Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering

Oleg Raitskin[1], Christian Schudoma[1], Anthony West[1,2] and Nicola J. Patron[1†]

[1] Earlham Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UZ, UK

[2] Current address: Isogenica Ltd. The Mansion, Chesterford Research Park, Little Chesterford, Saffron Walden, CB10 1XL, UK

[†]Corresponding author: nicola.patron@earlham.ac.uk

Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering

Abstract

Molecular tools adapted from bacterial CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats) systems for adaptive immunity have become widely used for plant genome engineering, both to investigate gene functions and to engineer desirable traits. A number of different Cas (CRISPR-associated) nucleases are now used but, as most studies performed to date have engineered different targets using a variety of plant species and molecular tools, it has been difficult to draw conclusions about the comparative performance of different nucleases. Due to the time and effort required to regenerate engineered plants, efficiency is critical. In addition, there have been several reports of mutations at sequences with less than perfect identity to the target. While in some plant species it is possible to remove these so-called 'off-targets' by backcrossing to a parental line, the specificity of genome engineering tools is important when targeting specific members of closely-related gene families, especially when recent paralogues are co-located in the genome and unlikely to segregate. Specificity is also important for species that take years to reach sexual maturity or that are clonally propagated. Here, we directly compare the efficiency and specificity of Cas nucleases from different bacterial species together with engineered variants of Cas9. We find that the nucleotide content correlates with efficiency and that Cas9 from *Staphylococcus aureus* is comparatively most efficient at inducing mutations. We also demonstrate that 'high-fidelity' variants of Cas9 can reduce off-target mutations in plants. We present these molecular tools as standardised DNA parts to facilitate their re-use.

Keywords: genome engineering, genome editing, gene editing, targeted mutagenesis, CRISPR/Cas

Introduction

Components of bacterial CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats) systems for adaptive immunity have been repurposed for engineering the genomes of eukaryotic organisms (Cong *et al.*, 2013; Mali *et al.*, 2013; Zetsche *et al.*, 2015). These molecular tools have been rapidly and successfully applied in many organisms, including plants (for recent reviews see Gao, 2018; Ricroch *et al.*, 2017; Yin *et al.*, 2017), primarily due to the ease at which they can be programmed to recognise new genomic targets. The majority of plant genome engineering studies have utilised Cas9 (CRISPR Associated Protein 9) from *Streptococcus pyogenes*, a monomeric nuclease found in the Type II CRISPR system of that species. The Cas9 protein can be directed to selected genomic targets by an engineered RNA moiety known as the single guide RNA (sgRNA) (Cong *et al.*, 2013; Mali *et al.*, 2013). One or more sgRNAs can be programmed to recognise new genetic targets by recoding the ~19 base pairs at the 5' end of the molecule, known as the spacer (Barrangou *et al.*, 2007). The sgRNA forms a ribonuclease complex with Cas9 enabling it to scan DNA, pausing when it encounters a cognate sequence known as the protospacer adjacent motif (PAM) (Sternberg *et al.*, 2014). On recognition of a cognate PAM, the ribonuclease complex probes the adjacent sequence to determine identity to the spacer region. If complementary, the spacer region of the sgRNA forms a Watson-Crick base-pair with the target DNA forcing the Cas9 protein to undergo a conformational change that enables the nuclease domains to cleave each of the DNA strands (Sternberg *et al.*, 2014). The most common application in plants has been targeted mutagenesis achieved following transgenic expression of sgRNA-guided Cas9 to introduce double strands breaks (DSBs) at selected genomic loci. These induced breaks are predominantly repaired by the endogenous mechanism of non-homologous end-joining (NHEJ), which sometimes introduces errors (Hu *et al.*, 2016; Schiml and Puchta, 2016). Since a perfect repair will continue to be recognised by the Cas9/sgRNA complex, and therefore cut again, constitutive expression eventually results in mutations, typically small insertions or deletions at the target. In many transgenic events, these mutations occur sufficiently early in the development and regeneration of the transgenic plant that all cells of the plant contain the same mutant genotype. A smaller number of studies have successfully leveraged the induction of targeted DSBs to increase the efficiency of targeted integration via the mechanism of homology directed repair (HDR), enabling transgenes to be inserted at a precise locus or for genomic DNA sequences to specifically recoded (for example, Baltes *et al.*, 2014; Begemann *et al.*, 2017; Cermak *et al.*, 2017; Čermák *et al.*, 2015; Gil-Humanes *et al.*, 2017; Li *et al.*, 2013).

Additional Cas proteins from other bacterial species have been adapted for genome engineering in eukaryotes and have also been applied to plants. These include Cas9 from the *Staphylococcus aureus* Type II CRISPR system (Kaya *et al.*, 2016; Ran *et al.*, 2015; Steinert *et al.*, 2015) and Cas12a (previously Cpf1) from the Type V CRISPR systems found in *Francisella novicida, Acidaminococcus sp. and Lachnospiraceae bacterium* (Begemann *et al.*, 2017; Tang *et al.*, 2017; Xu *et al.*, 2017; Zetsche *et al.*, 2015). Cas proteins from different species generally show preferences for different PAMs and, therefore, these additional Cas proteins have increased the number of genomic sites that can be targeted for engineering. While SpCas9 most efficiently cleaves DNA adjacent to NGG PAMs, SaCas9 is reported to show preference for NNGRRT (Friedland *et al.*, 2015; Kleinstiver, *et al.*, 2015a; Xie *et al.*, 2018). Cas12a-RNA complexes efficiently cleave target DNA preceded by a short T-rich PAM, with data suggesting a preference for TTTV (Yamano *et al.*, 2017; Zetsche *et al.*, 2015). In addition, engineered versions of Cas9 and Cas12a with mutations in their PAM recognition domains have further expanded the repertoire of target sites to include sequences adjacent to NGAG and NGCG PAMS (with variants of SpCas *et al.*, 2015b) and NNNRRT PAMs (with variants of SaCas9 (Kleinstiver, *et al.*, 2015a)).

Cas-mediated genome engineering in plants has often utilised established transformation methods (e.g. *Agrobacterium*-mediated) to deliver DNA molecules encoding expression cassettes for the Cas9 protein, one or more guide RNAs and, typically, a plant selectable marker cassette, to the plant cell aiming for integration of all components at a single genetic locus. To achieve this, several systems have been developed for the assembly of multigene constructs, many of which employ Type IIS restriction cloning, also known as Golden Gate assembly (Cermak *et al.*, 2017; Engler *et al.*, 2008; Ma *et al.*, 2015; Vazquez-Vilar *et al.*, 2016; Werner *et al.*, 2012; Xing *et al.*, 2014). Following the recovery of transgenic plants, most studies report the 'efficiency' of targeted mutagenesis as the percentage of transgenic plants in which mutations are found at the intended target. Most published studies have utilised different genetic targets, construct designs and regulatory elements making it difficult to draw conclusions about the comparative efficiencies and specificities of each nuclease. A few recent studies have attempted to compare Cas9 and Cas12a nucleases in similar experimental conditions (e.g. Lee *et al.*, 2018). However, they have necessarily used different targets with each type of nuclease, which are known to exert significant influence on efficiency (Moreno-Mateos *et al.*, 2017).

Another area of interest has been the specificity of Cas nucleases. With careful bioinformatic analysis it is often possible to identify target sites in the genomic region of interest that are unique within the genome, thereby reducing the likelihood of mutations being induced at

alternative loci. However, a number of different studies, including several in plants, have reported so-called "off-target" mutations at sites in the genome with less than perfect identity to the spacer (Lawrenson *et al.*, 2015; Li *et al.*, 2015; Zhang *et al.*, 2018). In some cases, small numbers of off-target mutations are of no concern, especially as mutations are known to occur during the process of tissue culture and regeneration (Phillips *et al.*, 1994; Zhang *et al.*, 2014). Further, many species can be back-crossed to the parent to remove unwanted mutations. However, when the aim is to induce mutations in specific members of a gene family, specificity is desirable, especially for recent paralogues that may be in close proximity and therefore impossible to segregate. In addition, the introduction of additional unwanted mutations in lineages that are not typically sexually propagated, such as cultivars of *Solanum tuberosum* (potato), is particularly undesirable.

To increase the specificity of Cas9 proteins for their target, researchers have engineered the nuclease and PAM-recognition domains (Hu *et al.*, 2018; Kleinstiver *et al.*, 2016; Slaymaker *et al.*, 2015). The resulting proteins were initially tested in mammalian cell cultures and reported to maintain similar levels of efficiency to the wild-type protein at sequences with an exact match to the spacer, but limited activity at sequences with less than perfect identity to the spacer. However, the efficiency and specificity of these 'high fidelity' variants engineered by different laboratories have not been compared and their use has not yet been reported in plants. Here we report the outcomes of experiments that compare the efficiency and specificity of multiple wild type and engineered variants of Cas nucleases. All constructs used exactly the same regulatory elements and are assembled into identical vectors for delivery. Importantly, the efficiency and specificity of each nuclease are initially compared at the same target and, by quantifying the frequency of mutagenesis induced during transient expression, we avoid the influence of transgene insertion location on transgene expression. We also present an analysis of all potential targets as well as their off-targets in the coding exons of Arabidopsis and test our tools at a larger number of targets to identify factors that influence efficiency.

Methods

Analysis of targets in Arabidopsis coding sequences

Coding sequences were extracted from the Arabidopsis thaliana TAIR10 annotated whole chromosome datasets (ftp://ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes/ and https://www.arabidopsis.org/download_files/Genes/TAIR10_genome_release/TAIR10_gff3/TAIR10_GFF3_genes.gff). Jellyfish (v.2.2.6) (Marçais and Kingsford, 2011) was combined

with a script to extract all possible legitimate candidate target sequences for each Cas nuclease [N_{20}NGG (SpCas9), N_{21}-NNGGGT (SaCas9), N_{20}-NGAG (SpCas9-VQR), N_{20}-NGCG (SpCas9-VRER), TTTV-N_{23} (Lb/AsCas12a)]. Potential off-targets (target sequences that are either identical to another target or differ by, at most, one base pair in the spacer region) were detected by mapping all identified targets against the TAIR10 genomic sequences using bbmap (v.38.06) with the following parameters (k=8, minid=0.75, ambig=all, mappedonly=t, secondary=t, sssr=0.75, ssao=t, saa=f, mdtag=t, nhtag=t, xmtag=t, amtag=t, nmtag=t, xstag=t, indelfilter=0, subfilter=4, editfilter=4). Alignments were filtered by Hamming distance allowing at most one mismatch in the spacer (plus mismatches at the ambiguous positions of the respective PAM sequences). Any potential off-targets located in coding sequences were identified with Bedtools intersect (v. 2.26.0; Quinlan and Hall, 2010) requiring an overlap equal to the length of the target including the PAM. All identified targets are provided in Supplementary Data 1. All scripts and a snakemake pipeline (Köster and Rahmann, 2012) containing the whole workflow are available at https://github.com/cschu/CRISPRanto.

Selection of targets and off-targets for assessment of targeted mutagenesis

To remove any potential variability that might be associated with gene expression, we selected targets/off-target pairs located in genes expressed in leaves. To do this, we compared our candidate target/off-target pairs to gene expression data from the Expression Atlas (https://www.ebi.ac.uk/gxa/home, experiment E-GEOD-38612), retaining those targets within genes expressed in leaves. We then selected candidate target/off-target pairs that differed within the first 3 bp distal to the PAM. Finally, we either selected target/off-target pairs in which the expected cleavage site overlapped with the recognition site for a Type II restriction endonuclease, or searched 1000 bp of sequence each side of the target/off-target pairs to identify the presence of an additional target, identical in both sequences) that would enable the creation of a deletion when a second sgRNA recognising this target was delivered.

Assembly of constructs

Constructs were assembled using the Plant Modular Cloning (MoClo) plasmid toolkit (Engler *et al.*, 2014), a gift from Sylvestre Marrillonet (Addgene Kit # 1000000044). New Level 0 parts were made according to the standards described in (Patron *et al.*, 2015). All Cas9 proteins were fused at the C-terminus to yellow fluorescent protein (YFP) and nuclear localisation (NLS) KKRKVKKRKVKKRV) tags. Cas12a proteins were fused to the same C-terminal NLS tag, followed by a 3xHA tag. Level 0 parts were assembled into transcriptional units in Level 1 acceptor plasmids in a one-step digestion-ligation reaction, except for

sgRNAs, in which a U6 promoter L0 part was assembled into Level 1 acceptor plasmid together with a PCR amplicon of the complete sgRNA (spacer and scaffold), as described in (Lawrenson *et al.*, 2015). Subsequently, Level 1 transcriptional units were assembled into the Level 2 acceptor plasmid pICSL4723 (Addgene #86172). The digestion-ligation reactions were set up either manually or at nanoscale using laboratory automation: For manual assembly, 15 fmol of each DNA part was combined with 7.5 fmol of the acceptor plasmid in a final volume of 5 µL dH_20. This was combined with 5 µL of reaction mix (3 µL of dH_20, 1 µL of T4 DNA ligase buffer 10x (NEB, Ipswich, MA, USA), 0.5 µL of 1 mg/mL purified bovine serum albumin (1:20 dilution in dH_20 of BSA, Molecular Biology Grade 20 mg/mL, NEB), 0.25 µL of T4 DNA ligase at 400 U/µL (NEB) and 0.25 µL of *BsaI* or *BpiI* restriction enzyme at 10 U/µL (ThermoFisher, Waltham, MA, USA)) and incubated in a thermocycler for 26 cycles of 37°C for three minutes followed by 16°C for four minutes and a final incubation at 37°C followed for 5 minutes and 80°C for five minutes. A 2 µL aliquot of each reaction was transformed into 20 µL electrocompetent cells and plated on selective LB-agar plates. Automated reactions were scaled down to a final reaction volume of 1 µ using the Echo 550 liquid handler (Labcyte Ltd. San Jose, SA, USA), transformed into 2 µL XL10-Gold® Ultracompetent Cells (Agilent Technologies, Santa Clara, CA, USA) and plated onto eight-well selective LB-agar plates on a Hamilton® STARplus platform. The sequences of assembled plasmids were verified by complete sequencing using 150 base pair paired-end reads on an Illumina MiSeq platform. Libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA, USA) with a modified 2 µL total volume protocol using a one in 25 dilution of components. A complete list of all 132 plasmids, comprising Level 0 DNA parts, Level 1 transcriptional units and Level 2 constructs, used in this study is given in Supplementary Data 2. Samples, together with complete annotated sequence files for the 128 new plasmids generated for this study have been deposited at the AddGene plasmid repository.

Transient expression in protoplasts

For each experiment, a sufficient number of protoplasts were prepared to enable the delivery of four replicates of all constructs to be compared. Protoplasts were prepared from the leaf tissues of *Nicotiana benthamiana* or *A. thaliana* as previously described (Yoo *et al.*, 2007). Protoplasts were quantified and divided into aliquots of 200 µL in transfection buffer (0.4M mannitol, 15mM $MgCl_2$, 4 mM MES, pH 5.6), each containing approximately 1 x 10^4/ml intact protoplasts, such that four separate aliquots of protoplasts from the same preparation were available for each of the plasmids to be compared. Plasmid DNA for delivery to protoplasts was prepared using the Plasmid Plus Midi kit (Qiagen, Hilden, Germany) with a modified protocol incorporating three additional wash steps prior to elution from the column. Freshly

made PEG (2 g of PEG (Mn 4000 (Sigma, 81240)) in 2 mL of 500 mM mannitol and 0.5 mL of 1M $CaCl_2$) was mixed with10 µg of purified DNA and added to each aliquot of protoplasts. Subsequently, protoplasts were washed and resuspended in 300 µL of washing buffer (154 mM NaCl, 125 mM, $CaCl_2$, 5 mM KCl, 2 mM MES; pH5.6) and incubated for 24 hours at 24ºC in an illuminated incubator with light intensity of approximately 70 µmol/m^2/s. Transformation efficiency was estimated by quantification of protoplasts in which YFP fluorescence was visible in the nuclei using an inverted fluorescence microscope (Zeiss Axio Observer Z1 or ThermoFisher Evos).

Detection and quantitation of targeted mutagenesis

DNA was extracted from the protoplasts using a cetyltrimethylammonium bromide (CTAB) extraction protocol: pellets of protoplasts were resuspended in 100 µL of extraction buffer (0.2M Tris-HCl, pH7.5; 0.05M EDTA; 2M NaCl; 2% CTAB, pH7.4) and incubated at 65^0C for 1 hour prior to addition of 45 µL chloroform. Following centrifugation, the upper aqueous phase was precipitated with an equal volume of isopropanol. DNA pellets were washed with 70% w/v ethanol, dried and resuspended in sterile distilled water with 5 µg/µL RNAse A. Each target was amplified using a pair of primers, specific to the locus of interest (Supplementary Data 3). PCR reactions were performed using 70 ng DNA and Q5 High-Fidelity DNA Polymerase (NEB) according to the instructions provided by the manufacturer. Mutations at the targets were identified by either Illumina or Sanger sequencing. For preparation for Illumina sequencing, amplicons were purified using Agencourt AMPure XP (Beckman Coulter) and indexed using the Nextera XT Library Preparation kit (Illumina) according to the manufacturer's instructions. Reactions were analysed by microfluidic gel fractionation (LabChip GXII, Perkin Elmer) and pooled. Primers were removed by fractionation (BluePippin, SAGE Science). The concentration and quality of DNA was analysed by QUBIT (ThermoFisher) and qPCR (Kapa Library Quantification Kit, Illumina). phiX Sequencing Control V3 (Illumina) was added to final concentration of 1.75 pM. Sequencing was performed on an Illumina MiSeq using the MiSeq Reagent Kit v2 Micro. Adapter sequences were removed and quality trimmed using bbduk (v.37.24) (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/) (parameters ktrim=r k=21 mink=11 hdist=2 qtrim=lr trimq=3 maq=10 ftr=250). Quantification of mutations at the target was performed using CRISPRESSO (http://crispresso.rocks/) (Pinello *et al.*, 2016) and automated with a custom script (available at https://github.com/cschu/CRISPRanto). The quantity of mutations in each sample was then normalised to the quantified transfection efficiency.

For analysis by Sanger sequencing, amplicons were purified (Qiaquick PCR purification, Qiagen) and incubated with a restriction enzyme for which the recognition sequence overlapped the expected cleavage site (see selection of targets) prior to reamplification. This reduced the amount of wild-type sequence in the sample, enabling detection of low-abundance amplicons. These amplicons were sequenced directly (Eurofins) and evidence of mutagenesis was conferred by the presence of multiple peaks, representing the different DSB-repair events across the population of cells, visible after the expected cut site (see Supplementary Data 4). These chromatogram signals were analysed using the ICE software that determines rates of CRISPR-Cas9 editing at a specific, sgRNA directed genomic location within a cell population (Synthego, https://ice.synthego.com/).

Results

Codon optimisation and sgRNA structure have minimal effects on the efficiency of targeted mutagenesis in plants

Prior to comparison with other Cas nucleases, we first assessed several variables for RNA-guided Cas9. We assessed the effect of codon-optimisation, comparing human (SpCas-h) versus plant (SpCas9-p), as well as variations in the single guide RNA sequence, comparing two sgRNAs: sgRNA and an sgRNA with extended step-loops (sgRNA-ES; Chen *et al.*, 2013). We also compared the use of a previously extended endogenous terminator for sgRNA expression cassettes (Wang *et al.*, 2015; Xu *et al.*, 2014). In these and subsequent experiments, all constructs were assembled similarly using identical regulatory sequences (Figure 1A). All sgRNAs contained a spacer to direct Cas9 to a target in the phytoene desaturate gene of *Nicotiana benthamiana* (*NbPDS1*). Following a quantitative assessment of the frequency of mutagenesis by Illumina sequencing, we found no significant differences in the number of sequencing reads with mutations at the target (Supplementary Data 5), indicating that neither human codon optimisation, the shorter stems found in the original sgRNA, or the minimal terminator significantly impaired the efficiency of mutagenesis in plants. In all subsequent experiments, we used SpCas9-h together with its original sgRNA as we have had previous success with these sequence in other species (Lawrenson *et al.*, 2015).

Cas proteins from different bacterial species show varied efficiencies of targeted mutagenesis at the same target in identical experimental conditions

To enable a direct comparison of the ability of four Cas proteins, SaCas9, SpCas9, AsCas12 and LbCas12a to induce targeted mutations, without the confounding influence of varied efficiency across targets, we designed three variants of a synthetic target such that the same recognition sequence was adjacent to the preferred PAM for each protein (Figure 1B). We

then co-delivered the plasmid with the cognate PAM to protoplasts together with plasmid DNA encoding each of the four nucleases and an appropriate sgRNA (Cas9) or crRNA (Cas12a) with a spacer to the synthetic target. We observed significantly more mutations at the target with SaCas9 than with SpCas9-h, AsCas12 or LbCas12a (Figure 1C).

Efficiency of mutagenesis correlates to GC content of the spacer

Analysis of the Arabidopsis genome for potential target sequences identified 3,853,090 potential targets for SpCas9, SaCas9 or Lb/AsCas12a in coding exons (Table1). Of the 2,695,798 targets recognised by SpCas9, 61,739 (2.29 %) have at least one potential off-target (identical or differing at only a single base) also in a coding exon, 7,201 of which differed by a single base in the first three positions. The total number of targets is expectedly lower for SaCas9 and Lb/AsCas12a, which recognise longer PAMs, (Table 1). We filtered these targets for those in genes previously shown to be expressed in leaves (see methods) and selected sequences for functional analysis in Arabidopsis leaf-derived protoplasts.
We were able to detect evidence of induced mutations at 14 out of 33 candidate targets for SpCas9. We were able to detect evidence of induced mutations at seven out of nine candidate targets for SaCas9. The sample set for SaCas9 is too small for meaningful analysis, however, analysis of nucleotide composition of the 33 SpCas9 targets found that the %GC content of spacers used where mutagenesis was detected was significantly greater than those for which no activity was detected (single-tailed T-test, $P<0.05$) (Figure 2). The data indicates that GC content of spacers should, for maximal efficiency, be greater that 40%. In contrast, we found no differences in the GC content within the seed region.

'High-fidelity' variants of Cas9 show reduced efficiency at targets with less than perfect identity to the spacer

To directly compare the efficiency and specificity, *in planta*, of five variants of SpCas9 nucleases, we used a similar experimental process as reported by (Kleinstiver, *et al.*, 2016). We designed a set of five sgRNAs each with a mutation in a different base of a spacer designed to target the *NbPDS* gene (Figure 3A). An sgRNA with an exactly matching spacer, or one of the five variants was delivered to plants cells in combination with each of the five variants of SpCas9 and the number of targeted mutations was quantified using Illumina sequencing. While the number of mutations induced by wild type Cas9 was significantly reduced when the spacer contained a mutation in the region close to the PAM, the presence of a mismatch between the spacer and target in the distal region had minimal effects (Figure 3B). In contrast, the frequency of mutations induced by the variants eCas9 1.0 and 1.2 was significantly reduced by a mismatch in any region of the sgRNA (Figure 3B). This experiment, although quantitative, was difficult to scale across a larger number of

targets. To compare the efficiency of mutagenesis across a larger number of targets and to observe the performance of spacers at non-identical endogenous targets, we conducted an analysis of the Arabidopsis genome to identify pairs of targets that differed by a single base pair. For each target at which mutagenesis was detected using SpCas9-h or SaCas9, the same sgRNA were also tested with protein variants (eCas9 1.0, eCas9 1.1, eSaCas9 and the recently reported xCas9 3.7) described to have reduced activity at targets with a less than perfect match to the spacer. As well as analysis at the targets, we also analysed the identified off-target locus for evidence of mutagenesis (Table 2). Of eight targets at which SpCas9-h induced mutations at the target, mutations were also detected at a second target that differed by one nucleotide in the first three positions. Of these eight, eCas9 1.0 was only able to induce mutations at three targets but off-target activity was detected at two of these. Similarly, eCas9 1.1 was able to induce mutations at five targets but off-target activity was detected at four; xCas9 was able to induce mutations at three targets but off-target activity was detected at two of these (Table 2 and Supplementary Data 4). In general, the signal of mutagenesis using the protein variants was less than for the equivalent wild-type protein sequence (Table 2 and Supplementary Data 4).

Discussion

Components of bacterial CRISPR/Cas systems have been applied to a wide variety of model and economically-important plant species including dicotyledonous fruit crops such as strawberries (Zhou *et al.*, 2018), tomatoes (Brooks *et al.*, 2014) and oranges (Jia and Wang, 2014) as well as monocotyledonous grain crops such as barley (Lawrenson *et al.*, 2015), wheat (Liang *et al.*, 2018) and maize (Char *et al.*, 2017). In most cases, engineered plants are regenerated via somatic embryogenesis in tissue culture; a process that is both laborious and time-consuming. Consequently, researchers are interested in the efficiency of the molecular tools they deliver. Since the first reports of Cas9 from *S. pyogenes* being applied as tool for genome engineering, there has been much interest in novel properties associated with Cas nucleases from addition CRISPR systems (Friedland *et al.*, 2015; Zetsche *et al.*, 2015). These include comparative efficiency and the potential for engineering at an increased number of genomic targets due to the recognition of additional PAMs. Because of this, Cas nucleases have often, necessarily, used different targets. Together with other differences introduced by different construct designs and experimental conditions, it is difficult to draw conclusions about relative efficiencies. Our experimental strategies allowed us to compare the efficiency of different components for genome engineering on the same target and indicate that SaCas9 may have the highest efficiency of targeted mutagenesis in plants (Figure 1). The drawback of our experimental design, however, is that that the same experimental conditions are unlikely to be optimal for each Cas nuclease. For

example, recent work has shown that temperature can increase the activity of Cas12a (Moreno-Mateos *et al.*, 2017). It is also important to note that our experimental system, necessarily, used the same regulatory elements in order to compare different nucleases. Since our experiments were performed in leaf-derived protoplasts of *N. benthamiana* and *A. thaliana*, we utilised the oft-used 35s promoter from Cauliflower Mosaic Virus (CaMV-35s), which expresses strongly in this cell type, to drive expression of all nucleases. This promoter will not be optimal for all species and the choice of promoters is likely to contribute to efficiency in different species. Even in Arabidopsis, where stable transgenesis utilises a floral-dip method, it is preferential to use a promoter that expressed strongly in developing embryos to minimise the recovery of plants that are genetically-chimeric due to expression of the Cas-nuclease in somatic cells during development (Wang *et al.*, 2015; Xu *et al.*, 2018; Yan *et al.*, 2015).

Much of the work on sequence determinants of spacers performed to date has been performed in mammalian cell cultures (Doench *et al.*, 2014, 2016; Gagnon *et al.*, 2014; Xu *et al.*, 2015). These systems provide the advantage of being able to delivery libraries of sgRNAs targeting genes involved in a particular process followed by the selection of cells with the expected phenotype correlating to disruption of those genes to be selected (e.g. by staining and FACS). In an analysis of a number of datasets, including one of 4,000 sgRNA targeting 17 genes, Doench *et al.* (2016) concluded that the patterns of activity are complex, and are not likely to be apparent by examining smaller numbers of sgRNA:DNA interactions. Current methods of DNA delivery to plant cells include the preparation of separate strains of *Agrobacterium tumefaciens* for delivery to callus or tissues, or the direct delivery of plasmids, either to protoplasts or to tissues (using biolistic delivery). Delivery of a large-scale combined library of constructs using these methods would result in populations of cells with multiple induced mutations from which it would not be possible to identify or separate cells with individual genotypes. In our study, we delivered individual constructs to aliquots of protoplasts prepared from the same batch of leaf tissue. This allowed us to work with larger number of constructs, however, since cultures of protoplasts cannot be perpetuated (they reform cell walls and form masses of cells if allowed or induced to divide), it is not possible to select populations of cells beyond the first 24-48 hours. The advantages of this system, however, allowed us to compare the impact of a number of different components by allowing similar levels of expression across different experiments that are difficult to achieve with integrated transgenes, the expression of which is subject to copy number and the location of integration (Kohli *et al.*, 2006). Using this method, we attempted to induce mutations at over 45 targets in the Arabidopsis genome, using sgRNAs with spacers with both perfect and imperfect identity. This enabled us to compare the features of functional spacers (those that

were able to induce targeted mutations) with non-functional spacer noting a correlation with GC content (Figure 2), broadly in agreement with that reported for mammals (Doench *et al.*, 2014). We did not, however, note any determinants for sequence composition in the seed region or at specific bases (data not shown), likely because of the comparatively small size of our dataset. The correlation of efficiency with GC content may be related to nucleosome occupancy, previously reported to influence the ability of Cas9 to recognise its targets (Horlbeck *et al.*, 2016); nucleosome-depleted regions in Arabidopsis tend to have a higher GC content (Liu *et al.*, 2015). As this negative correlation between nucleosome occupancy and GC content is not known to be common across all organisms, further studies in a wider range of plant species will need to be performed to determine if efficiency can be predicted by GC content alone.

Analysis of Arabidopsis coding sequences revealed that many targets for Cas-mediated genome engineering have one or more potential off-targets in other coding sequences that differ by just a single base (Table 1, Supplementary Data 1). We provide experimental evidence that SpCas9 and SaCas9 are readily able to induce mutations at targets with less than perfect identity to the spacer (Table 2 and Figure 3). Different laboratories have engineered variants of SpCas9 and SaCas9, reportedly reducing their activity at targets to which the spacer does not have 100% identity (Hu *et al.*, 2018; Kleinstiver *et al.*, 2016; Slaymaker *et al.*, 2015). To compare these proteins and determine their function in plants, we measured their efficiency and specificity across a number of targets (Table 2 and Figure 3). In general, the variants showed a reduction in efficiency at all targets, however this was more substantial at sites where the target and spacer did not share 100% identity. In cases where specificity to a single target is critical e.g. targeting a specific member of a gene family, particularly when recent paralogues are located in tandem, these variants are likely to be useful.

This study utilised the ease of standardized modular cloning to create a suite of comparable constructs to enable direct comparison of multiple different tools for genome engineering. Type IIS mediated assembly methods have been widely utilised to facilitate the construction of the complex plasmids required for multiplexed genome editing (Cermak *et al.*, 2017; Ma *et al.*, 2015; Vazquez-Vilar *et al.*, 2016; Xing *et al.*, 2014). In addition to standardising our experimental process, we are able to contribute an expanded toolkit of modular, reusable parts for plant genome engineering that will facilitate their application in new studies (Figure 4). The basic (Level 0) parts are flanked with inverted *BsaI* sites that will release parts with overhangs in the common genetic syntax for plants (Patron *et al.*, 2015), making them amenable for reuse with a number of different assembly toolkits for plants.

Acknowledgements

This work was funded by the UK Biotechnological and Biological Sciences Research Council (BBSRC) and Engineering and Physics Research Council (EPSRC) Synthetic Biology for Growth program [OpenPlant Synthetic Biology Research Centre; grant number BB/L014130/1]. Next-generation sequencing and library construction was delivered via the BBSRC National Capability in Genomics at the Earlham Institute [grant number BB/CCG1720/1] by members of the Genomics Pipelines Group and automated DNA assembly was performed in the Earlham DNA Foundry by AW.

Contributions

NJP and OR designed the study. OR executed all experiments. AW designed, optimised and executed automated DNA assembly. CS conducted all bioinformatic analysis of the Arabidopsis genome. OR, CS and NP analysed data. NJP wrote the manuscript with contributions from AW, OR and CS. NJP was responsible for supervision and fundraising.

Figure Legends

Figure 1. Comparison of Cas9 and Cas12 nucleases at a single target. (A) All constructs were assembled in the same backbone, using the same regulatory elements. Asterisks (*) represent elements compared in this study. The 'dummy' consists of 15 random nucleotides and enables the Cas9 and sgRNA expression cassettes to be reassembled with, e.g. a selectable marker cassette, in Position 1. (B) An identical target sequence was used to test all Cas proteins. The target was flanked by the preferred cognate protospacer adjacent motif (PAM) for SpCas9 (red box), SaCas9 (blue box) or Cas12a (black box). The spacer sequence used in each sgRNA (Cas9) or crRNA (Cas12a) are indicated by dotted lines. (C) Under identical experimental conditions, SaCas9 induced more mutations at the target than SpCas9, AsCas12a or LbCas12a. Error bars = standard error of the mean; n=4. Post hoc comparisons using Tukey's Honest Significant Difference test indicted significant differences: * = $p<0.05$, ** = $p<0.01$.

Figure 2 Nucleotide content analysis of 33 spacer sequences. All spacers have cognate targets in the coding sequences of leaf-expressed genes tested in sgRNAs with SpCas9-h. Each spacer was incorporated into a sgRNA and tested in Arabidopsis protoplasts. Blue bars indicate mutations were detected at target; red bars indicate no mutations detected at target. Seed = six base pairs adjacent to PAM. *= p-value 0.044591. The result is significant at $p <0.05$.

Figure 3. Assessment of Cas9 specificity at an identical target. (A) Schematic showing transversions in four regions of the spacer sequences of sgRNAs targeting *NbPDS1* **(B)** The efficiency of targeted mutagenesis is impacted by transversions in the spacer. Shaded bars represent a perfect match of the spacer to the target, black bars show sgRNAs with spacers with transversions relative to the target. Error bars = standard error of the mean; n=4. Post hoc comparisons using Tukey's Honest Significant Difference test indicted significant differences: * = $p<0.05$, ** = $p<0.01$.

Figure 4. An expanded toolkit for Cas-mediated genome engineering in plants. Regulatory elements and coding sequences are cloned as Level 0 parts, enabling one-step assembly into transcriptional units mediated by *BsaI* (e.g. using the Plant MoClo Toolkit, Engler et al., 2014) and subsequently into multigene constructs. Numbers in parentheses represent catalogue number at Addgene. **For simplicity, all cassettes are shown assembled on the forward strand. However, the orientation of any cassette in the final construct can be altered by use of a reverse Level 1 acceptor. '+S'= with stop codon '-S'= no stop codon.

Table 1 Numbers of Cas9 and Cas12 targets and potential off-targets in Arabidopsis coding exons.

Cas nuclease	Target	Total number of targets in coding exons	Number and similarity of potential off-targets in coding exons		
			Identical to target	One SNP in positions 1, 2 or 3 (distal to PAM)	One SNP at any other position
SaCas9	N_{21}-NNGGGT	118,048	1,556	228	557
SpCas9	N_{20}-NGG	2,695,798	45,707	7,201	16,334
Lb/As Cas12a	TTTV-N_{23}	1,039,244	13,644	1,903	13,644

Table 2. Assessment of the specificity of Cas9 variants at targets in Arabidopsis coding sequences. Differences between target (bold text) and off-target pairs are shown in red text. The Protospacer Adjacent Motif (PAM) is shown in blue text. Numbers in parentheses are scores from Inference of CRISPR Editing (ICE) software.
"*" indicates assessment made by detection of deletion induced with a second sgRNA. "|" indicates expected point of cleavage.

Gene Model Identifier	Target (bold) and off-targets	Cas9 nuclease			
		SpCas9-h	xCas9 3.7	eCas9 1.0	eCas9 1.1
AT5G58420	**TGGCAAAGGAACAAAGC|CATGGG**	Y (41)	N (0)	Y (37)	Y (15)
AT5G07090	CGGCAAAGGAACAAAGC|CATGGG	Y (67)	N (1)	N (7)	Y (3)
AT2G32730	**GGGAAGTATCAACAAGC|CATGGG**	Y (53)	N (0)	Y (39)	Y (58)
AT1G04810	GGAAAGTATCAACAAGC|CATGGG	Y (39)	N (0)	Y (24)	Y (30)
AT3G14720	**CAGAATGATCCACAGAG|CTCAGG**	Y (55)	Y (35)	Y (48)	N (1)
AT1G53510	AAGAATGATCCACAGAG|CTCAGG	Y (13)	N (0)	Y (18)	N (2)
AT5G10620	**CTGATGATTCAGGACCA|TGGAGG**	Y (24)	Y (23)	Y (32)	N (7)
AT5G10605	ATGATGATTCAGGACCA|TGGAGG	Y (39)	Y (34)	Y (26)	N (0)
AT5G19360	**CCATACTTGCCATTCCA|TGGGGG**	Y (47)	Y (30)	N (1)	Y (36)
AT5G12180	GCATACTTGCCATTCCA|TGGGGG	Y (45)	Y (41)	Y (39)	Y (33)
AT5G04770	**GTCGAGACGACTAGCAC|GGCCGG**	Y*	N*	N*	N*
AT3G10600	GTAGAGACGACTAGCAC|GGCCGG	Y*	N*	N*	N*
AT3G54020	**CGACAGTGATAGTTAGG|GCCAGG**	Y*	N*	N*	N*
AT2G37940	CGGCAGTGATAGTTAGG|GCCAGG	Y*	N*	N*	N*
AT2G68260	**TGAGCCGGCTGAGCCAT|GGAAGG**	Y*	N*	N*	N*
AT1G68260	TGTGCCGGCTGAGCCAT|GGAAGG	Y*	N*	N*	N*
		SaCas9	eSaCas9		
AT5G04160	**TATGACCTTTAAGAGAGA|AGCTTGGGT**	Y	Y		
AT3G10290	CATGACCTTTAAGAGAGA|AGCTTGGGT	Y	N		
AT1G62600	**GACGCTCGAATGAACAAC|GGATCGGGT**	Y*	Y*		
AT1G63390	AACGCTCGAATGAACAAC|GGATCGGGT	Y*	Y*		
AT1G73460	**TTCATGGTGATCAGGACT|TCCACGGGT**	Y*	N*		
AT1G73450	ATCATGGTGATCAGGACT|TCCACGGGT	Y*	N*		
AT1G14910	**GCCCTTGAAATATACAAG|CGTGCGGGT**	Y*	Y*		
AT2G01600	TCCCTTGAAATATACAAG|CGTGCGGGT	Y*	N*		

List of Supplementary Data

Supplementary Data 1. Details of genomic targets for Cas nucleases in *Arabidopsis thaliana* coding exons

Supplementary Data 2. Details of 135 plasmid constructs used in this study.

Supplementary Data 3. Table of oligonucleotide primers used to amplify genomic targets

Supplementary Data 4. Chromatograms and gel images showing evidence of Cas-induced targeted mutations and deletions.

Supplementary Data 5. Effect of Cas9 codon-optimisation and sgRNA scaffold on the efficiency of targeted mutagenesis at *PDS1* in *Nicotiana benthamiana* protoplasts (h = human; p= plant; ES = extended stem; T= terminator). Error bars = standard error of the mean; n=4.

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.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007) *CRISPR provides acquired resistance against viruses in prokaryotes. Science*, **315**, 1709–12.

Begemann, M.B., Gray, B.N., January, E., Gordon, G.C., He, Y., Liu, H., et al. (2017) *Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases*. *Sci. Rep.*, **7**, 1–6.

Brooks, C., Nekrasov, V., Lippman, Z.B., Van Eck, J., and Eck, J. Van (2014) *Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System*. *Plant Physiol.*, **166**, 1292–1297.

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., and Voytas, D.F. (2015) *High-frequency, precise modification of the tomato genome*. *Genome Biol.*, **16**, 232.

Cermak, T., Curtin, S., Gil-Humanes, J., Cegan, R., Starker, C.G., Kono, T.J.Y., et al. (2017) *A multi-purpose toolkit to enable advanced genome engineering in plants*. *Plant Cell*.

Char, S.N., Neelakandan, A.K., Nahampun, H., Frame, B., Main, M., Spalding, M.H., et al. (2017) *An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted*

mutagenesis in maize. *Plant Biotechnol. J.*, **15**, 257–268.

Chen, B., Gilbert, L. a, Cimini, B. a, Schnitzbauer, J., Zhang, W., Li, G.-W., et al. (2013) *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell*, **155**, 1479–91.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013) *Multiplex genome engineering using CRISPR/Cas systems. Science*, **339**, 819–23.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., et al. (2016) *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9*. *Nat. Biotechnol.*, **34**, 184–191.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., et al. (2014) *Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation*. *Nat. Biotechnol.*, **32**, 1262–7.

Engler, C., Kandzia, R., and Marillonnet, S. (2008) *A one pot, one step, precision cloning method with high throughput capability. PLoS One*, **3**, e3647.

Engler, C., Youles, M., and Grüetzner, R. (2014) *A Golden Gate modular cloning toolbox for plants*. *ACS Synth. Biol.*, **3**, 839–843.

Friedland, A.E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., et al. (2015) *Characterization of Staphylococcus aureus Cas9: A smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications*. *Genome Biol.*, **24**, 257.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., et al. (2014) *Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs*. *PLoS One*, **9**, e106396.

Gao, C. (2018) *The future of CRISPR technologies in agriculture*. *Nat. Rev. Mol. Cell Biol.,* ***19****, 275-276*

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., et al. (2017) *High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9*. *Plant J.*, **89**, 1251–1262.

Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., et al. (2016) *Nucleosomes impede cas9 access to DNA in vivo and in vitro*. *Elife*, **5**, e12677.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., et al. (2018) *Evolved Cas9 variants with broad PAM compatibility and high DNA specificity*. *Nature*., **556**, 57-63

Hu, Z., Cools, T., and De Veylder, L. (2016) *Mechanisms Used by Plants to Cope with DNA Damage*. *Annu. Rev. Plant Biol.*, **67**, 439–462.

Jia, H. and Wang, N. (2014) *Targeted genome editing of sweet orange using Cas9/sgRNA. PLoS One*, **9**, e93806.

Kaya, H., Mikami, M., Endo, A., Endo, M., and Toki, S. (2016) *Highly specific targeted*

mutagenesis in plants using Staphylococcus aureus Cas9. *Sci. Rep.*, **6**, 26871.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016) *High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects*. *Nature*, **529**, 490–495.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V. V., Zheng, Z., and Joung, J.K. (2015a) *Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition*. *Nat. Biotechnol.*, **33**, 1293–1298.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V. V., Nguyen, N.T., Zheng, Z., et al. (2015b) *Engineered CRISPR-Cas9 nucleases with altered PAM specificities*. *Nature*, **523**, 481–485.

Kohli, A., González-Melendi, P., Abranches, R., Capell, T., Stoger, E., and Christou, P. (2006) *The quest to understand the basis and mechanisms that control expression of introduced transgenes in crop plants*. *Plant Signal. Behav.*, **1**, 185–95.

Köster, J. and Rahmann, S. (2012) *Snakemake-a scalable bioinformatics workflow engine*. *Bioinformatics*, **28**, 2520–2522.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N.J., et al. (2015) *Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease*. *Genome Res.*, **30**, 258.

Lee, K., Zhang, Y., Kleinstiver, B.P., Guo, J.A., Aryee, M.J., Miller, J., et al. (2018) *Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize*. *Plant Biotechnol. J.*, 1–11.

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

Li, Z., Liu, Z., Xing, A., Moon, B.P., Koellhoffer, J.P., Huang, L., et al. (2015) *Cas9-Guide RNA Directed Genome Editing in Soybean*. *Plant Physiol.*, **169**, 960–970.

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J.L., and Gao, C. (2018) *Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 in vitro transcripts or ribonucleoproteins*. *Nat. Protoc.*, **13**, 413–430.

Liu, M.-J., Seddon, A.E., Tsai, Z.T.-Y., Major, I.T., Floer, M., Howe, G.A., and Shiu, S.-H. (2015) *Determinants of nucleosome positioning and their influence on plant gene expression*. *Genome Res.*, **25**, 1182–1195.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015) *A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants*. *Mol. Plant*, **8**, 1274–1284.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., et al. (2013) *RNA-Guided Human Genome Engineering via Cas9 Prashant*. *Science*, **339**, 823–826.

Marçais, G. and Kingsford, C. (2011) *A fast, lock-free approach for efficient parallel counting of occurrences of k-mers*. *Bioinformatics*, **27**, 764–70.

Moreno-Mateos, M.A., Fernandez, J.P., Rouet, R., Vejnar, C.E., Lane, M.A., Mis, E., et al. (2017) *CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing*. *Nat. Commun.*, **8**, 2024.

Patron, N.N.J., Orzaez, D., Marillonnet, S., Warzecha, W., Matthewman, C., Youles, M., et al. (2015) *Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts*. *New Phytol.*, **208**, 13–19.

Phillips, R.L., Kaeppler, S.M., and Olhoft, P. (1994) *Genetic instability of plant tissue cultures: breakdown of normal controls. Proc. Natl. Acad. Sci. U. S. A.*, **91**, 5222–6.

Pinello, L., Canver, M.C., Hoban, M.D., Orkin, S.H., Kohn, D.B., Bauer, D.E., and Yuan, G.C. (2016) *Analyzing CRISPR genome-editing experiments with CRISPResso*. *Nat. Biotechnol.*, **34**, 695–7.

Quinlan, A.R. and Hall, I.M. (2010) *BEDTools: A flexible suite of utilities for comparing genomic features*. *Bioinformatics*, **26**, 841–2.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., et al. (2015) *In vivo genome editing using Staphylococcus aureus Cas9*. *Nature*, **520**, 186–191.

Ricroch, A., Clairand, P., and Harwood, W. (2017) *Use of CRISPR systems in plant genome editing: toward new opportunities in agriculture*. *Emerg. Top. Life Sci.*, **1**, 168–192.

Schiml, S. and Puchta, H. (2016) *Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas*. *Plant Methods*, **12**, 8.

Slaymaker, I.M., Linyi, G., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2015) *Rationally engineered Cas9 nucleases with improved specificity*. *Science*, **10**, 1126.

Steinert, J., Schiml, S., Fauser, F., and Puchta, H. (2015) *Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and Staphylococcus aureus*. *Plant J.*, **84**, 1295–1305.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014) *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature*, **507**, 62–7.

Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., et al. (2017) *A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants*. *Nat. Plants*, **3**, 17018.

Vazquez-Vilar, M., Bernabé-Orts, J.M., Fernandez-del-Carmen, A., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D. (2016) *A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard*. *Plant Methods*, **12**, 10.

Wang, Z.-P., Xing, H., Dong, L., Zhang, H., Han, C., Wang, X., and Chen, Q. (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, 144.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012) *Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system*. *Bioeng. Bugs*, **3**, 38–43.

Xie, H., Tang, L., He, X., Liu, X., Zhou, C., Liu, J., et al. (2018) *SaCas9 Requires 5'-NNGRRT-3' PAM for Sufficient Cleavage and Possesses Higher Cleavage Activity than SpCas9 or FnCpf1 in Human Cells*. *Biotechnol. J.*, **14**, e1700561.

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., et al. (2014) *A CRISPR/Cas9 toolkit for multiplex genome editing in plants*. *BMC Plant Biol.*, **14**, 327.

Xu, H., Xiao, T., Chen, C.H., Li, W., Meyer, C.A., Wu, Q., et al. (2015) *Sequence determinants of improved CRISPR sgRNA design*. *Genome Res.*, **25**, 1147–57.

Xu, P., Su, H., Chen, W., and Lu, P. (2018) *The Application of a Meiocyte-Specific CRISPR/Cas9 (MSC) System and a Suicide-MSC System in Generating Inheritable and Stable Mutations in Arabidopsis*. *Front. Plant Sci.*, **9**, 1–12.

Xu, R., Li, H., Qin, R., Wang, L., Li, L., Wei, P., and Yang, J. (2014) *Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice*. *Rice*, **7**, 5.

Xu, R., Qin, R., Li, H., Li, D., Li, L., Wei, P., and Yang, J. (2017) *Generation of targeted mutant rice using a CRISPR-Cpf1 system*. *Plant Biotechnol. J.*, **15**, 713–717.

Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017) *Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1*. *Mol. Cell*, **67**, 633–645.

Yan, L., Wei, S., Wu, Y., Hu, R., Li, H., Yang, W., and Xie, Q. (2015) *High-Efficiency Genome Editing in Arabidopsis Using YAO Promoter-Driven CRISPR/Cas9 System*. *Mol. Plant*, **8**, 1820–1823.

Yin, K., Gao, C., and Qiu, J.-L. (2017) *Progress and prospects in plant genome editing*. *Nat. Plants*, **3**, 17107.

Yoo, S.D., Cho, Y.H., and Sheen, J. (2007) *Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis*. *Nat. Protoc.*, **2**, 1565–72.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Regev, A., Koonin, E. V., Zhang, F., et al. (2015) *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*. *Cell*, **163**, 759–771.

Zhang, D., Wang, Z., Wang, N., Gao, Y., Liu, Y., Wu, Y., et al. (2014) *Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications*. *PLoS One*, **9**, e96879.

Zhang, Q., Xing, H.L., Wang, Z.P., Zhang, H.Y., Yang, F., Wang, X.C., and Chen, Q.J. (2018) *Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention*. *Plant Mol. Biol.*, **96**, 445–456.

Zhou, J., Wang, G., and Liu, Z. (2018) *Efficient genome editing of wild strawberry genes, vector development and validation*. *Plant Biotechnol. J.* e-pub ahead of print.

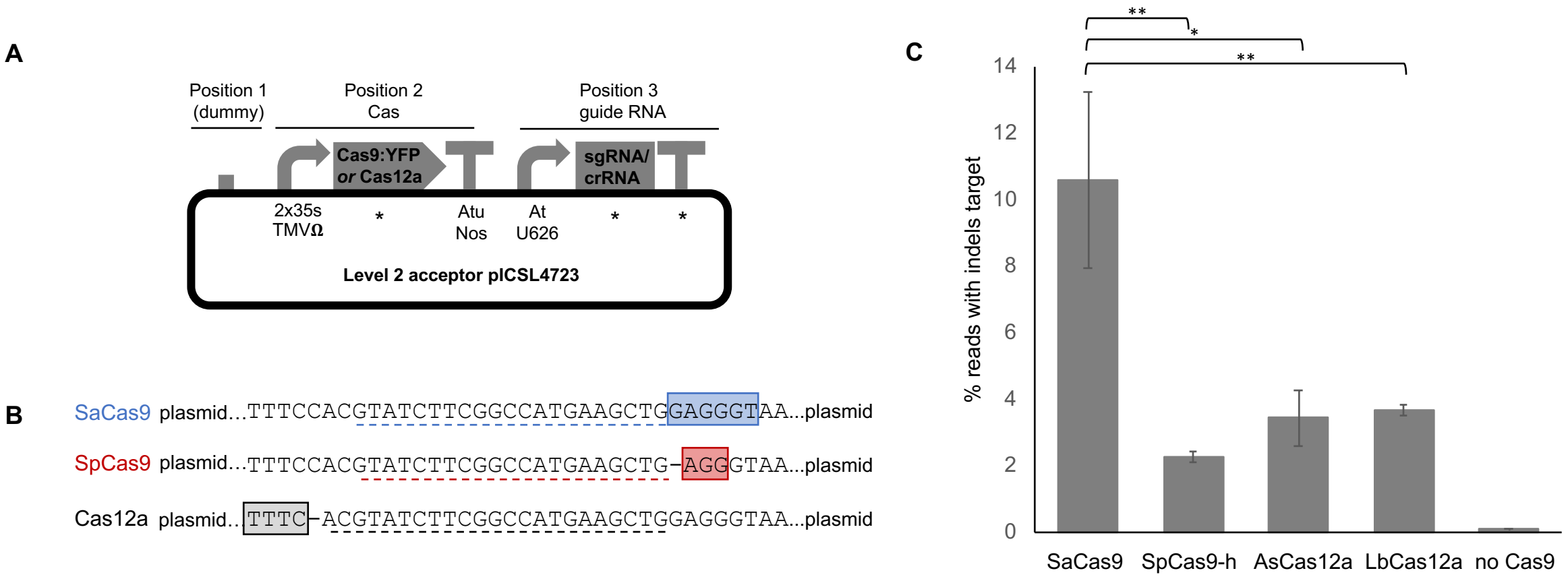

Figure 1. Comparison of Cas9 and Cas12 nucleases at a single target. (**A**) All constructs were assembled in the same backbone, using the same regulatory elements. Asterisks (*) represent elements compared in this study. The 'dummy' consists of 15 random nucleotides and enables the Cas9 and sgRNA expression cassettes to be reassembled with, e.g. a selectable marker cassette, in Position 1. (**B**) An identical target sequence was used to test all Cas proteins. The target was flanked by the preferred cognate protospacer adjacent motif (PAM) for SpCas9 (red box), SaCas9 (blue box) or Cas12a (black box). The spacer sequence used in each sgRNA (Cas9) or crRNA (Cas12a) are indicated by dotted lines. (**C**) Under identical experimental conditions, SaCas9 induced more mutations at the target than SpCas9, AsCas12a or LbCas12a. Error bars = standard error of the mean; n=4. Post hoc comparisons using Tukey's Honest Significant Difference test indicted significant differences: * = $p<0.05$, ** = $p<0.01$.

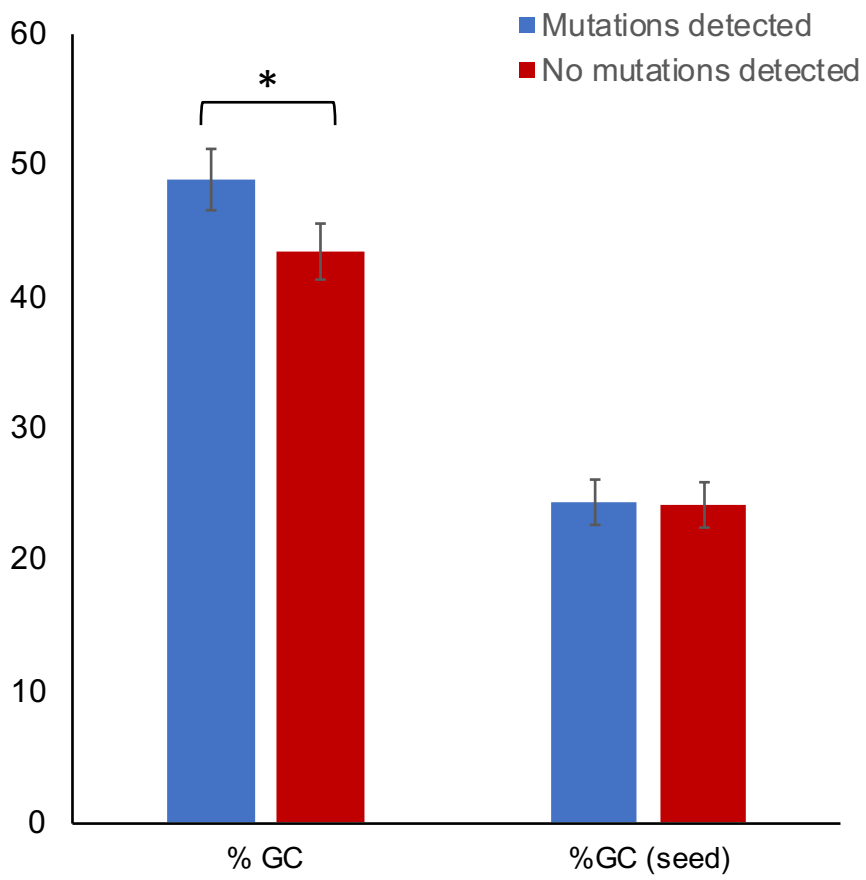

Figure 2 Nucleotide content analysis of 33 spacer sequences. All spacers have cognate targets in the coding sequences of leaf-expressed genes tested in sgRNAs with SpCas9-h. Each spacer was incorporated into a sgRNA and tested in Arabidopsis protoplasts. Blue bars indicate mutations were detected at target; red bars indicate no mutations detected at target. Seed = six base pairs adjacent to PAM. *= p-value 0.044591. The result is significant at $p < 0.05$.

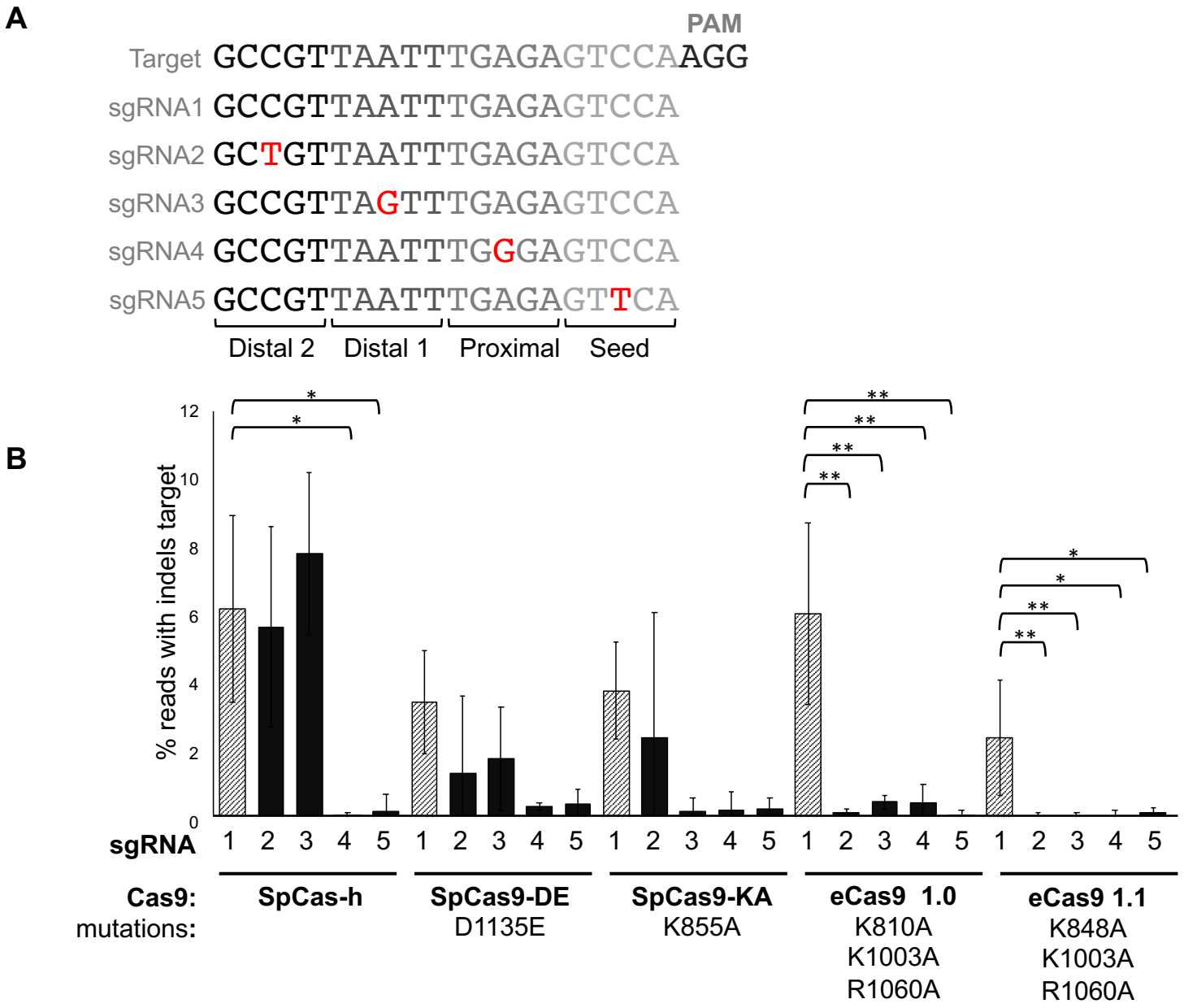

Figure 3. Assessment of Cas9 specificity at an identical target. (A) Schematic showing transversions in four regions of the spacer sequences of sgRNAs targeting *NbPDS1* **(B)** The efficiency of targeted mutagenesis is impacted by transversions in the spacer. Shaded bars represent a perfect match of the spacer to the target, black bars show sgRNAs with spacers with transversions relative to the target. Error bars = standard error of the mean; n=4. Post hoc comparisons using Tukey's Honest Significant Difference test indicted significant differences: * = p<0.05, **=p<0.01.

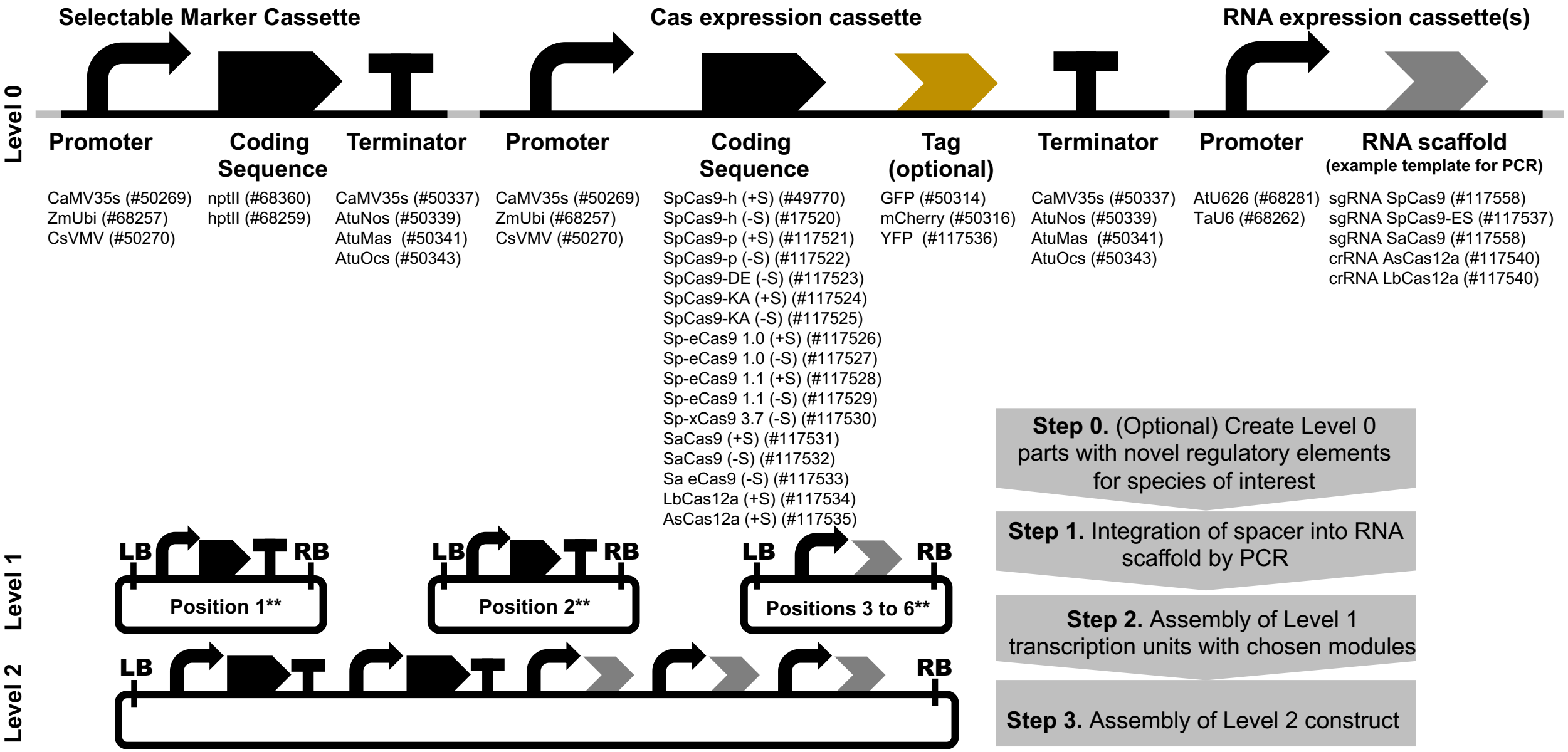

Figure 4. An expanded toolkit for Cas-mediated genome engineering in plants. Regulatory elements and coding sequences are cloned as Level 0 parts, enabling one-step assembly into transcriptional units mediated by *BsaI* (e.g. using the Plant MoClo Toolkit, Engler et al 2014) and subsequently into multigene constructs. Numbers in parentheses represent catalogue number at Addgene. **For simplicity, all cassettes are shown assembled on the forward strand. However, the orientation of any cassette in the final construct can be altered by use of a reverse Level 1 acceptor. '+S'= with stop codon '-S'= no stop codon.