant band by the total intensity of all bands in a given lane (17).

High resolution melting

High resolution melting (HRM) analyses were performed on PCR clones of undigested DNA from T2 seedlings of wild-type lines Cas9-CRU #2 and Cas9-PPO #7 using Precision Melt Supermix (Bio-Rad), containing EvaGreen saturated dye, and the Bio-Rad C1000 Touch thermal cycler (Bio-Rad). Melt curves were analyzed using the Bio-Rad Precision Melt Analysis software. For the *CRU* target primers SP492 and SP563 were used and for the *PPO* target primers SP560 and SP561 (Table S2). Samples with various melt curves were sequenced by Macrogen Europe (Amsterdam, The Netherlands).

Data availability

Plasmids and plant lines are available upon request. Figure S1 contains phenotypic data. Figure S2 contains sequences of resistant target sites. Table S1 contains primer sequences. Table S2 contains deletion length distributions. Table S3 contains the number of insertions and templated insertions.

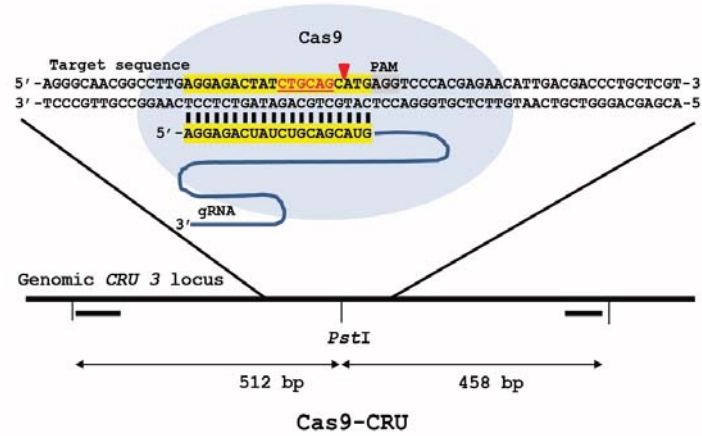
Results

DSB-mediated mutagenesis by CRISPR/Cas9 at the CRU3 and PPO loci

In order to investigate repair of induced DSBs, sequence-specific nucleases were designed and expressed in *Arabidopsis*. Wild-type plants were transformed with CRISPR/Cas9 expression constructs via the *Agrobacterium*-mediated floral dip method (15) and T2 transformants were used for further analysis. Nuclease target sites in the *CRU3* and *PPO* genes were selected. The *CRU3* gene encodes a seed storage protein. The *PPO* gene encodes an essential enzyme that is involved in the final step of chlorophyll biosynthesis, and mutagenesis of the *PPO* gene is therefore toxic to plants. Plants expressing nucleases targeted at *CRU3* showed a phenotype similar to wild-type, but T2 seedlings of some plant lines expressing Cas9-PPO showed a stunted growth phenotype indicative of homozygous inactivation of the essential *PPO* gene in many cells (Figure S1). To detect mutagenesis caused by nuclease activity and subsequent erroneous NHEJ-mediated DSB repair at the molecular level, T2 seeds of several individual transformants were germinated on phosphinotricin and pooled 10 days-old T2 seedlings were used for DNA isolation and analysis for the presence of NHEJ-induced indels. In order to discriminate DNA molecules with a mutation, PCR products from the region containing the target site were digested with restriction enzymes having a recognition site near the DSB site (*Pst*I for Cas9-CRU and *Fau*I for Cas9-PPO) (Figure 1). Loss of the restriction site as a consequence of erroneous repair resulted in restriction digest-resistant PCR products. After gel electrophoresis, the relative band intensities were measured to estimate the number of mutations in the target sites (Figure 2). Digestion of the PCR products from untransformed wild-type plants left some of the PCR products undigested, probably due to incomplete digestion. However, a distinguishably higher fraction of the PCR products from plant lines transformed with CRISPR/Cas9 nucleases were resistant to enzyme digestion (Figure 2A, B).

To get a better insight into the mutations induced by the CRISPR/Cas9 nucleases, the restriction digest-resistant PCR products of the *CRU3* and *PPO* targets were cloned and sequenced. Pre-digested genomic DNA was used for PCR to enrich for mutated sequences. Sequencing revealed mainly deletions and some insertions and substitutions in CRISPR/Cas9 lines (Figure 2C, Figure S2). Short homologous sequences on either the left or right side flanking the deletion were often also present, suggesting MMEJ may have been involved in DSB repair.

A



B

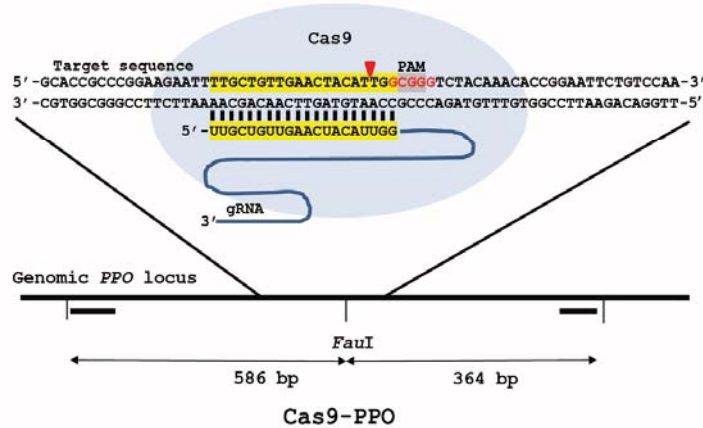


Figure 1. CRISPR/Cas9 endonucleases for DSB induction in *CRU3* and *PPO*. Cas9-CRU (A) with its protospacer in the *CRU3* locus and Cas9-PPO (B) with its protospacer in the *PPO* locus are shown. sgRNA DNA binding sequences are highlighted in yellow, the PAM sequence is highlighted in black and the *Pst*I and *Fau*I restriction sites are shown in red lettering. The primers (—) used to amplify the target regions and the sizes are indicated. Red arrows indicate the position of DSB induction.

In our experimental design for both Cas9-CRU and Cas9-PPO, the *Pst*I and *Fau*I restriction sites are nearby but do not overlap the DSB site. Therefore, the preselection of loss of a restriction site has the caveat of neglecting mutations that occur outside of the restriction site. To get a more precise insight into the types and frequencies of DSB-mediated mutations in *CRU3* and *PPO*, we used undigested DNA of T2 seedlings for qPCR followed by high resolution melting (HRM) for footprint analysis. Then indeed, also footprints outside of the restriction site were detected. For Cas9-CRU, HRM was performed on 142 PCR clones from T2 seedlings of line Cas9-CRU #2, which revealed one deletion of 32 bp outside of the *Pst*I site (results not shown). No footprints were detected in the remaining 141 clones, indicating a very low number of mutations in this plant line. For Cas9-PPO, HRM was performed on 48 PCR clones from T2 seedlings of line Cas9-PPO #7, which has a severe phenotype (Figure S1). Nine clones with footprints ranging from 1 bp insertion to 5 bp deletions were found outside the *Fau*I site (Figure 3). None of the 48 clones contained wild-type sequences, indicating a high number of mutations in this plant line.

Taken together, these results show that our CRISPR/Cas9 nuclease constructs are able to induce mutations at the target sites.

Increased DSB-mediated mutagenesis by CRISPR/Cas9 in c-NHEJ deficient mutants

To compare mutagenesis in wild type and NHEJ-compromised genetic backgrounds, T-DNA insertion lines *ku80*, *parp1 parp2*, and *ku80 parp1 parp2* as described previously (11, 18) were transformed with Cas9-CRU and Cas9-PPO nuclease constructs, and several independent primary transformants were obtained. The target region was PCR amplified using total genomic DNA from pooled T2 10 days-old seedlings of several plant lines as a template, followed by restriction enzyme digestion of the PCR product. The relative band intensity was measured to estimate the amount of mutations for each line. The resistant bands in most of the Cas9-CRU lines were hardly detectable. Clear resistant bands were, however, observed in Cas9-PPO lines (Figure 2B). Plants appeared normal. Apparently, DSB repair of the induced DSBs was still efficient (but less precise) even in the triple mutant *ku80 parp1 parp2*.

Larger deletions are predominant in c-NHEJ deficient mutants

To assess the outcomes of DSB repair at the nucleotide level in wild type and mutant lines, genomic DNA was pre-digested to enrich for mutated sequences and the resistant bands were purified, cloned and sequenced. The results showed that there were deletions, insertions and

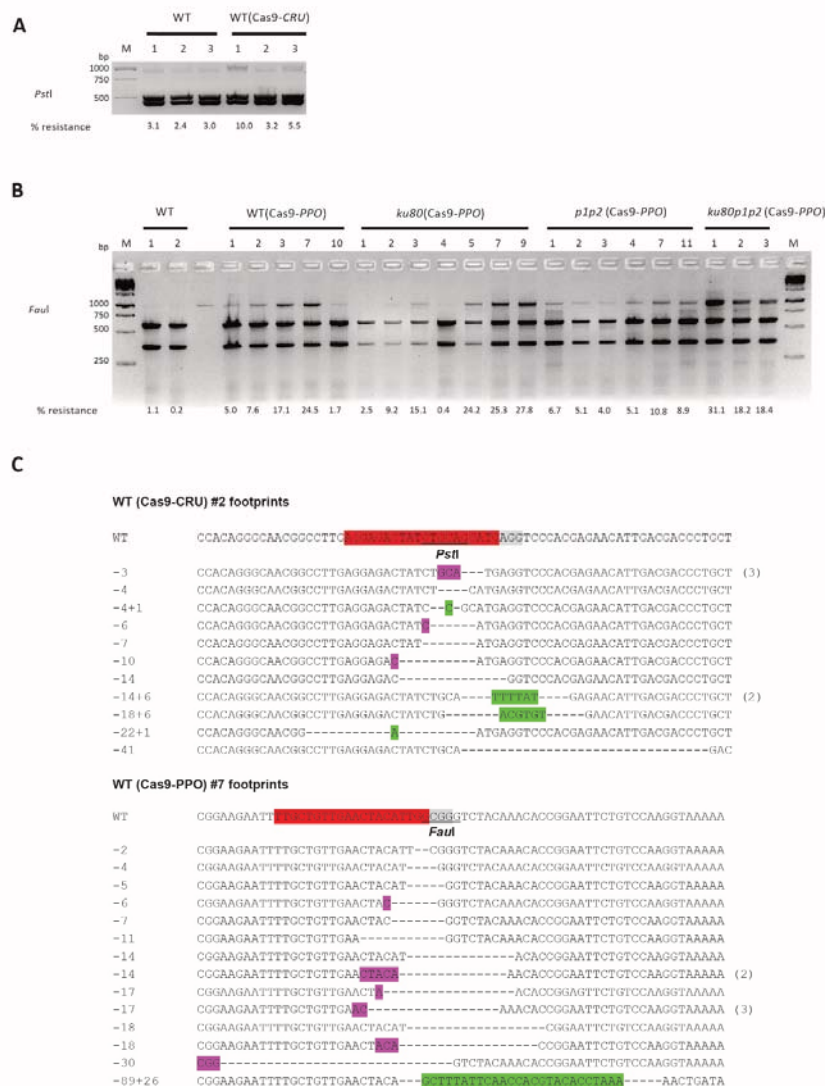


Figure 2. CRISPR/Cas9 endonucleases induced mutagenesis. **A.** The CRU3 target site was amplified using undigested genomic DNA from untransformed wild type seedlings and Cas9-CRU transformed T2 seedlings and digested with *PstI*. **B.** The *PPO* target site was amplified from untransformed wild type seedlings and Cas9-PPO transformed T2 seedlings of wild type and *ku80*, *parp1 parp2* (*p1p2*), and *ku80 parp1 parp2* (*k80p1p2*) mutant plant lines and digested with *FauI*. M is the 1 kb DNA marker, with sizes of the bands at the left, and the % resistant bands is shown below the lanes. **C.** Sequences of *CRU3* and *PPO* targets from Cas9-CRU transformant #2 and Cas9-PPO transformant #7. The sgRNA protospacer is in red, PAM sequence is in grey, restriction sites are underlined, deletions are shown by dashes, insertions are in green, microhomologies used for repair are in purple. Number of multiple clones with the same sequence are shown at the right. Numbers are length of deletions (-) and insertions (+).

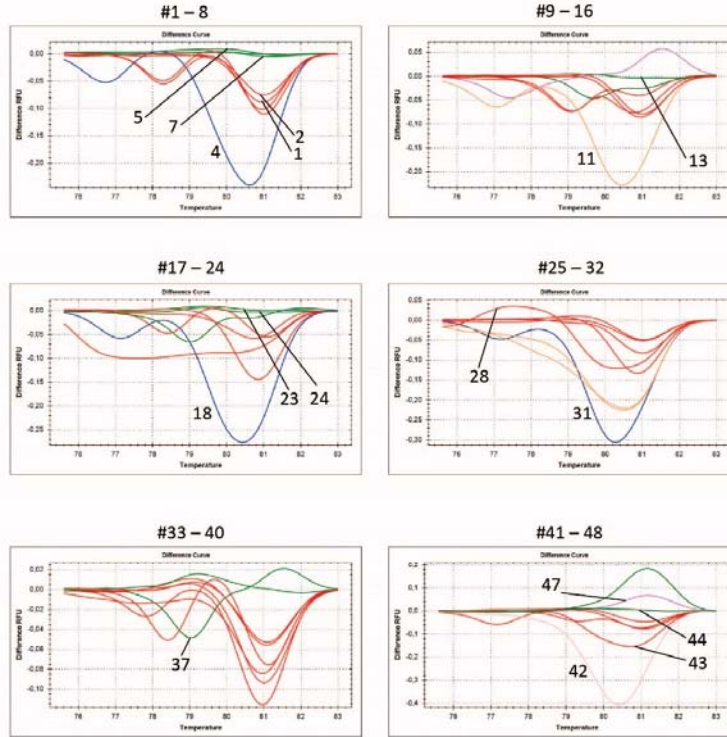
substitutions at the *CRU3* and *PPO* target sites (Figure S2). The majority of mutations recovered were deletions. Substitutions seem to be very rare events based on the sequenced data and these might be PCR artefacts.

We examined the length of deletions from all genotypes expressing Cas9-CRU or Cas9-PPO (Figure 4A, Figure S2, Table S3). Due to the loss of restriction site method, 1 to 2 bp deletions could not be detected in our experimental approach, and only larger deletions that overlap the restriction site were scored. Moreover, each mutation event was scored once for calculating the deletion size average. In wild-type Cas9-CRU transformants, 57% of deletions were <10 bp and about 23% ranged from 10 to 19 bp, 15% ranged from 20 to 49 bp and 5% were ≥50 bp. The *parp1 parp2* mutant lines showed somewhat longer deletions; about 25% of deletions ranged from 20 to 49 bp and 7% were ≥50 bp. Larger deletions were predominant in the *ku80* and *ku80 parp1 parp2* mutant lines. In *ku80* lines 62% of the deletions were larger than 20 bp (22% were ≥50 bp), and in the *ku80 parp1 parp2* lines 61% of the deletions were larger than 20 bp (12% were ≥50 bp).

Deletion lengths in Cas9-PPO transformants were also examined (Figure 4A). Similar to Cas9-CRU, there were no big differences in deletion length between the wild type and *parp1 parp2* mutant lines. In the wild type about 33% of deletions were <10 bp, 32% ranged from 10 to 19 bp, 22% ranged from 20 to 49 bp and 13% were ≥50 bp. In *parp1 parp2* lines 27% of deletions were <10 bp, 27% ranged from 10 to 19 bp, 24% ranged from 20 to 49 bp and 21% were ≥50 bp. Larger deletions of the *PPO* target were however, again predominant in *ku80* and *ku80 parp1 parp2* mutant lines. About 75% of deletions in *ku80* lines were larger than 20 bp, and about 73% of deletions in *ku80 parp1 parp2* lines were larger than 20 bp.

We performed statistical analysis using a two-tailed Mann-Whitney test, to find out whether the observed differences were significant. For the Cas9-CRU and Cas9-PPO nucleases, comparison of deletion lengths in wild type to *ku80* and *ku80 parp1 parp2* lines showed a statistically significant difference, whereas comparison of deletion lengths in wild type to *parp1 parp2* lines did not (Figure 4B). These results indicate that imprecise end-joining after loss of the c-NHEJ key component KU80 resulted in substantial increases in deletion length and suggests a shift to a more error-prone repair mechanism of DSB repair in absence of KU80.

A



B

WT GGAAGAATT GGGCTTAACTAC GGGCTTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAGC
FauI

1 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG +1
2 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG +1
4 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG +1
5 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
7 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
11 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -14+1
13 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
18 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -14
23 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
24 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
28 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
31 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -14
37 GGAAGAATTTTGTGTTGAACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -5
42 GGAAGAATTTTGTGTTGAACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -17
43 GGAAGAATTTTGTGTTGAACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -9
44 GGAAGAATTTTGTGTTGAACACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -1
47 GGAAGAATTTTGTGTTGAACATTTGGCGGGTCTACAAACACCGGAATCTGTCCAAGGTAAAAACAG -4

Figure 3. HRM analysis of the *PPO* target. HRM analyses were performed on 48 PCR clones from undigested DNA of a pool of 10 T2 seedlings of wild-type Cas9-*PPO* transformant #7. **A.** Melt curves of samples 1 to 48 measured in relative fluorescence units (RFU). Numbers indicated in the graph refer to the sequences below. **B.** Sequences of representative *PPO* targets. *PPO* sgRNA protospacer (red), the PAM sequence (gray) and *FauI* restriction site (underlined) are indicated in the WT sequence. Footprints included deletions (dashed lines) and insertions (green). Microhomologies used for repair are shown in purple.

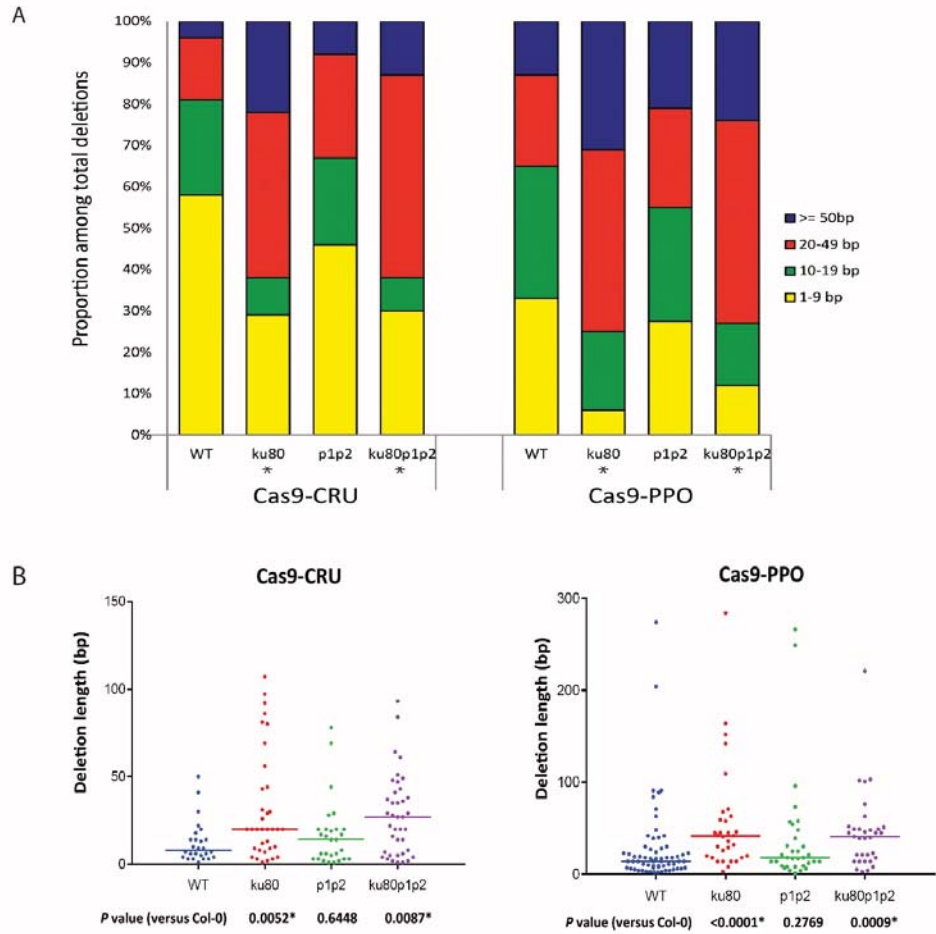
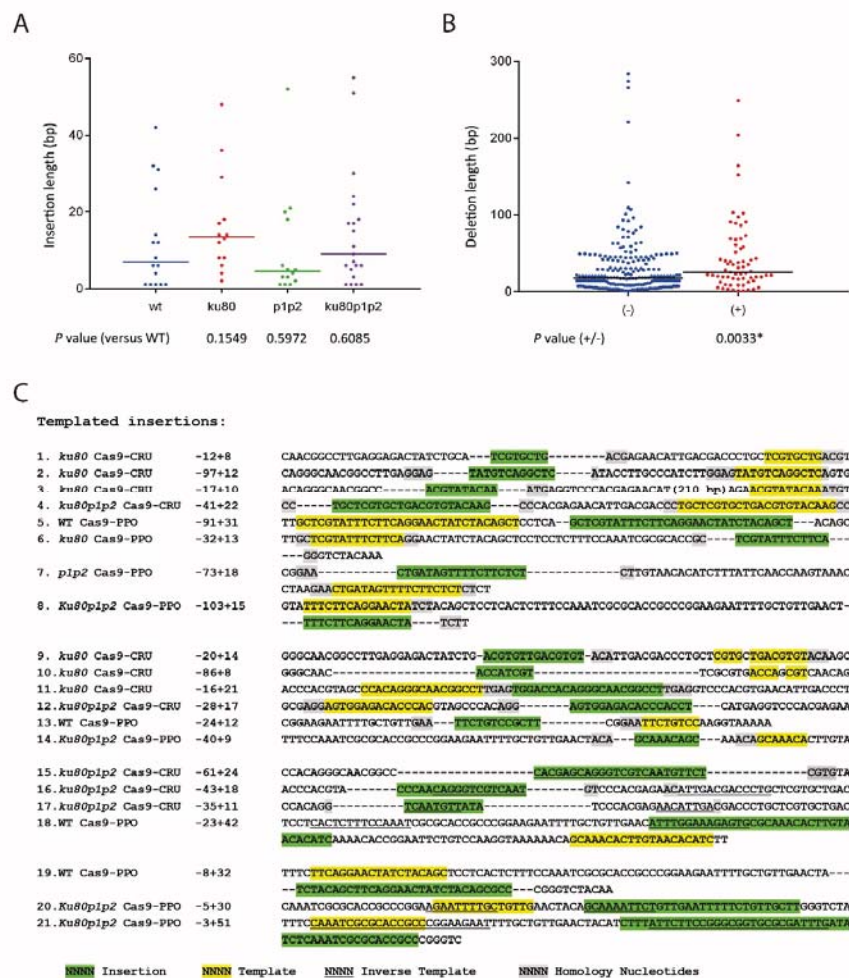


Figure 4. Analysis of deletion length. **A.** Distribution of deletion lengths of mutated sequences obtained for the indicated genotypes with Cas9 nucleases. **B.** Scatter plot of deletion lengths of the sequences used in (A). Median deletion lengths are indicated by horizontal lines. P-values are derived from a two-tailed Mann-Whitney test. Asterisks (*) indicate a statistically significant difference from wild-type ($P < 0.05$).

Templated insertions in wild type and NHEJ mutants

Insertion events, sometimes accompanied by deletions, were observed at the target loci in Cas9-CRU and Cas9-PPO transformed wild type and mutant lines, although less frequently than deletions. The number of insertion events in mutants was comparable to or higher than the number of insertion events in the wild type. More than half of the insertions were smaller than 10 bp. A maximum insertion length of 60 bp was observed. Furthermore, insertion lengths in NHEJ mutant lines were not significantly different from those in wild type when insertion data of both targets were combined, indicating that the insertion mechanism may be independent of KU80 and PARP (Figure 5A). In addition, from the combined data of both targets it can be deduced that the deletion length of junctions with insertions were significantly larger than junctions without insertions (Figure 5B).

Interestingly, many inserted sequences have at least one match to DNA within 100 bp of the repaired DSB. Some insertions have complex compositions with multiple stretches of identity, including reverse complementary homology. These results suggest that polymerase θ may be involved in the repair of these DSBs (19). Another signature of Pol θ -mediated DSB repair is the presence of sequence identity between the 3' end that generated the junction (the primer) and the sequence immediately upstream of the template that is used for DNA synthesis. Such sequence identity is present in about 50 % of the inserted sequences (Figure 5C). The *ku80* and *ku80 parp1 parp2* mutant lines appeared to have more templated-insertion events than wild type and *parp1 parp2* lines, although such insertions were found in all four genotypes (Table S4). Therefore, the templated insertions probably resulted from a KU80- and PARPs-independent alternative end-joining mechanism, such as that mediated by the recently discovered Pol θ (20).



Discussion

In this study, we demonstrated that CRISPR/Cas9-induced DSBs can be successfully repaired in *Arabidopsis*, even after loss of key components of the NHEJ repair pathway. In general, plants expressing these nucleases look healthy and develop normally (unless an essential gene was targeted). As precise repair restores the break site leading to an intact substrate for the nuclease, a cycle of DSB induction and repair continues until mutations in the target sequence prevent the action of the nuclease. We showed that especially the Cas9-PPO nuclease was a very efficient tool for targeted mutagenesis, with close to 100% mutation of the target site in one line. Since the *PPO* gene is an essential gene, this resulted in stunted growth of the seedlings (Figure S1). This phenotype was not observed in earlier targeted mutagenesis experiments of *PPO* using ZFNs (20), indicating a higher activity of CRISPR/Cas9 on the *PPO* gene compared to the ZFNs. We noticed that the sgRNA recognized a sequence in the *PPO* gene with GG just 5' of the PAM sequence, which was recently shown to promote higher rates of mutagenesis (22).

As a result of imperfect end-joining, various mutations in the target sites were found in each line. The DSB sites of our CRISPR/Cas nucleases did not overlap the restriction sites that were used. Therefore, very small deletions or insertions could not be detected in our experimental approach. In the HRM approach we did observe several 1 bp deletions and insertions that were missed by the loss of restriction site method. Such mutations have been found to be present in high frequency in other studies (14). However, our method is useful to detect differences between wild type and mutants. In the analysis of DSB repair outcome in NHEJ mutants, we observed a statistically significant increase in the median deletion length at the repair junction in the *ku80* and *ku80 parp1 parp2* mutants compared to wild type, but not in the *parp1 parp2* mutant. This suggests that, in the absence of KU, cells shift to more error-prone end-joining mechanisms. KU is known to competitively bind to DSB ends and protects break ends from end processing (23, 24). Thus, when KU is absent, DNA ends are exposed to end resection proteins which would promote the generation of larger deletions. Similar results have been described previously with ZFNs-induced DSB repair in a *ku80* mutant and in *ku70* and *lig4* mutants (25, 26).

We previously showed that PARP1 and PARP2 are involved in the MMEJ repair pathway by an *in vitro* end-joining assay in *Arabidopsis* (11). In the *in vivo* end-joining experiments described here, however, we did not observe a role for PARP1 and PARP2 in MMEJ, and therefore there must be another repair pathway independent of PARP1 and PARP2 that uses microhomology. It is still elusive whether b-NHEJ is a single pathway or a category containing multiple mechanisms (1). The similar mutation characteristics observed in the *parp1*

parp2 mutant and the wild type supports the dominant role of the KU-dependent c-NHEJ pathway rather than the PARP-dependent b-NHEJ pathway. Notably, we did not observe much differences in junction characteristics between *ku80* and *ku80 parp1 parp2* mutants, indicating other repair pathways (independent of PARP) with similar characteristics become active when c-NHEJ is absent. However, we cannot rule out that other factors, for example PARP3 (27, 28), could slip into the b-NHEJ pathway without disturbing its outcome.

Insertions were found at both c-NHEJ-proficient and -deficient repair junctions, although most junctions were repaired without an insertion. *ku80* mutants had more insertion events than the wild type. However, the median insertion length in break junctions is a few base pairs and no statistically significant difference was observed among mutants and wild type. Besides, larger deletions (of more than 20 bp) were found with insertions compared to those in repair products without insertions during Cas9-induced repair. Templated insertions were observed both in c-NHEJ efficient and c-NHEJ deficient mutants in animal cells. The current models of templated mutagenesis are based on a MMEJ mechanism involving DNA Polymerase θ (19, 29, 30). In plants, templated insertions were also observed after DSB induced repair in *Arabidopsis* and tobacco (31–35). Furthermore, a recent study showed that the *Arabidopsis* Pol θ ortholog *Tebichi* (*Teb*) is essential for T-DNA integration (20). Templated insertions were found at the repair junctions of T-DNA inserts, and it was shown that *teb* mutants were resistant to T-DNA integration and very sensitive to the DNA damaging agents bleomycin and MMS. Our results indicate that nuclease-induced DSBs may be repaired by Ku-dependent NHEJ, or a backup pathway in the absence of Ku, leading to larger deletions in the latter case. Templated insertions, which have the hallmarks of Pol θ -mediated repair, may be formed in both cases, but with a higher frequency in the absence of Ku, revealing a complex interplay of repair factors during DSB repair in *Arabidopsis*.

Acknowledgements

We thank the Puchta lab for supplying the pEn-Chimera and pDe-Cas9 vectors. We also like to thank BSc internship students Manoah van de Velde and Kübra Kontbay, and MSc internship student Kelly van Kooperen for their technical assistance. 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 (12428). HS was supported by a scholarship from the Chinese Scholarship Council (CSC) and PJJH by the fellowship associated with the appointment as KNAW Academy Professor.

References

1. Deriano L, Roth DB (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet* 47:433–55.
2. Walker JR, Corpina RA, Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412:607–614.
3. Wang M, et al. (2006) PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res* 34:6170–6182.
4. Luijsterburg MS, et al. (2016) PARP1 links CHD2-mediated chromatin expansion and H3.3 deposition to DNA repair by non-homologous end-joining. *Mol Cell* 61(4):547–562.
5. Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. *Int J Dev Biol* 57(6–7–8):629–637.
6. Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. *Annu Rev Plant Biol* 64(1):327–350.
7. Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V (2015) Editing plant genomes with CRISPR/Cas9. *Curr Opin Biotechnol* 32:76–84.
8. Jinek M, et al. (2012) A programmable dual-RNA – guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–822.
9. Jinek M, et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997.
10. Nishimasu H, et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156(5):935–949.
11. Jia Q, den Dulk-Ras A, Shen H, Hooykaas PJJ, de Pater S (2013) Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in *Arabidopsis thaliana*. *Plant Mol Biol* 82(4–5):339–51.
12. Puchta H, Fauser F (2013) Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J* 78(5):727–741.
13. Alonso JM (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301(5633):653–657.

14. 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.
15. Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–743.
16. 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.
17. Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31(8):691–693.
18. Jia Q, Bundock P, Hooykaas PJJ, de Pater S (2012) *Agrobacterium tumefaciens* T-DNA integration and gene targeting in *Arabidopsis thaliana* non-homologous end-joining mutants. *J Bot* 2012:1–13.
19. Roerink SF, Schendel R, Tijsterman M (2014) Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res* 24(6):954–962.
20. van Kregten M, et al. (2016) T-DNA integration in plants results from Polymerase Theta-mediated DNA repair. *Nat Plants* 2:1–6.
21. 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.
22. Farboud B, Meyer BJ (2015) Dramatic enhancement of genome editing by CRISPR/cas9 through improved guide RNA design. *Genetics* 199(4):959–971.
23. Downs J a, Jackson SP (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* 5:367–378.
24. Fell VL, Schild-Poulter C (2015) The Ku heterodimer: Function in DNA repair and beyond. *Mutat Res - Rev Mutat Res* 763:15–29.
25. Osakabe K, Osakabe Y, Toki S (2010) Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. *Proc Natl Acad Sci U S A* 107(26):12034–9.
26. Qi Y, et al. (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. *Genome Res* 23(3):547–54.
27. Rulten SL, et al. (2011) PARP-3 and APLF function together to accelerate nonhomologous end-joining. *Mol Cell* 41:33–45.
28. Langelier MF, Riccio A a., Pascal JM (2014) PARP-2 and PARP-3 are selectively activated by 5' phosphorylated DNA breaks through an allosteric regulatory mechanism shared with PARP-1. *Nucleic Acids Res* 42:7762–7775.
29. McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24(11):529–38.
30. Koole W, et al. (2014) A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun* 5:3216.
31. Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. *Plant Cell* 4(3):333–347.

32. Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res* 25(22):4650–7.
33. Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* 17(20):6086–95.
34. Lloyd AH, Wang D, Timmis JN (2012) Single molecule PCR reveals similar patterns of non-homologous DSB repair in tobacco and arabidopsis. *PLoS One* 7(2). doi:10.1371/journal.pone.0032255.
35. Vu GTH, et al. (2014) Repair of site-specific DNA double-strand breaks in barley occurs via diverse pathways primarily involving the sister chromatid. *Plant Cell* 26(5):2156–2167.

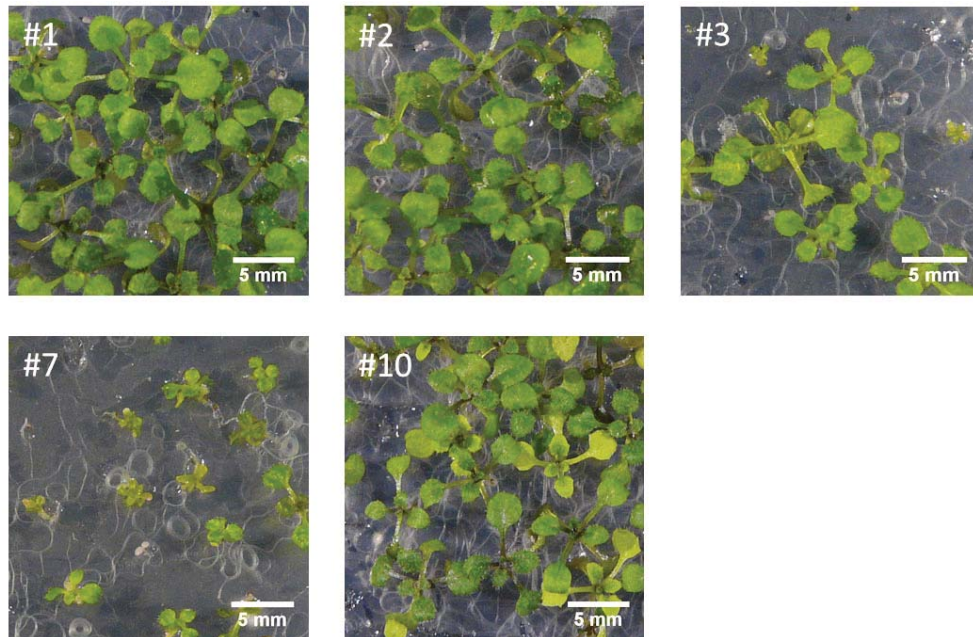


Figure S1. Phenotypes of 16-day-old T2 seedlings of 5 independent Cas9-*PPO* transformants. A stunted growth phenotype is observed in some seedlings of lines #3 and #7. The other lines have a phenotype similar to wild-type.

1. CRISPR/Cas9-CRU

WT Cas9-CRU

plant line 1

```

-3 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-3 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-6 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-7 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-9 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-13 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-14 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-30 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-50 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)

```

plant line 2

```

-3 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (3)
-4 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-4+1 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-6 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-7 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-10 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-14 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-14+6 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-18+6 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-22+1 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-41 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC

```

plant line 6

```

-1 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-3 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-5 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-6 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-9 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC

```

ku80 Cas9-CRU

plant line 1

```

-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-29 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-20+14 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-43 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)
-92 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC

```

plant line 2

```

-86+8 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-2 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-3 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-4 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-8 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-10 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-12+8 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-56+36 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC

```

plant line 3

```

-9 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-30 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-81 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC

```

plant line 4

```

-1 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-3 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-7 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-31+6 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC

```

plant line 6

```

-26 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (6)

```

plant line 9

```

-20 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (2)

```

plant line 11

```

-61+48 GACTATCTCCAGTCCATGAGGTTGATAAAGGCTGTAGATAGTTGGCTTCATGGAATAACGCTG (4)

```

plant line 12

```

-80 CGTGAGGCCACCTGATGAGTCCACGAGAACATTGACGACCTGCTCGTGTGAC
-97+12 CCACAGGGCAACGGCCTTGAGGAGACTATCTGAGGTTCCACGAGAACATTGACGACCTGCTCGTGTGAC (3)

```

Figure S2. Sequences of resistant target sites. DNA from several plant lines of wild type and NHEJ mutants with different CRISPR/Cas9 constructs was predigested with the appropriate restriction enzyme (*PsiI* for Cas9-CRU and *FauI* for Cas9-PPO), used for PCR, digested with the same enzyme and resistant products were cloned and sequenced. Footprints included deletions (dashed lines), insertions (green) and substitutions (blue). Microhomologies used for repair are shown in purple. Number of multiple clones with the same sequence are shown at the right. Numbers are length of deletions (-) and insertions (+).

<i>pip2</i> Cas9-CRU		
plant line 1	-3	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---TGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-6	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---ATGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-8	CCACAGGGCAACGGCCTTGAGGAGACTAT---TGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-17	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---AGAACATTGACGACCCCTGCTCGTCTGAC
	-20	CCACAGGGCAACGGCCTTGAGG---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (3)
	-16+21	CCACAGGGCAACGGCCTTGACGTGACCCACAGGGCAACGGCTTGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTG
	-69+52	TATCT---ACAGTCCATGAGGTTCTATAAGGTCTGTAGATAGTTGGCTTCTATGGAGTA---GCCTGA (3)
plant line 2	-1	CCACAGGGCAACGGCCTTGAGGAGACTATCTG-AGCATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-2	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---ATGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-20	CCACAGGGCAACGGCCTTGAGG---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-17+10	CCACAGGGCAACGGCC---ACGTATACAA---ATGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-78	ACCTCTAAGACAGCAGC---78---CACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
plant line 4	-3	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---TGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-6	CCACAGGGCAACGGCCTTGAGGAGACTAT---ATGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (3)
	-44	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---44---ACGTGTAC (4)
plant line 8	-3	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---TGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-6	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---ATGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-20	CCACAGGGCAACGGCCTTGAGG---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-28	CCACAGGGCAAC---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
plant line 11	-1	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA-CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-3	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---TGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-5+2	CCACAGGGCAACGGCCTTGAGGAGACTAT---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
plant line 15	-19+5	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---CTTAA---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-2	CCACAGGGCAACGGCCTTGAGGAGACTATCTGC---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-14	CCACAGGGCAACGGCCTTGAGG---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-14	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---CCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-20	CCACAGGGCAACGGCCTTGAGG---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-29	CCACAGGGCAACGG---CACGAGAACATTGACGACCCCTGCTCGTCTGAC
<i>ku80p1p2</i> Cas9-CRU		
plant line 1	-3	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA---TGAGGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-4	CCACAGGGCAACGGCCTTGAGGAGACTATCT---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (3)
	-8	CCACAGGGCAACGGCCTTGAGGAGACTATCT---AGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-22	CCACAGGGCAACGGCCTTGAGGAGACTAT---GAACATTGACGACCCCTGCTCGTCTGAC (4)
	-29	CCACAGGGCAACGG---CACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-35+11	CCACAGG---TCAATGTTATA---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-84	TAAGACAGCCCTACGAGAGCGA---84---CCCTGCTCGTCTGAC
plant line 2	-93	TTTCAGGTT---93---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-1	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA-CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-2+1	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA-ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-5	CCACAGGGCAACGGCCTTGAGGAGACTATCT---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (2)
	-16	CCACAGGGCAACGGCCTTGAGGAGACTATCT---ACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-20	CCACAGGGCAACGGC---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-35+3	CCACAGGGCAACGGCCTTGAGGAGAC---GAC---ACCTGCTCGTCTGAC
	-47	GGAGACACCCACG---47---CCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
plant line 3	-49	CCACAGGGCAACGGCCTTGAGGAGACTATCTG-49---TACAAAGCCAGCCCTAGCTCGCTGAC
	-4	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-6	CCACAGGGCAACGGCCTTGAGGAGACTATCT---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-27	CCACAGGGCAACGGCCTTGAGGAGACTAT---TGACGACCCCTGCTCGTCTGAC
	-36	CCAC---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
plant line 6	-5	CCACAGGGCAACGGCCTTGAGGAGACTATC---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-14	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---CCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-20	CCACAGGGCAACGGCCTTGAGG---TCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-41+22	CC---TGCTGCTGACGCTGACAG---CCACGAGAACATTGACGACCCCTGCTCGTCTGAC
plant line 7	-61+24	CCACAGGGCAACGGCC---CACAGCAGGGTCTCAATGTTCT---CCTGTA
	-145	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCAACGGCATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-14	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---CCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-38+6	CCACAGGGCAACGGCCTTGAGGAGA---CTTGA---CCTGCTGAC (2)
	-48	CCACAGGGCAACG---ACCTGCTCGTCTGAC
plant line 13	-37	CCAC---GTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC (5)
plant line 15	-27	CCACAGGGCAACGGCCTTGAGGAGA---ACATTGACGACCCCTGCTCGTCTGAC (2)
	-28+17	CCACAGG---AGTGGAGACACCCACCC---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-51+6	CCACAGGGCAACGGCCTTGAGGAGACTATCTG---ACGTG---CAAGCCACGCC
	-64	GGAGGAGTCTGAGA---64---ACATTGACGACCCCTGCTCGTCTGAC
plant line 16	-2+1	CCACAGGGCAACGGCCTTGAGGAGACTATCTGCA-ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-7	CCACAGGGCAACGGCCTTGAGGAGACTAT---CATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-20	CCACAGGGCAACGGC---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-29	CCACAGG---ATGAGTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC
	-43+18	ATCCACGTA---GCCAACAGGGTCTGCAAT---GTCCCACGAGAACATTGACGACCCCTGCTCGTCTGAC

Figure S2. Continued

2. CRISPR/Cas9-PPO

WT Cas9-PPO

plant line 1

```

-5 CGGAAGAATTTTGGCTGTTGAACATACATTG-----TCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-3 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA (2)
-7 CGGAAGAATTTTGGCTGTTGAACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-10 CGGAAGAATTTTGGCTGTTGAACAT---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATACA-----AACACCGGAATTCGTCCAAGGTAAAAA (5)
-18 CGGAAGAATTTTGGCTGTTGAACATACA-----CGGAATTCGTCCAAGGTAAAAA
-19+4 CGGAAGAATTTTGGCTGTTGAACATACA-----CACG---CGGAATTCGTCCAAGGTAAAAA
-19+1 CGGAAGAATTTTGGCTGTTGAACATACA-----A---CGGAATTCGTCCAAGGTAAAAA
-24+12 CGGAAGAATTTTGGCTGTTGAA---TCTGTGCGCTT---CGGAATTCGTCCAAGGTAAAAA
-27 CGGAAGAATTTTGGCTGTTGAA---ACAAACACCGGAATTCGTCCAAGGTAAAAA

```

plant line 2

```

-63 CGGAAGAATTTTGGCTGTTGAACATACAT---AACACCGGAATTCGTCCAAGGTAAAAA
-84 GCTCGTATTCTCTG---AACACCGGAATTCGTCCAAGGTAAAAA
-3 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-4 CGGAAGAATTTTGGCTGTTGAACATACA---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-6 CGGAAGAATTTTGGCTGTTGAACATAC---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA (6)
-9 CGGAAGAATTTTGGCTGTTGAACATAC---TCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA (2)
-28 CGGAAGAATTTTGGCTGTTGAA---ATTCTGTCCAAGGTAAAAA
-30 CGG---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA (2)
-40 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA
-48 CGGAAGAATTTTGGCTGTTGAACATAC---AACACTGTGAACATCTTTA
CTCGTATTCTTCAGGAACAT---71---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
TTTCTTCAGGAACATCTACAGCTCCCTCACTTTTCCAAATCGGGCAGCGCCGGAAGAATTTTGGCTG
TTGAACAT---TGTGAGGCTGCGGAAGATCTACAGGCG---CGGGTCTACAA
TCCTCACTCTTTCCAAATCGGCACCGCCGGAAGAATTTTGGCTGTTGAACTTTGGAAGAGTGC
-8+32 -23+42

```

plant line 3

```

-3+1 CGGAAGAATTTTGGCTGTTGAACATACATTG---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-4 CGGAAGAATTTTGGCTGTTGAACATACAT---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-6 CGGAAGAATTTTGGCTGTTGAACATACAT---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA (3)
-9 CGGAAGAATTTTGGCTGTTGAA---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-11 CGGAAGAATTTTGGCTGTTGAA---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-11 CGGAAGAATTTTGGCTGTTGAA---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACTAC---AACACCGGAATTCGTCCAAGGTAAAAA (6)
-14 CGGAAGAATTTTGGCTGTTGAACATACAT---AACACCGGAATTCGTCCAAGGTAAAAA
-17 CGGAAGAATTTTGGCTGTTGAACTAC---CTACAAACACCGGAATTCGTCCAAGGTAAAAA
-21 CGGAAGAATTTTGGCTGTTGAACTAC---AACCGGAATTCGTCCAAGGTAAAAA
-22 CGGAAGAATTTTGGCTGTTGAA---ACAAACACCGGAATTCGTCCAAGGTAAAAA
-23 CGGAAGAATTTTGGCTGTTGAACTAC---CGGAATTCGTCCAAGGTAAAAA
-42 CGGAAGAATTTTGGCTGTTGAACTAC---TCCAAGGTAAAAA
-42+8 CGGAAGAATTTTGGCTGTTGAACATAC---CGGAATTC---AACAGCAACACTTGTG
CAGGAATCTACAGCTCTCTCA---GCTCGTATTCTTCAGGAACATCTACAGC---ACAGCA
CGGAAGAATTTTGGCTGTTGAACATAC---CGGAATTCGTGTC---GATCCACT
-274 TTTCCAA---AAGCCATTCTCAGTTCTTAGTGGTCA
CGGAAGAATTTTGGCTGTTGAACATACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-4 CGGAAGAATTTTGGCTGTTGAACATACAT---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-5 CGGAAGAATTTTGGCTGTTGAACATACAT---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-6 CGGAAGAATTTTGGCTGTTGAACATAC---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-7 CGGAAGAATTTTGGCTGTTGAACATAC---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-11 CGGAAGAATTTTGGCTGTTGAA---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATACAT---AACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA (2)
-17 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAGTTCGTCCAAGGTAAAAA
-17 CGGAAGAATTTTGGCTGTTGAACTAC---AACACCGGAATTCGTCCAAGGTAAAAA (3)
-18 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGAATTCGTCCAAGGTAAAAA
-18 CGGAAGAATTTTGGCTGTTGAACATAC---CGGAATTCGTCCAAGGTAAAAA
-30 CGG---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-89+26 CGGAAGAATTTTGGCTGTTGAACATAC---GCTTATTCAGGCGGGAAGTAA---AATGATA

```

plant line 7

```

-2 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-4 CGGAAGAATTTTGGCTGTTGAACATACAT---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-5 CGGAAGAATTTTGGCTGTTGAACATACAT---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-6 CGGAAGAATTTTGGCTGTTGAACATAC---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-7 CGGAAGAATTTTGGCTGTTGAACATAC---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-11 CGGAAGAATTTTGGCTGTTGAA---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATACAT---AACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA (2)
-17 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAGTTCGTCCAAGGTAAAAA
-17 CGGAAGAATTTTGGCTGTTGAACTAC---AACACCGGAATTCGTCCAAGGTAAAAA (3)
-18 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGAATTCGTCCAAGGTAAAAA
-18 CGGAAGAATTTTGGCTGTTGAACATAC---CGGAATTCGTCCAAGGTAAAAA
-30 CGG---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-89+26 CGGAAGAATTTTGGCTGTTGAACATAC---GCTTATTCAGGCGGGAAGTAA---AATGATA
-2 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
CGGAAGAATTTTGGCTGTTGAACATACAT---GGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-4 CGGAAGAATTTTGGCTGTTGAACATACAT---CTACAAACACCGGAATTCGTCCAAGGTAAAAA
-9 CGGAAGAATTTTGGCTGTTGAACATAC---GGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA (4)
-19+1 CGGAAGAATTTTGGCTGTTGAACTAC---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-39 CGGAAGAATTTTGGCTGTTGAACTAC---42---AACAGCAACCA
-91+12 CGTATTCTTCAGGAAC---91---GAGACCCGTGAG---CTGTCCAAGGTAAAAA

```

Ku80 Cas9-PPO

plant line 1

```

-20 CGGAAGAATTTTGGCTGTTGAACATAC---CGGAATTCGTCCAAGGTAAAAA
-32+13 CTACAGCTCTCTCTCTTTCCAAATCGGCACCGC---TGGTATTCTCA---GGGTCTACAA
CGGAAGAATTTTGGCTGTTGAACATACAT---CGGGTCTACAAACACCGGAATTCGTCCAAGGTAAAAA
-45 CGC---TTCTGTCCAAGGTAAAAA (6)

```

plant line 2

```

-14 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA (6)
-20 CGGAAGAATTTTGGCTGTTGAACATAC---CGGAATTCGTCCAAGGTAAAAA (6)
-26+2 CGGAAGAATTTT---CAACACCGGAATTCGTCCAAGGTAAAAA
-45 CGC---TTCTGTCCAAGGTAAAAA

```

plant line 3

```

-8 CGGAAGAATTTTGGCTGTTGAACATAC---CTACAAACACCGGAATTCGTCCAAGGTAAAAA (2)
-14 CGGAAGAATTTTGGCTGTTGAACATACAT---AACACCGGAATTCGTCCAAGGTAAAAA
-14 CGGAAGAATTTTGGCTGTTGAACATAC---AACACCGGAATTCGTCCAAGGTAAAAA (2)
-18 CGGAAGAATTTTGGCTGTTGAACATACAT---CGGAATTCGTCCAAGGTAAAAA
-18 CGGAAGAATTTTGGCTGTTGAACATAC---CTACAAACACCGGAATTCGTCCAAGGTAAAAA
-30 CGC---GTCTACAAACACCGGAATTCGTCCAAGGTAAAAA (2)
-41 CGGAAGAATTTTGGCTGTTGAACATAC---42---AACAGCAACCA
-59 CGGAAGAATTTTGGCT---59---AACACTGTGAAC (2)
-109 TTTCTCAGGAACATCTAC---CTTGAACACATCTTTA

```

Figure S2. Continued

plant line 4	-45	CGG CGGGA -----TTCTGTCCAAGTAAAA		
	-58+17	CGGAAGAATTTTGTCTGTGAAC-----CTACAAACACCTAATG-----TTGTAAACAT	(2)	
	-63	CGGAAGAATTTTGTCTGTGAAC TACA -----CTTTATTCAACCAAT		
	-71	TCTTCAGGAAC TA -----71-----AACACCGGAATTCGTGCCAAGTAAAA		
	-152+4	ATCTTGTAGTTAAATCTGCA-----GTCA-----AAGGTAAAA	(2)	
	-284	CAGCTCTCAGCTCTT CTCA -----284-----TTCTCAGTTCTAGTGT	(3)	
	-14	CGGAAGAATTTTGTCTGTGAAC TACA -----AACACCGGAATTCGTGCCAAGTAAAA	(4)	
	-29	CGGAAGAATTTTGTCTGT TA -----TTCTTCCAAGTAAAA		
	-36+18	TTGAA-----CTCAAAACACCTGTAGT-----CAAGGTAAAAACAGCAAACTTGT	(2)	
	-42	CGGAAGAATTTTGTCT TA -----GTAAAA		
plant line 5	-68+29	CGGAAGAATTTTGTCT TA AAACATATGAACCGGATGAGGTCAAC-----CACATCTTTAT		
	-142	GGTATATCTTGTAGT TA -----142-----CAACACCGGAATTCGTGCCAAGTAAAA		
	-164+14	CGGAAGAATTTTGTCTGTGAAC TAC -----TAGGCCAACTAGCT-----ACAGAGATTTGAGGAA		
	pip2 Cas9-PPO			
	plant line 1	-1+1	CGGAAGAATTTTGTCTGTGAAC TAC CGGGTCTACAAACACCGGAATTCGTGCCAAGTAAAA	
		-4	CGGAAGAATTTTGTCTGTGAAC TACA -----GTCTACAAACACCGGAATTCGTGCCAAGTAAAA	
		-14	CGGAAGAATTTTGTCTGTGAAC TACA -----AACACCGGAATTCGTGCCAAGTAAAA	(3)
		-25+3	CGGAAGAATTTTGTCT-----ATG-----AACACCGGAATTCGTGCCAAGTAAAA	
		-30	CGG-----GTCTACAAACACCGGAATTCGTGCCAAGTAAAA	
		-6	CGGAAGAATTTTGTCTGTGAAC TACA -----TCTACAAACACCGGAATTCGTGCCAAGTAAAA	(4)
-8		CGGAAGAATTTTGTCTGTGAAC TACA -----TCTACAAACACCGGAATTCGTGCCAAGTAAAA		
-9+1		CGGAAGAATTTTGTCTGTGAAC TACAT -----TACAAACACCGGAATTCGTGCCAAGTAAAA		
-14		CGGAAGAATTTTGTCTGTGAAC TACA -----AACACCGGAATTCGTGCCAAGTAAAA		
-18		CGGAAGAATTTTGTCTGTG-----ACAAACACCGGAATTCGTGCCAAGTAAAA		
plant line 2	-21	CGGAAGAATTTTGTCTGT TA -----ACCGGAATTCGTGCCAAGTAAAA		
	-31	CGG-----31-----TCTACAAACACCGGAATTCGTGCCAAGTAAAA		
	-54	CTCTTT CAAA -----54-----CACCGGAATTCGTGCCAAGTAAAA		
	-58	CGC CC -----AAGGTAAAA		
	-249+4	TCCGAACAGAAATGTT-----CAAA-----GTAAAA		
	-266	CGCAAGGAGTTGAACATTAGGTATATC TA -----266-----CGCAGTCTGAA		
	-4	CGGAAGAATTTTGTCTGTGAAC TACAT -----GGGTCTACAAACACCGGAATTCGTGCCAAGTAAAA		
	-9	CGGAAGAATTTTGTCTGTGAAC-----GGGTCTACAAACACCGGAATTCGTGCCAAGTAAAA		
	-8+20	CGGAAGAATTTTGTCTGTGAAC TACAT AGTACTATAGTCTGCTGCTG-----CTACAAACACCGGAATTCGTGCCAAGTAAAA		
	-12+5	CGGAAGAATTTTGTCTGTGAAC TACA ACAG-----CAACACCGGAATTCGTGCCAAGTAAAA		
plant line 3	-14	CGGAAGAATTTTGTCTGTGAAC CTACA -----AACACCGGAATTCGTGCCAAGTAAAA	(3)	
	-17	CGGAAGAATTTTGTCTGTG-----TACAAACACCGGAATTCGTGCCAAGTAAAA		
	-18	CGGAAGAATTTTGTCTGTGAAC-----AACACCGGAATTCGTGCCAAGTAAAA		
	-25	CGGAAGAATTTT-----ACAAACACCGGAATTCGTGCCAAGTAAAA		
	-73+18	CGGA-----CTGTAGTTTCTCTCTCT-----CTTGTAAACATCTTTATTCA		
	-96	CGGAAGAATTTTGTCTGT TAAT -----96-----GATAGTTT		
	-7	CGGAAGAATTTTGTCTGTGAAC TACA -----GGGTCTACAAACACCGGAATTCGTGCCAAGTAAAA		
	-14	CGGAAGAATTTTGTCTGTGAAC TACA -----AACACCGGAATTCGTGCCAAGTAAAA		
	-17+3	CGGAAGAATTTTGTCTGTGAAC TAC -----CACCGGAATTCGTGCCAAGTAAAA		
	-21+1	TCGCCAC-----6-----CAACACCGGAATTCGTGCCAAGTAAAA		
plant line 4	-39+6	CGGAAGAATTTTGTCTGT TA ACCGCT-----TACAAACACCGGAATTCGTGCCAAGTAAAA		
	-48	CGGAAGAATTTTGTCTGT TA -----48-----AGCAACACCTTGT		
	-57	CGCACCG CA -----AAGGTAAAA		
	Ku80pip2 Cas9-PPO			
	plant line 1	-14	CGGAAGAATTTTGTCTGTGAAC TACA -----AACACCGGAATTCGTGCCAAGTAAAA	(3)
		-21+1	CGGAAGAATTTTGTCT-----E-----CAACACCGGAATTCGTGCCAAGTAAAA	
		-39	CGGAAGAATTTTGT CTG -----CCAAGTAAAA	
		-40+9	CGGAAGAATTTTGTCTGTGAAC TACA SCAAACAG-----AAACAGCAACACTTGT	
		-45	CGGAAGAATTTTGTCTGTGAAC TACA -----45-----GCAAACTTGT	
		-48	CGGAAGAATTTTGTCTGT TA -----48-----AGCA	
-49		CGCG CAACG -----GAATTCGTGCCAAGTAAAA	(2)	
-102+17		CGGAAGAATTTTGTCTGTGAAC TACAT -----TCTTTCCAAGTAAAA-----TCTTT		
-221		CGGAAGAATTTTGTCTGT TA -----221-----TTAGGAGTTAGGATATGCC		
-4		CGGAAGAATTTTGTCTGTGAAC TACAT GGGTCTACAAACACCGGAATTCGTGCCAAGTAAAA	(2)	
plant line 2	-14	CGGAAGAATTTTGTCTGTGAAC CTACA -----AACACCGGAATTCGTGCCAAGTAAAA	(3)	
	-18	CGGAAGAATTTTGTCTGTGAAC TACAT -----CGGAATTCGTGCCAAGTAAAA		
	-21	CGGAAGAATTTTGTCTGTG-----AACACCGGAATTCGTGCCAAGTAAAA		
	-21+1	CGGAAGAATTTTGTCTGTG-----AACACCGGAATTCGTGCCAAGTAAAA		
	-23+7	CGCCCGGAAGAA T -----GTCTACAAACACCGGAATTCGTGCCAAGTAAAA		
	-45	CGC CGGAA -----TTCTGTCCAAGTAAAA		
	-46	CTCTTCCAA AC -----46-----TACAAACACCGGAATTCGTGCCAAGTAAAA		
	-49	CTCTTCCAAATCGCG CAACG -----49-----GAATTCGTGCCAAGTAAAA		
	-51	CGGAAGAATTTTGTCTGTGAAC TACAT -----51-----CTTGTAAACAT		
	-63	CGGAAGAATTTTGTCTGTGAAC CTACA -----63-----CTTTATTCAACC	(2)	
plant line 3	-76	CGGA-----76-----TAACATCTTTATTCAA		
	-101	CTTCATGCTGTGCTGTAT CT -----101-----GTCCAAGTAAAA		
	-3+51	CTCTTCCAAATCGCGCACCGCCCGGAAGAA TTTGTCTGTGAAC TACATCTTTATTCTTCGGGCGG TCGGCGGATTTGATATCTCAAACTCGGCAACCGCCCGGGTCTACAACACCG		
	-5+30	CGGAAGAATTTTGTCTGTGAAC TACA CAAAATCTGTGTAATTTCTGTGTTCTGGGTCTACAAC AC	(2)	
	-8+55	CTCTTCCAAATCGCGCACCGCCCGGAAGAA TTTGTCTGTGAAC CTCACTTTATTCTTCGGGCGG TCGGCGGATTTGATATCTCAAACTCGGCAACCGCCCGGGTCTACAACACCG		
	-14	CGGAAGAATTTTGTCTGTGAAC TACAT -----ACACCGGAATTCGTGCCAAGTAAAA		
	-14	CGGAAGAATTTTGTCTGTGAAC TACA -----AACACCGGAATTCGTGCCAAGTAAAA	(4)	
	-39+6	CGGAAGAATTTTGTCTGTGAAC TACA -----41-----AAACAGCAAAACA		
	-41	CGGAAGAATTTTGTCTGTGAAC CTACA -----41-----AAACAGCAAAACA		
	-45	CGGAAGAATTTTGTCT TA -----45-----AAAAACAGCAAAACA		
plant line 4	-49	CGCG CAAC -----49-----GAATTCGTGCCAAGTAAAA		
	-52	CGGAAGAATTTTGTCTGTGAAC TACAT -----52-----TGTAAACATC		
	-103+15	CGCACCGCCCGGAAGAA TTTGTCTGTGAAC TACTTTCTCAGGAACTA-----TCTTCTCTCT		

Figure S2. Continued

Table S1. Primers used for cloning and PCR reactions.

Primer	Sequence	Used for
SP509	ATTGAGGAGACTATCTGCAGCATG	sgRNA cloning CRU3
SP510	AAACCATGCTGCAGATAGTCTCCT	sgRNA cloning CRU3
SP512	ATTGTTGCTGTTGAACTACATTGG	sgRNA cloning PPO
SP513	AAACCCAATGTAGTTCAACAGCAA	sgRNA cloning PPO
SP492	GCTTCAGAACCAACAAGACAGC	<i>CRU3</i> target sense (HRM)
SP492	TGAGCCTGACATACTCCAAG	<i>CRU3</i> target antisense (HRM)
SP245	TGCCAACACTCCAGGCTCTG	<i>CRU3</i> target sense (CRISPR/Cas9)
SP248	CAAGTGGTCAACGACAACGG	<i>CRU3</i> target antisense (CRISPR/Cas9)
SP392	CACTTTGACAGATTAGGTAG	<i>PPO</i> target sense (CRISPR/Cas9)
SP538	CTTCCACTAACTCACCTTC	<i>PPO</i> target antisense (CRISPR/Cas9)
SP560	CTCCTCACTCTTTTCCAAATCG	<i>PPO</i> target sense (HRM)
SP561	AGATGTGTTACAAGTGTGTGCTG	<i>PPO</i> target antisense (HRM)
SP563	ACCTCTAAGACAGCCCTACG	<i>CRU3</i> target sense (HRM)

Table S2. Distribution of deletion length for the target site derived from the indicated genotypes.

Nuclease mutants	Deletion length				Total deletions
	1-9 bp	10-19 bp	20-49 bp	≥50 bp	
WT Cas9-CRU	15 (57.7%)	6 (23.1%)	4 (15.4%)	1 (3.8%)	26
<i>ku80</i> Cas9-CRU	10 (28.6%)	3 (8.6%)	14 (40%)	8 (22.8%)	35
<i>parp1 parp2</i> Cas9-CRU	13 (46.4%)	6 (21.4%)	7 (25%)	2 (7.2%)	28
<i>ku80 parp1 parp2</i> Cas9-CRU	12 (30.8%)	3 (7.7%)	19 (48.7%)	5 (12.8%)	39
WT Cas9-PPO	21 (33.3%)	20 (31.7%)	14 (22.2%)	8 (12.8%)	63
<i>ku80</i> Cas9-PPO	2 (6.3%)	6 (18.7%)	14 (43.7%)	10 (31.3%)	32
<i>parp1 parp2</i> Cas9-PPO	9 (27.3%)	9 (27.3%)	8 (24.2%)	7 (21.2%)	33
<i>ku80 parp1 parp2</i> Cas9-PPO	4 (12.1%)	5 (15.2%)	16 (48.5%)	8 (24.2%)	33

Table S3. Insertions and templated-insertions.

Nuclease mutants	Total mutations	Insertions	Templated-insertions >3 bp
WT Cas9-CRU	26	4 (15.4%)	0
<i>ku80</i> Cas9-CRU	35	7 (20.0%)	6 (17.1%)
<i>parp1 parp2</i> Cas9-CRU	28	5 (17.9%)	0
<i>ku80 parp1 parp2</i> Cas9-CRU	39	11 (28.2%)	5 (12.8%)
WT Cas9-PPO	63	12 (19.1%)	4 (6.3%)
<i>ku80</i> Cas9-PPO	32	7 (21.9%)	4 (12.5%)
<i>p1p2</i> Cas9-PPO	33	10 (30.3%)	1 (3.0%)
<i>ku80p1p2</i> Cas9-PPO	33	10 (30.3%)	6 (18.2%)

Chapter 4

***In planta* gene targeting during meiosis**

**Gary D. Strunks^{1,2}, Bart J. P. M. Klemann^{1,3},
Paul J. J. Hooykaas¹ and Sylvia de Pater¹**

¹Molecular and Developmental Genetics, Institute of Biology,
Leiden University, 2333 BE, the Netherlands

²Current affiliation: Naktuinbouw, 2371 GD Roelofarendsveen,
the Netherlands

³Current affiliation: Dümmer Orange, 2678 PS De Lier,
the Netherlands

Abstract

Over the past 30 years, increasing the efficiency of gene targeting (GT) in plants has been an important goal in plant biotechnology. Improvements have been made using sequence-specific nucleases to induce DNA double strand breaks (DSBs) at target loci and activate repair via homologous recombination (HR). GT can then be achieved by HR-mediated integration of an artificial repair template, sharing homology with the target locus. Further improvements have been made with the *in planta* GT method, in which the repair template is pre-inserted in the genome and can be excised by nucleases. Although these improvements led to substantial increases in GT efficiency, GT is still not efficient enough to be feasible for crop biotechnology. A major roadblock for GT is the low efficiency of HR in somatic cells. Directing GT to cells with an intrinsically higher HR frequency might overcome this obstacle. In this study, we developed an *in planta* GT system targeting the seed-specific *cruciferin 3* (*CRU3*) gene. We directed GT to meiocytes, which undergo meiosis and should therefore have higher HR frequency, by expressing Cas9 with the *RPS5a*, *SPO11-1* and *SPL* promoters. A pre-inserted, *CRU3*-GT repair template was used, containing a promoterless GFP-sequence flanked by 5' and 3' *CRU3* homology arms, which allows for detection of GT events as GFP-fluorescent seeds. Cas9 activity at *CRU3* was found for all promoters used, indicated by targeted mutations due to imperfect NHEJ-mediated DSB repair. Crossing two plant lines harboring either the Cas9-expression cassette T-DNA or the *CRU3*-GFP repair template, resulted in GFP-fluorescent F1 seeds. Surprisingly, GFP-fluorescence did not coincide with GT events. True GT events were detected with PCR analysis in F2 plants, although seeds derived from these plants did not show fluorescence. Furthermore, GT events found with PCR analysis could not be confirmed with Southern blot analysis. These results suggest that GT events may have taken place only in a limited number of the plant's cells. Taken together, our results have not yet led to an improvement in GT efficiency using the meiosis-specific *in planta* GT system. However, based on targeted mutagenesis experiments, a good candidate promoter for improved GT efficiency may be the egg cell-specific *EC1.2* promoter.

Introduction

Targeted mutagenesis and gene targeting (GT) are important tools for plant genome editing. They exploit the plant's endogenous repair mechanisms of DNA double-strand breaks (DSBs), nonhomologous end-joining (NHEJ) and homologous recombination (HR), to induce targeted mutations or precisely introduce novel DNA sequences (1). Targeted DSBs can be induced by sequence-specific nucleases, of which nowadays those based on the CRISPR/Cas system are the most efficient and user friendly (2).

In plants and other higher eukaryotes, DSBs will mostly be repaired via NHEJ in somatic cells, which can occasionally result in mutations at the repair junction. Efficient targeted mutagenesis at loci of interest can be realized when using an effective nuclease. This has been described in various organisms, including plants using the CRISPR/Cas9 system (for reviews, see: 3–7). Cas9-mediated targeted mutagenesis is readily feasible when using constitutive Cas9 expression, and leads to the appearance of targeted mutations in many cells of the plant. Unfortunately, only in rare cases mutations are inherited and only few heterozygotes, bi-alleles and homozygotes are found in the next generation (8). Therefore, in order to increase the chances of inherited events, germline-specific Cas9 expression systems have been developed, and this indeed has led to a higher number of inherited mutations and a larger variety of mutations in the next generation (9).

Contrary to targeted mutagenesis, high-frequency GT is even much more challenging to realize in higher eukaryotes, especially in plants. In early GT experiments in plants, GT frequencies remained very low with approximately one GT event per $10^4 - 10^5$ transformation events (10–12). These low numbers are mainly due to the low intrinsic frequency of HR-mediated DSB repair in somatic cells, compared to NHEJ.

A breakthrough came with the use of sequence-specific nucleases. GT frequencies could be enhanced substantially using various sequence-specific nucleases (13–21). However, in the classical GT approach the GT frequency is dependent on the transformation frequency of the repair template T-DNA and GT events predominantly occur during transformation. This is especially a roadblock for plants that cannot be efficiently transformed. An approach called *in planta* GT has been developed to overcome this issue (22–24). With the *in planta* GT system the repair template T-DNA is stably integrated into the plant genome. A sequence-specific nuclease is used to induce a DSB at both the target locus and at targets flanking the GT repair template, which releases the repair template to be freely available for HR-mediated repair. As each plant cell in principle contains the necessary components for GT (a sequence-specific

nuclease, a GT repair template and HR machinery), GT events can occur in each cell during the plant's lifetime, independent of transformation frequency. GT events that occur in germline cells will be transmitted to the next generation. In a recent study, *in planta* GT has been shown to occur without releasing the pre-inserted GT repair template, suggesting that this is not a prerequisite (21).

The low frequency of HR in somatic cells is another obstacle for efficient GT, and even if GT events occur in a number of cells only a few will be passed on to the germline. A potential strategy to overcome this issue is to direct GT to cells or tissue that have a higher intrinsic HR frequency. Considering that HR is a major repair pathway in cells that undergo meiosis, the meiocytes, directing GT to these cells by expressing Cas9 in a meiosis-specific manner and at the same time supplying a pre-inserted repair template, might result in a higher frequency of heritable GT events.

In this study we describe the development of a germline-specific CRISPR/Cas9 expression system and an *in planta* gene targeting system directed to meiocytes for the *Arabidopsis cruciferin 3* (*CRU3*) gene. *CRU3* encodes a seed storage protein and is predominantly expressed during the seed developmental stage. Construction of the *CRU3* GT repair template was based on an earlier study in which a promoterless GFP sequence was used, resulting in a *CRU*-GFP open reading frame and seed-specific GFP fluorescence upon successful GT (25). Previously, it was shown that germline-specific Cas9 expression using the *sporocyteless* (*SPL*) promoter resulted in more heritable mutations compared to constitutive Cas9 expression (9). For our meiosis-specific *in planta* GT system we used the *SPL* promoter and also the promoters from the *ribosomal protein 5A* (*RPS5a*) and *sporulation 11-1* (*SPO11-1*) genes, which are active during meiosis. The *SPO11-1* gene is of particular importance during meiotic recombination, because it encodes the endonuclease for programmed DSB induction and initiation of meiotic recombination (26, 27). Here, we sought to increase GT frequency by combining the *in planta* GT approach using a pre-inserted repair template that could be released, and the use of Cas9 expression under different promoters to direct GT to the meiosis stage.

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 $\mu\text{mol m}^{-2} \text{s}^{-1}$ 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 $\mu\text{mol m}^{-2} \text{s}^{-1}$ during 16h/day photoperiod at 20°C and 70% humidity.

Vector construction and plant transformation

The constitutive Cas9 expression vector pDE-Cas9-CRU2 (pSDM3908) was constructed as described in chapter 2. Cas9 expression vectors containing the RPS5a (pRPS5a), SPO11-1 (pSPO11), SPL (pSPL) or EC1.2 (pEC1.2) promoters were created by replacing the ubiquitin promoter (pUbi) in pDE-Cas9-CRU2. The ubiquitin promoter was removed from pDE-Cas9-CRU2 with *EcoRI*. Promoters pRPS5a, pSPO11, pSPL and pEC1.2 were PCR-amplified from genomic DNA using primers SP505/SP506 (pRPS5a, 1669 bp), GS23/GS24 (pSPO11, 1088 bp), GS25/GS26 (pSPL, 1094 bp) and GS27/GS28 (pEC1.2, 1016 bp) (Table S1). The blunt-ended PCR products were cloned into vector pJET1.2 (ThermoFisher Scientific). The promoter sequences were verified by sequencing (Macrogen, Amsterdam, The Netherlands) using the pJET forward and reverse primers. Promoter pRPS5a from pJET1.2-pRPS5a was digested with *BglIII* and blunted with Klenow fragment (ThermoFisher Scientific). pJET1.2-SPO11 and pJET1.2-SPL were digested with *EcoRV*. The promoter fragments were ligated into the blunted *EcoRI*-digested pDE-Cas9-CRU2 to create expression vector pDE-pRPS5a-Cas9-CRU2 (pSDM3909), pDE-pRPS5a-Cas9-CRU2 (pSDM3910) and pDE-pSPL-Cas9-CRU2 (pSDM3911). pJET- pEC1.2 was digested with *EcoRI* and the promoter fragment was ligated into *EcoRI* digested pDE-Cas9-CRU2, creating expression vector pDE-pEC1.2-Cas9-CRU2 (pSDM3912). To identify clones with the correct orientation of the promoter sequences, restriction digestions were performed with *AflIII* + *BstEII* (pRPS5a), *NcoI* + *HindIII* (pSPO11), *NcoI* (pSPL) and *HindIII* (pEC1.2).

The CRU-GFP repair templates were constructed in several steps. First, the 5' and 3' CRU3 homology arms were amplified by PCR from wild-type genomic DNA using primers GS1/GS2 (5' homology arm, 1140 bp) and GS3/GS4 (3' homology arm, 3497 bp) (Table S1).

The PCR products were purified using the Qiagen PCR purification kit (Qiagen) and ligated into pJET1.2. The 5' CRU3 arm was sequenced with pJET FW/Rev primers and 3' CRU3 arm was sequenced with pJET FW/Rev and GS5, GS6, GS7 and GS8 sequencing primers (Table S1). Sequencing was performed by Macrogen. After sequence validation, the 5' CRU3 homology arm was digested from pJET1.2 with *Bam*HI, sites present in primers GS1/GS2, and cloned in frame with GFP-tNOS in vector pART7-GFP-tNOS (25, 28). pART7-CRU-GFP-tNOS was digested with *Hind*III and *Eco*RV to validate the correct orientation of the 5' CRU3 homology arm. The 3' CRU3 homology arm was digested from pJET1.2 with *Pst*I sites present in primers GS3 and GS4 and cloned in the *Pst*I-digested pART7-CRU-GFP-tNOS backbone, creating pART7-CRU-GFP-tNOS-CIFERIN. pART7-CRU-GFP-tNOS-CIFERIN was digested with *Eco*RI and *Eco*RV to validate the correct orientation of the 3' CRU3 homology arm. To create the binary vector for the CRU3 gene targeting repair template, the CRU-GFP-tNOS-CIFERIN sequence was isolated from pART7-CRU-GFP-tNOS-CIFERIN with *Hind*III and *Eco*RI and cloned in binary vector pCambia 2300 (www.cambia.org), creating 2300-CRU ((pSDM3913).

To construct the *in planta* gene targeting constructs, recognition sites for CRISPR/Cas9-CRU2 were added flanking the 5' and 3' homology arms. The CRISPR/Cas9-CRU2 target region was first PCR-amplified with primers GS9 and GS10 (Table S1), containing *Asc*I and *Afl*III restriction sites, resulting in a 255 bp fragment. The target region at the 5' end was cloned into pART7-CRU-GFP-tNOS-CIFERIN flanking the 5' homology arm with *Afl*III and the target region at the 3' end was cloned into pART7-5'target-CRU-GFP-tNOS-CIFERIN with *Asc*I, resulting in pART7-5'target-CRU-GFP-tNOS-CIFERIN-3'target. The sequences of the complete repair templates were verified by sequencing. Subsequently, the complete 5'target-CRU-GFP-tNOS-CIFERIN-3'target was cloned into vector pCambia 2300 and using *Hind*III and *Eco*RI, resulting in binary vector pCambia 2300-CRU-IP, where IP stands for *in planta*. To construct binary vector pMDC100-LSL-CRU-IP, the LIR and LIR + SIR sequences were first PCR amplified using primers SP524/SP522 (LIR) and SP519/SP523 (LIR + SIR) (Table S1) and cloned into pJET1.2. Subsequently, LIR was cloned into pMDC100 with *Spe*I and *Sac*I and LIR + SIR was cloned into pMDC-LIR with *Hind*III and *Sbf*I. The 5'target-CRU-GFP-tNOS-CIFERIN-3'target was removed from pART7-5'target-CRU-GFP-tNOS-CIFERIN-3'target using *Hind*III and *Eco*RI, and cloned into the pEntry-GW vector. The 5'target-CRU-GFP-tNOS-CIFERIN-3'target was inserted between the LSL repeats of pMDC100-LSL by a Gateway LR reaction, resulting in pMDC100-LSL-CRU-IP (pSDM3915).

Plant binary vectors pDe-Cas9-CRU2, pCambia 2300-CRU, pCambia 2300-CRU-IP and pMDC 100-LSL-CRU-IP were introduced into the *Agrobacterium* strain AGL1 by electroporation (den Dulk-Ras and Hooykaas 1995).

Generation of gene targeting plant lines

Plants were transformed with binary vectors via the floral dip method (29). Cas9-CRU2 T-DNAs were selected with phosphinothricin and the *CRU3* repair template T-DNAs with kanamycin. Homozygous plant lines for the pUbi/pRPS5a/pSPO11/pSPL-Cas9-CRU2 T-DNA and the 2300-CRU, 2300-CRU-IP and 100-LSL-CRU-IP T-DNA were generated and were crossed in different combinations to generate *CRU3* gene targeting plant lines. The selected homozygous plant lines used for crossings corresponding with the transformed T-DNA constructs are described in Table 1. F1 seeds derived from these crossings were analyzed for GFP fluorescence and gene targeting events at the molecular level were analyzed with PCR in F1 seedlings. Seeds and seedlings from several F2 plant lines were used for screening for in planta GT events with PCR and Southern blot.

Root transformation

Root transformation was performed as described (30). Root segments of plant lines CRU-GFP-3-4, CRU-GFP-3-5, CRU-GFP-IP-3-1 and CRU-GFP-IP-4-5 were infected with *A. tumefaciens* strain AGL1 harboring the pDe-Cas9-CRU2 binary vector of which the T-DNA contains the phosphinothricin selection cassette. Roots were separated from seedlings and co-cultivated on callus induction medium containing 100 μ M acetosyringone for 48 hours at 24 °C and 35 μ mol m⁻² s⁻¹ light intensity. Root segments were washed and incubated on shoot-induction medium containing 30 μ g/ml phosphinothricin, 100 μ g/ml vancomycin and 100 μ g/ml timentin and incubated at 24 °C with 75 μ mol m⁻² s⁻¹ light intensity. After 20 days the transformation efficiencies were counted as number of calli per plate. Calli were transferred to fresh shoot-induction medium and developing shoot tissue was used for genomic DNA isolation. Samples were screened for in planta GT events with PCR and Southern blot analysis.

Footprint analysis

T2 seeds with a single copy of the T-DNA insert or T3 seeds homozygous for the T-DNA insert, that were derived from independently selected T1 plants, were germinated on ½ MS supplemented with 10 μ g/ml phosphinothricin for CRISPR/Cas9 T-DNA selection. After 10

days of growth, pools of 10 seedlings per plant line were disrupted to a powder under liquid N₂ in a tissue lyser (Retch, Haan, Germany). The CTAB method was used to extract genomic DNA (14). DNA was first subjected to predigestion with DdeI and then used for PCR with Phusion polymerase (Thermo Scientific) to amplify the nucleases target site using primers SP491/SP492 (Table S1). PCR products were digested with DdeI and separated in agarose gels. DdeI digestion-resistant fragments were visualized and then extracted from agarose gels, cloned into pJet1.2 (ThermoFisher Scientific) and sequenced by MacroGen Europe (Amsterdam, the Netherlands). The spectrum of mutations of some genomic DNA samples was determined using the TIDE web tool (31, <https://tide.deskgen.com>).

Detection of gene targeting events with PCR

F1 or F2 seeds derived from different crossings of lines harboring a CRU3 repair template, and lines harboring the Cas9-CRU2 under control of various promoters were screened for GFP fluorescence. A number of GFP-positive seeds were germinated on ½ MS medium without selection. Tissue from shoots derived from root transformation, seedling leaves or flower heads was disrupted to a powder under liquid N₂ in a tissue lyser (Retch, Haan, Germany). Genomic DNA was isolated using the fast extraction method (32) or CTAB method (14). Detection of gene targeting events was done with PCR with Phusion polymerase (ThermoFisher Scientific), using primers SP261, SP262 for 5' GT and primers GS43 and SP608 for 3' GT (Table S1). Detection of the repair template T-DNA was done with PCR using primers GS1/SP262 for the 5' end and GS43/GS4 for the 3' end (Table S1).

Southern blot analysis

Genomic DNA (5 – 10 µg) from PCR-positive GT shoots or plants was digested with PstI (5' GT) or HindIII (3' GT). Digested DNA was separated on 0.7% agarose gels overnight and, after depurination, denaturation and neutralization steps, blotted overnight on Hybond-N+ blotting paper (GE Healthcare). The blot was washed in 2xSSC and baked at 120 °C. Blots were hybridized at 42 °C in DIG easy hyb (Roche Diagnostics) and 50 µg/ml herring sperm DNA with the 5' and 3' probes that were labelled in a PCR reaction using primers SP623/SP624 (5' probe) and SP625/SP626 (3' probe) using DIG-labelling mix (Roche Diagnostics). After hybridizing for 16 – 20h, blots were washed two times in 2xSSC; 0.1% SDS at room temperature, and three times in 0.2xSSC; 0.1% SDS at 65 °C. Bands were detected using the DIG wash and block buffer set with DIG-antibody and CDP-STAR (Roche Diagnostics). Blots were exposed on Fujifilm medical X-Ray film for 20 – 30 min.

Microscopy

GFP-fluorescent seeds were visualized using the Leica MZ16 FA Fluorescence stereomicroscope (Leica Microsystems). Exposure was adjusted to 1.1 seconds on the bright field channel and 4 seconds on the GFP channel, with 1x gain, 1.00 saturation and 1.21 gamma levels and zoom adjusted to 7.17. Images from the bright field and GFP channels were merged.

WT		TCCAGGTCGTGAGG	CCA	CCT	CTAAG	CAGCCCTACGA	GAGCGAGGAGTGG
pUbi-Cas9-CRU2							
T2 plant line 1	+1	TCCAGGTCGTGAGGCCACCTCTA	A	A	A	A	A
	-1	TCCAGGTCGTGAGGCCACCTC	A	A	A	A	A
	-3	TCCAGGTCGTGAGGCCACC	A	A	A	A	A
	-5	TCCAGGTCGTGAGGCCACC	A	A	A	A	A
	-8	TCCAGGTCGTGAGGCC	A	A	A	A	A
T2 plant line 3	+1	TCCAGGTCGTGAGGCCACCTCTA	A	A	A	A	A
	-2	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-8	TCCAGGTCGTGAGGCCACC	A	A	A	A	A
	-13	TCCAGGTCGTGAGGCCACC	A	A	A	A	A
T2 plant line 6	-4	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-5	TCCAGGTCGTGAGGCCACCTC	A	A	A	A	A
pRPS5a-Cas9-CRU2							
T3 plant line 3-1	-12	TCCAGGTCGTGAGGCCACCTCTA	A	A	A	A	A
	-13	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
T3 plant line 6-6	+1	TCCAGGTCGTGAGGCCACCTCTA	A	A	A	A	A
	-1	TCCAGGTCGTGAGGCCACCTC	A	A	A	A	A
	-9	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-26	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
pSPO11-Cas9-CRU2							
T3 plant line 6-7	+1	TCCAGGTCGTGAGGCCACCTCTA	A	A	A	A	A
	-1	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-5	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-7	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	+23-11	GAGGAAACATCGTTCGTGTTAAGGGACCTTTCCAGGTCGTGAGG	A	A	A	A	A
	-23+5	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
pSPL-Cas9-CRU2							
T2 plant line 1	-1	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
T2 plant line 7	+1	TCCAGGTCGTGAGGCCACCTCTA	A	A	A	A	A
	-1	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-4	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-5	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-11	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-12	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A
	-15	TCCAGGTCGTGAGGCCACCT	A	A	A	A	A

Figure 1. CRISPR/Cas9-mediated mutagenesis at the *CRU3* locus in T2 and T3 seedlings. Part of the *CRU3* wild type locus is shown at the top with the protospacer shown in yellow. The PAM is shown in black and the *DdeI* restriction site is red underlined. Insertions (green), deletions (dashed lines) and substitutions (blue) were found in plants expressing Cas9 constitutively under the *UBI* promoter, or meiosis/germline-specific under the *RPS5a*, *SPO11-1*, *SPL*. Microhomologies are shown in purple. Numbers are the length of deletions (-) or insertions (+). The number of clones with the same sequence are shown at the right. One insertion in pSPO11-Cas9-CRU2 is a partial templated insertion, indicated by the underlined template sequence upstream of the insertion, which is part of the wild-type sequence not shown in the other sequences. Homologies (grey) between sequences flanking the template and insertion are probably used as primer.

Results

Mutagenesis at the CRU3 locus with constitutive and meiosis or germline-specific CRISPR/Cas9 systems

Previously, we described targeted mutagenesis using a constitutively expressed pUbi-Cas9-CRU2 that effectively generated mutations at the *CRU3* locus (chapter 2 of this thesis). To examine if the Cas9 expression constructs with the *RPS5a*, *SPO11-1*, *SPL* promoters are suitable for gene targeting, we first investigated if Cas9 would be expressed at all under these promoters and could induce targeted mutations at *CRU3*. Assessment of targeted mutagenesis was done by a loss of restriction site assay (see chapter 2 and 3 of this thesis). A *DdeI* restriction site in the vicinity of the Cas9-CRU2 target site allows for selection of DNA fragments that contained a mutation at the *DdeI* site due to erroneous NHEJ-mediated DSB repair. *Arabidopsis* wild type plants were transformed with CRISPR/Cas9 expression constructs using the floral dip method (33) and several primary transformants were selected on phosphinothricin. Single copy T2 plant lines for pSPL-Cas9-CRU2, and homozygous T3 lines for pRPS5a-Cas9-CRU2 and pSPO11-Cas9-CRU2 were analyzed and compared to the previously analyzed T2 plant lines harboring the pUbi-Cas9-CRU2 construct (chapter 2). Genomic DNA was extracted from pools of 10 T2 or T3 seedlings. To enrich for mutations at *CRU3* and to be able to readily obtain footprints, genomic DNA was subjected to a predigestion with *DdeI*. A region of 272 bp around the Cas9 cleavage site was PCR-amplified using primers SP491/SP492 and PCR fragments were again subjected to *DdeI* digestion. A non-mutated sequence should result in two bands of 199 and 73 bp, whereas absence of a *DdeI* site should result in one band of around 272 bp. *DdeI*-resistant PCR fragments were cloned and sequenced. All CRISPR/Cas9 expression cassettes tested were expressed, as footprints were identified in plants harboring them. Mutations, consisting of various insertions or deletions ranging from +23 to -26 bp, were generated at the *CRU3* target locus (Figure 1).

An in planta gene targeting system for the CRU3 locus

The *in planta* gene targeting system created for this study required two components: the CRISPR/Cas9 expression cassette Cas9-CRU2 and a *CRU3* repair template (RT). Several Cas9 expression constructs were made that could express Cas9 using the *RPS5a*, *SPO11-1* and *SPL* promoters. The Cas9-CRU2 nuclease can induce a DSB in the *CRU3* gene and simultaneously in target sites flanking the RT (2300-CRU-GFP-IP and 100-LSL-CRU-GFP-IP). In a second version of the RT construct these target sites were not present and this RT could thus not be excised. The *CRU3* repair template (2300-CRU-GFP) consisted of a promoterless GFP sequence and the nopaline synthase terminator (tNOS), flanked by 5' and 3' *CRU3* homology arms of 1080 bp and 3471 bp, respectively (Figure 3A). A TAG stop codon was placed at the start of the 5' *CRU3* homology arm and the ATG start of the endogenous *CRU3* sequence was omitted from the homology arm to reduce the possibility that random T-DNA integration would lead to expression of a CRU-GFP fusion protein. Furthermore, the homology arms were chosen so that they did not contain the CRISPR/Cas9 target site. DSBs would therefore only be induced in the *CRU3* target locus and in the target sites flanking the *CRU3* homology arms on the repair template.

Wild-type *Arabidopsis* plants were transformed with the various CRISPR/Cas9-CRU2 T-DNA constructs and the RT T-DNAs 2300-CRU-GFP, 2300-CRU-GFP-IP and 100-LSL-CRU-GFP-IP independently. T-DNA 2300-CRU-GFP contains a GT repair template that cannot be excised from the genome, while T-DNAs 2300-CRU-GFP-IP and 100-LSL-CRU-GFP-IP, contain *in planta* GT repair templates that can be excised from the genome via Cas9-CRU2-mediated DSB induction at sequences flanking the repair templates. Homozygous lines for each T-DNA were selected in the next generation (Table 1). Homozygous plant lines were crossed to give rise to F1 plants which contain both the Cas9-CRU2 and the RT constructs. Due to Cas9 activity and erroneous NHEJ-mediated DSB repair in parental lines homozygous for the Cas9-CRU2 construct, F1 plants may inherit one *CRU3* allele containing mutations at the target site. However, the parent homozygous for the *CRU3* repair template should still provide an intact *CRU3* locus that can undergo gene targeting. F1 progeny contained the necessary components for GT to occur in meiotic or germline tissue, and inherited GT events in the F2 should be visible as fluorescent seeds when expressing the *CRU3*-GFP fusion protein under the endogenous *CRU3* promoter (Figure 2, 3A).

To see if any GFP fluorescence was already visible in the F1 progeny, we screened F1 seeds derived from several crossings with the fluorescent stereo microscope. To our surprise,

F1 seeds derived from most crossings were already GFP fluorescent (Figure S1). GFP signals were seen in seeds derived from plants in which the repair template can be excised from the genome (*in planta* lines) as well as in plants in which this is not possible. However, the GFP signal was stronger in seeds derived from the *in planta* lines. Also, the strongest fluorescent signals were seen in plants derived from crossings CRU-GFP-IP-4-5 x pRPS5a-Cas9-CRU2-6-7 and CRU-GFP-IP-4-5 x pSPO11-Cas9-CRU2-6-7, which harbor excisable repair templates and in which Cas9 can be expressed during meiosis.

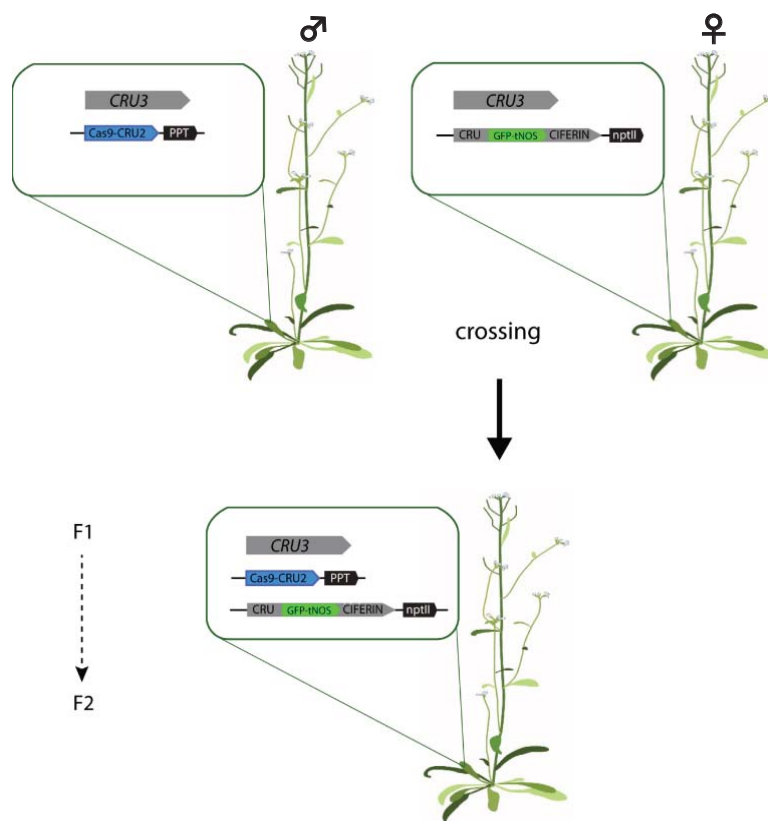


Figure 2. Outline of the experimental approach for *in planta* GT. Plant lines homozygous for either the Cas9-CRU2 T-DNA or the CRU-GFP repair template T-DNA were crossed to obtain F1 progeny harboring both the Cas9-CRU2 and CRU-GFP repair template T-DNA constructs and at least one intact endogenous *CRU3* locus. F1 plants were selected on phosphinotricin (PPT) and kanamycin (nptII) for both constructs and propagated to obtain F2 plants for further analysis for GT.

Table 1. Overview of selected homozygous plant lines and their accompanied T-DNA constructs containing the CRISPR/Cas9 expression cassettes with various promoters and the CRU-GFP repair templates. CRU-GFP-3-4 and 3-5 are repair templates that cannot be excised from the genome. CRU-GFP-IP-3-1 and 4-5 are repair templates that can be excised from the genome, where IP stands for *in planta*.

	T-DNA construct	Selected homozygous T3 plant lines
CRISPR/Cas9 constructs	pUbi-Cas9-CRU2	pUbi-Cas9-CRU2-1-12
	pSPO11-Cas9-CRU2	pSPO11-Cas9-CRU2-6-2, 6-7
	pRPS5a-Cas9-CRU2	pRPS5a-Cas9-CRU2-3-1, 6-6
	pSPL-Cas9-CRU2	pSPL-Cas9-CRU2-1-1
CRU-GFP-RT constructs	2300-CRU-GFP	CRU-GFP-3-4, 3-5
	2300-CRU-GFP-IP	CRU-GFP-IP-3-1
	100-LSL-CRU-GFP-IP	CRU-GFP-IP-4-5

To investigate whether the observed GFP fluorescence coincided with a GT event, several GFP fluorescent F1 seeds were germinated on medium without selection and genomic DNA was extracted from flower heads using the CTAB protocol, as this gives sufficient amounts of good quality DNA. PCR analysis was performed to detect recombination at the 5' and 3' region. For detection of 5' recombination forward primer SP261, which is located in the endogenous CRU3 promoter region, and reverse primer SP262 located in the GFP sequence on the repair template were used that should result in a 1133 bp fragment. For detection of 3' recombination, forward primer GS43, which is located in the tNOS sequence on the repair template, and reverse primer SP608, which is located downstream on the endogenous *CRU3* locus, were used which should give a 3567 bp fragment (Figure 3A). No GT was detected with PCR in plants derived from GFP fluorescent seeds. As seeds from the parental plant lines containing the CRU-GFP repair template T-DNA alone were not GFP-fluorescent, GFP expression was not due to activation and expression of GFP due to promoter fusion during T-DNA integration. The switch to GFP expression only occurred after crossing the parental lines.

Screening for gene targeting events at the CRU3 locus in F2 plants

As F1 plants contained all necessary equipment for GT events to occur and to be transmitted to the next generation, we screened for GT events using PCR in F2 plants derived from various F1 crossings. F2 seeds derived from F1 plants that developed from GFP-fluorescent seeds were first screened for GFP fluorescence. None of the F2 seeds were GFP fluorescent. Assuming that the CRU-GFP repair template was functional and GT thus would result in CRU-GFP fusion protein expression visible in seeds, this suggested that no GT events were transmitted to the F2 generation. However, as we cannot fully exclude that after GT a CRU-GFP fusion protein would not be properly expressed, we analyzed the F2 progeny also by the same PCR reaction as described earlier. To quickly screen a larger number of plants, we first performed PCR on genomic DNA from leaves of individual F2 seedlings using the fast DNA extraction protocol. Seedlings that showed either a 5' GT (5' recombination event) or both 5' and 3' GT (true GT event, TGT) with PCR were transferred to soil and grown further. Genomic DNA was isolated from flower heads using the CTAB protocol and PCR was again performed (Figure 3B).

Both 5' GT and TGT were identified in young seedlings from lines 6 (CRU-GFP-IP-4-5 x pUbi-Cas9-1-12), 10 (CRU-GFP-IP-4-5 x pUbi-Cas9-1-12), 40 (CRU-GFP-IP-4-5 x pSPO11-Cas9-6-2) and 235 (CRU-GFP-IP-4-5 x RPS5a-Cas9-6-6) (Table 2). These are all plant lines in which the repair template could be excised, indicating that GT occurred using the excised repair template for DSB repair. When performing the PCR again on flowerheads of mature plants, 5' GT events found in line 6 and 10 could now be classified as TGT events. This might be due to the fact that the 3' GT fragment is a longer fragment that may be more difficult to PCR-amplify on fast extracted genomic DNA. The 5' GT found in the fast-extracted samples for lines 40 and 235 could not be reproduced in the CTAB samples of the mature plants. Based on these results, TGT events were detected in the plants that constitutively express Cas9, while only one TGT event was detected in plants that express Cas9 under control of the *RPS5a* promoter, which is mainly active in the early embryo and the meristems, as well as during meiosis. Neither 5' GT nor TGT events were detected in plant lines in which Cas9 was expressed with the *SPO11-1* and *SPL* promoters, suggesting that there is no increase in GT events at *CRU3* when using these promoters. Furthermore, no GT events were detected in seedlings with RTs without flanking target sites for release from the genome, indicating that this is required for GT.

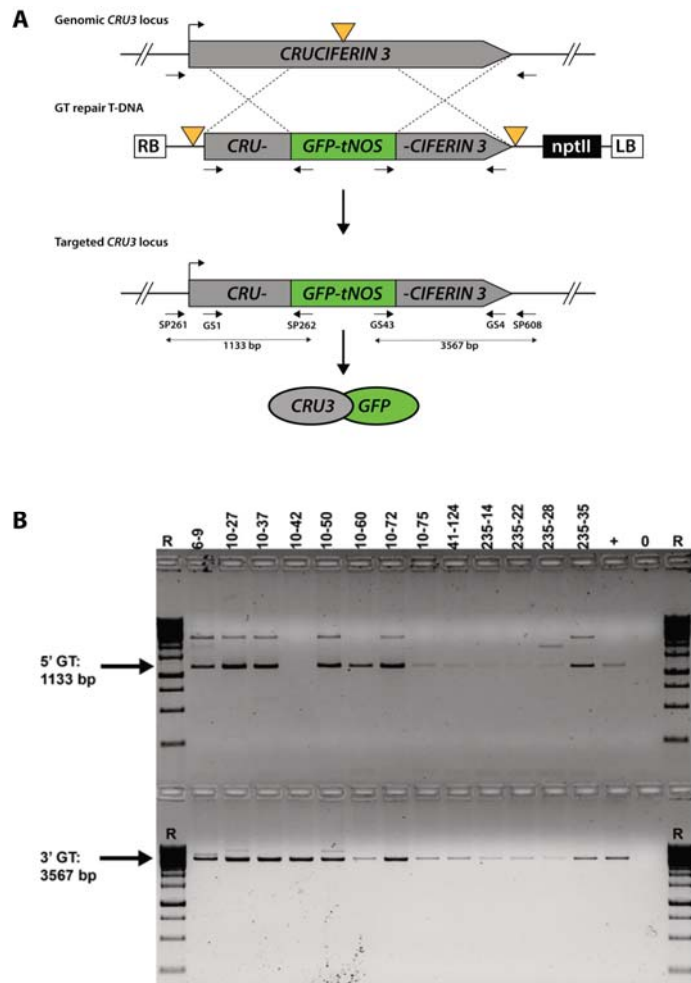


Figure 3. *In planta* gene targeting at the *CRU3* locus. **A.** Schematic overview of the genomic *CRU3* target locus, the *CRU3* GT repair template T-DNA and the targeted *CRU3* locus after a GT event. A CRISPR/Cas9 nuclease can induce a DSB at targets on the *CRU3* genomic locus (grey bar) and flanking the repair template (orange triangles). The *CRU3* GT repair template T-DNA contains a repair template with a GFP sequence and NOS terminator sequence (green bar), flanked by 5' and 3' *CRU3* homology arms of 1080 bp and 3471 bp, respectively. The 5' homology arm is missing 39 bp 5' region, including the ATG start codon. The nptII selection gene is used for T-DNA selection using kanamycin. Position of primers to detect the T-DNA and GT events with PCR are shown (horizontal arrows). Sizes of PCR fragments indicative of a 5' and 3' GT event are indicated. Correct integration of the GFP sequence should result in a CRU3-GFP fusion protein (grey and green ovals). **B.** Detection of GT events with PCR on F2 samples. A 5' GT event should give a 1133 bp fragment and a 3' GT should give a 3567 bp fragment. R is the 1 kb ladder (ThermoFisher Scientific). + is a putative positive control. 0 is a water control sample. Sample numbers above each lane designate the F1 line (first number) and F2 line (second number).

Table 2. Overview of GT events in F2 plants derived from crossings of homozygous CRISPR/Cas9 plants and CRU-GFP-GT repair template plants.

Promoter	F2 plant line	Fast extraction			CTAB extraction	
		#PCRs	5' GT	TGT	5' GT	TGT
pUbi	1 (CRU-GFP-3-5 x pUbi-Cas9-1-12)	55	0	0	-	-
	6 (CRU-GFP-IP-4-5 x pUbi-Cas9-1-12)	19	1	0	0	1
	7 (CRU-GFP-IP-4-5 x pUbi-Cas9-1-12)	4	0	0	-	-
	10 (CRU-GFP-IP-4-5 x pUbi-Cas9-1-12)	101	4	1	0	5
pSPO11	24 (CRU-GFP-IP-4-5 x pSPO11-Cas9-6-7)	34	0	0	-	-
	30 (CRU-GFP-IP-4-5 x pSPO11-Cas9-6-7)	23	0	0	-	-
	40 (CRU-GFP-IP-4-5 x pSPO11-Cas9-6-2)	51	1	0	0	0
	50 (CRU-GFP-3-5 x pSPO11-6-7)	39	0	0	-	-
pRPS5a	235 (CRU-GFP-IP-4-5 x pRPS5a-Cas9-6-6)	41	3	1	0	1
	271 (CRU-GFP-3-5 x pRPS5a -Cas9-3-1)	34	0	0	-	-
pSPL	148 (CRU-GFP-IP-4-5 x pSPL-Cas9-1-1)	51	0	0	-	-
	252 (CRU-GFP-3-5 x pSPL-Cas9-1-1)	59	0	0	-	-

In order to exclude that the PCR products indicative for GT were PCR derived recombination events, Southern blot analysis was performed. Genomic DNA from GT plants was digested with either *Pst*I (5' GT) or *Hind*III (3' GT) and probes were used to detect the GT bands (Figure 4). Correct HR-mediated integration of the CRU-GFP repair template at the *CRU3* locus should result in a 6180 bp fragment for the 5' probe and a 7375 bp fragment for the 3' probe, while a wild-type sequence should show fragments for the 5' and 3' probes of 4802 bp and 6067 bp, respectively. None of the samples showed the expected GT bands, suggesting that GT did not occur. However, it cannot be excluded that GT events occurred in a small number of plant cells and that only PCR amplification was sensitive enough to detect these events.

Screening for somatic gene targeting events at CRU3 in shoots

To investigate if our gene targeting system for *CRU3* is triggered during transformation, we performed root transformation. Roots of plants from homozygous plant lines CRU-GFP-3-4, CRU-GFP-3-5, CRU-GFP-IP-3-1 and CRU-GFP-IP-4-5 were transformed with *Agrobacterium* strains harbouring the pUbi-Cas9-CRU2 T-DNA. Root segments were selected for the presence of pUbi-Cas9-CRU2 T-DNA and genomic DNA was extracted with the CTAB method from individual shoots. PCR was performed to detect GT. This time, first a PCR of 15 cycles was performed spanning the *CRU3* target using primers SP261 and SP608. Product from this PCR was used as a template for a second PCR of 30 cycles for 5' GT using primers SP261/SP262 and for 3' GT using primers GS43/SP608. PCR to detect the presence of the CRU3-GFP T-DNA was performed with primers GS1/SP262 (5') and GS43/GS4 (3') (Figure 3A).

A total of 10 shoots of line CRU-GFP-IP-3-1, 20 shoots of line CRU-GFP-3-4 and 14 shoots of CRU-3-5 were recovered. Two true GT events were found for line CRU-GFP-IP-3-1 and one true GT event was found for line CRU-GFP-3-4 (Figure 5A).

Southern blot analysis was performed to validate the results found with PCR (Figure 5B). Together with the samples that showed true GT, a number of negative samples and a wild-type sample were analysed. Southern blot analysis was performed in the same way as described earlier. All samples showed wild-type fragment sizes for the 5' and 3' probe, indicating that also here GT did not occur in the entire shoots. However, like with our *in planta* GT approach, a difference in sensitivity between the PCR method and Southern blot method to detect GT might also be an explanation for these contradictory results.

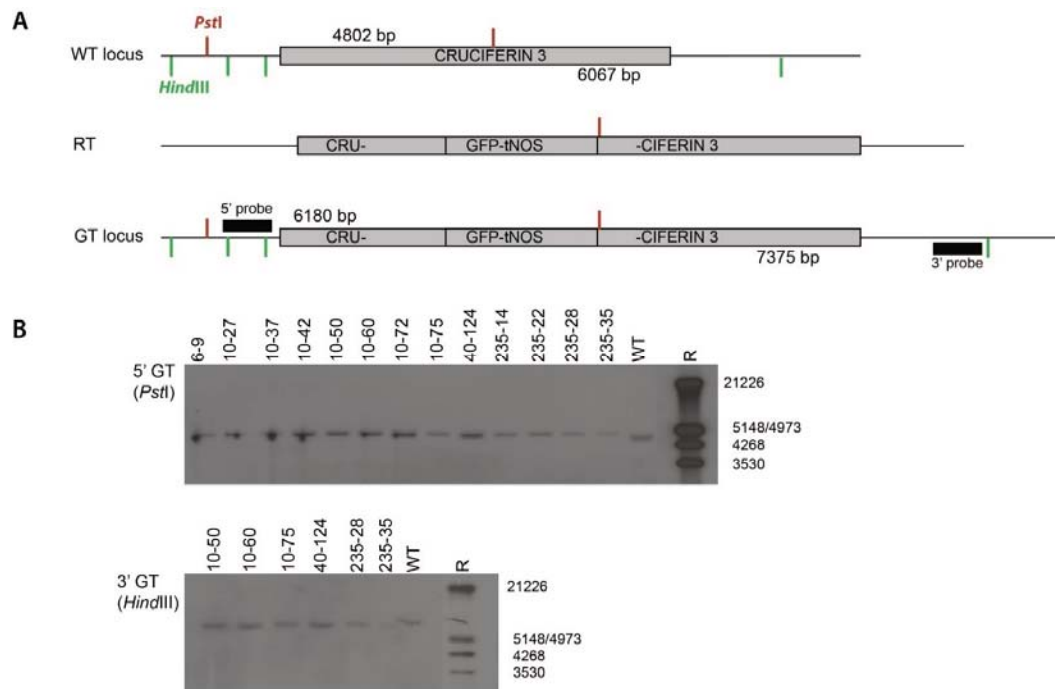


Figure 4. Southern blot analysis of samples that showed GT in PCR screening (Figure 3). **A.** Strategy to detect GT with Southern Blot analysis. For 5' GT a 6180 bp fragment (WT, 4802 bp) should be detected after digestion of genomic DNA with *PstI* (red). For 3' GT a 7375 bp fragment (WT, 6067 bp) should be detected after digestion of genomic DNA with *HindIII* (green). **B.** Blots shown for 5' and 3' GT. At the, right sizes of bands in base pairs of the DIG-labeled Molecular Weight Marker III (Roche) are shown.

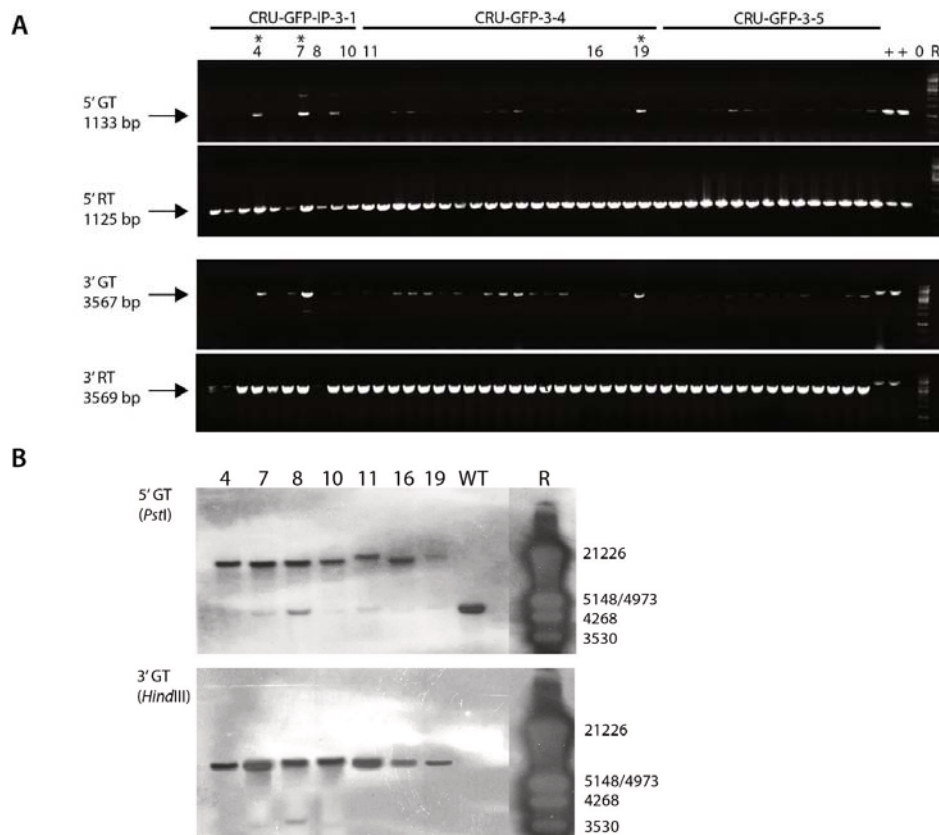


Figure 5. Screening for GT at *CRU3* in shoots from plant lines CRU-IP-3-1, CRU-3-4 and CRU-3-5 transformed with the pUbi-Cas9-CRU2 construct in root transformation experiments. **A.** PCR to detect GT at *CRU3* using genomic DNA from shoots. True GT events were detected in samples 4, 7 and 19 (*). +, positive control. 0, water control. R, 1 kb ruler (ThermoFisher Scientific). **B.** Southern blot analysis to validate PCR results. The large fragments of ~20 kb seen on the 5' GT blot are probably probe-labeled undigested fragments. At the right, sizes of bands in base pairs of the DIG-labeled Molecular Weight Marker III (Roche) are shown.

Heritable mutations at CRU3 with germline-specific Cas9 expression

We did not detect any GT events in plants that expressed Cas9 under the SPL promoter. However, previous studies have shown that germline-specific Cas9 expression with the *SPL* promoter reduces chimerism and increases the inheritance of mutations in the next generation (9). Such results were also obtained using another germline-specific promoter, namely the Arabidopsis egg cell 1.2 (*EC1.2*) promoter (9, 34). We investigated if germline-specific Cas9 expression under the *SPL* and *EC1.2* promoters in the T2 generation resulted in a higher frequency of stably inherited *CRU3* mutations and reduced chimerism in the T3 generation compared to plants that constitutively expressed Cas9.

To see if mutagenesis could be driven by the pEC1.2 at all, footprint analysis in T2 seedlings was performed as described for the promoters used in the GT experiment. Various footprints could be readily detected (Figure S2). Subsequently, footprint analysis was performed on individual T3 plants harbouring the pUbi-Cas9-CRU2, pSPL-Cas9-CRU2 and pEC1.2-Cas9-CRU2 expression constructs. Genomic DNA was extracted from leaves of T3 seedlings that were grown on ½ MS medium with PPT as selection marker to select for the Cas9-CRU2 T-DNA constructs. The *CRU3* target was PCR-amplified and the PCR product was digested with *DdeI*. After separation of the fragments by gel electrophoresis, the intensity of the *DdeI* resistant bands were measured with ImageJ. In case of an inherited mutation the PCR product would be at least 50% resistant to digestion by *DdeI* (heterozygous mutant) or 100% resistant (homozygous or bi-allelic mutant). For pUbi-Cas9-CRU2 T2 plants, in only 5 out of 200 analyzed plants from two T2 plant lines a small amount of the PCR fragment was *DdeI* resistant (<50% *DdeI* resistance), indicating these mutations were not inherited, but present in some somatic cells (Figure 6A). In agreement, in none of the 190 T3 plants a *DdeI* resistant PCR fragments was detected. For pSPL-Cas9-CRU2 9 plants with <50% *DdeI* resistance from a total of 176 T3 plants were found, indicating that these mutations were also chimeric. However, among the T3 plants harboring the pEC1.2-Cas9-CRU2 construct, two heterozygous mutants (>50% *DdeI* resistance) and one homozygous mutant (100% *DdeI* resistance) were found from a total of 101 plants. The *DdeI* resistant fragments of the 9 pSPL-Cas9-CRU2 samples and of the 3 pEC1.2-Cas9-CRU2 samples were cloned and sequenced. Chimeric mutations in pSPL-Cas9-CRU2 and heterozygous mutations in pEC1.2-Cas9-CRU2 were confirmed using the TIDE web tool, which can decompose a mixture of sequences into individual sequences based on sequencing peak patterns (31). Indeed, chimeric mutations in pSPL-Cas9-CRU2 were characterized by a mixture of multiple alleles with insertions or

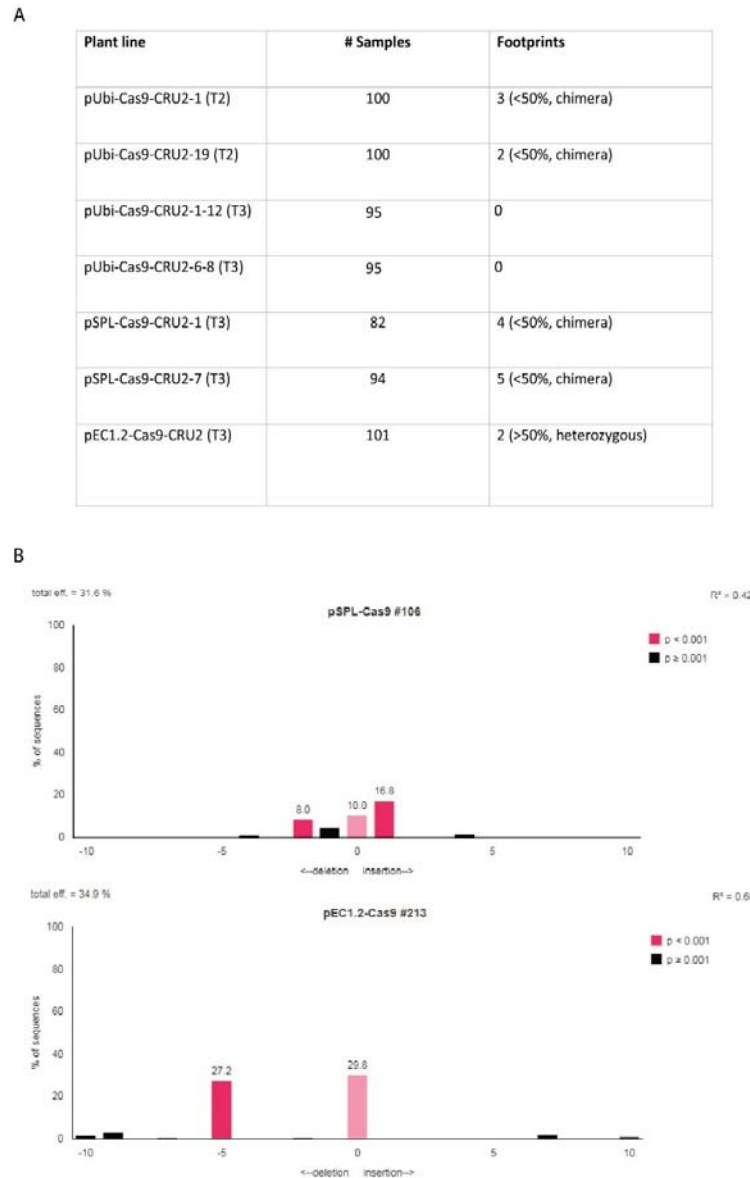


Figure 6. Footprints at *CRU3*. **A.** Overview of footprints in T2 and T3 plants with pUbi-Cas9-CRU2, and T3 plants with pSPL-Cas9-CRU2 and pEC1.2-Cas9-CRU2 expression constructs. **B.** Decomposition of sequence mixture using the TIDE software tool reveals individual sequences with mutations. Shown are a pSPL-Cas9 sample (#106) with multi allelic mutations consisting of a 1 bp insertion and a 2 bp deletion, and a pEC1.2-Cas9 sample (#213) with a heterozygous 5 bp deletion. The R-squared value represents a goodness-of-fit measure of the expected versus the observed data. Total efficiency is calculated as the R-squared value minus the percent of wild-type sequence (pink bar, no indel). Values above the bars indicate the estimated abundance of each indel in the mix of sequences. Red bars indicate statistically significant indels ($p < 0.001$, two-tailed t -test). Black bars indicate indels that are not statistically significant ($p \geq 0.001$, two-tailed t -test)

deletions, while heterozygous mutations in pEC1.2-Cas9-CRU2 were confirmed by a 50:50 ratio of sequences of the wild type allele and the mutated allele (Figure 6B). Surprisingly, the one mutation characterized by complete *DdeI* resistance and therefore expected to be homozygous, could not be confirmed with TIDE and we disregarded this sample. All in all, these results confirmed that Cas9 expression with the *EC1.2* promoter can indeed lead to inherited mutations, and this promoter therefore compares favourably in this respect to pUbi and pSPL, which makes pEC1.2 the most suitable promoter for Cas9-mediated heritable targeted mutagenesis.

Discussion

We developed a GT system for the seed-specific Arabidopsis *CRU3* gene based on the GT system developed by Shaked *et al.* (25). We used a GFP sequence in the GT repair template allowing for detection of GT events as GFP-fluorescent seeds. We tested if GT frequency at *CRU3* could be increased by directing GT towards meiocytes. As meiocytes undergo meiosis to produce four haploid spores, meiotic recombination in these cells might make them better equipped for GT, compared to the low efficiency of somatic recombination in vegetative tissue. To accomplish this, Cas9 expression systems were developed using the *RPS5a*, *SPO11-1* and *SPL* promoters. Additionally, we used pre-inserted GT repair templates that could be excised to see if this can increase GT efficiency further (22).

Previous studies have shown that *in planta* GT leads to enhancement of GT on artificial and endogenous loci, and that *in planta* GT is possible with both excised (22, 23) and non-excised repair templates (21). These approaches relied on constitutive Cas9 expression, which would mostly lead to GT events in somatic tissue, but also to mutation of the target site. Consequently, this would reduce the chance of GT occurring in the germline. Our hypothesis for *in planta* gene targeting during meiosis was that meiosis-specific Cas9 expression would concentrate DSB repair in meiotic tissue, and that the increased HR frequency during meiosis in meiocytes would provide a better environment for further enhancement of GT. Although not exclusively meiosis-specific, the *SPO11-1*, *RPS5a* and *SPL* promoters were used to drive Cas9 expression, and results were compared to those obtained with the constitutively expressed pUbi-Cas9. The SPO11-1 protein is the endogenous initiator of meiotic HR through programmed DSB induction (35), and the SPO11-1 promoter is active during meiosis (26, 36, 43). Activity of the *RPS5a* promoter has been shown in male meiocytes during meiosis using a GFP reporter system (43), although it is also active in the early embryo and in meristems (37). We also

decided to use the *SPL* promoter for GT, because it is predominantly active in male meiocytes (38, 39).

By crossing two parental plant lines homozygous for either the CRISPR/Cas9 T-DNA or the CRU-GFP GT repair template T-DNA, F1 progeny were generated containing all the necessary components for GT to occur. GFP-fluorescence was already detected in F1 seeds derived from all crossings. This was surprising, as we hypothesized that *in planta* GT events would take place mostly during meiosis in flowers of the F1 plants, thus only detectable among the F2 seeds. Further analysis showed that fluorescent F1 seeds did not have the molecular signature of GT events, and that the GFP signal was absent in F2 seeds derived from plants germinated from GFP-fluorescent F1 seeds. One possible explanation for this result can be that the CRU-GFP GT repair template was integrated somewhere in the genome in frame with a different seed-specific promoter. However, as we could not detect fluorescent seeds derived from the mother plant line only containing the CRU-GFP GT repair template T-DNA, this is not the case and it must be due to an event that occurred after crossing the two plant lines. Another possibility might be that a CRU-GFP fusion protein was expressed transiently through elongation of the CRU-GFP repair template at one side using the endogenous *CRU3* gene as a template, thereby obtaining a promoter sequence. However, when such an elongated RT is not integrated the recombined RT will be lost in the next generation. It could also be argued that absence of fluorescence in F2 seeds might be due to lower expression or silencing of Cas9/sgRNA; however, this is unlikely to have happened in all of the lines. The intensity of the GFP signal varies between crossings, and seemed higher in some cases when the CRU-GFP repair template can be excised and when repair template plant lines were crossed with plant lines containing pSPO11-Cas9 or pRPS5a-Cas9 T-DNAs, in which Cas9 can be expressed during meiosis. It remains to be determined why and to what extent Cas9 expression or the location in the genome of the CRU-GFP repair template influences the strength of the observed GFP signal.

Although no GFP signal was detected in F2 seeds, true GT events were detected in F2 plants with PCR analysis. GT events were detected in plants derived from the pUbi-Cas9 x CRU-GFP-IP, pRPS5a-Cas9 x CRU-GFP-IP and pSPO11-Cas9 x CRU-GFP-IP crossings. Interestingly, all GT events were detected in plants in which the CRU-GFP repair template could be excised, suggesting that repair template excision can benefit GT. This has also been shown by others (22–24). However, we could not confirm GT events with Southern blot analysis. On one hand, this may indicate that the GT events detected with PCR analysis were

in fact PCR derived recombination events, and that Southern blot analysis confirmed the absence of real GT events. On the other hand, it may be that true GT events took place in somatic cells, and were detectable with PCR, but the number of cells were too low to be detectable with Southern blot analysis. This has also been noted by others (8), and would explain why most GT events were detected with PCR in the constitutively expressed pUbi-Cas9 x CRU-GFP-IP crossing. However, the total number of events was low, and more plants need to be analyzed to give a conclusive answer.

Cas9 was able to induce DSBs at the *CRU3* target locus after being expressed by all three promoters used for GT, as well as the *EC1.2* promoter, as indicated by the detection of a variety of footprints in T2 and T3 plants. To evaluate if any mutations were passed on to the next generation, we analyzed mutation types in T2 or T3 plants. Mutations that were found in T2 plants with the pUbi-Cas9 construct were chimeras not only consisting of cells with the wild-type alleles, but also of cells with various different allelic mutations. No inherited hetero- or homozygous mutations were found. Similar results were previously found in our lab when using pUbi-Cas9 constructs targeting the *PPO* gene and the *CRU3* gene with a different sgRNA, which also rendered little to no mutations in the next generation. Although using strong constitutive Cas9 expression with the 35S or *UBI* promoter can result in heritable mutations, most mutations will occur in somatic plant tissue, resulting in chimeras that have a mixture of cells with different allelic mutations (8, 9). Cas9 expression with the *SPL* promoter only yielded chimeric mutations in our experiments, in contrast to the results described by Mao *et al.* 2016 (9), in which chimeric mutations in the next generation only constituted approximately 25% of the total mutations and mainly heterozygous mutations were found (9). This may be explained by the fact that the pSPL-Cas9 construct used by Mao *et al.* (9) mimicked the native *SPL* gene as much as possible, consisting of a sequence of 5.7 kb from 5' regulatory region to the 3' termination region, with the Cas9 coding sequence replacing the first exon, while we used a sequence of 1094 bp upstream of the *SPL* start codon to drive Cas9 (9).

Cas9 expression with the *EC1.2* promoter resulted in two heterozygous mutations in T3 plants. The pEC1.2-Cas9-NOST construct described by Mao *et al.* 2016 (9) and Wang *et al.* 2015 (34) resembled ours, with a ~1 kb EC1.2 promoter region to drive Cas9 expression. However, only Mao *et al.* could already detect hetero- and homozygous mutations at a low frequency in T1 plants, and up to almost 70 percent in T2 plants, while Wang *et al.* did not detect any hetero- and homozygous mutations in T1 plants and only at low frequency in T2 plants (9, 34). Beside a pEC1.2-Cas9-NOST construct, Wang *et al.* (34) also used a pEC1.2-

Cas9 construct with the ribulose biphosphate carboxylase small subunits E9 terminator (*rbcS-E9t*) of *Pisum sativum*, which overall performed much better in T1 and T2 plants compared to the pEC1.2-Cas9-NOS_t construct, indicating that the use of different terminators can enhance Cas9 efficiency, possibly by providing better Cas9 mRNA stability in egg cells (34). Thus, it may be worthwhile to optimize our pEC1.2-Cas9 construct with this terminator in future experiments.

We detected mutations using the loss of restriction enzyme activity assay, which can only detect mutations if footprints occur in the restriction enzyme recognition sequence. Mutations that occur outside of the restriction enzyme recognition sequence were therefore not detected, thus underestimating the actual number of mutations and omitting such homo- and heterozygous mutations.

Mao *et al.* 2016 (9) and Wang *et al.* 2016 (34) directly sequenced PCR products of the target loci to detect heritable mutations, which gives a less biased result compared to the loss of restriction enzyme activity assay.

Previously, Shaked *et al.* (25) showed that GT at the *CRU3* locus was possible using a similar GFP-RT construct as we used. With our *in planta* GT system, we could not conclusively show evidence for GT at *CRU3*. There can be several reasons for this. Firstly, it can be that Cas9 expression with the promoters used was not specific and strong enough for DSB induction in the correct time frame of meiotic HR. To solve this, promoters need to be identified and used that are specific to the recombination stage of meiosis, and also adequately expressed for the nuclease to be able to induce DSBs. We did not yet use the *EC1.2* promoter for GT. Given that this promoter performed best in the generation of heritable targeted mutations, it may be worthwhile to also test this promoter in GT experiments. Furthermore, a recent study showed that expression of the *Staphylococcus aureus* Cas9 ortholog (SaCas9) with the *EC1.1* promoter fused with an enhancer sequence of pEC1.2 resulted in *in planta* GT at the Arabidopsis *ALS* gene with a GT frequency up to 6% (24). It thus seems that pEC1.1-Cas9 expression is able to induce GT in cells that contribute to the germline, and it is therefore worthwhile to investigate if this approach also leads to high frequency GT at *CRU3* and other genes. Secondly, *CRU3* is predominantly expressed in the seed development stage and the seed. It might be that the seed-specificity of *CRU3* makes it less suitable for GT to occur in other plant tissues because of a condensed chromatin state. Finally, the blunt-ended DSB induced by Cas9 might not be the best option for GT. Producing a DSB with overhangs using, e.g. a double nickase such as Cas9-D10A, or the Cas12a nuclease that produces a 4 bp overhang, may be attractive alternatives

(24, 40–42). The sticky ends provided by overhangs may result in more efficient targeted integration of complementary DNA sequences.

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 BSc internship students Bram Smoorenburg, Jiyang Chan and Jun Lerou and MSc internship student Kate McAleer and PhD exchange student Sidra Akhtar for their technical assistance. 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. Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. *Int J Dev Biol* 57:629–637.
2. Jinek M, et al. (2012) A Programmable Dual-RNA – Guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–822.
3. Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. *Annu Rev Plant Biol* 64(1):327–350.
4. Baltes NJ, Voytas DF (2015) Enabling plant synthetic biology through genome engineering. *Trends Biotechnol* 33(2):120–131.
5. Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V (2015) Editing plant genomes with CRISPR/Cas9. *Curr Opin Biotechnol* 32:76–84.
6. Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* 33(1):41–52.
7. Osakabe Y, Osakabe K (2015) Genome editing with engineered nucleases in plants. *Plant Cell Physiol* 56(3):389–400.
8. Feng Z, et al. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc Natl Acad Sci U S A* 111(12):4632–7.
9. Mao Y, et al. (2016) Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in *Arabidopsis*. *Plant Biotechnol J* 14(2):519–32.
10. Paszkowski J, Baur M, Bogucki A, Potrykus I (1988) Gene targeting in plants. *EMBO J* 7(13):4021–6.
11. 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.
12. Lee KY, Lund P, Lowe K, Dunsmuir P (1990) Homologous recombination in plant cells after *Agrobacterium*-mediated transformation. *Plant Cell* 2(5):415–425.

13. Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc Natl Acad Sci U S A* 93(10):5055–5060.
14. 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.
15. 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.
16. Shukla VK, et al. (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459(7245):437–441.
17. Townsend J a, et al. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459(7245):442–5.
18. Qi Y, et al. (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. *Genome Res* 23(3):547–54.
19. Zhang Y, et al. (2013) Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. *Plant Physiol* 161(1):20–27.
20. Endo M, Mikami M, Toki S (2016) Biallelic Gene Targeting in Rice. *Plant Physiol* 170(2):667–677.
21. 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.
22. Fauser F, et al. (2012) In planta gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
23. 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.
24. 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*. *Plant J* 94(4):735–746.
25. Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102(34):12265–12269.
26. Chen C, et al. (2010) Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. *BMC Plant Biol* 10:280.
27. Yang H, Lu P, Wang Y, Ma H (2011) The transcriptome landscape of Arabidopsis male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J* 65(4):503–516.
28. Gleave AP (1992) A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol Biol* 20:1203–1207.
29. Clough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16:735–743.
30. Vergunst AC, et al. (2000) VirB / D4-Dependent Protein Translocation from Agrobacterium into Plant Cells. *Science* 290:979–982.
31. Brinkman EK, Chen T, Amendola M, Van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42(22):1–8.

32. 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.
33. Clough SJ, Bent a F (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735–43.
34. Wang ZP, et al. (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol* 16(1):1–12.
35. Hartung F, et al. (2007) The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in *Arabidopsis*. *Plant Cell* 19(10):3090–3099.
36. Li J, et al. (2012) Characterization of a set of novel meiotically-active promoters in *Arabidopsis*. *BMC Plant Biol* 12(1):104.
37. 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.
38. Yang W-C, Ye D, Xu J, Sundaresan V (1999) The SPOROCTELESS gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev* 13(16):2108–2117.
39. Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu Rev Plant Biol* 56(1):393–434.
40. Ran FA, et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380–1389.
41. 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* 79:348–359.
42. Zetsche B, et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas System. *Cell* 163:1–13.
43. van Tol N, et al. (2019) Two novel strategies to assess in vivo meiotic protein expression in *Arabidopsis thaliana* (version 1; peer review: awaiting peer review). *F1000 Research* 8:1–21

Supplementary information

Table S1. Primers used for cloning and PCR

Primer	Sequence (5' – 3')	Used for
GS1	GCGGATCCAAGCTTCTTAAGTAGGTTCTCCTCGTCCTTAACGG	Cloning 5' CRU homology arm FW, PCR 5' RT T-DNA
GS2	CGGGATCCTGTCTTGTGGTTCTGAAGC	Cloning 5' CRU homology arm Rev
GS3	GCCTGCAGACCAGCGTCAACAGCTATACC	Cloning 3' CRU homology arm FW
GS4	GGCTGCAGGAATTTCGGCGCGCCAAACTACTTGGATCATGCACGAG	Cloning 3' CRU homology arm Rev, PCR 3' repair template T-DNA
GS9	ATTAGGCGCGCCCTTAAGGCAGAGGAAACATCGTTTCG	Nuclease target linker
GS10	TATAGGCGCGCCCTTAAGCAAGGTATAGCTGTTGACGC	Nuclease target linker
GS21	ATTGTCGTAGGGCTGTCTTAGAGG	sgRNA cloning <i>CRU3</i> (Cas9-CRU-2 construct)
GS22	AAACCCTCTAAGACAGCCCTACGA	sgRNA cloning <i>CRU3</i> (Cas9-CRU-2 construct)
GS23	GCGATATCAGCTTCACCAAACTTGACTAAGTATGAGACC	pSPO11-1 cloning FW
GS24	GCGATATCCTCTTTTCGAGTTTCAAACTGAAAAATGC	pSPO11-1 cloning Rev
GS25	GCGATATCGTTACTATTGGACAGTGTTTCTCACC	pSPL cloning FW
GS26	GCGATATCTGATGATGATCTTCTTCTCGGAAC	pSPL cloning Rev
GS27	ATTAGAATTCAAATGTTCTCTCGCTGACGTAAG	pEC1.2 cloning FW
GS28	TATAGAATTCGTGTTAGAAGCCACTATTCTTTCTTTTGG	pEC1.2 cloning Rev
GS43	GACAGTGACGACAAATCGTTGG	PCR 3' GT/3' RT T-DNA
SP261	CTCAGCAATCTCCTCGTTG	PCR 5' GT
SP262	TCGCCCTTGCTCACCAT	PCR 5' GT/5' RT T-DNA
SP491	GCTTCAGAACCAACAAGACAGC	<i>CRU3</i> target FW
SP492	TGAGCCTGACATACTCCAAG	<i>CRU3</i> target Rev
SP505	GCAGATCTTTTGATTGCTATTTGCAGTGCACC	pRPS5a cloning FW
SP506	CGGGTACCGGCTGTGGTGAGAGAAACAGAG	pRPS5a cloning Rev
SP519	CGGAAGCTTTAGCAGAAGGCATGTTGTTGTG	LIR + SIR sequence cloning Rev
SP522	GGCGAGCTCGAGGGTCGTACGAATAATTGG	LIR sequence cloning FW
SP523	GCGCCTGCAGGATCCCTTCTATAATTCTTTGCAATCC	LIR + SIR sequence cloning FW
SP524	GCGGATCCCGGGAAC TAGTTAGCAGAAGGCATGTTGTTGTG	LIR sequence cloning Rev
SP608	CAGAAACAGAGCACCAAAATGGG	PCR 3' GT

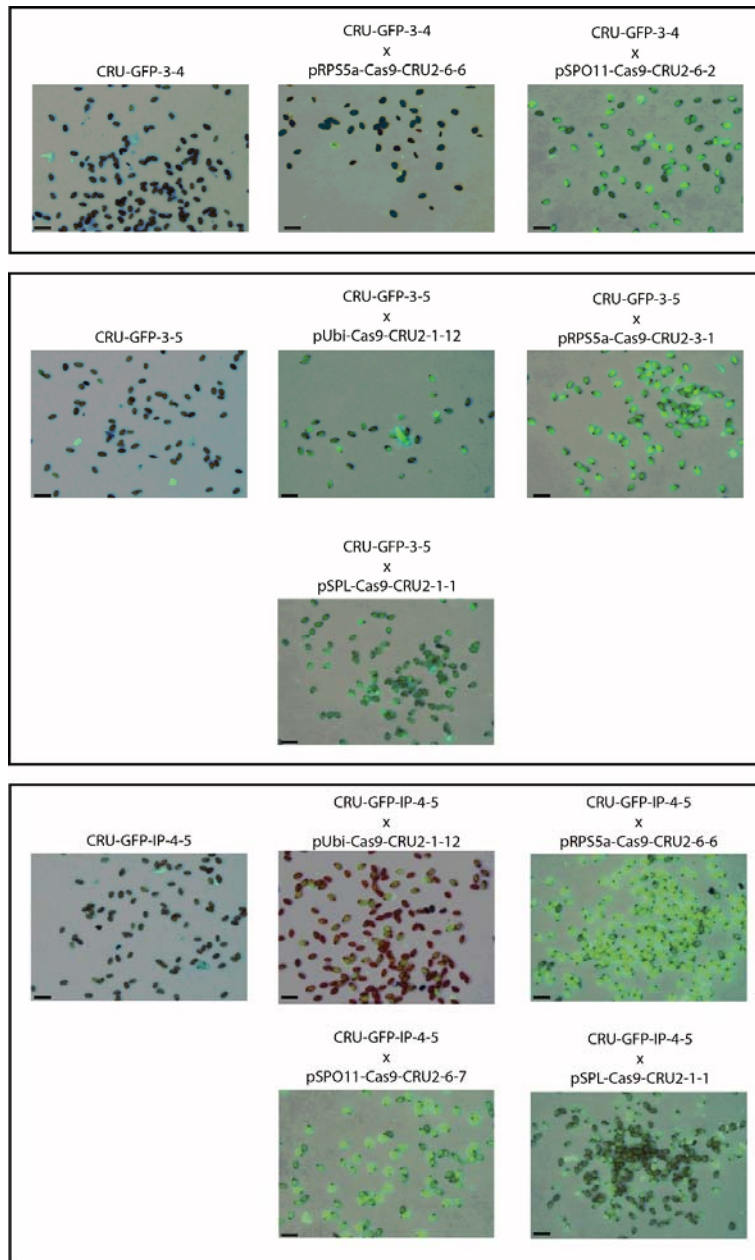


Figure S1. Fluorescent stereomicroscopy images of GFP-fluorescent F1 seeds derived from crossing homozygous plant lines containing a CRU-GFP repair template or the Cas9-CRU2 construct driven by constitutive or meiosis/germline-specific promoters. Seeds derived from homozygous parental lines containing only the CRU-GFP repair template T-DNA are shown on the left side of each panel as a control. Size bar (black line) represents 1 mm.

WT		TCCAGGTCGTGAGG	CCA	CCT	CTAAG	ACAGCCCTACGA	GAGCGAGGAGTGG
pEC1.2-Cas9-CRU2							
T2 plant line 1	+1	TCCAGGTCGTGAGGCCACCTCTAA					
	sub	TCCAGGTCGTGAGGCCACCTCT					
	sub	TCCAGGTCGTGAGGCCACCT					
	-1	TCCAGGTCGTGAGGCCACCT					
	-8	TCCAGGTCGTGAGGCCACCT					
	-13	TCCAGGTCGTGAGGCCACCTC					
	-13	TCCAGGTCGTGAGGCCA					

Figure S2. CRISPR/Cas9-mediated mutagenesis at the *CRU3* locus in one T2 plant line using the *EC1.2* promoter to drive Cas9 expression. Numbers are the length of deletions (-) or insertions (+). Part of the *CRU3* wild type locus is shown at the top with the protospacer shown in yellow. The PAM is shown in black and the *DdeI* restriction site is red underlined. Shown are insertions (green), deletions (dashed lines) and substitutions (blue). Microhomologies are shown in purple.

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. Hooykaas¹ and Sylvia de Pater¹**

¹Molecular and Developmental Genetics, Institute of Biology,
Leiden University, 2333 BE, the Netherlands

²Current affiliation: Naktuinbouw, 2371 GD Roelofarendsveen,
the Netherlands

³Current affiliation: Dümmer 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 were 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 $\mu\text{mol m}^{-2} \text{s}^{-1}$ 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 $\mu\text{mol m}^{-2} \text{s}^{-1}$ 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 $\mu\text{g/mL}$), nystatin (100 $\mu\text{g/mL}$) 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 $\mu\text{g/mL}$), nystatin (100 $\mu\text{g/mL}$) and phosphinothricin (PPT) (15 $\mu\text{g/mL}$). 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 *KpnI*,

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 *Eco*RI and blunted with Klenow fragment, 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 *Eco*RV. The 3' *ADH1* homology arm was cloned into linearized pJET-5'ADH-35S-PPT-tNOS with *Bam*HI. Correct orientation of 3' *ADH1* was confirmed with *Eco*RI 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.

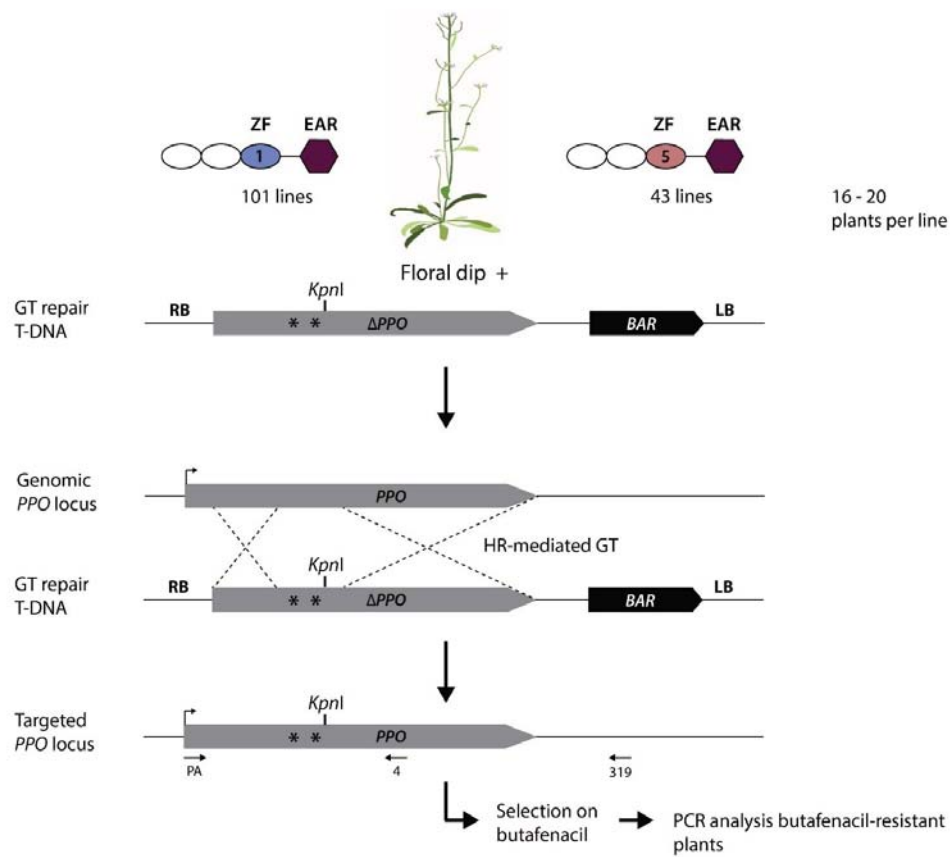


Figure 1. Overview of screening strategy for GT at *PPO* in ZF-EAR plant lines. Plant lines harboring ZF-EAR constructs from ZF-EAR pool 1 (blue oval) and pool 5 (red oval) were transformed with a GT repair template T-DNA for GT at *PPO* using floral dip. A total of 101 plant lines from pool 1 and 43 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 *KpnI* 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 *KpnI* 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 *KpnI* site in the *PPO* repair template, plants rendered butafenacil-resistant through a GT event should have incorporated this new *KpnI*-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 *KpnI* 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 *KpnI*-sensitive and *KpnI*-resistant PCR product. Eleven true GT events were detected, indicated by *KpnI*-sensitivity

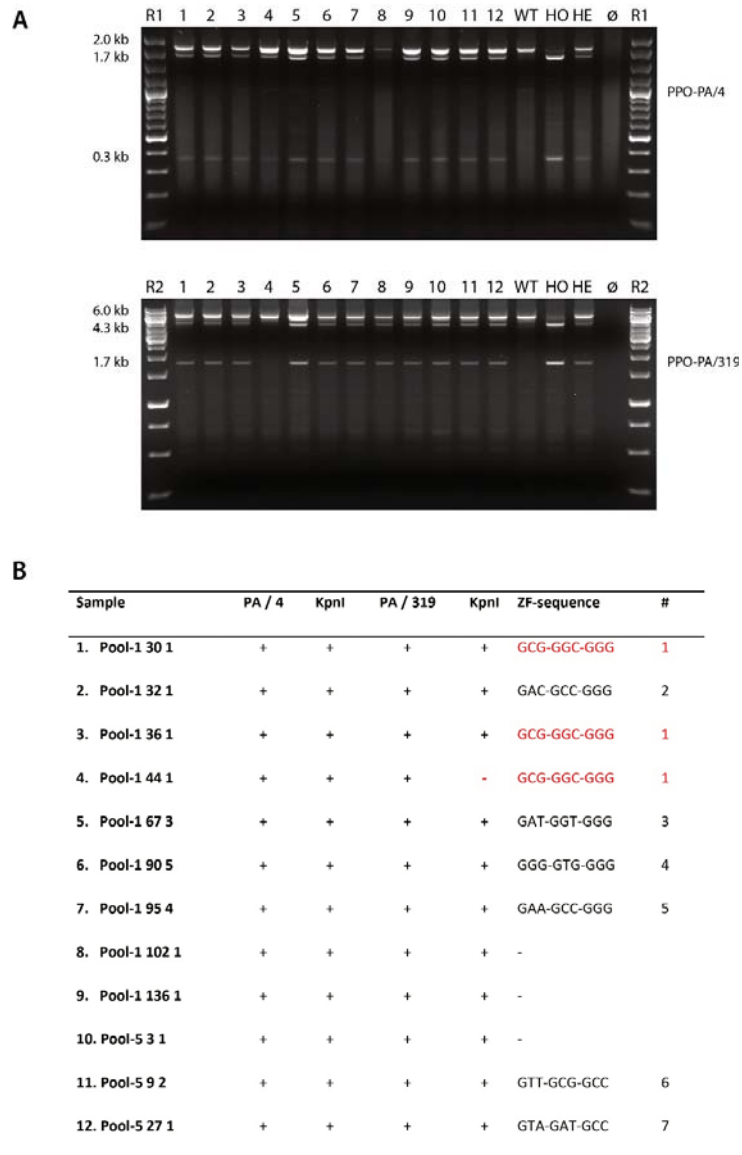


Figure 2. GT events detected with initial screening of ZF-EAR lines. **A.** PCR was performed on 12 butafenacil-resistant plant lines transformed with ZF-EAR constructs and as controls on wild-type (WT), homozygous GT (HO) and heterozygous (HO) plant lines. Primers PPO-PA and PPO-4 were used for 5' GT (A) and primers PPO-PA and SP319 for true GT (B), followed by *KpnI* 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 *KpnI*-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 a 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

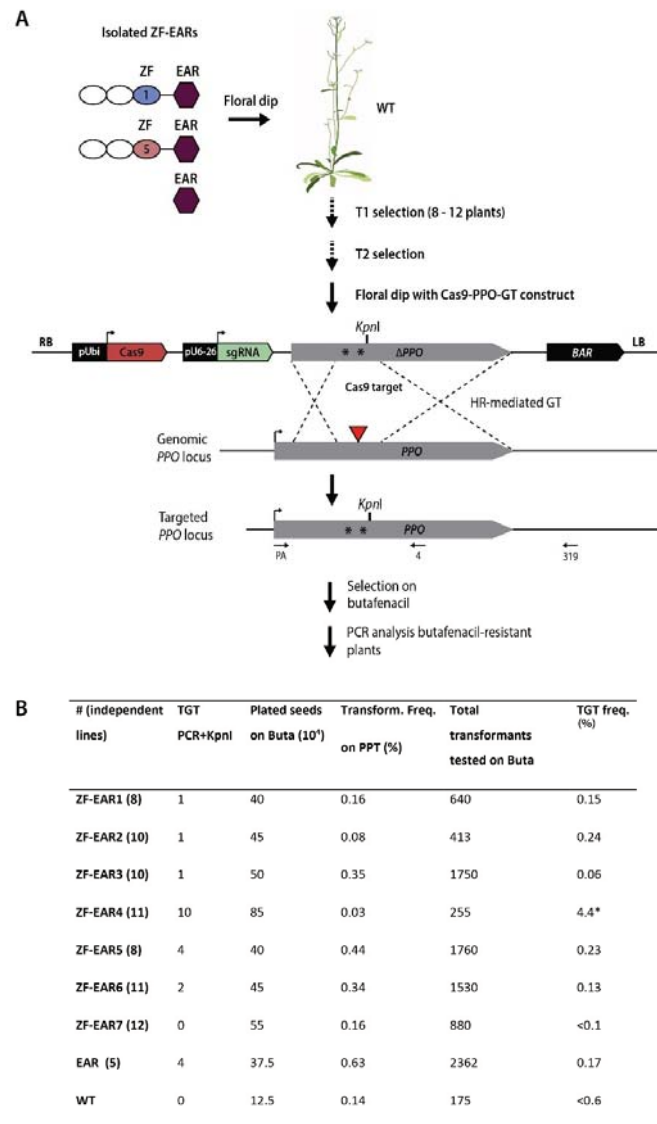


Figure 3. Screening for GT at *PPO* in reconstituted ZF-EAR plant lines. **A.** The 3F-EAR encoding fragments, designated 3F-EAR1-7, were isolated from plants with GT events from the initial screening, and reconstituted ZF-EAR T-DNA constructs were generated. Col-0 plants were transformed with these 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 cassettes. Primary transformant seeds were plated on medium containing butafenacil to select for GT events at *PPO*. PCR analysis and *KpnI* 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 *KpnI* 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 *KpnI*-sensitivity and *KpnI*-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 butafenacil 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.

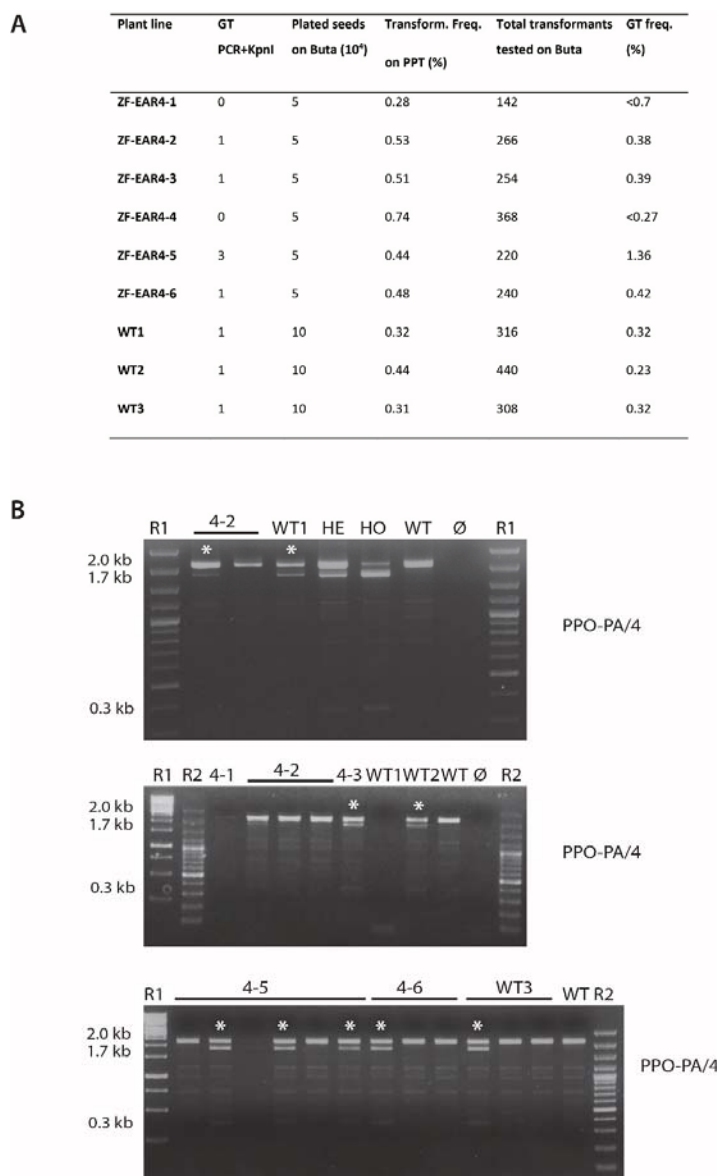


Figure 4. Overview of GT events at *PPO* in reconstituted ZF-EAR4 lines. **A.** Number of GT events and GT frequencies at *PPO* in T2 3F-EAR4-1 to 4-6 reconstituted plant lines, and wild type plants. **B.** GT events in butafenacil resistant 3F-EAR4-2, 4-3, 4-5, 4-6 and wild type plants found with PCR 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 butafenacil-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 *KpnI*. 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 *KpnI*-sensitivity and *KpnI*-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 *CRU3* 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.

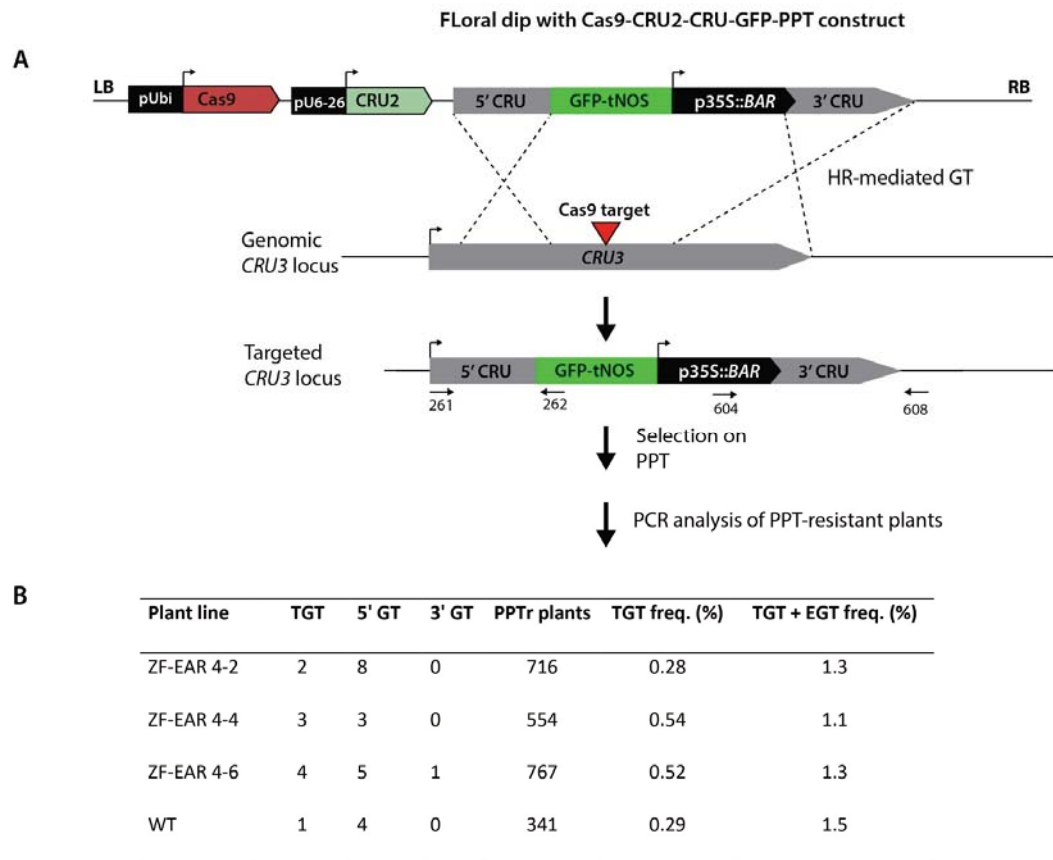


Figure 5. Screening for GT at *CRU3* in reconstituted ZF-EAR4 plant lines **A**. T2 plant lines 4-2, 4-4 and 4-6 harboring ZF-EAR4 transformed with the Cas9-CRU2-CRU-GFP-PPT construct, containing the CRU-GT repair template with BAR as a selection marker, conferring PPT resistance, and Cas9-CRU2 expression cassettes. 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.

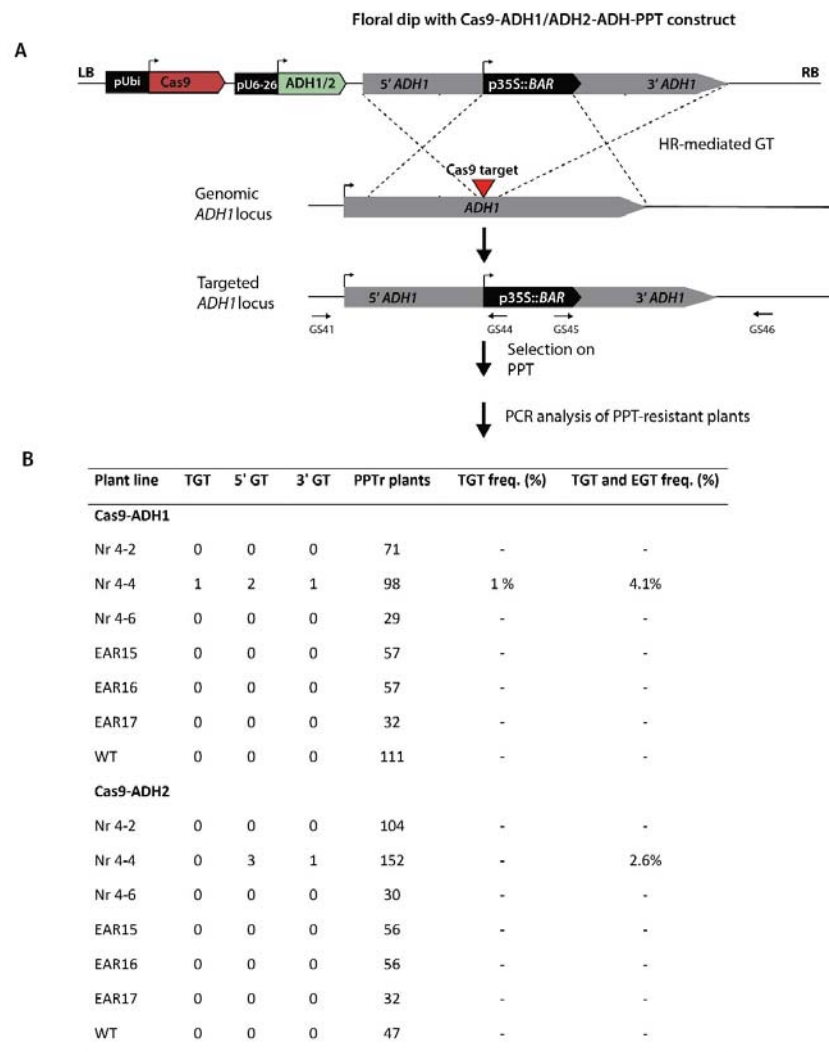


Figure 6. Screening for GT at *ADH1* in reconstituted ZF-EAR4 plant lines **A**. T2 plant lines 4-2, 4-4 and 4-6 harboring ZF-EAR4 transformed with the Cas9-ADH1-ADH-PPT or Cas9-ADH2-ADH-PPT constructs, containing the *ADH1* GT repair template with *BAR* as a selection marker, conferring PPT resistance, or Cas9-ADH1 and Cas9-ADH2 expression cassettes, 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 1/5th 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, Triezenberg 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	GCCCAAACCTAAATTTCATC	3F-EAR sequence, forward
NOSr	CAAGACCGGCAACAGGAT	3F-EAR sequence, reverse
PPO-PA	GTGACCGAGGCTAAGGATCGT	<i>PPO</i> GT forward
PPO-4	CATGAAGTTGTTGACCTCAATC	<i>PPO</i> GT reverse (5' GT)
SP319	CTATCAAAGAGCACAGACAGC	<i>PPO</i> GT reverse (true GT)
SP261	CTCAGCAATCTCCTCGTTG	5' GT <i>CRU3</i> forward
SP262	TCGCCCTTGCTCACCAT	5' GT <i>CRU3</i> reverse
SP604	ATGAGCCCAGAACGACGCCC	3' GT <i>CRU3</i> forward
SP608	CAGAAACAGAGCACCAAATGGG	3' GT <i>CRU3</i> reverse
SP585	GGATCCTGCAGTAATTCGGGGGATCTGGATTTTAG	35S-PPT-Tnos forward
SP586	CACTGCAGGCTAGAGCAGCTTGCAAC	35S-PPT-Tnos reverse
GS41	ATTCCAACCTTGATGACCAAG	5' GT <i>ADH1</i> forward
GS44	GCAATGATGGCATTGTAGG	5' GT <i>ADH1</i> reverse
GS45	GTGTGAGTAGTTCCCAGATAAGG	3' GT <i>ADH1</i> forward
GS46	TGAGTCTTGAAGGCATCGTC	3' GT <i>ADH1</i> reverse

3F-EAR3: GAT-GGT-GGG

GAT sequence: T-deletion

correct GCGGAGAAGCCTTATGCCTGTCCTGAGTGCAGGAAGTCTTTTAGCACCTCGGGTAATCTCGTCCGTCACCAACGTACGCATACC
-T GCGGAGAAGCCTTATGCCTGTCCTGAGTGCAGGAAGTCTT-TAGCACCTCGGGTAATCTCGTCCGTCACCAACGTACGCATACC

Resulting amino acid sequence 3F-EAR3:

Correct sequence:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPECGKSFSTSGNLVRHQRTHTGEKPYACPECGKSFSTSGHLVRHQRTHTGE
KPYACPECGKSFSSRDKLVRHQRTHTGELGGSGEKP GKKTSGAQQLDLDLELR LGFA*

-T:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPECGKSLAPRVISSVTNVRIPARSPMRALSVASPLAPRVIQVCTNVHIPAK
NRMLVQNVVSPFLVAISSFATKGLIPASWVAVPEKSPGRKPRARPANWIWIWN*

3F-EAR4: GGG-GTG-GGG

1st GGG sequence: T-deletion

correct GCGGAAAAACCGTATGCTTGTCAGAAATGTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAGGACTCATACC
-T GCGGAAAAACCGTATGCTTGTCAGAA-GTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAGGACTCATACC

Resulting amino acid sequence 3F-EAR4:

Correct sequence:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPECGKSFSSRDKLVRHQRTHTGEKPYACPECGKSFSSRDELVRHQRTHTGE
KPYACPECGKSFSSRDKLVRHQRTHTGELGGSGEKP GKKTSGAQQLDLDLELR LGFA*

-T:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPEVVSFPLVAISSFATKGLIPAKSRMRAPNVGSPFLVATSLFATKGPIPAK
NRMLVQNVVSPFLVAISSFATKGLIPASWVAVPEKSPGRKPRARPANWIWIWN*

3F-EAR5: GAA-GCC-GGG

GGG sequence: G/A substitution

correct GCGGAAAAACCGTATGCTTGTCAGAAATGTGGTAAGTCCTTTTCTCGTAGCGATAAGCTCGTTCGCCACCAAGGACTCATACC
G/A GCGGAAAAACCGTATGCTTGTCAGAAATGTGGTAAGTCCTTTTCTCGTAGCAATAAGCTCGTTCGCCACCAAGGACTCATACC

Resulting amino acid sequence 3F-EAR5:

Correct sequence:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPECGKSFSSQSSNLVRHQRTHTGEKPYACPECGKSFSDCRDLARHQRTHTGE
KPYACPECGKSFSSRDKLVRHQRTHTGELGGSGEKP GKKTSGAQQLDLDLELR LGFA*

G/A:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPECGKSFSSQSSNLVRHQRTHTGEKPYACPECGKSFSDCRDLARHQRTHTGE
KPYACPECGKSFSSRDKLVRHQRTHTGELGGSGEKP GKKTSGAQQLDLDLELR LGFA*

3F-EAR7: GTA-GAT-GCC

GTA sequence: G-deletion

correct GCGGAGAAGCCTTACGCTTGCCCCGAGTGTGGCAAATCGTTCTCCAGTCTCTCGTTGGTCCGTCATCAACGGACGCATACC
-T GCGGAGAAGCCTTACGCTTGCCCCGAGT-TGGCAAATCGTTCTCCAGTCTCTCGTTGGTCCGTCATCAACGGACGCATACC

Resulting amino acid sequence 3F-EAR7:

Correct sequence:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPECGKSFSSQSSNLVRHQRTHTGEKPYACPECGKSFSTSGNLVRHQRTHTGE
KPYACPECGKSFSDCRDLARHQRTHTGELGGSGEKP GKKTSGAQQLDLDLELR LGFA*

-G:
MADYKDDDDKRPLEPPKKKKRKVELAGTEAQAALPTGEKPYACPELANRSPSPRWSVINGRIPARSLMPVLSAGSLAPRVISSVTNVRIPAR
SLTLARSVASHSATAVISQGINAPIASWVAVPEKSPGRKPRARPANWIWIWN*

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.

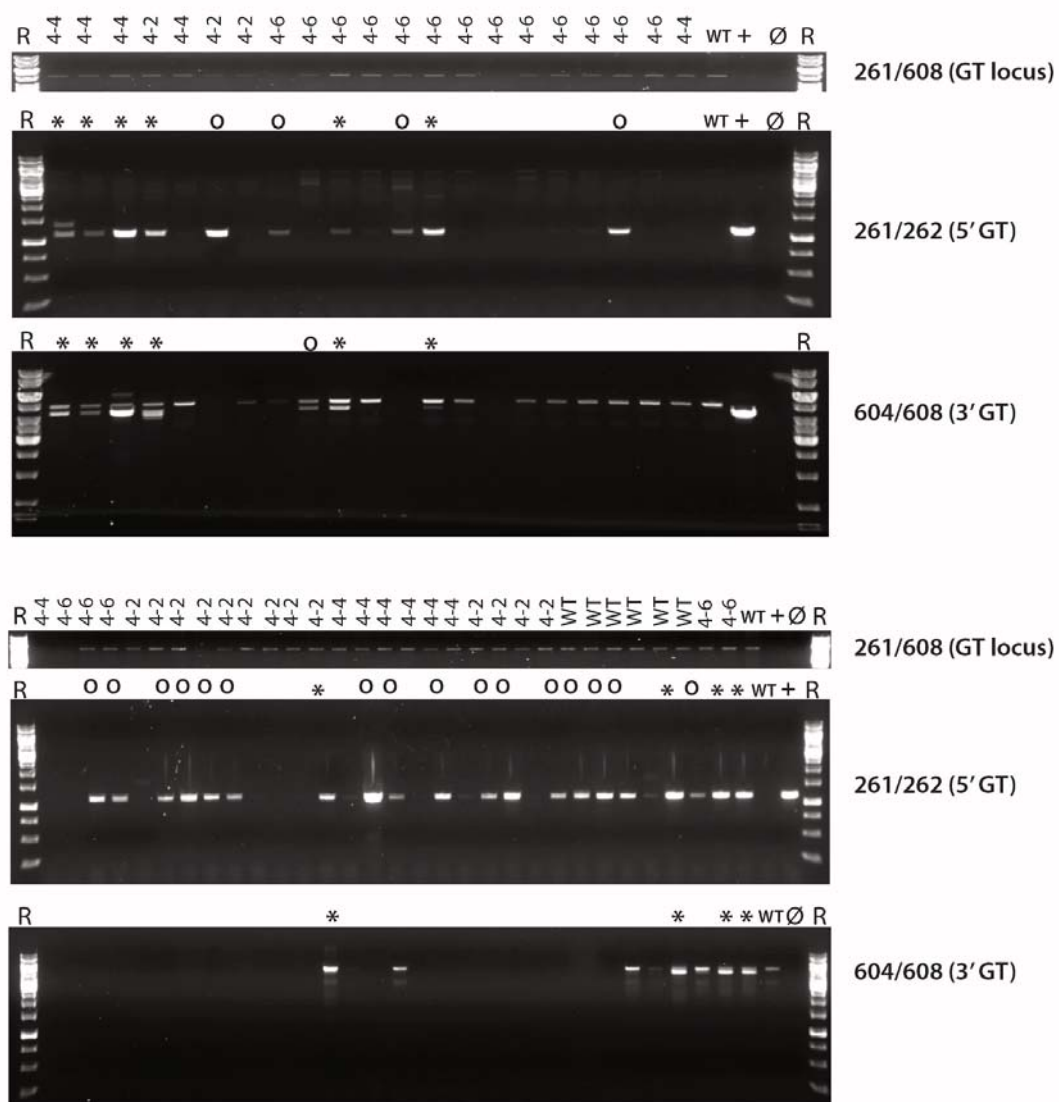


Figure S2. *CRU3* GT events detected with PCR in ZF-EAR4 lines 4-2, 4-4 and 4-6 and WT Col-0. The *CRU3* target was first amplified using primers SP261/SP608. This PCR product was used as a template for 5' GT and 3' GT detection using primers SP261/SP262 and SP604/SP608, respectively. True GT events 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).

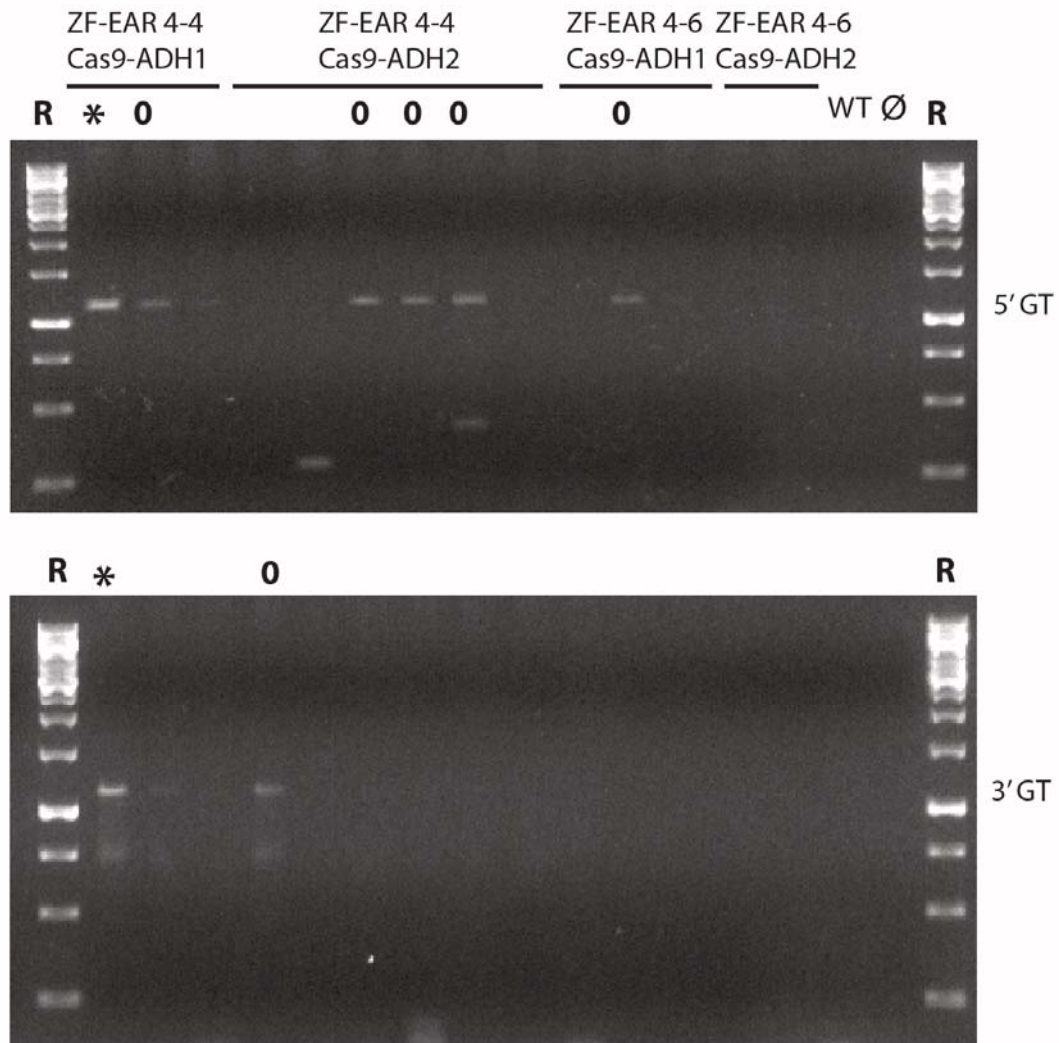


Figure S3. *ADH1* GT events detected with PCR in ZF-EAR4 lines 4-4 and 4-6. 5' GT and 3' GT was detected using primers GS41/44 and GS45/46, respectively. True GT events are indicated with a star. 5' or 3' ectopic GT events are indicated with a circle. WT is wild-type Col-0, Ø is PCR control without template.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

Chapter 6

Summarizing discussion

Gary Strunks

DNA double-strand breaks (DSBs) are harmful lesions that can cause genomic instability and mutations if not repaired. Over the course of evolution, cells have evolved complex and highly conserved mechanisms to detect and repair DSBs in order to maintain genome integrity. DSBs can be repaired by two main pathways: nonhomologous end-joining (NHEJ) and homologous recombination (HR). NHEJ-mediated DSB repair can occasionally lead to mutations (insertions or deletions) at the break site, while HR ensures accurate DSB repair by utilizing homologous DNA sequences. In higher eukaryotes, the majority of DSBs are repaired through NHEJ in somatic cells, as this mechanism acts throughout the cell cycle. HR, on the other hand, is dependent on the availability of homologous DNA sequences at specific points during the cell cycle (1).

Understanding the NHEJ and HR pathways has been of great importance for the application of targeted mutagenesis and gene targeting in molecular biology and biotechnology. In **chapter 1**, the current understanding of DSB repair via NHEJ and HR is described in detail, as well as the exploitation of these pathways for plant genome editing. The advent of artificial sequence-specific nucleases (SSNs) such as zinc-finger nucleases (ZFNs), TALENs and CRISPR/Cas systems have substantially increased the feasibility of NHEJ-mediated targeted mutagenesis, which can now routinely be used as a plant biotechnology tool. However, the development and applicability of GT as a plant biotechnology tool has been much more challenging due to two roadblocks: the low frequency of HR in somatic cells and low transformation efficiencies in plants that impact GT frequency. Although several approaches led to increased GT efficiencies, including the use of SSNs and manipulation of the NHEJ and HR machinery, GT frequencies generally remained below the 1% range (2, 3). Therefore, new methods have to be developed in order to make GT an effective and feasible plant biotechnology tool.

To attempt to overcome the two roadblocks of low HR frequency in somatic cells and low transformation efficiencies in plants, we developed an *in planta* GT system aimed at performing GT in meiocytes, cells that by nature are prone to perform HR. The first step in developing this system was the construction and selection of efficient SSNs. In **chapter 2**, we tested the targeted mutagenesis efficiencies of two classes of SSNs: TALENs and CRISPR/Cas9, designed to target the Arabidopsis *CRU3* and *ADH1* genes. The CRISPR/Cas9 system performed noticeably better at both genes and performed best when a slightly different protospacer sequence was used ending with a GG sequence. These results were in accordance with other studies performed in *C. elegans* and human cells (4, 5), and may point towards a

more optimal protospacer design in general. The CRISPR/Cas9 constructs designed for the *CRU3* and *ADH1* genes were satisfactorily efficient for targeted mutagenesis. The Cas9 version used in all of our experiments was derived from *Streptococcus pyogenes* (SpCas9), but it has been reported that targeted mutagenesis efficiencies can get even higher when using Cas9 variants derived from different bacterial species. For example, average targeted mutagenesis frequencies at the Arabidopsis *ADH1* gene as high as 76% were obtained when using Cas9 derived from *Staphylococcus aureus* (SaCas9) (6), showing that SaCas9 can be an attractive alternative to SpCas9 for targeted mutagenesis and GT.

Elucidating the interplay between NHEJ and HR pathways in plants can give more insight in how these pathways can be further exploited for genome engineering. In **chapter 3**, we performed CRISPR/Cas9-mediated targeted mutagenesis at the Arabidopsis *CRU3* and *PPO* genes and analysed the repair outcomes in plants deficient in classical NHEJ (C-NHEJ) factor KU80 and alternative end-joining (A-EJ) factors PARP1 and PARP2. Absence of KU80 resulted in larger deletions, confirming its role in preventing DSB end resection. Absence of PARP1/2, however, did not seem to have a significant effect on repair outcomes, which suggested the activity of a KU- and PARP-independent a-EJ mechanism. We suggested that this activity could possibly involve Polymerase θ (POLQ), because POLQ is the factor essential for random T-DNA integration at DNA breaks in *Arabidopsis thaliana* (7). Indeed, POLQ was recently identified as the A-EJ key factor by Mara *et al.* 2019, in the moss *Physcomitrella patens* (8). In the absence of POLQ, CRISPR-induced mutation rate and A-EJ-based DSB repair was strongly reduced. Absence of POLQ also led to a dramatic increase in GT frequency by both reducing random integration of the repair template and increasing HR frequency (8). However, compared with flowering plants, GT rates in *P. patens* are high and it remains to be determined what effect disruption of POLQ has on GT efficiency in flowering plant such as Arabidopsis.

As GT frequency in higher plants is low, high transformation frequencies for the repair template are necessary for GT experiments. However, many crops cannot be transformed easily. To overcome this limitation, the group of Holger Puchta developed the *in planta* GT strategy (9). This strategy relies on the simultaneous induction of a DSB at a target gene and DSB induction at sequences flanking a stably integrated repair template, resulting in its release from the genome to be utilized for HR-mediated repair. GT events that occur in the germline will be transmitted to the next generation. With GT frequencies of about 0.14%, there was still room for further improvements (10). In **chapter 4** we describe an attempt to improve GT efficiency with the *in planta* strategy by directing GT to meiocytes, which should have an intrinsically

higher HR rate and could subsequently lead to more heritable GT events. We developed our *in planta* GT system for the seed-specific Arabidopsis *CRU3* gene based on the GT system developed by Shaked *et al.* (11), and used the efficient Cas9-CRU construct described in chapter 2 for DSB induction. A GFP sequence in the GT repair template was used to allow for detection of GT events as GFP-fluorescent seeds. To direct GT to meiocytes, Cas9 was expressed under the *SPO11-1*, *RPS5a* and *SPL* promoters, which are active during meiosis. Plants harbouring a CRU-GFP repair template were crossed with plants harbouring meiosis/germline-specific Cas9 expression cassettes, and GT events that arise during the development of F1 progeny should be visible in F2 seeds as inherited events.

Surprisingly, GFP-fluorescence was already detected in F1 seeds derived from all crossings. Molecular analysis did not reveal signatures of GT events at *CRU3* in plants derived from these seeds, and fluorescence was absent in F2 seeds derived from plants germinated from GFP-fluorescent F1 seeds. One possible explanation might be that the observed GFP signal was the result of a transiently expressed CRU-GFP fusion protein, created by one sided elongation of the CRU-GFP repair template, thereby obtaining a promoter sequence. Such an RT will be lost in the next generation when it is not integrated in the genome. It remains to be determined if this is indeed the case.

GFP-fluorescence was absent in F2 seeds, but true GT (TGT) events were detected in F2 plants with PCR analysis. All GT events were detected in plants in which the repair template could be excised from the genome, suggesting that repair template excision could benefit GT, which has also been shown in other studies (3, 9, 10). However, it is not a prerequisite (12). GT events that were detected with PCR analysis could not be confirmed with Southern blot analysis. This might indicate that the detected GT events were somatic events that occurred in a small number of somatic cells, and were therefore only detectable with PCR and not with Southern blot. This has been observed by others (13), and is also in line with the observation that most GT events occurred in F2 plants that constitutively expressed Cas9 and were not transmitted to the germline, which is supported by the lack of heritable mutations using pUbi-Cas9 and absence of GFP fluorescence in F3 seeds.

Heritable mutations could not be detected when using the *SPL* promoter for targeted mutagenesis, which is in contrast with a study by Mao *et al.* 2016 (14). However, heritable mutations were detected using the egg cell 1.2 (*EC1.2*) promoter, although at lower levels than described by others (14, 15). It thus seems that this promoter is a suitable candidate for germline-specific Cas9 expression for *in planta* GT, but we were not yet able to test this

promoter for *in planta* GT. Recently, however, the *EC1.1* promoter fused with the *EC1.2* enhancer was used to drive SaCas9 expression for *in planta* GT at the Arabidopsis *ALS* gene, which resulted in a substantial increase in GT frequency up to 6% (3). This promoter combined with an efficient Cas9 thus seems to be very suitable for gene targeting, and it may be worthwhile to test this approach in our *CRU3* *in planta* GT system. That being said, the seed-specificity of *CRU3* might hamper the occurrence of GT in other plant tissue due to a condensed chromatin state.

Besides *in planta* GT, other ways to enhance GT efficiency were explored. **Chapter 5** describes attempts to identify plants with higher GT frequency using the genome interrogation method. Previously, genome interrogation was successfully applied in *Arabidopsis* to uncover novel phenotypes, including increased somatic HR, high salinity tolerance and enhanced growth characteristics (16–18). We used the 3F-EAR genome interrogation library instead of the library based on 3F-VP16 fusions, as plants transformed with 3F-EAR fusions previously showed larger phenotypic variation compared to plants transformed with 3F-VP16 fusions (18). Initial screening of plants transformed with the 3F-EAR library, using the *PPO* GT system, yielded true GT events, and 3F-EAR sequences were retrieved from these plants. Although four out of seven retrieved 3F sequences contained mutations, we decided to continue with the analysis of GT frequencies in reconstituted 3F-EAR lines, as these 3F-EAR constructs might still be able to have an effect on GT.

When screening for GT at *PPO* in reconstituted 3F-EAR plants we detected an increase in GT frequency in plants harbouring the 3F-EAR4 construct. However, we could not detect a significant GT enhancement in reconstituted 3F-EAR4 plants for the *CRU3* and *ADH1* genes. Based on these results, we cannot rule out that 3F-EAR4 could indeed trigger GT enhancement, but in order to find statistically significant numbers of GT events, this experiment should be performed on a larger scale. Although genome interrogation has proven to be a powerful method to uncover novel phenotypes, using this method in combination with GT experiments on a larger scale would be time consuming and logistically challenging. In addition, the ZF-VP16 collection could be screened for GT mutants.

In conclusion, this thesis describes the development of novel tools for the improvement of plant genome editing, in particular gene targeting. Although we could not conclusively show improvement in GT efficiency, the tools developed could be further improved in order to be able to induce higher GT frequencies in future experiments. For example, the *S. aureus* Cas9 ortholog (SaCas9) has been shown to be more efficient than the widely used SpCas9 for both

targeted mutagenesis and GT (2, 3). Germline-specific expression of Cas9 with an efficient promoter turns out to be a crucial step in generating both heritable targeted mutations and GT events. Germline-specific SaCas9 expression by the *EC1.1* promoter fused to the *EC1.2* enhancer resulted in feasible GT frequencies at the Arabidopsis *ALS* gene (3). SaCas9 expression with the *EC1.1* promoter or other efficient germline-specific promoters may also help to increase *in planta* GT efficiencies at other genes, including *CRU3*. Beside optimizing tools for DSB induction, inactivation or repression of POLQ in higher plants might help to enhance GT frequencies even further, preferably via transient methods (7, 8). Combining such a method with efficient germline-specific Cas9 expression and DSB induction could eventually lead to GT frequencies that facilitate to apply routinely gene modification and gene replacement in crop plants.

References

1. Mao Z, et al. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* 7(18):2902–2906.
2. Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. *Plant Cell Rep* 35(7):1–10.
3. 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*. *Plant J* 94(4):735–746.
4. Farboud B, Meyer BJ (2015) Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* 199:959–971.
5. Malina A, et al. (2015) PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. *Nat Commun* 6:10124.
6. Steinert J, Schiml S, Fauser F, Puchta H (2015) Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *Plant J* 84(6):1295–1305.
7. van Kregten M, et al. (2016) T-DNA integration in plants results from Polymerase Theta-mediated DNA repair. *Nat Plants* 2:1–6.
8. Mara K, et al. (2019) POLQ plays a key role in the repair of CRISPR / Cas9-induced double-stranded breaks in the moss *Physcomitrella patens*. *New Phytol* 222:1380–1391.
9. Fauser F, et al. (2012) *In planta* gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
10. 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.

11. Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102(34):12265–12269.
12. 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.
13. Feng Z, et al. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc Natl Acad Sci U S A* 111(12):4632–7.
14. Mao Y, et al. (2016) Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol J*:1–14.
15. Wang ZP, et al. (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol* 16(1):1–12.
16. 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.
17. 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.* doi:10.1111/pce.12805.
18. Tol N Van, et al. (2017) Enhancement of Arabidopsis growth characteristics using genome interrogation with artificial transcription factors. *PLoS One* 12(3):e0174236.

Chapter 7

Samenvatting

Gary Strunks

Dubbelstrengs breuken (DSB-en) zijn één van de meest schadelijke vormen van DNA schade die, indien niet gerepareerd, genomische instabiliteit en mutaties kunnen veroorzaken. Cellen hebben complexe en zeer geconserveerde mechanismen geëvolueerd om DSB-en te kunnen repareren en zo de genomische integriteit in stand te houden. Er zijn twee hoofdroutes voor DSB-reparatie: Niet-Homologe End Joining (NHEJ) en Homologe Recombinatie (HR). Reparatie via NHEJ kan soms resulteren in mutaties (inserties of deleties) op of rond de plaats van de breuk. HR zorgt voor een accurate DSB-reparatie door gebruik te maken van homologe DNA sequenties. In hogere eukaryoten wordt het grootste deel van somatische DSBs gerepareerd via NHEJ, omdat dit mechanisme in de gehele celcyclus beschikbaar is. HR is daarentegen afhankelijk van de beschikbaarheid van homologe DNA sequenties op specifieke momenten in de celcyclus (1).

Begrip van de NHEJ en HR-herstelroutes is van groot belang voor toepassingen van gerichte mutagenese en gen targeting (GT) in de moleculaire biologie en biotechnologie. **Hoofdstuk 1** geeft een gedetailleerd overzicht van de huidige kennis van DSB-reparatie via NHEJ en HR, alsook van het gebruik van deze herstelroutes in “genoomediting” in planten. De komst van kunstmatige sequentie-specifieke nucleasen (SSN-en) zoals zinc-finger nucleasen (ZFN-en), TALENs en het CRISPR/Cas systeem heeft ervoor gezorgd dat gerichte mutagenese routinematig toegepast kan worden in de plantenbiotechnologie. De ontwikkeling en toepasbaarheid van GT als genoomediting methode is echter veel lastiger gebleken. Dit komt door twee obstakels: de lage HR frequentie in somatische cellen en de lage transformatie efficiëntie in planten. Hoewel het via een aantal benaderingen was gelukt om de efficiëntie van GT te verhogen, zoals door het gebruik van SSNs en door manipulatie van de NHEJ en HR machinerie, bleven GT frequenties in het algemeen onder de 1% (2, 3). Het is daarom van belang dat er nieuwe methoden ontwikkeld worden om GT een efficiënte en haalbare toepassing te maken voor plantenbiotechnologie.

Dit proefschrift beschrijft de ontwikkeling van een *in planta* GT systeem als poging om de twee obstakels van lage HR frequentie in somatische cellen en lage transformatie efficiëntie in planten te overwinnen. Het *in planta* GT systeem is gericht op het uitvoeren van GT in meiocyten, cellen met een van nature hogere HR efficiëntie. De eerste stap in de ontwikkeling van dit systeem was de constructie en selectie van efficiënte SSNs. **Hoofdstuk 2** beschrijft het testen van twee klassen SSNs: TALENs en CRISPR/Cas9, ontwikkeld voor het maken van gerichte mutaties in de Arabidopsis genen *CRU3* en *ADH1*. Het CRISPR/Cas9 systeem presteerde duidelijk beter voor beide genen dan de TALENs, en presteerde het beste met een

GG sequentie aan het 3' uiteinde van de protospacer. Deze resultaten waren in overeenstemming met studies in *C. elegans* en humane cellen (4, 5) en kunnen wijzen op een algemene regel voor meer optimaal protospacerontwerp. De ontworpen CRISPR/Cas9 constructen voor *CRU3* en *ADH1* waren voldoende efficiënt voor gerichte mutagenese.

Het ophelderen van de interactie tussen NHEJ en HR in planten kan meer inzicht geven in hoe deze routes kunnen worden benut voor “genome editing”. In **hoofdstuk 3** hebben we gerichte mutagenese uitgevoerd met CRISPR/Cas9 voor de Arabidopsis *CRU3* en *PPO* genen en de hersteluitkomsten geanalyseerd in planten die deficiënt zijn voor de klassieke NHEJ factor KU80 en de alternatieve end joining (A-EJ) factoren PARP1 en PARP2. Afwezigheid van KU80 resulteerde in grotere deleties en bevestigde een rol voor KU80 in de preventie van DSB-uiteinde resectie. De afwezigheid van PARP1/2 leek echter geen significant effect te hebben op de hersteluitkomsten, wat duidde op een activiteit van een KU- en PARP-onafhankelijk A-EJ mechanisme. We stelden voor dat Polymerase θ (POLQ) betrokken is bij deze activiteit, omdat POLQ een essentiële factor is voor willekeurige T-DNA integratie bij DNA breuken in *Arabidopsis thaliana* (6).

Omdat de GT frequentie in hogere planten laag is, zijn hoge transformatiefrequenties voor de reparatiesjabloon nodig voor GT experimenten. Veel gewassen kunnen echter niet gemakkelijk worden getransformeerd. Om deze beperking te ondervangen, ontwikkelde de groep van Holger Puchta de *in planta* GT strategie (7). Deze strategie is gebaseerd op de gelijktijdige inductie van een DSB op een gen van interesse en DSB-inductie op sequenties flankerend aan een stabiel geïntegreerd reparatiesjabloon. Dit resulteert in het vrijkomen van het reparatiesjabloon uit het genoom voor het gebruik van DSB herstel via HR. GT events die in de kiembaan plaatsvinden, worden doorgegeven aan de volgende generatie. Met GT frequenties van 0,14% (8) was er echter nog steeds behoefte aan verdere verbetering. In **hoofdstuk 4** beschrijven we een poging om de GT efficiëntie te verbeteren met de *in planta* strategie door GT te laten plaatsvinden in meiocyten, welke een intrinsiek hogere HR frequentie zouden moeten hebben en vervolgens zouden kunnen leiden tot meer overerfbare GT-events. We hebben ons *in planta* GT systeem ontwikkeld voor het zaad specifieke Arabidopsis *CRU3* gen op basis van het GT systeem ontwikkeld door Shaked *et al.* (9), en gebruikten het efficiënte Cas9-CRU construct beschreven in hoofdstuk 2 voor DSB inductie. Een GFP-sequentie in het GT-reparatiesjabloon werd gebruikt om “GT events” te detecteren als GFP-fluorescerende zaden. Om GT in meiocyten te laten plaats vinden, werd Cas9 tot expressie gebracht onder de *SPO11-1*-, *RPS5a*- en *SPL* promotors, die actief zijn tijdens meiose. Planten die een CRU-GFP-

reparatiesjabloon bevatten werden gekruist met planten die meiose/kiemlijn-specifieke Cas9-expressiecassettes bevatten. GT zou nu kunnen plaatsvinden in de meiocyten van deze F1 nakomelingen, en fluorescente F2 zaden moeten opleveren.

Opmerkelijk genoeg werd GFP-fluorescentie al gedetecteerd in F1 zaden afkomstig van alle kruisingen. Moleculaire analyse van planten verkregen uit deze zaden liet echter een onveranderd *CRU3* gen zien in deze planten. Zaden gevormd door deze F1 nakomelingen na zelfbevruchting (F2 zaden) waren ook niet fluorescent, maar echte “GT events” werden met PCR analyse wel gedetecteerd in F2 planten. Dergelijke “GT events” werden met name gedetecteerd in planten waarin de reparatiesjabloon uit het genoom kon worden geknipt, wat suggereert dat excisie van het reparatiesjabloon GT bevordert. Dit is ook in eerder onderzoek aangetoond (3, 7, 8), maar het is geen vereiste (10). “GT events” die werden gedetecteerd met PCR analyse konden niet worden bevestigd met Southern blot analyse. Dit kan erop duiden dat de gedetecteerde recombinaties somatische gebeurtenissen waren die slechts in een klein aantal somatische cellen voorkwamen, en daarom alleen detecteerbaar waren met PCR en niet met Southern blot. Dit is ook waargenomen door anderen (11). Het is ook in overeenstemming met de waarneming dat de meeste somatische mutaties, die plaatsvonden in F2 planten die Cas9 constitutief tot expressie brachten niet overerfbaar waren.

Naast *in planta* GT werden ook andere manieren onderzocht om de GT efficiëntie te verbeteren. **Hoofdstuk 5** beschrijft pogingen om planten met een hogere GT frequentie te identificeren met behulp van de “genome interrogation” methode. Eerder werd deze methode met succes toegepast in *Arabidopsis* om nieuwe fenotypen te ontdekken, waaronder verhoogde somatische HR, hoge zouttolerantie en verbeterde groeikenmerken (12–14). We gebruikten in onze GT experimenten een collectie plantenlijnen met 3F-EAR “genome interrogation” effectoren. GT experimenten (voor het *PPO* locus) met deze planten leverde een aantal “GT events” op, waaruit de 3F-EAR sequenties konden worden geïsoleerd. Met de geïsoleerde 3F-EAR sequenties werden opnieuw planten getransformeerd, zodat gereconstitueerde 3F-EAR planten werden verkregen.

Bij screening voor GT (opnieuw op het *PPO* locus) in deze gereconstitueerde 3F-EAR planten, detecteerden we een toename in GT frequentie in planten die het 3F-EAR4 construct bevatten. We konden echter geen significante GT verbetering vinden in gereconstitueerde 3F-EAR4 planten wanneer GT was gericht op de *CRU3* en *ADH1* genen. Op basis van deze resultaten kunnen we niet uitsluiten dat 3F-EAR4 inderdaad voor een GT verbetering kan zorgen, maar om statistisch significante GT frequenties te vinden moet dit experiment op

grotere schaal worden uitgevoerd. Hoewel “genome interrogation” een krachtige methode is gebleken voor het vinden van nieuwe fenotypen, bleek het gebruik van deze methode in combinatie met GT experimenten op grotere schaal tijdrovend en logistiek uitdagend te zijn.

Concluderend beschrijft dit proefschrift de ontwikkeling van nieuwe tools voor de verbetering van “genome editing” in planten, in het bijzonder GT. Het combineren van kiemlijn-specifieke Cas9 expressie en inactivatie of repressie van NHEJ sleutelcomponenten zoals POLQ (6, 15) zou uiteindelijk kunnen leiden tot GT frequenties die het mogelijk maken om genmodificatie en genvervanging routinematig toe te passen in land- en tuinbouwgewassen.

Referenties

1. Mao Z, et al. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* 7(18):2902–2906.
2. Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. *Plant Cell Rep* 35(7):1–10.
3. 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*. *Plant J* 94(4):735–746.
4. Farboud B, Meyer BJ (2015) Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* 199:959–971.
5. Malina A, et al. (2015) PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. *Nat Commun* 6:10124.
6. van Kregten M, et al. (2016) T-DNA integration in plants results from Polymerase Theta-mediated DNA repair. *Nat Plants* 2:1–6.
7. Fauser F, et al. (2012) In planta gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
8. 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.
9. Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102(34):12265–12269.
10. 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.
11. Feng Z, et al. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc Natl Acad Sci U S A* 111(12):4632–7.
12. 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.
13. 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*. doi:10.1111/pce.12805.

14. Tol N Van, et al. (2017) Enhancement of Arabidopsis growth characteristics using genome interrogation with artificial transcription factors. *PLoS One* 12(3):e0174236.
15. Mara K, et al. (2019) POLQ plays a key role in the repair of CRISPR / Cas9-induced double-stranded breaks in the moss *Physcomitrella patens*. *New Phytol* 222:1380–1391.

Curriculum Vitae

Curriculum Vitae

Gary Strunks was born on the 18th of March, 1986, in Amsterdam, the Netherlands. He grew up in Alkmaar, and received his high school diploma from the Jan Arentsz comprehensive school in Alkmaar in 2003. In September 2003, he started his Bachelor studies in Electrical Engineering at the Inholland University of Applied Sciences in Alkmaar, receiving his BEng degree in July 2007. In September 2007, he decided to follow his true passion and enrolled in the BSc Biology studies at the VU University in Amsterdam, graduating in July 2010. He continued his education by entering the 'Biomolecular Sciences' MSc program at the VU University in Amsterdam. After having received his MSc degree in February 2013, he started his PhD research in July 2013 on the STW-funded project "Efficient gene targeting during meiosis", in the group of Molecular and Developmental Genetics at the Institute of Biology Leiden under supervision of dr. Sylvia de Pater and prof. dr. Paul J.J. Hooykaas, the results of which are described in this thesis. Currently, he is working as a junior researcher at the R&D department of the Netherlands Inspection Service for Horticulture (Naktuinbouw) in Roelofarendsveen, the Netherlands.

