



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

## Chapter 2

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

**Gary D. Strunks<sup>1,2</sup>, Bart J. P. M. Klemann<sup>1,3</sup>,  
Paul J. J. Hooykaas<sup>1</sup> and Sylvia de Pater<sup>1</sup>**

<sup>1</sup>Molecular and Developmental Genetics, Institute of Biology,  
Leiden University, 2333 BE, the Netherlands

<sup>2</sup>Current affiliation: Naktuinbouw, 2371 GD Roelofarendsveen,  
the Netherlands

<sup>3</sup>Current 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*III (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 N<sub>2</sub> 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 <b>HD</b>
TALE-CRU-1 right	HD HD NI HD NG HD HD NG HD NN HD NG HD NG HD NN NG <b>NI</b>
TALE-CRU-2 left	NI NN HD HD HD NI HD NI NN NN NN HD NI NI HD NN NN HD HD <b>NG</b>
TALE-CRU-2 right	NG HD NG HD NN NG NN NN NN NI HD HD NG HD <b>NI</b>

**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	ACCACCGGACAGATTATTCG	<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**

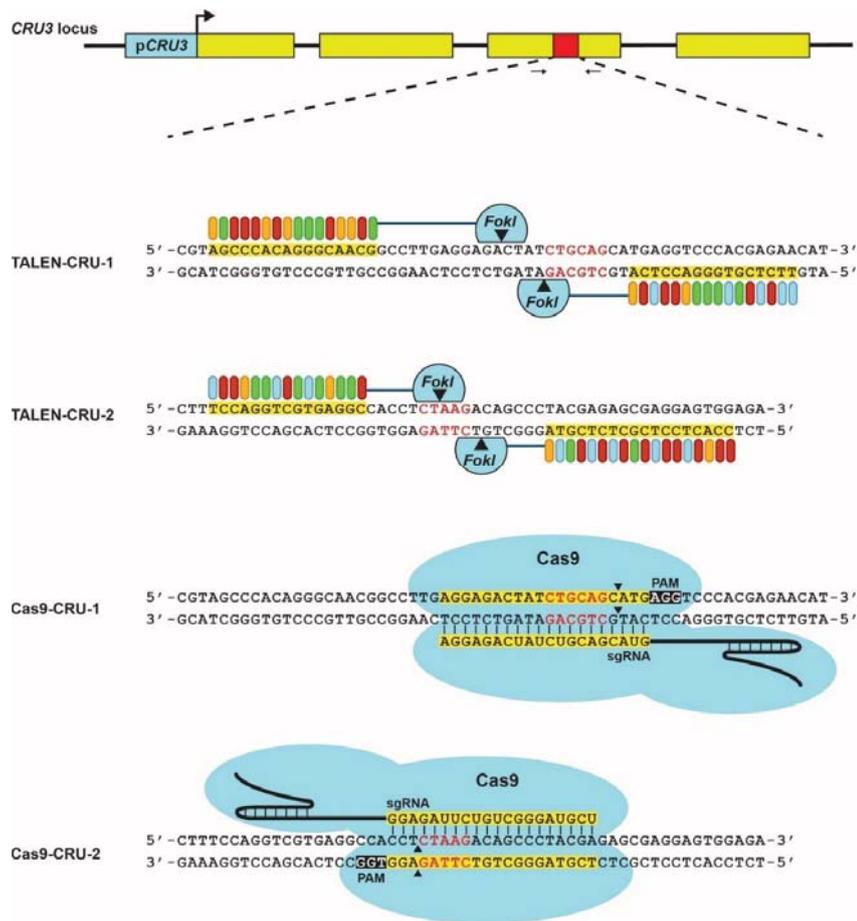
### ***Mutagenesis at the CRU3 locus with TALENs and CRISPR/Cas9***

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, *PstI* was used for the TALEN-CRU-1 and Cas9-CRU-1 assays and *DdeI* 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 *PstI* or *DdeI* 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 *PstI* (TALEN-CRU-1 and Cas9-CRU-1) resulting in a 147 bp and a 125 bp fragment or *DdeI* (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** AGCCACAGGGCAACGGCCTTGAGGAGACTATCTGCAGCATGAGGTCCCACGAGA

**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** TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG

**TALEN-CRU-2**

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

**WT** AGCCACAGGGCAACGGCCTTGAGGAGACTATCTGCAGCATCAGGTCCCACGAGA

**Cas9-CRU-1**

plant line 1 -3 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTBCA-----TGCAGGTCCCACGAGA (2)  
-6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT-----TGCAGGTCCCACGAGA  
-13 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTCA-----CGAGA (2)  
-20 AGCCACAGGGCAACGGCCTTGAGGAG-----TCCCACGAGA (2)  
-50 AGCCACAGGG-----//-----ACGACCTGCTC (2)  
plant line 2 -10 AGCCACAGGGCAACGGCCTTGAGGAG-----ATGAGGTCCCACGAGA  
-14 AGCCACAGGGCAACGGCCTTGAGGAG-----GGTCCCACGAGA  
-14+6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTGCA-----LIIIIAT-----GAGA (2)  
-18+6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTG-----ACGTG-----GA  
plant line 6 -5 AGCCACAGGGCAACGGCCTTGAGGAGACTATCTG-----GAGGTCCCACGAGA  
-6 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT-----ATGAGGTCCCACGAGA  
-9 AGCCACAGGGCAACGGCCTTGAGGAGACTATCT-----GTCCCACGAGA

**WT** TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG

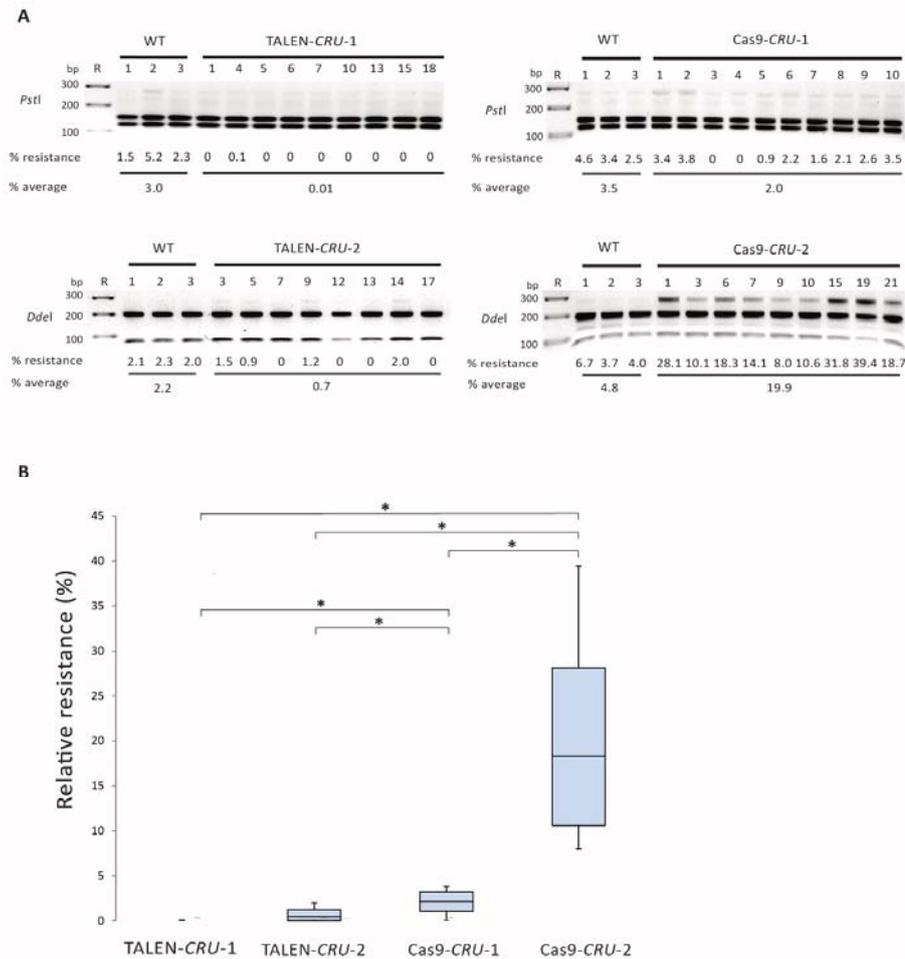
**Cas9-CRU-2**

plant line 1 +1 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-1 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-3 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-5 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-8 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
plant line 3 +1 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-2 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-8 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-13 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
plant line 6 -4 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG  
-5 TCCAGGTCGTGAGGCCACCTCTAAGACAGCCCTACGAGAGCGAGGAGTGG

**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 insertion 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).



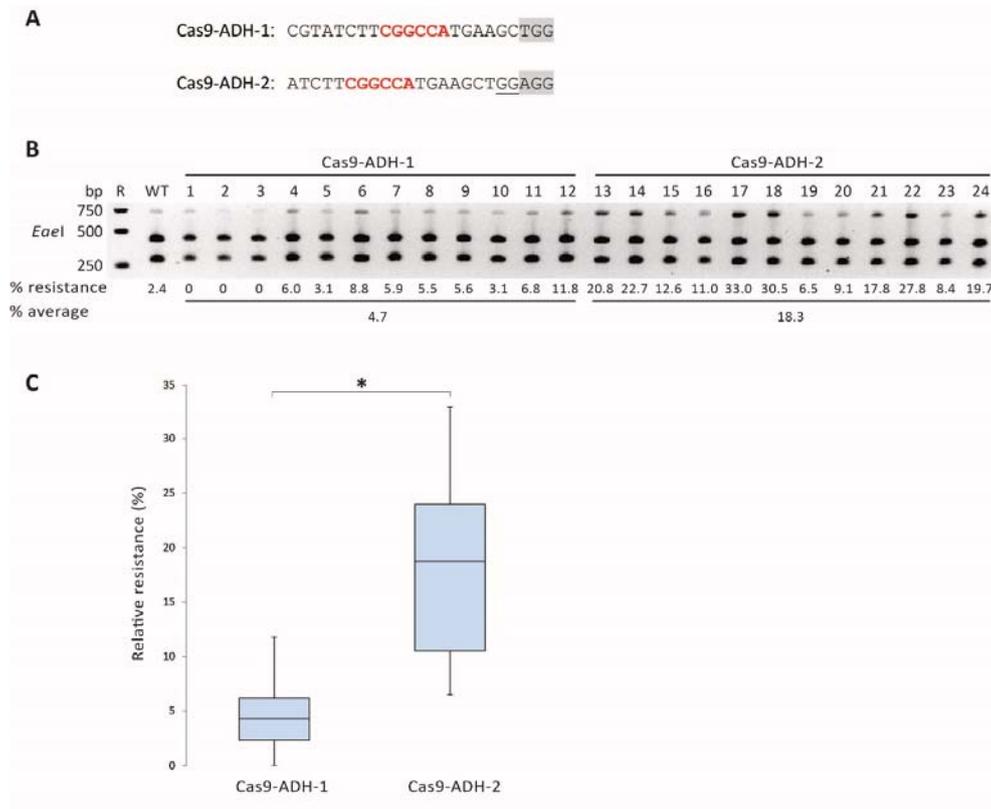
**Figure 3.** Semi-quantification of TALEN and CRISPR/Cas9 endonuclease-induced mutagenesis. **A.** The *CRU3* target was PCR-amplified from undigested genomic DNA from untransformed wild-type seedlings and T2 seedlings transformed with TALEN-CRU-1, TALEN-CRU-2, Cas9-CRU-1 and Cas9-CRU-2 constructs, resulting in a 272 bp fragment. PCR products were digested with *PstI* (TALEN-CRU-1 and Cas9-CRU-1) resulting in a 147 bp and a 125 bp fragment or *DdeI* (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 *PstI*- and *DdeI*-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 *PstI* and *DdeI*-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 ( $p = 4.1 \times 10^{-5}$ ), 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  $\leq 0.05$  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 ( $p = 1,7 * 10^{-5}$ , 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  $\leq 0.05$  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.