the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) **84,** 1295–1305 doi: 10.1111/tpj.13078

TECHNICAL ADVANCE

Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*

Jeannette Steinert, Simon Schiml, Friedrich Fauser† and Holger Puchta*

Botanical Institute II, Karlsruhe Institute of Technology, POB 6980, Karlsruhe 76049, Germany

Received 22 September 2015; revised 28 October 2015; accepted 5 November 2015; published online 18 November 2015.
*For correspondence (e-mail holger.puchta@kit.edu).
†Present address: Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA, 94305, USA.

SUMMARY

The application of the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system of *Streptococcus pyogenes* (SpCas9) is currently revolutionizing genome engineering in plants. However, synthetic plant biology will require more complex manipulations of genomes and transcriptomes. The simultaneous addressing of different specific genomic sites with independent enzyme activities within the same cell is a key to this issue. Such approaches can be achieved by the adaptation of additional bacterial orthologues of the CRISPR/Cas system for use in plant cells. Here, we show that codon-optimised Cas9 orthologues from *Streptococcus thermophilus* (St1Cas9) and *Staphylococcus aureus* (SaCas9) can both be used to induce error-prone non-homologous end-joining-mediated targeted mutagenesis in the model plant *Arabidopsis thaliana* at frequencies at least comparable to those that have previously been reported for the *S. pyogenes* CRISPR/Cas system. Stable inheritance of the induced targeted mutations of the *ADH1* gene was demonstrated for both St1Cas9- and SaCas9-based systems at high frequencies. We were also able to demonstrate that the SaCas9 and SpCas9 proteins enhance homologous recombination via the induction of double-strand breaks only in the presence of their species-specific single guide (sg) RNAs. These proteins are not prone to inter-species interference with heterologous sgRNA expression constructs. Thus, the CRISPR/Cas systems of *S. pyogenes* and *S. aureus* should be appropriate for simultaneously addressing different sequence motifs with different enzyme activities in the same plant cell.

Keywords: double-strand break repair, homologous recombination, non-homologous end-joining, gene editing, clustered regularly interspaced short palindromic repeats, technical advance.

INTRODUCTION

Nearly 20 years ago it was shown that genome engineering can be achieved in plant genomes by the induction of site-specific double-strand breaks (DSBs) with different types of nucleases (Puchta *et al.*, 1996; Salomon and Puchta, 1998). Engineered nucleases, such as meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and most recently the CRISPR/Cas system (for 'clustered regularly interspaced short palindromic repeats'; 'CRISPR-associated'), have become major tools for targeting DSBs to pre-defined genomic sites (for reviews see Voytas, 2013; Puchta and Fauser, 2014; Bortesi and Fischer, 2015; Weeks *et al.*, 2015). Currently, the CRISPR/Cas system is predominantly used in plant biology as a highly efficient nuclease. Its specificity can easily be adapted to any target site of interest simply by replacing 20 nucelotides of the respective single-guide RNA (sgRNA) with 20 DNA oligonucleotides (protospacers) in the expression vector.

This commonly used tool is derived from the type II CRISPR/Cas system and is based on the RNA-guided nuclease Cas9 from *Streptococcus pyogenes* (SpCas9) (for review see Doudna and Charpentier, 2014). For correct target recognition, the nuclease forms a complex with a chimeric sgRNA that includes a portion that is complementary to the target protospacer. To bind its target sequence, the nuclease also requires a so-called protospacer-adjacent motif (PAM). This PAM is located downstream of the target sequence, and consists primarily of the nucleotides 'NGG'

in the case of SpCas9. Notably, the PAM sequences and sizes of the Cas9 proteins vary in different bacterial strains, making them attractive candidates for targeting new genomic loci that cannot be targeted with conventional SpCas9 due to the described PAM specificities (Ran *et al.*, 2015).

Although it is possible to simultaneously modify several target sequences using a SpCas9 multiplex system (Cong *et al.*, 2013; Li *et al.*, 2013; Xie *et al.*, 2015), this approach is only able to simultaneously perform a single catalytic function at multiple sites. Thus, it is not possible, for example, to induce a DSB as well as to regulate transcriptional activities using the same Cas9 protein. The most promising possibility for overcoming this problem is the use of a combination of different Cas9 orthologues that are able to perform different catalytic functions in a single cell. Thus far, only *S. pyogenes*-based CRISPR/Cas systems have been established for applications in model and crop plants (Li *et al.*, 2013, 2014, 2015; Nekrasov *et al.*, 2013; Shan *et al.*, 2013; Fauser *et al.*, 2014; Schiml *et al.*, 2014; Piatek *et al.*, 2015).

In the present study, we demonstrate that different CRISPR/Cas orthologues derived from *Streptococcus thermophilus* and *Staphylococcus aureus* can be used for plant genome engineering. These orthologous systems have different requirements concerning the interaction between crRNA (CRISPR RNA) and tracrRNA (trans-activating crRNA), the target sequence and the PAMs. Considering these facts, these systems should work independently within a single cell.

Both of these orthologues are smaller than *S. pyogenes* Cas9, which makes them useful alternatives, particularly if plant viral replicons are used for expression (Baltes *et al.*, 2014). Regarding *S. thermophilus*, two Cas9 open reading frames (ORFs) orthologous to SpCas9 have been identified, CRISPR3 and CRISPR1; and the latter orthologue is smaller and therefore easier to apply to viral transfection. The orthologous CRISPR1 system of *S. thermophilus* has been shown to work in bacterial and in human cells and also requires different PAMs and a unique sgRNA compared with SpCas9 (Jinek *et al.*, 2012; Esvelt *et al.*, 2013; Kleinstiver *et al.*, 2015; Ma *et al.*, 2015; Ran *et al.*, 2015). Recently, Ran *et al.* tested six smaller Cas9 orthologues and demonstrated that the Cas9 protein of *S. aureus* seems to be one of the most promising orthologues for application in mammalian cells; this orthologue has different PAM and sgRNA requirements and mutagenesis frequencies that are comparable to those of SpCas9 (Ran *et al.*, 2015).

Here, we cloned plant codon-optimised versions of the Cas9 orthologues St1Cas9 and SaCas9, and investigated whether these orthologues could be used for efficient genome engineering in *Arabidopsis thaliana* via non-homologous end-joining (NHEJ) events. For both the SpCas9 and SaCas9 proteins, we also examined whether the correct combination of nucleases with specific sgRNAs led to species-specific PAM sequence recognition for induction of homologous recombination (HR) or whether interference with orthologues of the other species occurred.

RESULTS

Engineering orthologous CRISPR/Cas-based nucleases

For expression, the ORFs of both tested orthologues, St1Cas9 and SaCas9, were codon-optimised for *A. thaliana* and cloned into a binary vector for transformation with *Agrobacterium tumefaciens* as previously described (Fauser *et al.*, 2014). All constructs were driven by the constitutive Ubiquitin 4–2 promoter from *Petroselinum crispum* (PcUbi4-2, Kawalleck *et al.*, 1993). The orthologous sgRNA chimeras are under the control of the Arabidopsis U6-26 promoter and can be customised for any chosen target sequence (Fauser *et al.*, 2014). Figure 1 depicts both of the described T-DNA constructs. In this study, we used orthologous CRISPR/Cas systems to generate targeted mutagenesis events at an endogenous target site (alcohol dehydrogenase 1, *ADH1*, At1G77120) and to detect induction of HR and possible cross-species interference with SaCas9 constructs in a reporter line (DGU.US).

Detection of targeted mutagenesis events using the *S. thermophilus* Cas9 nuclease by amplicon deep sequencing

The efficiency of the *S. thermophilus* CRISPR/Cas system in inducing targeted mutagenesis events via NHEJ was determined by amplicon deep sequencing. Two different

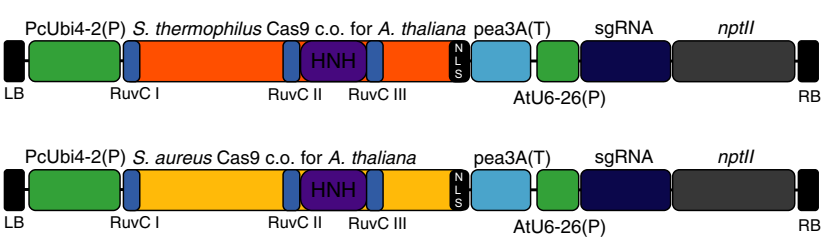

Figure 1. Orthologous CRISPR/Cas expression constructs.
The constructs for *Streptococcus thermophilus* and *Staphylococcus aureus* Cas9 expression are under control of the PcUbi4-2 promoter and the *Pisum sativum* pea3A terminator. Expression of single-guide (sg) RNA is driven by the Arabidopsis U6-26 promoter. For plant selection, the constructs harbour an *npt II* resistance cassette. Double-strand break-inducing activities of the nucleases are based on the nuclease domains (RuvC-like domains, HNH motif).

sgRNAs were designed to target the *ADH1* locus; each construct targeted a different PAM sequence (see Figure 2). Each sgRNA was assembled in the pEn-St1_Chimera vector and subsequently combined with the pDe-St1_Cas9 destination vector via Gateway® cloning. After *Agrobacterium*-mediated transformation, 30 transgenic primary transformants were pooled, and DNA was extracted after 2 weeks of growth. The PCR products spanning the protospacer sequence were amplified using multiplex identifier (MID)-labelled primers and analysed via amplicon deep sequencing on a Roche 454 system, as performed previously (Fauser *et al.*, 2014; Schiml *et al.*, 2014). Figure 2 depicts the results for both constructs as the relative numbers of mutated reads by position. Each position displays the mutations grouped into deletions, insertions and substitutions. For both sgRNAs, the mutation rates at the respective target sites were strongly increased and led to targeted mutagenesis frequencies of up to 12.5% 3 bp away from the PAM for the St1Cas9 nuclease targeting the PAM sequence 'NNAGAA'. These mutations consisted of 8.6% insertions (mostly of a single nucleotide) and 3.8% deletions. Regarding the second sgRNA targeting the PAM sequence 'NNGGAA', frequencies up to 15.4% within 4 bp of the PAM were detected; 11.6% were insertions primarily of a single nucleotide and 3.8% were deletions.

Generation of heritable targeted mutagenesis events with *S. thermophilus* Cas9 constructs

Heritable targeted mutagenesis events with St1Cas9 constructs were detected by targeting the *ADH1* marker gene. The two constructs targeting different PAMs were transformed via floral dipping into *A. thaliana*. Mendelian segregating T_2 lines were selected for allyl alcohol treatment. For the St1Cas9 construct targeting the 'NNAGAA' PAM sequence, approximately 11 500 seeds from four independent primary transformants were treated with allyl alcohol and the resistant seedling frequencies were determined by counting viable versus dead plants, leading to heritable targeted mutation frequencies (Table 1). In wild-type plants, alcohol dehydrogenase 1, which is encoded by the *ADH1* gene, catalyses the reaction of allyl alcohol to acryl aldehyde (Jacobs *et al.*, 1988). This highly toxic product leads to cell death; thus, phenotypic determination of targeted mutagenesis events in the *ADH1* locus is possible by treatment with allyl alcohol and subsequent counting of viable, resistant plants. Confirmation of targeted mutagenesis events was performed by Sanger sequencing of 71 resistant T_2 plants with St1Cas9 constructs targeting the PAM 'NNAGAA', revealing that 44% of the plants exhibited either homozygous or biallelic mutations in *ADH1* (Figure 3a). Allyl alcohol-resistant seedlings were counted for the 'NNGGAA' PAM-targeting constructs, and 23 individual T_2 plants were sequenced. A number of heritable homoallelic mutations that were found in the T_2 plants are shown in Figure 3(b). In total, approximately 7500 seeds of three independent primary transformants were analysed, and the results revealed a mutagenesis frequency of 23% including homozygous and biallelic mutations of the *ADH1* locus. A summary of all primary transformants and their progeny used for targeted mutagenesis in *ADH1* with St1Cas9 is given in Table S2 in the Supporting Information.

Detection of targeted mutagenesis of the *S. aureus* Cas9 nuclease via amplicon deep sequencing

We also tested the SaCas9 constructs for specific induction of targeted mutagenesis. We designed two sgRNAs targeting a different PAM sequence in the *ADH1* locus (Figure 4). For these constructs, each protospacer was cloned into the pEn-Sa_Chimera vector following combination of the sgRNA with the corresponding SaCas9 via a Gateway® reaction in the pDe-Sa_Cas9 destination vector. After *Agrobacterium*-mediated transformation, deep sequencing was performed with 30 primary transformants each, as described above for St1Cas9. Figure 4 depicts the results for both constructs as relative numbers of mutated reads by position. As expected, the highest mutation frequencies were observed 4 bp upstream of the PAM for both constructs. The sgRNA targeting the 'NNGGGT' PAM led to 76.3% mutated reads at that position, while the construct for the 'NNGAA' PAM showed 58.3% mutated reads. Remarkably, the kinds of mutation varied between the two constructs, being mostly insertions for 'NNGAA' (52.1%, predominantly a single nucleotide), while for 'NNGGGT' our data showed a higher number of deletions (46.7% deletions and 21.6% insertions), and both insertions and deletions appeared to be larger than for 'NNGAA'.

Heritable targeted mutagenesis with *S. aureus* Cas9 nuclease

Determination of heritable targeted mutagenesis events due to SaCas9 constructs was also performed by targeting the *ADH1* locus, including different PAM sequences. Similar to the St1Cas9 constructs, the SaCas9 constructs were transformed into *A. thaliana* via *Agrobacterium*-mediated transformation, and the selection of the primary transformants was performed in the T_1 generation. After cultivation, the T_1 progeny (T_2 generation) was tested for Mendelian segregation on selection media as an indication of single-locus T-DNA integration events. Determination of homozygous mutations in the T_2 generation was performed by allyl alcohol selection. Therefore, for the SaCas9 construct targeting the 'NNGGGT' PAM sequence, the progeny of 11 primary transformants was treated with allyl alcohol. A total of approximately 11 000 seeds were analysed. The targeted mutagenesis frequency was 89%, and the mutated plants included both homozygous and biallelic mutations in *ADH1* (Table 2). For the construct targeting

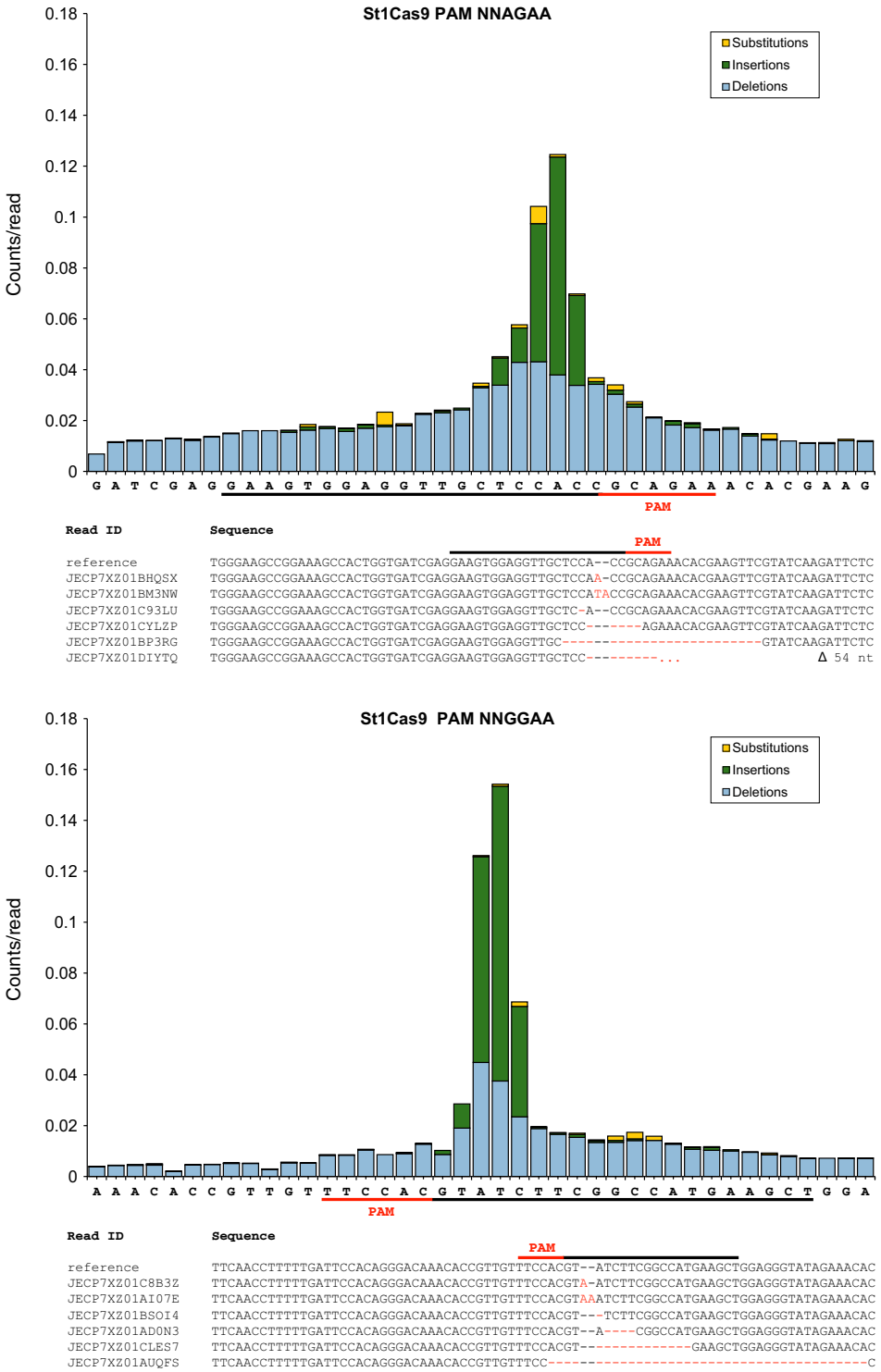

Figure 2. Deep sequencing analysis of *Streptococcus thermophilus* Cas9 constructs.
The figure shows mutations detected by deep sequencing analysis of St1Cas9 constructs for two protospacers with different protospacer-adjacent motif (PAM) sequences. Relative numbers of reads indicate mutation frequencies and types of mutation. For the nuclease targeting the PAM sequence 'NNAGAA', 12.5% of the reads were mutated, and most of these mutations were small insertions and deletions. The mutation frequency for the nuclease targeting the PAM sequence 'NNGGAA' was 15.4%, and these mutations also included insertions and deletions. For the nuclease targeting the 'NNAGAA' PAM, mutation frequencies were greatest 3-bp upstream of the respective PAM sequences. Regarding the second nuclease, frequencies were greatest 4-bp upstream of the PAM 'NNGGAA' sequence. Subsets of mutations generated by both St1Cas9 constructs are also depicted. The reference lane represents the wild-type sequence; most of the sequences detected with deep sequencing revealed insertions. Deletions ranging from 1 bp to tens of bp were also observed for both nucleases.

Table 1 Heritable targeted mutagenesis (TM) in *ADH1* using St1Cas9 nuclease constructs

Construct	Plant line	Plants tested	Viable plants	Dead plants	TM frequency (%)
St1Cas9 PAM NNAGAA	#2	3056	185	2871	6.1
	#4	2202	884	1318	40.2
	#9	3750	1916	1834	51.1
	#13	2635	2048	587	77.7
	Total	11634	5033	6601	43.8
St1Cas9 PAM NNGGAA	#2	2358	246	2112	10.4
	#7	2327	411	1916	17.7
	#14	2797	1137	1660	40.7
	Total	7482	1794	5688	22.9

PAM, protospacer-adjacent motif.

the PAM sequence 'NNGAA', approximately 10 000 seeds derived from progeny of eight independent primary transformants were analysed via allyl alcohol treatment, and resistant seedling frequencies were determined. On average, 34% of the tested plants were either homozygous or biallelic for a mutation in the *ADH1* gene. Disrupted *ADH1* sequences were confirmed by sequencing of a subset of the resistant plants; for constructs targeting the PAM sequence 'NNGGGT' 84 plants were analysed and for constructs targeting the PAM sequence 'NNGAA' 87 plants. An excerpt of homoallelic mutations inherited in the T_2 generation for both tested constructs is illustrated in Figure 5. Table S2 gives an overview of all the lines used in this experiment.

Testing the efficiency and specificity of the Cas9 nucleases of *S. aureus* and *S. pyogenes*

Homologous recombination is induced by DSBs as well as by NHEJ. Therefore, site-specific nucleases are also efficient tools for inducing HR in plant genomes (Puchta *et al.*, 1996; Siebert and Puchta, 2002). Thus, HR can also be used as a means for estimating the efficiencies with which nucleases induce genomic DSBs compared with other enzymes when the same type of recombination substrate is used (Fauser *et al.*, 2014). This system also allowed us to easily and reliably test whether the respective Cas9 protein only acted as a nuclease with its own 'species-specific' sgRNA or whether it might also function with sgRNAs from different species. The simultaneous application of two different Cas9 proteins for complex manipulations of plant cells would only be possible in the former case. The assay system used here is a stably integrated reporter line that enables the visualisation of somatic recombination events in Arabidopsis (Orel *et al.*, 2003). This system is based on β-glucuronidase activity (*uidA*, GUS), which converts the artificial substrate X-GlucA into an easily visualised indigo-coloured substrate. A short spacer sequence disrupts the GUS ORF, leading to a non-functional GUS gene (Puchta *et al.*, 1996; Orel *et al.*, 2003). The induction of a DSB in the spacer sequence results in the restoration of the GUS gene using homologous recombination via a single-strand annealing (SSA) pathway (Figure 6a).

Constructs for the reporter line DGU.US were designed and cloned for the SaCas9 nuclease using two protospacers with different PAM sequences that targeted the interrupting spacer sequence in the GUS ORF. Cross-species constructs were also cloned to test the specificities of the respective nucleases. Via *Agrobacterium*-mediated transformation, the constructs were transformed into the reporter line, and T_1 primary transformants were stained with X-GlucA. Large stained sectors were obtained for the organism-specific constructs but not for the cross-species constructs or the control line without DSB induction, demonstrating that efficient induction of HR was possible for all of the tested nucleases and that each nuclease–sgRNA combination was specific for its respective species (Figure 6b).

To quantify the HR efficiencies, each Cas9–sgRNA combination was measured with a fluorescence assay. Here, the substrate 4-methylumbelliferyl-β-D-glucuronide (4-MUG) was applied; this substrate can be converted to the fluorescent 4-methylumbelliferone (4-MU) via β-glucuronidase activity. For each tested line, 10 primary transformants were treated with 4-MUG after 2 weeks of growth, and each experiment was performed in triplicate as previously reported (Fauser *et al.*, 2014). All of the species-spe-

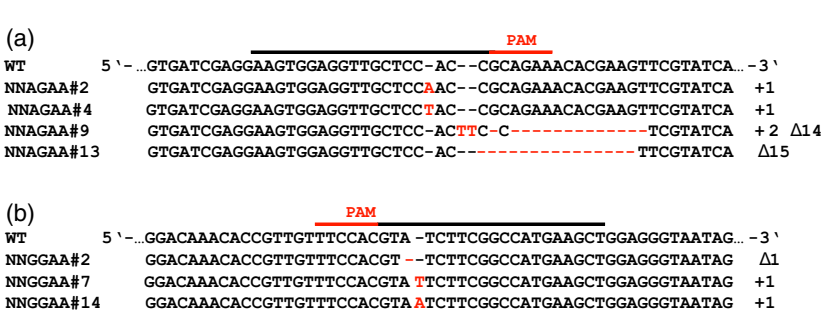

Figure 3. Heritable targeted mutagenesis in *ADH1* via *Streptococcus thermophilus* Cas9 constructs.
Subsets of the Sanger sequencing results for the constructs targeting the protospacer-adjacent motif (PAM) sequence 'NNAGAA' (a) or the PAM sequence 'NNGGAA' (b) are displayed. Only homozygous mutations (insertions, deletions) are depicted. WT, wild type.

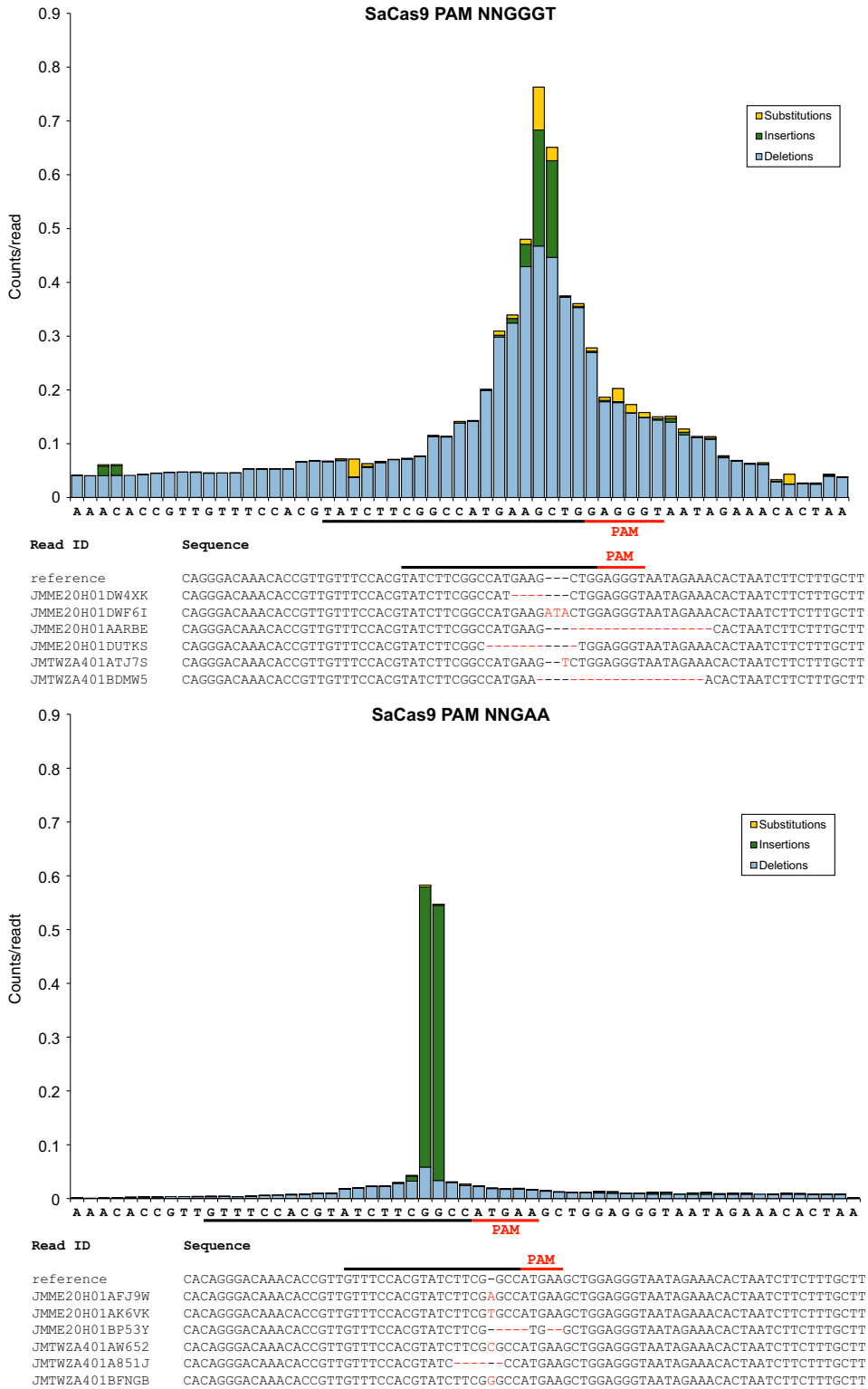

Figure 4. Deep sequencing analysis of *Staphylococcus aureus* Cas9 constructs.
This figure shows the mutation analysis of the SaCas9 protospacer regions. The relative read counts from the whole dataset are depicted according to sequence position and separated by the type of mutation (deletion, insertion or substitution). For both constructs, the mutation frequencies were highest in the protospacer region and peak 4-bp upstream of the respective protospacer-adjacent motif (PAM) sequence. Also depicted are individual kinds of mutations induced by the SaCas9 constructs. The reference sequence shows the wild-type allele. Most of the detected mutations for the construct targeting the 'NNGAA' PAM were 1-bp insertions or small deletions. For the nuclease targeting the 'NNGGGT' sequence, more deletions are observed and both deletions and insertions appear larger.

Table 2 Heritable targeted mutagenesis (TM) in *ADH1* using SaCas9 nuclease constructs

Construct	Plant line	Plants tested	Viable plants	Dead plants	TM frequency (%)
SaCas9 PAM NNGGGT	#1	429	404	25	94.2
	#3	1002	689	313	68.8
	#5	989	971	18	98.2
	#6	648	638	10	98.5
	#7	1405	856	549	61.0
	#8	802	753	49	93.9
	#10	756	710	46	93.9
	#11	2034	1856	178	91.3
	#12	544	516	28	94.9
	#13	1216	1194	22	98.2
	#14	1023	925	98	90.4
	Total	10 848	9512	1336	89.4
SaCas9 PAM NNGAA	#2	1150	70	1080	6.1
	#3	930	204	726	22.0
	#5	959	445	514	46.4
	#6	759	502	257	66.1
	#8	2619	634	1985	24.2
	#10	1845	1597	248	86.6
	#13	939	170	769	18.1
	#14	766	29	737	3.8
	Total	9967	3651	6316	34.2

PAM, protospacer-adjacent motif.

cific constructs exhibited high fluorescence levels compared with the control lines. The SaCas9-specific construct exhibited DSB induction that was at least as efficient as that of the SpCas9 construct. The cross-species constructs exhibited a fluorescence that was similar to that of the uninduced control (Figure 7, Table S3). Thus, under our experimental conditions, not only was the CRISPR/Cas system of *S. aureus* able to efficiently induce HR, but there was also no indication of the occurrence of any cross-species interference that would make the simultaneous application of both Cas9 orthologues in a single plant cell impossible.

DISCUSSION

The CRISPR/Cas system of *S. pyogenes* has become the major tool for genome engineering in eukaryotes. However, it is not possible to use multiple Cas9 proteins that perform different catalytic functions within a single cell utilising only a single type of CRISPR/Cas system. Here, we tested different bacterial type II CRISPR/Cas systems for their applicability to plant genome engineering. The bacterial type II CRISPR/Cas system consists primarily of the Cas9 protein itself and the mature crRNA–tracrRNA complex. This type II system is subdivided into types II-A to II-C, and this subdivision is defined by the harbouring of only a minimal set of cas genes in the CRISPR loci (type II-C) or by the presence of an additional signature csn2 or cas4 gene (type II-A and type II-B, respectively) (Hsu *et al.*, 2014). These subtypes are also divided into different clusters by their distinct locus characteristics, by cas gene composition and by the direction of transcription of the CRISPR array relative to the respective cas operon (Fonfara *et al.*, 2014). All Cas9 proteins tested in the present work belong to the type II-A CRISPR/Cas system but are subdivided into different clusters (Fonfara *et al.*, 2014; Louwen *et al.*, 2014). These proteins differed in their respective sgRNA sequences (for detailed sequences see Ran *et al.*, 2015); hairpins and additional stem loops in the sgRNA sequences seem to be responsible for the correct connection between the sgRNA and the corresponding Cas9 (Briner *et al.*, 2014).

To achieve various catalytic functions in a single cell, different Cas9 orthologues have to be adopted for genome engineering. Recently, Cas9 orthologues from different bacteria were tested for their applicability in mammalian cells (Ma *et al.*, 2015). Furthermore, Ran *et al.* (2015) tested six Cas9 orthologues and demonstrated that Cas9 from *S. aureus* leads to gene editing frequencies that are comparable to those of *S. pyogenes* Cas9 in human and mouse cells.

In this report we have successfully used type II CRISPR/Cas systems from *S. thermophilus* and *S. aureus* in plants to achieve targeted mutagenesis via NHEJ. We obtained similar high mutagenesis frequencies with SpCas9.

St1Cas9 was used to achieve the targeted induction of the DSBs at the *ADH1* locus using two different PAM sequences. These sequences led to nearly identical induction frequencies, indicating that both the 'NNAGAA' and

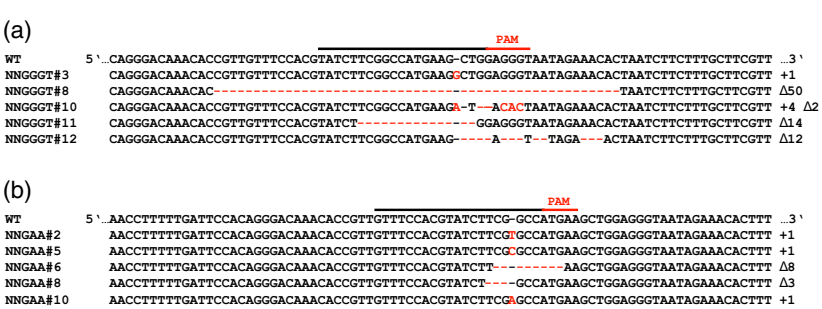

Figure 5. Heritable targeted mutagenesis in *ADH1* via *Staphylococcus aureus* Cas9 constructs.
Sanger sequencing results for constructs targeting either the protospacer-adjacent motif (PAM) sequence 'NNGGGT' (a) or 'NNGAA' (b). A subset of sequencing results is depicted, and only homozygous mutations (insertions and deletions) are shown. WT, wild type.

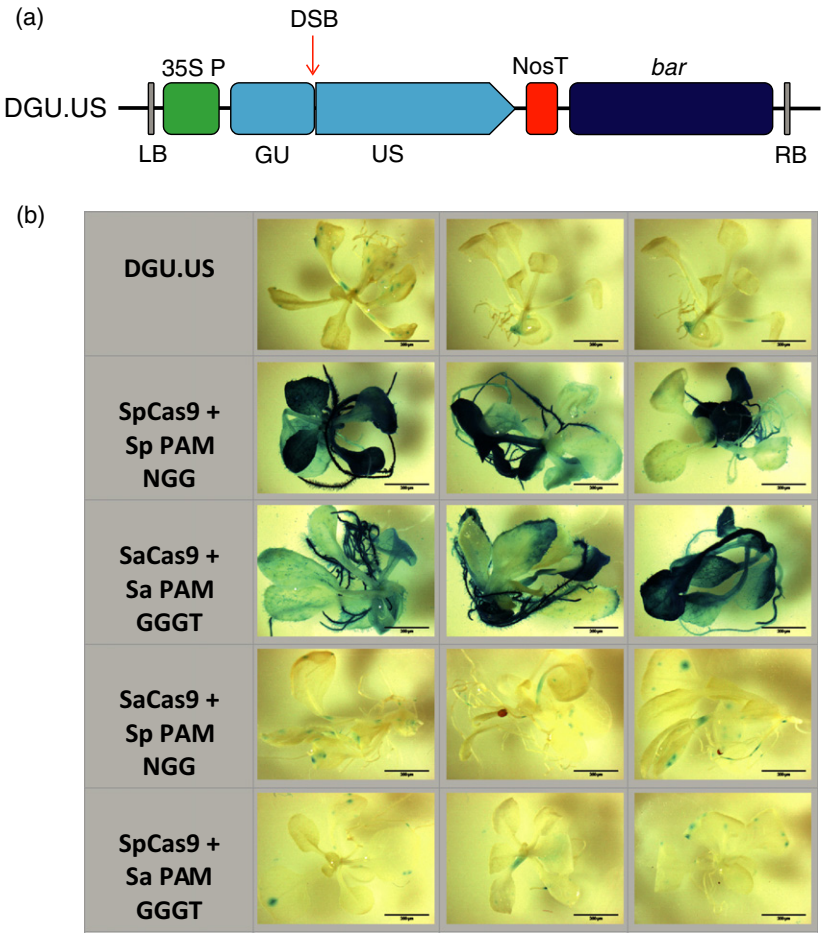

Figure 6. Induction of homologous recombination (HR) in a plant reporter line using *Staphylococcus aureus* constructs.
(a) The HR reporter construct DGU.US. The CRISPR protospacer design depends on the respective Cas9 species used (SaCas9 or SpCas9). Following induction of a double-strand break (DSB) in the spacer region, the break is repaired via HR by the single-strand annealing pathway, which leads to a restored GUS open reading frame. The GUS activity was measured on the basis of histochemical staining.
(b) Histochemical staining of T_1 plants. A representative selection of plants is shown and indicates the induction of HR in organism-specific constructs leading to the staining of large blue sectors, whereas cross-species constructs exhibited staining levels similar to those of the control lines (DGU.US).

'NNGGAA' PAMs are equally applicable for genome engineering. Interestingly, the repair pattern observed here is consistent with the results from Ran *et al.* (2015). These authors demonstrated that in human embryonic kidney cells, St1Cas9 cuts either the second or third nucleotide upstream of the PAM, whereas all other tested orthologues only cleaved the third nucleotide upstream of the PAM.

Targeted mutagenesis events at the *ADH1* locus were also generated using SaCas9 constructs. Induction frequencies were even higher than those achieved with St1Cas9 and SpCas9 (Fauser *et al.*, 2014) and reached over 90% for specific lines.

Our experiments also revealed the PAM specificities of SaCas9. The constructs targeting the PAM sequence 'NNGGGT' led to greater mutation yields in the target sequence than the constructs that targeted the PAM 'NNGAA' sequence. For the second PAM 'NNGAA', the observed repair patterns were similar to those we have previously obtained with SpCas9 (Fauser *et al.*, 2014), which indicated an identical cleavage mechanism at the third nucleotide upstream of the PAM. However, the construct targeting the PAM sequence 'NNGGGT', led to mutation yields of almost 80%, and most of these were larger deletions. Single base pair insertions were not found as often as for the construct targeting the 'NNGAA' sequence or the previously reported SpCas9 constructs. These deep sequencing results achieved here are consistent with the Sanger sequencing results after allyl alcohol treatment. The lines targeting the PAM 'NNGGGT' show mostly deletions at the sequence level, whereas the lines for the 'NNGAA' PAM show similar patterns to the St1Cas9 constructs or the previously described SpCas9 constructs. Although PAM specificities were not thoroughly analysed here, our results indicate that DSB induction efficiencies and mutagenesis outcomes might differ depending on the selected PAM. We cannot exclude that the genomic context and not the PAM sequences themselves influence the repair pattern. However, we regard this explanation as unlikely, as all our constructs target sites within exons of the *ADH1* locus. For gene editing with SaCas9 constructs, we recommend the 'NNGGGT' PAM, as it led to the highest mutagenesis frequencies in our experiments. Further detailed investigations will have to be performed in plant cells to reveal preferences for all different possible PAMs which follow the originally described NNGRR(T) structure (Ran *et al.*, 2015).

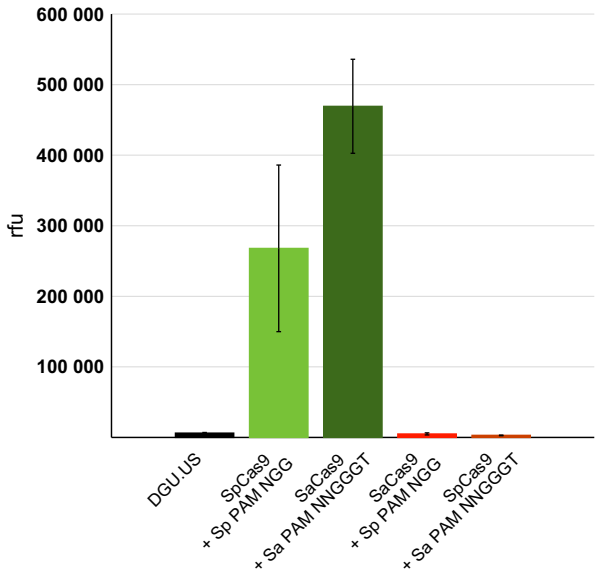

Figure 7. Quantification of homologous recombination (HR) events using a fluorescent assay.
Activity of the GUS gene due to HR-induced restoration events was measured via fluorescence in the reporter line DGU.US. High fluorescence levels were achieved in lines expressing organism-specific constructs. In contrast, lines with cross-species constructs exhibited low fluorescence similar to control lines (DGU.US). Further information is given in Table S3.

We also found out that SaCas9 was able to induce HR with an efficiency that was similar to that of SpCas9 in the reporter line used. Testing the PAM sequence 'NNGGGT' resulted in HR frequencies that were similar to those of SpCas9. This finding was not surprising, because the same PAM was also very efficient in inducing heritable mutations in the *ADH* gene by NHEJ. For both HR and NHEJ, the induction of nuclease-mediated DSBs is the initial rate-limiting step in the repair reaction (Puchta and Fauser, 2014). However, the most important aspect of the HR analysis was that it enabled us to determine whether the Cas9–sgRNA interaction was 'species specific' or whether some type of cross-species interference could activate the Cas9 protein via heterologous sgRNAs in plant cells. Because the restoration of the β-glucuronidase marker in the HR assay is a sensitive measure of DSB induction, we were able to demonstrate that the correct pairing of sgRNA and corresponding Cas9 orthologue is necessary for target recognition.

Unfortunately, we were not able to test HR induction elicited by St1Cas9 because our HR reporter line did not cover a suitable PAM sequence. Nevertheless, there are strong indirect indications that *S. thermophilus* CRISPR/Cas systems can be used without interference from the Cas9 proteins of *S. aureus* and *S. pyogenes* in plant cells. In bacteria, Fonfara *et al.* (2014) demonstrated that only Cas9 proteins from closely related CRISPR/Cas systems can substitute for the function of *S. pyogenes* Cas9 in deletion mutants lacking SpCas9. Because the *S. thermophilus* Cas9 protein belongs to another subcluster of the type II-A CRISPR/Cas system no processing was observed, which indicated that there is also likely no cross-activation in *S. thermophilus* CRISPR/Cas systems involving Cas9 from either *S. pyogenes* or *S. aureus* in plants.

The orthologous CRISPR/Cas systems described here were able to mediate efficient editing in *A. thaliana* plants and could also be applied to induce targeted DSBs at any desired locus. Applications for which these orthologues will be useful include target genes in which SpCas9 PAM sequences are not present and approaches in which vector sizes are limited. Both orthologues described here are 1 kb smaller than the commonly used SpCas9 (3.4 kb for St1Cas9 and 3.2 kb for SaCas9 compared with 4.2 kb for SpCas9). The use of these orthologues in complex genetic approaches, such as simultaneous applications of different CRISPR/Cas-associated or -fused enzyme activities, will be of special importance in the future.

We demonstrated that the SpCas9 nuclease can undergo a single-amino-acid substitution to become a nickase that is highly active in plant cells (Fauser *et al.*, 2014). Thus, it is not possible to induce SSBs and DSBs at the same time at different sequence motifs within a single plant cell. Previously, we developed a specific type of *in planta* gene targeting technique that relies on the induction of a DSB at the target locus and on the simultaneous excision of the target vector from the plant genome (Fauser *et al.*, 2012; Schiml *et al.*, 2014). The advantage of this technique is that targeting can occur during the entire life cycle of a plant and that targeted events can be collected as seeds in the next generation. Thus, labour-intensive, crop-plant-specific transformation procedures can be minimized. It is tempting to speculate that the excision of the vector might not only be achievable by a nuclease but also by a nickase. Based on the use of the SaCas9 orthologue acting as a nuclease that induces DSBs and the SpCas9 orthologue serving as a nickase that induces SSBs, it is now feasible to test this approach. Additionally, Ma *et al.* (2015) demonstrated that it is possible to label chromosomal loci in human cells using dCas9 constructs (deadCas9, a catalytically inactive nuclease) of SpCas9 and St1Cas9 constructs and to simultaneously label various pairs of chromosomal loci with different colours. Similar approaches might now also be tested for plants using different orthologous dCas9 constructs.

With the plant-adopted Cas9 orthologues developed in this report, the toolbox for genome and transcriptome engineering in plants has been dramatically increased, and many new approaches for manipulating cells by targeting different types of enzyme activities that simultaneously target specific genomic sites are now possible. It has also been demonstrated in plants that dSpCas9 can be used to

activate or repress transcription (Piatek *et al.*, 2015). Using different Cas9 orthologues, it should be possible to simultaneously switch some classes of genes on and others off; thus, the complex reprogramming of the transcriptomes of a plant cell is now within reach.

An interesting question to consider is the extent of the off-target effects of the St1 and SaCas9 orthologues. Both orthologues require longer PAMs than the SpCas9; six nucleotides ('NNAGAA' or 'NNGGAA') are required for St1Cas9, and five to six nucleotides ('NNGGGT' or 'NNGAA') are required for SaCas9, whereas only three nucleotides ('NGG') are required for SpCas9 target recognition, which should significantly reduce the frequency of off-target cutting compared with SpCas9. The downside to this higher specificity is the resulting decrease in potential target sites. To exclude off-target effects beforehand, one could also apply the paired nickase approach to the orthologous Cas9 proteins, as we were able to do in plants using the SpCas9 CRISPR/Cas system (Schiml *et al.*, 2014).

EXPERIMENTAL PROCEDURES

Strains

All lines used in this study were in the Columbia-0 background. Seeds were sown on agar plates containing germination medium or on substrate containing 1:1 Floraton 3 (Floragard, https://www.floragard.de/) and vermiculite (Deutsche Vermiculite Dämmstoff GmbH, http://www.vermiculite.de/).

T-DNA constructs

The vectors used in this study were based on our previously described CRISPR/Cas system (Fauser *et al.*, 2014; Schiml *et al.*, 2014). These Gateway® compatible vectors contain either a destination cassette (pDe-St1_Cas9, pDe-Sa_Cas9) or an entry cassette (pEn-St1_Chimera, pEn-Sa_Chimera). Therefore, orthologous Cas9 ORFs from *S. thermophilus* and *S. aureus* were codon-optimised for *A. thaliana,* synthesized by GeneArt® (Life Technologies Corporation, www.lifetechnologies.com) and flanked by *Asc*I recognition sites. These ORFs were transferred into the pDe-Cas9-D10A vector by exchanging the *S. pyogenes* Cas9 ORF with St1Cas9 or St1Cas9 using *Asc*I to create pDe-St1_Cas9 and pDe-Sa_Cas9, respectively. The destination vectors harbour a kanamycin resistance cassette (*npt* II) as previously described for pDe-Cas9-D10A (Fauser *et al.*, 2014).

The entry vectors were also assembled by GeneArt® and harboured species-specific sgRNAs. Spacers can be introduced via *Bbs*I as previously described (Fauser *et al.*, 2014).

For analysis of β-glucuronidase activity with SaCas9 constructs, the oligonucleotides JS 233 and JS 234 were used for oligo-annealing and subsequent cloning into pEn-Sa_Chimera (all oligonucleotides used in this study are listed in Table S1). Cross-species constructs for β-glucuronidase activity assays were assembled by pairing destination vectors of one species with entry vectors of another species, for example pDe-Sa_Cas9 with pEn_Chimera or pDe_Cas9 with pEn-Sa_Chimera vectors.

For *ADH1* mutagenesis, oligonucleotides JS 187, JS 188 and JS 191, JS 192 were used for St1Cas9-specific constructs, and oligonucleotides JS 239 to JS 242 were used for SaCas9-specific constructs.

Subsequently the chimeras were transferred into the corresponding destination vectors pDe-St1_Cas9 or pDe-Sa_Cas9 via single-site Gateway® LR reactions (Thermo Fisher Scientific Inc., http://www.thermofisher.com/).

Sequence information is provided in Figures S1–S3. All vectors described here are available upon request.

Plant transformation

Arabidopsis plants were transformed via *Agrobacterium*-mediated transformation as previously described (Clough and Bent, 1998).

Analysis of β-glucuronidase activity

The GUS staining was performed as previously described (Orel *et al.*, 2003) 14 days after sowing the seeds on agar plates containing germination medium (GM) with a corresponding selection marker and cefotaxime.

For detailed quantification of Cas9 activity, a 4-MUG assay was performed as previously described (Fauser *et al.*, 2014).

Amplicon deep sequencing

T_1 plants were grown on GM containing kanamycin and cefotaxime. After 2 weeks, 30 plantlets from each group were pooled, and DNA was extracted as previously described (Salomon and Puchta, 1998).

The PCR products were amplified with MID-containing primers (JS 221, JS 222, JS 225, and JS 226 for St1Cas9 constructs and JS 226 and JS 269 to JS 275 for SaCas9 constructs) using a proofreading polymerase with 100 ng of genomic DNA. Purification of amplicons was performed using the Peqlab Cycle Pure Kit (PEQLAB Biotechnologie GmbH, http://www.peqlab.de/). Subsequent Roche 454 Sequencing was performed by Eurofins Genomics (Eurofins Genomics GmbH, http://www.eurofinsgenomics.eu/). Data analysis was performed using the Galaxy web server (Giardine *et al.*, 2005; Blankenberg *et al.*, 2010; Goecks *et al.*, 2010) and the Integrative Genomics Viewer 2.3 (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013).

Evaluation of germinal mutations

T_1 seeds were sown on GM agar plates containing kanamycin and cefotaxime. After 2 weeks of incubation, primary transformants were selected and transferred to soil for further cultivation in the greenhouse.

T_2 plants were tested for 3:1 segregation on respective selection media to detect single-locus lines.

To determine whether the *ADH1* locus was altered, T_2 seeds were treated with allyl alcohol. After sterilising the seeds, they were treated with 25 mM allyl alcohol for 2 h at 21°C and were shaken at 900 r.p.m. and were subsequent washed three times with double-distilled water. After this treatment, the seeds were sown on GM and incubated for approximately 2 weeks. The growing plants were counted and genotyped by amplifying the *ADH1* locus using primers JS 257 and JS 258 and JS 261 and JS 262 for St1-Cas9 constructs and JS 292 and JS 293 for Sa-Cas9 constructs. Subsequent Sanger sequencing of these amplicons was performed using Primers JS 257 and JS 261 for St1Cas9 constructs and JS 292 for SaCas9 constructs.

ACKNOWLEDGEMENTS

We would like to thank Jana Bender, Carina Jülch and Waltraud Wehrle for excellent technical assistance in the laboratory and the botanical garden. We also thank Julian Röhrig and Zhengzhong

Yu for helping us with the plate reader. This work was funded by the European Research Council (advanced grant 'COMREC').

SUPPORTING INFORMATION

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

Figure S1. Sequence information for pDe-St1_Cas9.

Figure S2. Sequence information for pDe-Sa1_Cas9.

Figure S3. Sequence information for pEn-St1_Chimera and pEn-Sa_Chimera.

Table S1. Primers used in this study.

Table S2. Summary of primary transformants and their progeny used for targeted mutagenesis in *ADH1* with St1Cas9 and SaCas9.

Table S3. Raw data for the 4-methylumbelliferyl-β-D-glucuronide assay.

REFERENCES

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

Blankenberg, D., vonKuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. and Taylor, J. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19, Unit 19.10.1–21.

Bortesi, L. and Fischer, R. (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.*, **33**, 41–52.

Briner, A.E., Donohoue, P.D., Gomaa, A.A., Selle, K., Slorach, E.M., Nye, C.H., Haurwitz, R.E., Beisel, C.L., May, A.P. and Barrangou, R. (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell*, **56**, 333–339.

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

Cong, L., Ran, F.A., Cox, D. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, **339**, 819–823.

Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, **346**, 1258096.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J. and Church, G.M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods*, **10**, 1116–1121.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C. and Puchta, H. (2012) In planta gene targeting. *Proc. Natl Acad. Sci. USA*, **109**, 7535–7540.

Fauser, F., Schiml, S. and 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.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.-L., Bzdrenga, J., Koonin, E.V. and Charpentier, E. (2014) Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.*, **42**, 2577–2590.

Giardine, B., Riemer, C., Hardison, R.C. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.

Goecks, J., Nekrutenko, A. and Taylor, J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.

Jacobs, M., Dolferus, R. and Van den Bossche, D. (1988) Isolation and biochemical analysis of ethyl methanesulfonate-induced alcohol dehydrogenase null mutants of arabidopsis thaliana (L.) Heynh. *Biochem. Genet.*, **26**, 105–122.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, **337**, 816–821.

Kawalleck, P., Somssich, I.E., Feldbrügge, M., Hahlbrock, K. and Weisshaar, B. (1993) Polyubiquitin gene expression and structural properties of the ubi4-2 gene in Petroselinum crispum. *Plant Mol. Biol.*, **21**, 673–684.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q. *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523**, 481–485.

Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. and Sheen, J. (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotechnol.*, **31**, 688–691.

Li, J.-F., Zhang, D. and Sheen, J. (2014) Cas9-based genome editing in Arabidopsis and tobacco. *Methods Enzymol.*, **546**, 459–472.

Li, J.-F., Zhang, D. and Sheen, J. (2015) Targeted Plant Genome Editing via the CRISPR/Cas9 Technology. *Methods Mol. Biol.*, **1284**, 239–255.

Louwen, R., Staals, R.H.J., Endtz, H.P., van Baarlen, P. and van der Oost, J. (2014) The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.*, **78**, 74–88.

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S. and Pederson, T. (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl Acad. Sci. USA*, **112**, 3002–3007.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones Jonathan, D.G. and Kamoun, S. (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, **31**, 691–693.

Orel, N., Kyryk, A. and Puchta, H. (2003) Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *Plant J.*, **35**, 604–612.

Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M. and Mahfouz, M.M. (2015) RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnol. J.*, **13**, 578–589.

Puchta, H. and Fauser, F. (2014) Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J.*, **78**, 727–741.

Puchta, H., Dujon, B. and 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. USA*, **93**, 5055–5060.

Ran, F.A., Cong, Le, Yan, W.X. *et al.* (2015) *In vivo* genome editing using Staphylococcus aureus Cas9. *Nature*, **520**, 186–191.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

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

Schiml, S., Fauser, F. and 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.

Shan, Q., Wang, Y., Li, J. *et al.* (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.*, **31**, 686–688.

Siebert, R. and Puchta, H. (2002) Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell*, **14**, 1121–1131.

Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*, **14**, 178–192.

Voytas, D.F. (2013) Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.*, **64**, 327–350.

Weeks, D.P., Spalding, M.H. and Yang, B. (2015) Use of designer nucleases for targeted gene and genome editing in plants. *Plant Biotechnol. J.* DOI: 10.1111/pbi.12448 [Epub ahead of print].

Xie, K., Minkenberg, B. and Yang, Y. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl Acad. Sci. USA*, **112**, 3570–3575.

