Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2019) **17**, pp. 1971–1984 doi: 10.1111/pbi.13113

Assessment of Cas12a-mediated gene editing efficiency in plants

Joan Miquel Bernabé-Orts[1], Iván Casas-Rodrigo[1,†], Eugenio G. Minguet[1], Viola Landolfi[2], Victor Garcia-Carpintero[3], Silvia Gianoglio[2], Marta Vázquez-Vilar[1], Antonio Granell[1] and Diego Orzaez[1,*]

[1]*Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia, Valencia, Spain*
[2]*Department of Agricultural, Forest and Food Sciences, University of Torino, Grugliasco, Italy*
[3]*Institute for the Conservation and Breeding of Agricultural Biodiversity (COMAV-UPV), Universitat Politècnica de València, Valencia, Spain*

Received 6 February 2018;
revised 19 February 2019;
accepted 20 March 2019.
**Correspondence* (Tel +34 963879933;
fax +34 963877859;
email dorzaez@ibmcp.upv.es)
[†] Present address: Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH) Zurich, Basel, Switzerland.

Keywords: CRISPR/Cas9, CRISPR/Cas12a, GoldenBraid, plant gene editing, off-target.

Summary

The CRISPR/Cas12a editing system opens new possibilities for plant genome engineering. To obtain a comparative assessment of RNA-guided endonuclease (RGEN) types in plants, we adapted the CRISPR/Cas12a system to the GoldenBraid (GB) modular cloning platform and compared the efficiency of *Acidaminococcus* (As) and *Lachnospiraceae* (Lb) Cas12a variants with the previously described GB-assembled *Streptococcus pyogenes* Cas9 (SpCas9) constructs in eight *Nicotiana benthamiana* loci using transient expression. All three nucleases showed drastic target-dependent differences in efficiency, with LbCas12 producing higher mutagenesis rates in five of the eight loci assayed, as estimated with the T7E1 endonuclease assay. Attempts to engineer crRNA direct repeat (DR) had little effect improving on-target efficiency for AsCas12a and resulted deleterious in the case of LbCas12a. To complete the assessment of Cas12a activity, we carried out genome editing experiments in three different model plants, namely *N. benthamiana*, *Solanum lycopersicum* and *Arabidopsis thaliana*. For the latter, we also resequenced Cas12a-free segregating T2 lines to assess possible off-target effects. Our results showed that the mutagenesis footprint of Cas12a is enriched in deletions of −10 to −2 nucleotides and included in some instances complex rearrangements in the surroundings of the target sites. We found no evidence of off-target mutations neither in related sequences nor somewhere else in the genome. Collectively, this study shows that LbCas12a is a viable alternative to SpCas9 for plant genome engineering.

Introduction

CRISPR (clustered regularly interspaced short palindromic repeats) is an adaptive immune system, which defends prokaryotes against invader nucleic acids (Barrangou *et al.*, 2007; Makarova *et al.*, 2015; Marraffini, 2015). Components of Class 2 Type II CRISPR systems have been engineered into easily programmable RNA-guided endonucleases (RGENs), setting the stage for genome editing approaches that will be widely available to the research community (Hsu *et al.*, 2014). The arrangement of CRISPR-based tools typically involves two components: the Cas9 effector protein and a ~100-nt small guide RNA (sgRNA), which is a combination of the crRNA and tracrRNA originally encoded by the CRISPR loci (Jinek *et al.*, 2012). To date, the most widely used Cas9 orthologue has been SpCas9 (*Streptococcus pyogenes*). The tandem SpCas9-sgRNA targets genomic sequences of 20 nt upstream to a 5′-NGG-3′ protospacer adjacent motif (PAM) and introduces a double-strand break (DSB) in many organisms (Bentley *et al.*, 1998; Bortesi and Fischer, 2015; Tabebordbar *et al.*, 2016; Yu *et al.*, 2013). The DSB takes place near the PAM sequence and originates blunt ends that, in plants, will be repaired preferentially by nonhomologous end joining (NHEJ), occasionally introducing indel mutations (Ran *et al.*, 2013).

The canonical CRISPR/Cas9 RGEN system offers a variety of applications in plant biotechnology, from crop breeding to plant synthetic biology. The efficiency of SpCas9 in genome editing has been demonstrated in dozens of plant species, where it has been used to produce new traits of agronomic importance, including resistance to abiotic stress and herbicides (Khan *et al.*, 2017; Soda *et al.*, 2017; Yin *et al.*, 2017a). Despite its success, further expansion of the CRISPR toolbox with new RGEN tools will be necessary to overcome inherent limitations of SpCas9. One of these limitations is the strict PAM dependence, which constrains the availability of target sites. For instance, noncoding regions, which are important in breeding for the generation of artificial quantitative trait variability (Kumar *et al.*, 2017; Zsögön *et al.*, 2016), are relatively poor in 5′-NGG-3′ sites (Wang *et al.*, 2016). Other current limitations of SpCas9, like its large size for viral delivery or the low efficiency in gene targeting caused by blunt DSB, could eventually be overcome with the introduction of alternative editing tools.

Recently, a new putative Class 2 Type V CRISPR system has been identified (Zetsche *et al.*, 2015). This locus encodes for a different Cas effector protein, Cas12a (Cpf1), and a single crRNA of 42 nt; the CRISPR/Cas12a system displays different features from that of SpCas9 (Zaidi *et al.*, 2017), making it an interesting complement to SpCas9 in plant genome engineering. *Acidaminococcus sp.* Cas12a (AsCas12a) and *Lachnospiraceae bacterium* Cas12a (LbCas12a) are the most commonly used Cas12a orthologues, which have been proven to be effective in mammalian cells (Kim *et al.*, 2016, 2017b; Zetsche *et al.*, 2015). Both RGENs recognize a 5′-TTTN-3′ PAM sequence and introduce a staggered DSB at the end of the protospacer (PS) sequence, which has been reported to favour gene insertions (Zetsche *et al.*, 2015). Finally, Cas12a is smaller

than SpCas9, which might facilitate viral delivery (Zetsche *et al.*, 2015).

CRISPR/Cas12a engineering activity was recently demonstrated in plant cells. First, targeted mutagenesis of rice and tobacco genomes was achieved using the *Francisella novicida* Cas12a (FnCas12a) orthologue (Endo *et al.*, 2016). Most recently, T0 mutant lines in rice have been obtained using AsCas12a and LbCas12a (Tang *et al.*, 2017; Wang *et al.*, 2017; Xu *et al.*, 2017; Yin *et al.*, 2017b). Also, a DNA-free approach, delivering AsCas12a and LbCas12a loaded with crRNA, was validated in wild tobacco and soya bean protoplasts (Kim *et al.*, 2017a). However, the decision to use Cas9 or Cas12a is difficult to make since few studies have assessed the reliability and robustness of Cas12a as an alternative to Cas9 in plants. Here, we report a comprehensive assessment of Cas12a activity in plants through the analysis of more than 130 mutagenesis events in three different plant species. Our results indicate that Cas12a is a viable alternative to Cas9 for genome editing in plants, although further efforts will be required to make the efficiency of the editing activity more predictable.

Results

A GB-assisted cloning strategy for plant Cas12a constructs

GoldenBraid 3.0 (GB3.0) is a software-assisted multigene cloning system (https://gbcloning.upv.es) based on type IIS restriction enzymes (Sarrion-Perdigones *et al.*, 2011) and conforming to the phytobrick standard (Patron *et al.*, 2015). GB3.0 was earlier adapted to plant gene editing with SpCas9 (Vazquez-Vilar *et al.*, 2016). Here, all the necessary elements to carry out targeted mutagenesis with AsCas12a and LbCas12a were adapted to GB3.0. Figure 1a depicts the cloning workflow followed to create the so-called Cas12a gene editing module (CGEM). Cloning a functional crRNA requires the following three components: (i) an invariant 5′ element comprising the *At*U6-26 RNApol III promoter fused to either the AsCas12a or LbCas12a direct repeat (DR) (GB1442 or GB1443 standard phytobricks, respectively); (ii) two partially complementary oligonucleotides containing the PS sequence and flanked by 4 nt overhangs complementary to the adjacent elements; and (iii) a 3′ element comprising the signals for correct termination and processing of the crRNA (GB1444 or similar). All level 0 elements are combined into a level 1 composite part through a cyclic restriction–ligation Golden Gate-like multipartite BsaI reaction, to build the entire crRNA TU. Finally, the assembled crRNA cassette is joined with pre-assembled As/LbCas12a TUs (GB1440, GB1441) on level >1 with a BsmBI-mediated restriction–ligation reaction, producing the final CGEM. The CGEM can be directly used for transient expression or attached to a selection marker (e.g. kanamycin resistance gene GB1181) for stable transformation experiments. All the necessary elements to perform plant targeted mutagenesis with Cas12a will be deposited in Addgene and in the GB3.0 repository.

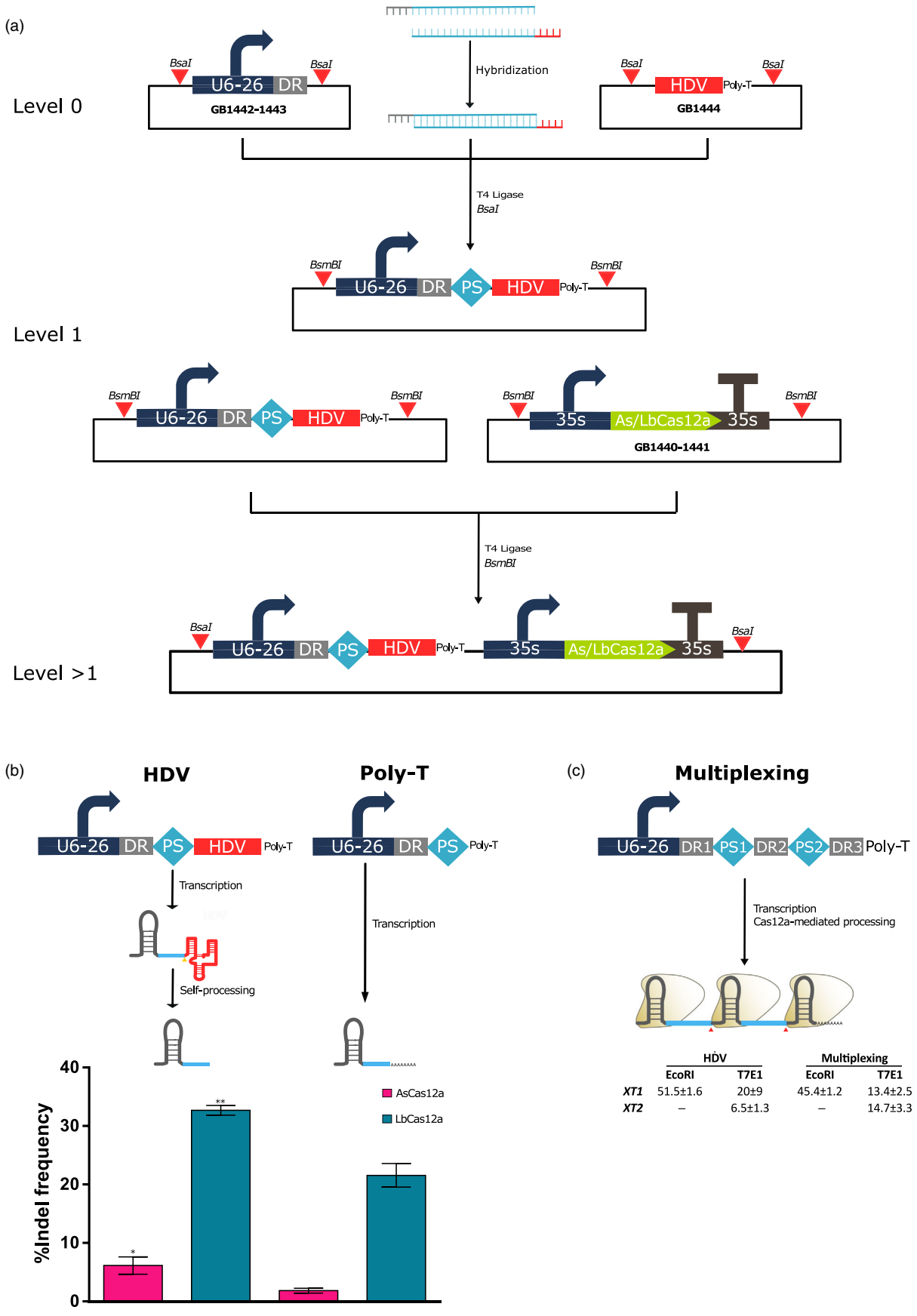
As a first optimization step, we analysed two design alternatives concerning 3′ RNA processing signals. Since superfluous nucleotides at the 3′ end of the crRNA derived from the poly-T U6 terminator could negatively affect the efficiency or the specificity of the Cas12a, we compared the efficiency of a self-processing hepatitis delta virus (HDV) ribozyme that removes the poly-A tail from the transcript leaving no additional nucleotides at the 3′ position of the crRNA with a classical construct carrying unprocessed U6 termination signal, which conserves a spurious tail of adenines at the 3′ position of the Cas12 crRNA (Figure 1b). The introduction of a self-cleavable ribozyme in the 5′ end of the transcript was unnecessary in our design, as we added a 5′-G to the Cas12a DR, which is compatible with the RNApol III promoter *At*U6-26 transcription start site. We evaluated the efficiency of both constructs on the *Nicotiana benthamiana XT1* target locus (*Niben101Scf04205g03008.1*) using a transient assay. Mutation at the *XT1* target site can be easily scored by the loss of an *Eco*RI site. Restriction enzyme (RE) analysis revealed that both AsCas12a and LbCas12a introduced mutations in *XT1* (Figure 1b); however, statistically significant differences were observed between the HDV and U6 terminator strategies for both nucleases, with higher activity observed for the HDV crRNA construct. For this target, LbCas12a was the most effective nuclease, exhibiting an activity sixfold (HDV set-up) and 11-fold (U6 terminator set-up) higher than AsCas12a.

Recently, a multiplexing strategy was described for Cas12a based on its ability to process multimeric crRNAs containing two or more PS sequences separated by DRs (Zetsche *et al.*, 2017). To test this function in plants, we designed a double-PS construct simultaneously targeting the *XT1* and *XT2* (*Niben101Scf04551g02001.1*) loci and tested its mutagenesis efficiency in *N. benthamiana* (Figure 1c). Because mutagenesis at the *XT2* target site does not produce a restriction enzyme polymorphism, editing efficiency for each target was estimated using the T7 endonuclease I (T7E1) mismatch cleavage assay. As shown, the self-processing strategy was successful in guiding Cas12a to both target loci and the efficiencies scored were similar with both strategies (HDV and multiplexing).

Transient expression of CGEM provides efficient targeted mutagenesis in *N. benthamiana* leaves

To assess to what extent the differences between AsCas12a and LbCas12a efficiency observed for the *XT1* locus could be generalized, similar comparisons were made at several positions of the *N. benthamiana* genome using the HDV strategy. Furthermore, SpCas9 constructs targeting the same locus were added for comparison. In total, we examined 8 chromosomal sites, four of them designed to have completely overlapping target sequences between As/LbCas12a and SpCas9 and the remaining four with a partially overlapping PS (Figure 2a).

Figure 1 Cas12a gene editing module (CGEM) assembly and strategy of the crRNA expression. (a) Level 1 cloning of the crRNA TU through BsaI restriction–ligation reaction requires three elements: *At*U6-26:DR (GB1442, GB1443) and partially complementary oligos of the PS and HDV termination signal (GB1444). BsmBI-mediated binary assembly into level >1 of the crRNA TU and the As/LbCas12a TU (GB1440, GB1441) to create the Cas12a gene editing module. (b) Schematic of the constructs used for the expression of a single crRNA and the final structure of the transcript. Efficiency comparison between HDV and U6 poly-T approach expressed as the percentage of undigested band after *Eco*RI digestion. Statistical *t*-test analysis showed difference for AsCas12a and LbCas12a (*P*-value <0.05). Error bars represent SEM; $n = 3$. (c) Representation of the construct used for the expression of multiple crRNAs in the same transcript, and table showing the comparison of the efficiencies between HDV and multiplexing.


(a)
Level 0
BsaI
U6-26
DR
BsaI
GB1442-1443
Hybridization
BsaI
HDV
Poly-T
BsaI
GB1444
T4 Ligase
BsaI
Level 1
BsmBI
U6-26
DR
PS
HDV
Poly-T
BsmBI
BsmBI
35s
As/LbCas12a
35s
BsmBI
GB1440-1441
T4 Ligase
BsmBI
Level >1
BsaI
U6-26
DR
PS
HDV
Poly-T
35s
As/LbCas12a
35s
BsaI
(b)
HDV
Poly-T
U6-26
DR
PS
HDV
Poly-T
U6-26
DR
PS
Poly-T
Transcription
Self-processing
Transcription
AsCas12a
LbCas12a
%Indel frequency
40
30
20
10
0
(c)
Multiplexing
U6-26
DR1
PS1
DR2
PS2
DR3
Poly-T
Transcription
Cas12a-mediated processing
HDV
Multiplexing
EcoRI
T7E1
EcoRI
T7E1
XT1
51.5±1.6
20±9
45.4±1.2
13.4±2.5
XT2
–
6.5±1.3
–
14.7±3.3

Mutagenesis efficiency for each RGEN-locus pair was scored in transient transformation assays in *N. benthamiana* using the T7EI mismatch cleavage assay. Each experimental point comprised three biological replicates for which three different leaves were infiltrated and pooled (see Figure S3a). Results plotted in Figure 2a reveal strong differences in the mutagenesis efficiency of RGENs depending on the target, ranging from undetectable levels to a maximum above 30% efficiency for LbCas12a at the *TFL1 14.1* locus. The Cas12a editing results obtained with T7EI assay for *XT1* locus were compatible with those obtained previously with the RE assay, although the T7E1 assay is less sensitive than restriction polymorphisms. Only one of the assayed targets (*XT2A*) was mutated by all three RGENs at similar levels (9%). On the contrary, the *XT2B* locus was only efficiently mutagenized by SpCas9. For the targets assayed, LbCas12a was the most reliable RGEN (Figure 2b), causing detectable indels in 7 out of 8 targets and displaying the higher mutagenesis average, as compared to SpCas9 and AsCas12a (SpCas9 = 5 ± 2%; AsCas12a = 3 ± 2%; LbCas12a = 17 ± 6%).

Modifications in the crRNA DR loop affect RGEN activity but cannot compensate for low-efficiency editing

The crRNA structure includes a DR region whose proper folding is important for nuclease activity (Kim *et al.*, 2017b; Lee *et al.*, 2018). The DR region of AsCas12a and LbCas12a is identical except in the distal loop, which contains a 5′-UCUU-3′ tetranucleotide in AsDR and a 5′-UAAGU-3′ pentanucleotide in the case of LbDR (Figure 3a). Cas12a variants are known to accept crRNAs from other species (Zetsche *et al.*, 2015) demonstrating a certain flexibility in the DR loop. A possible factor influencing Cas12a efficiency is the stability of the DR region. The stability can be affected by the presence of additional complementary regions between the DR and the PS. Therefore, we decided to test the possibility of reducing the size of the DR loop, thus minimizing interactions with PS.

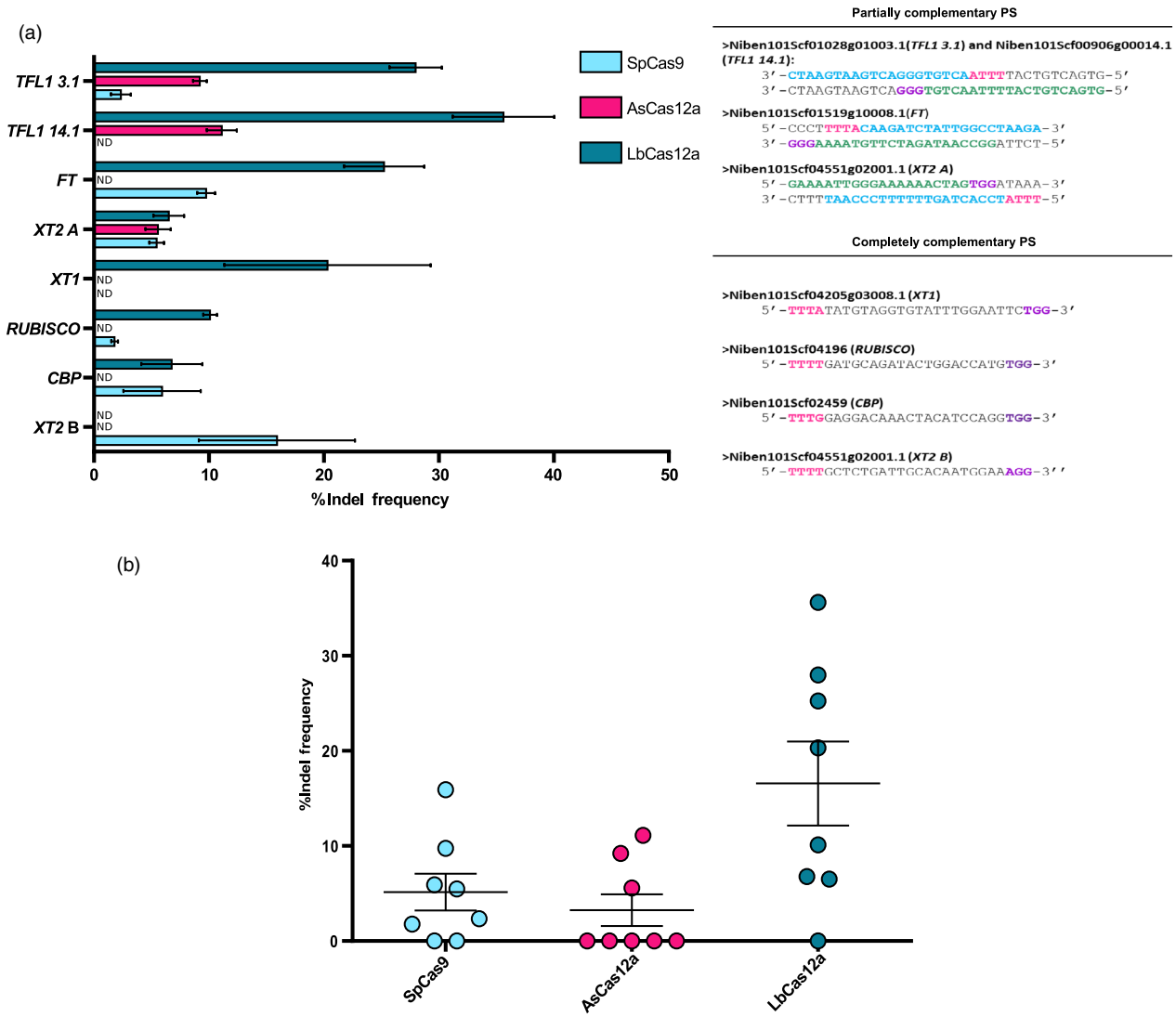


Figure 2 Comparison of the RGEN at several loci of *N. benthamiana*. (a) Mutation frequency for individual loci targeted with the three RGENs and determined by T7E1 mismatch cleavage assay. The sequence for each locus is detailed on the right part of the figure, with the Cas12a PAM highlighted in pink, the PS of Cas12a in light blue, the Cas9 PAM in purple and the PS in dark green. Error bars indicate SEM; $n = 3$. ND: not detected. (b) Summary of the mutated and nonmutated loci for each endonuclease. Mean indel frequencies ± SEM are shown.

Several simplified synthetic DR loops (synDRs) were designed with different abilities to disrupt/conserve the DR integrity, as assessed by the RNA secondary structure prediction software ViennaRNA (Lorenz *et al.*, 2011) and MFold (Zuker, 2003) that calculate the stability of DR loop-DR/PS interactions based on the predicted –ΔG (Table S4). As shown in Figure 3b, all crRNA structures (wtDRs and synDRs) exhibited detectable mutagenesis levels at *TFL1 3.1* and *TFL1 14.1* targets, whereas only LbCas12a was active at the *FT* locus. Notably, the AsCas12a-synDR 5′-UGU-3′ combination outperformed native AsCas12a-AsDR at the *TFL1 3.1* locus, reaching levels similar to those obtained with LbCas12a in its native combination. However, none of the synDR loops assayed was capable of increasing AsCas12a activity at the *FT* target to detectable levels. In general, LbCas12a performed better with AsDRs and LbDRs than with synDRs containing shorter loops, whereas AsCas12a activity was unaffected or slightly improved by trinucleotide loops.

Cas12a as a genome editing tool in *N. benthamiana*, tomato and Arabidopsis

As the next step in the evaluation of the mutagenesis activity in transient assays, we assessed the ability of GB3.0 Cas12a constructs to perform genome editing using standard stable transformation protocols in three different plant species. To this end, we used a generic gene editing module represented in Figure 4a, containing three TUs: the kanamycin selection marker, the crRNA expression cassette and the TU encoding SpCas9, AsCas12a or LbCas12a nucleases. Stable transformation of *N. benthamiana* was first used to confirm the results obtained in leaf transient experiments targeting the *XT1* locus using AsCas12a and LbCas12a (Figure 4b). Genotyping of the T0 transgenic generation revealed that 54% of the 37 kanamycin-resistant plants carrying the LbCas12a construct had mutations at the target site (Figure S1a). In contrast, no edited plants were found for the AsCas12a construct (see Figure S1b). TIDE analysis of the T0 edited lines revealed that the majority of the mutated lines were biallelic (60%); line 22 even showed 4 different mutations, consistent with an early chimera derived from two differentially edited progenitor cells (Figure 4b).

AsCas12a, LbCas12a and SpCas9 were also assayed in tomato plants in a parallel experiment where all three editing constructs were designed to target the same locus *Solyc01g079620* (*MYB12*). Figure 4c shows the precise targeting sites for each RGEN and the TIDE analysis for the mutated lines. Only one chimeric plant was recovered from the AsCas12a experiment, showing a 9-bp deletion present in approximately 10% of the copies of the genomic DNA samples analysed. In contrast, two edited plants were recovered for both LbCas12a and SpCas9 nucleases. Both SpCas9-edited plants were biallelic, while only one of the LbCas12-mutated plants had mutations in the two alleles.

Finally, we assayed LbCas12a activity in Arabidopsis. In this case, we also decided to determine the ability of LbCas12a to

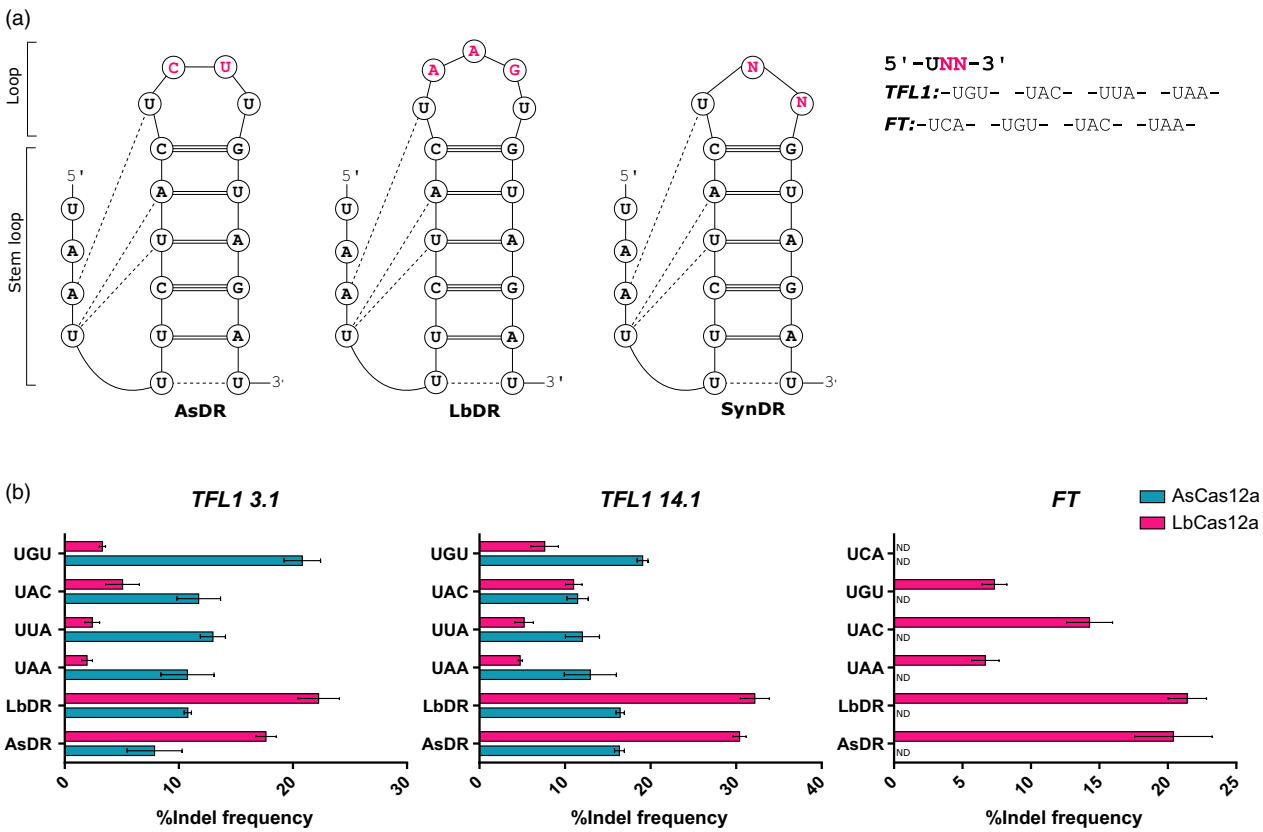
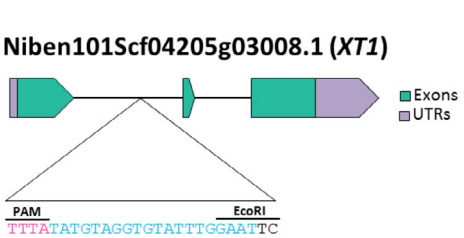


Figure 3 The DR loop sequence influences Cas12a activity. (a) Structure of the AsCas12a (AsDR), LbCas12a (LbDR) and synthetically engineered DR (synDR). Highlighted in pink are the different nucleotides of the three structures. The varying nucleotides of the synDR loops are also indicated (5′-UNN-3′) for each target. (b) Mutagenesis efficiency results of the wtDRs and synDRs at *TFL1* and *FT* targets. All the DR structures were evaluated for each endonuclease. Error bars represent SEM; $n = 3$. ND: not detected.

(a)

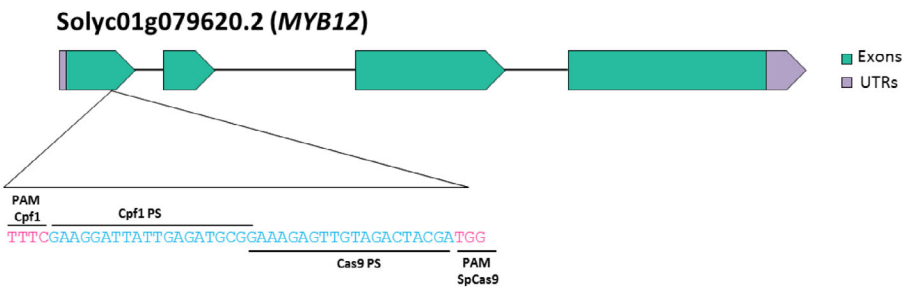
(b)



N. benthamiana

Line T0	Overall efficiency	Mutation efficiency
LbCas12a 2	98.1	–16 (57.9), –10 (40.2)
LbCas12a 4	94.4	–7 (49.5), –1 (42.4)
LbCas12a 6	60.4	–9 (56.7)
LbCas12a 7	94.1	–40 (56.7), –7 (36.6)
LbCas12a 10	54.7	–7 (54.2)
LbCas12a 13	99.4	–10 (98)
LbCas12a 14	97.3	–9 (45.9), –6 (50.5)
LbCas12a 17	94.8	–8 (42.7), –2 (50.1)
LbCas12a 18	51.2	–8 (50.5)
LbCas12a 20*	–	–83, –14, –7
LbCas12a 21	97.6	–10 (50.7), –8 (46.1)
LbCas12a 22	93.8	–12 (9.9), –9 (15.1), –8 (17.7), –6 (50.1)
LbCas12a 31	93.6	–17 (51.5), –7 (41.3)
LbCas12a 33	95.3	–5 (37.9), –30 (56.6)
LbCas12a 34*	–	–83, -14, –9, –7
LbCas12a 35	93.7	–10 (48.1), –8 (42.7)
LbCas12a 37	66.1	–7 (63.2)

(c)



S. lycopersicum

Line T0	Overall efficiency	Mutation efficiency
AsCas12a 2	8.8	–9 (7.1)
LbCas12a 3	82.7	–6 (38.3), –2 (43.5)
LbCas12a 5A	49.1	–2 (48.8)
SpCas9 6	95.5	–2 (49.6), –1 (45.7)
SpCas9 8	96.4	–3 (50.1), +1 (46)

Figure 4 Mutagenesis data extracted from stable transformations of different model plants. (a) Generic GEM encompasses three TUs: kanamycin selection marker (*NptII*), the crRNA expression cassette and the RGEN TU. (b) Representation of the *XT1* loci of *N. benthamiana* with the targeted sequence, and table showing TIDE data from the mutated lines. Marked lines (*) were not analysed with TIDE. (c) Genomic loci of the *S. lycopersicum MYB12* targeted gene, with the three RGENs, and results of the TIDE analysis of the edited plants. The 'overall efficiency' represents the sum of all the individual mutations traced by TIDE. The 'mutation efficiency' specifies the type of mutations associated with its efficiency in brackets. Only mutations whose *P*-value <0.001 have been considered in 'mutation efficiency'.

produce large deletions in the genome. To this end, we designed a CGEM with the LbCas12a under the transcriptional control of the *A. thaliana* Ubiquitin10 (UBQ10) promoter and two crRNAs directed towards two target sites flanking the Arabidopsis *PDS3* locus (see Figure 5a). Indels flanking noncoding regions are unlikely to produce loss-of-function mutations. Therefore, with this design, only editing events producing large deletions in both *PDS3* alleles are expected to generate the albino phenotype. The CGEM construct also contained a DsRED marker gene under the control of a seed-specific promoter to track the presence (DsRED (+)) or absence (DsRED(−)) of the transgene. Despite the strict constraints imposed, we could observe white somatic spots, consistent with biallelic loss-of-function chimeras, in 5 out of 50 T1 stably transformed plants (scored as DsRED(+);Figure S2b). The analysis of white somatic spots confirmed that they corresponded to different large deletions in the *PDS3* locus (Figure 5b). Three of the identified chimeric lines (PDS-1, PDS-2 and PDS-3) were grown to the T2 generation and, as expected, some of them (lines PDS-1 and PDS-3) produced offspring showing a completely albino phenotype, consistent with *PDS3* loss-of-function (Figure 5c). The albino phenotype was present in both DsRED(+) and DsRED(−) offspring (Figure 5c), indicating that, at this point, *PDS3* mutations segregated independently of the CGEM T-DNA. The presence of loss-of-function deletions in T2 DsRED(−) seedlings, thus regarded as germ-line-associated, was confirmed by PCR (Figure 5c). Interestingly, an allele with a full deletion of the *PDS3* gene (indicative of concomitant activity in targets 1 and 2) was detected in T2 pools derived from white individuals from line PDS-1, in addition to a partial deletion of the locus. Partial deletions with some additional rearrangements were also detected in pools derived from the T2 generation of lines PDS-2 and PDS-3 (PCR B; Figure 5c). The structure of deletion-containing PCR bands was analysed by Sanger sequencing and resolved (Figure S2c), revealing a number of rearrangements as a result of Cas12a mutagenesis, predominantly in target 1. In addition to large deletions, small indels ranging from 3 to 34 nt were also identified as expected in green seedling pools using TIDE analysis (Figure 5d).

The presence of deletion-containing alleles in DsRED(−) seedlings was clearly indicative of the presence of Cas12a mutations in the germ-line. An additional confirmation for this germ-line association was obtained with a segregation analysis of the T3 seeds from the PDS-1-118 T2 DsRED(−) line. Attending to the number of plants with the recessive albino phenotype, the expected 3 : 1 segregation hypothesis was initially discarded by X^2 statistic (Figure S2e). However, the PCR analysis of green T3 seedlings (heterozygous and wt) in the PDS-1-118 line revealed the presence of the loss-of-function deletion in a 1 : 2 proportion, a segregation which is compatible with germ-line inheritance (Figure S2e). Combined, these results indicate that the observed *PDS3* deletions are germ-line-associated and that the segregation of albino plants is probably distorted by reduced fertility and viability of the homozygous *PDS3* loss-of-function mutation, as suggested previously (Qin *et al.*, 2007).

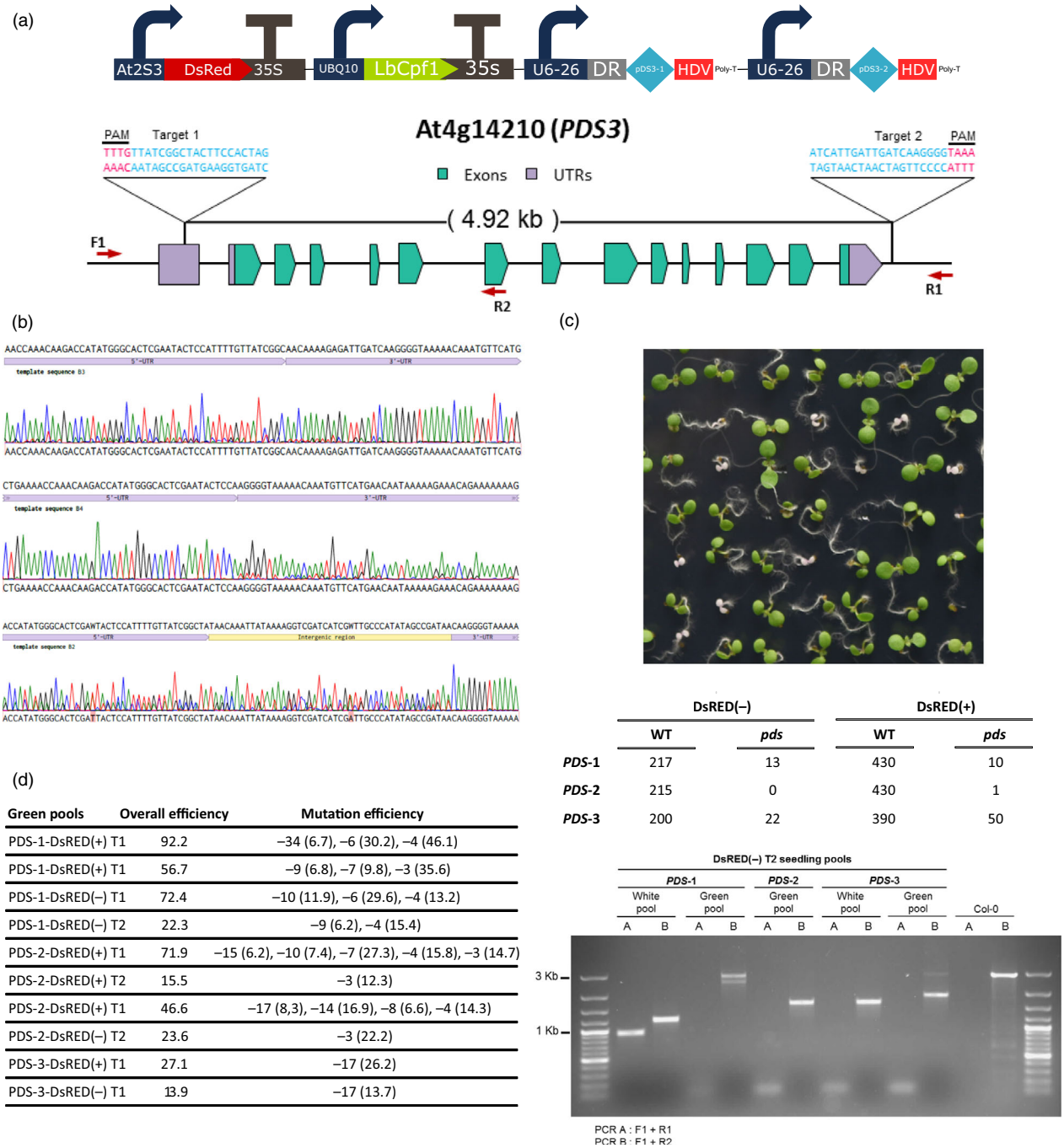
A meta-analysis of Cas12a mutagenesis profile shows a high frequency of small- to medium-size deletions

To get insight into the compared signatures of Cas12a and Cas9, we performed a meta-analysis collecting all TIDE mutagenesis data produced for *N. benthamiana* (stable and transient experiments), tomato and Arabidopsis (stable experiments) (Figure 6). A total of 272 mutagenesis events were compiled: 137 for Cas12a and 135 for Cas9. To represent the mutation landscape, we assigned an arbitrary value of one to each mutation event detected in a TIDE experiment. Then, we summed all the events with a certain indel size and divided it by the number of total events detected in order to calculate the mutation frequency of each indel size. The results are plotted in Figure 6. As observed, the Cas12a deletion profile was different from that of Cas9, showing a tendency to produce larger deletions. In contrast, Cas9 displayed small insertions (+1, +2 and +4 bp) and small deletions more frequently. Summarizing, our meta-analysis illustrates that Cas12a produces a deletion-enriched mutagenesis profile in comparison with that produced by SpCas9.

Whole-genome sequencing of LbCas12a-mutated *A. thaliana* plants revealed no appreciable off-target effects

An important feature for editing nucleases is the extent of off-target mutagenesis. Usually, this parameter is evaluated by examining candidate off-target loci by PCR. However, recent studies have reported large DNA rearrangements near the target loci, which remain undetected with traditional target-specific PCR analysis (Adikusuma *et al.*, 2018; Kosicki *et al.*, 2018). Also, indiscriminate single-strand DNA (ssDNA) DNase activity of Cas12a upon activation by target-specific cleavage has been observed *in vitro* (Chen *et al.*, 2018; Li *et al.*, 2018). These observations have led to the suggestion that increased mutagenesis rates could occur elsewhere in the genome, particularly in areas with transient ssDNA formation, such as active DNA replication and transcriptional sites. To carefully evaluate the potential off-target effects of LbCas12a, we performed whole-genome sequencing (WGS) of eight selected Cas12a-edited Arabidopsis lines.

Four green individuals from the PDS-3 line (PDS-3-111, 113, 114 and 115) and four of PDS-1 (PDS-1-117, 118, 119 and 123) DsRED(−) T2 plants bearing different mutations at target site 1, including large deletions in heterozygosis, were each resequenced at 20X coverage. It is important to note that these plants had segregated Cas12a, and therefore, the observed mutations are necessarily germ-line-associated. The genotype of the *PDS3* locus for each individual is highlighted in Figure S2d. As a reference, two DsRED(−) T1 plants (WT-104 106) and a pool of five DsRED(−) T1 plants (including WT-104 and WT-106), in which Cas12a was never present, were also included for resequencing; the pool was sequenced at a deeper coverage (60X) to assess the population variability (named as POOL).



(d)

Green pools	Overall efficiency	Mutation efficiency
PDS-1-DsRED(+) T1	92.2	–34 (6.7), –6 (30.2), –4 (46.1)
PDS-1-DsRED(+) T1	56.7	–9 (6.8), –7 (9.8), –3 (35.6)
PDS-1-DsRED(–) T1	72.4	–10 (11.9), –6 (29.6), –4 (13.2)
PDS-1-DsRED(–) T2	22.3	–9 (6.2), –4 (15.4)
PDS-2-DsRED(+) T1	71.9	–15 (6.2), –10 (7.4), –7 (27.3), –4 (15.8), –3 (14.7)
PDS-2-DsRED(+) T2	15.5	–3 (12.3)
PDS-2-DsRED(+) T1	46.6	–17 (8,3), –14 (16.9), –8 (6.6), –4 (14.3)
PDS-2-DsRED(–) T2	23.6	–3 (22.2)
PDS-3-DsRED(+) T1	27.1	–17 (26.2)
PDS-3-DsRED(–) T1	13.9	–17 (13.7)

Figure 5 LbCas12a is capable of inducing large deletions in *A. thaliana*. (a) CGEM used for *A. thaliana* stable transformation and locus chosen to study the ability of LbCas12a to produce large deletions, showing the two target sites on the UTRs flanking the *PDS3* gene. Arrows indicate primers used for the amplifications. (b) Alignments of three different complete deletions found in the white spots of the T1 generation visualized with Benchling. The complete genes have been deleted and the UTRs joined. (c) Analysis of albino phenotype (*pds* mutants, white seedlings) segregation in T2 seedlings grown in MS plates. Segregation analysis of DsRED(+) and DsRED(–) seeds from the three lines (PDS-1, PDS-2 and PDS-3) and PCR amplification of white or green DsRED (–) seedling pools from the three different lines. (d) TIDE mutation analysis of DsRED(–) and DsRED(+) from pools of green plants from PDS-1, PDS-2 and PDS-3 lines. Both target 1 (T1) and target 2 (T2) were analysed. 'Overall efficiency' and 'mutation efficiency' mean the same as in Figure 4.

Once all on-target mutations present in the sequenced lines were confirmed by WGS, we proceed with the off-target analysis. The first off-target study consisted in the analysis of homologous sites. Using Cas-OFFinder software (Bae *et al.*, 2014b) set for a maximum of four mismatches, three putative off-target sites for target 1 were found. Software tools failed to detect SNPs, indels and structural variations (SVs) at these positions in any of the analysed plants (Table 1a). Next, we searched for unspecific enrichment of mutation rates in the Cas12a-'treated' plants as compared to their relatives not subjected to Cas12a activity. Table 1b shows the results obtained with this approach. The columns SNPs, indels and SVs show the total number of

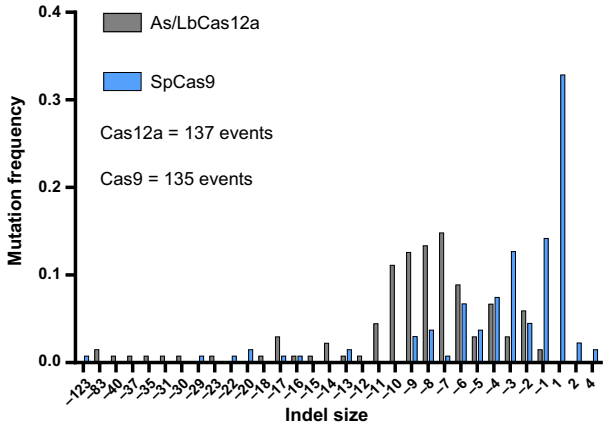


Figure 6 Meta-analysis of *N. benthamiana, S. lycopersicum* and *A. thaliana* mutation events. Compilation of all the TIDE data produced in the laboratory for Cas12a and Cas9 from transient and stable expression experiments of three plant species (*N. benthamiana, S. lycopersicum* and *A. thaliana*) to compare mutagenesis profile of both endonucleases.

Table 1 Off-target effects of LbCas12a in DsRED(−) T2 PDS-1 and PDS-3 *A. thaliana* mutants. (a) Off-target effects found at three different off-target sites predicted by Cas-OFFinder software. The PAM sequence is underlined, and the mismatches are shown in lowercase. (b) Variations found in the WGS analysis. The numbers outside the parentheses indicate the variations found when comparing with the reference genome, and those inside the parentheses indicate events obtained comparing each edited plant vs WT samples

(a)

Off-target sequence	SNPs	Indels	SVs
TTTCaTATgaGCTACTTtCACTAG	0	0	0
TTTATTATgaGCTACTTCCACaAG	0	0	0
TTTATTtTtGGCTACTcCCAtTAG	0	0	0

(b)

DsRED(−) plant	SNPs	Indels	SVs
PDS-3-111	1577 (24)	460 (30)	5 (3)
PDS-3-113	1535 (25)	474 (35)	6 (4)
PDS-3-114	1549 (24)	469 (36)	3 (1)
PDS-3-115	1566 (23)	471 (31)	2 (0)
PDS-1-117	1558 (8)	442 (7)	2 (1)
PDS-1-118	1569 (9)	441 (5)	8 (8)
PDS-1-119	1575 (4)	450 (23)	3 (1)
PDS-1-123	1579 (11)	452 (13)	3 (1)
WT-104	1533	437	2
WT-106	1565	423	2

SNPs were analysed with GATK-HC, VARSCAN2 and FREEBAYES; indels with GATK-HC, VARSCAN2 and PINDEL; and structural variations with (SVs) with Delly.

consensus variations found in each T2 plant compared with the reference genome (as detected by all programs used in the analysis; Figure S4). We also traced possible background mutations of PDS plants comparing them against WT and POOL samples. The data obtained discard massive increases in 'indiscriminate' mutagenesis possibly associated with Cas12a activity. Note, for instance, that in line PDS-1-117, only 16 of all detected variations could not be traced back to the genomes of its Cas12a-free T1 relatives (8 out of 1558 SNPs, 7 out of 442 indels and 1 out of 2 SVs). This low number of putative 'new' mutations is fully compatible with development-associated spontaneous events.

Interestingly, the WGS also revealed multiple structural variations surrounding the target site in the PDS-1-118 line, which would have remained undetected by classical PCR analysis. The reads observed in this line can be explained as the result of a duplication of approximately 13 kb and two large deletions of 2 and 4 kb (Figure S5). This event was not found in sibling lines, clearly suggesting an association with the Cas12a signature.

Discussion

Plant genome engineering facilitated by RGEN is expected to play an important role in the adaptation of agriculture to 21st-century challenges. After the first wave of tools based on Cas9, which was shown capable of producing edited mutants in virtually every species tested, a second generation of refined tools is being progressively incorporated into the genome editing toolbox. These new tools are emerging in parallel with a number of modular cloning methods that facilitate the customization of often multiplexed editing constructs (Engler *et al.*, 2009; Peccoud *et al.*, 2011; Sarrion-Perdigones *et al.*, 2013; Vazquez-Vilar *et al.*, 2017). Thus, following the same design principles adopted earlier to integrate CRISPR/Cas9 into GB modular cloning (Vazquez-Vilar *et al.*, 2016), we adapted Cas12a tools to the so-called phytobrick standard, adopting both crRNA and Cas12a TUs as level 1 structures to maximize the exchangeability while preserving the combinatorial potential.

Mutagenesis efficiency was evaluated using transient *N. benthamiana* assays, based on our previously described detailed standard procedures for the analysis of Cas9-based mutagenesis (see Vazquez-Vilar *et al.*, 2016). We chose the T7E1 assay for the estimation of the mutagenesis efficiency since it is a cost-effective method. It should be noted that this procedure has a lower sensitivity than RE polymorphism analysis; therefore, actual efficiencies below a certain threshold (approximately 5%) should not be discarded for some targets. Also, it has been previously described that T7E1 can underestimate mutagenesis rates in highly efficient targets or in DNA populations showing low mutation variability (e.g. homozygous mutations). However, our analyses are unlikely to be affected by this T7E1 bias as they were performed on complex cell populations with maximum estimated efficiency rates of 30%, a range of activities in which the T7E1 assay can be considered a reliable estimation method (Sentmanat *et al.*, 2018). Our results, far from showing a clear front-runner nuclease, indicated a strong target-dependent efficiency. Interestingly, all loci analysed were efficiently mutated with at least one RGEN indicating general accessibility of the selected loci. On the contrary, all three RGENs produced undetectable mutagenesis levels in at least one selected locus. In general, LbCas12a and SpCas9 showed more reliable results than AsCas12a, producing detectable mutation rates in 7 and 6 of the 8 assayed targets, respectively, while AsCas12a produced detectable mutation rates only in three targets. The average and maximum mutagenesis rates were displayed by LbCas12a, indicating that this enzyme can be a fair competitor or even outperform SpCas9 for certain applications. Our findings are consistent with previous

publications in mammalian cells reporting high variability among the evaluated targets, with SpCas9 as the most robust of the three RGENs in Kim *et al.* (2016) and LbCas12a in Kleinstiver *et al.* (2015). In plants, Endo *et al.* (2016) targeted four different positions of the *NtSTF1* gene with FnCas12a, but only one crRNA produced mutations.

In our experiments, SpCas9 showed lower dispersion in the scored activities, which grouped between 5% and 20%, whereas LbCas12a showed more variable results. This observation might reflect the larger size of the SpCas9-sgRNA (~100 bp) as compared to AsCas12a and LbCas12a crRNAs (~40 bp). This larger size might 'protect' the SpCas9-sgRNA against disruption of the secondary structure towards unproductive/unstable conformations that prevent mutagenesis (Xie *et al.*, 2014). In contrast, the smaller size of the Cas12a crRNAs might facilitate the adoption of unproductive/unstable conformations, explaining the broader range of editing efficiencies. We explored the possibility of creating improved structures by engineering DRs with a simplified loop of three nucleotides, trying to minimize the interactions between the DR and the PS that could impair crRNA function. Although certain improvements in mutagenesis activities were obtained with synthetic DR loops when combined with AsCas12a, in general the strategy failed to obtain higher efficiencies for LbCas12a or to 'turn on' inactive combinations as exemplified with *FT* and AsCas12a. Despite this, our data serve to demonstrate that engineering of DR loop can be used to modulate Cas12a efficiency, producing a range of activities that can be used in transcriptional regulation approaches, for example.

Although efficiency studies using transient assays are highly informative, they need to be confirmed by generating stably edited plants. We first assayed *N. benthamiana* as a way to corroborate the editing efficiencies observed in transient assays. LbCas12a efficiency for the *XT1* locus was surprisingly high compared with a previous report in a solanaceous plant (tobacco) using FnCas12a, where no biallelic mutants were recovered (Endo *et al.*, 2016). In contrast, we could not recover mutant lines from AsCas12a transformants. LbCas12a has also been reported to produce high mutation rates with a high ratio of biallelic plants in rice stable transformation (Endo *et al.*, 2016; Tang *et al.*, 2017; Wang *et al.*, 2017; Xu *et al.*, 2017; Yin *et al.*, 2017b). Together, there are now enough data to conclude that the robustness of LbCas12a is a general feature in plants. Regarding the comparison between LbCas12a and SpCas9 for stable genome editing, recent experiments in maize showed higher editing rates with SpCas9 as compared to LbCas12a targeted to the *glossy2* locus (Lee *et al.*, 2018), in line with the data obtained in our tomato experiment. However, these observations reflect only the results obtained with a very limited number of genome targets, and therefore, it would be premature to draw conclusions from these findings, as our transient expression data show a strong locus-dependent effect on RGENs activities, with LbCas12a producing higher mutagenesis rates in the majority of the loci assayed. It is also possible that *in vitro* culture conditions during stable transformation could account for some of the differences observed. Recently, a strong temperature dependency for AsCas12a has been reported, showing a severe decrease in activity below 37°C (Moreno-Mateos *et al.*, 2017). This effect could explain in part the lower performance of AsCas12a in stable transformation, as the transformation/regeneration processes take place at 22–25 °C. This would discard AsCas12a in practice for plant editing procedures involving *in vitro* culture, unless transformation conditions are adapted to maximize editing (rather than transformation) efficiency.

The meta-analysis plot shows that SpCas9 tends to produce small indels of few nucleotides in contrast to the Cas12a enzymes tested, which induce a broader range of deletions but no insertions. Deletions are often associated with microhomology-mediated end joining (MMEJ) repair (Bae *et al.*, 2014a), which depends on the specific target sequence that marks the formation of the microhomologies. It should be taken into account that the TIDE data used in this meta-analysis derive from a heterogenic collection of *N. benthamiana*, *S. lycopersicum* and *A. thaliana* targets. For this reason, biases due to the mutation signature of certain targets (determined by the local microhomologies) or by the organism considered can be discarded. In addition, the reported profiles are consistent with the literature, as other authors have shown a prevalence of deletions with Cas12a in comparison with SpCas9 in mammalian cells (Kim *et al.*, 2016) and plants (Endo *et al.*, 2016). The distinctive mutation signature of Cas12a has interesting functional implications, considering, for example, that larger deletions are more prone to generate loss-of-function mutants and remove regulatory operators in promoter regions, two features that can be exploited in breeding practices (Biłas *et al.*, 2016). Furthermore, whereas DSB in Cas9 takes place at a position proximal to the PS sequence, Cas12a cleaves at a distal position, thus allowing target conservation after cleavage, which allegedly promotes larger deletions by MMEJ or gene insertions via homology direct repeat (HDR; Moreno-Mateos *et al.*, 2017; Tóth *et al.*, 2016). Whether the Cas12a signature will enhance gene targeting (GT) efficiency in plants is a possibility that remains to be tested, either alone or in combination with other GT-enhancing methods, such as geminivirus replicons (Cermak *et al.*, 2015).

Finally, we searched for putative off-target effects not only in related homology sites, but also elsewhere in the genome using WGS analysis. We could not detect homologous off-targets in any of the plants under analysis. Moreover, we do not find evidences in any of the analysed lines of enhanced genome-wide mutagenesis to occur as a result of Cas12a activation as suggested previously (Chen *et al.*, 2018; Li *et al.*, 2018). Just the opposite, WGS data showed that all analysed lines had variations with the reference genome and with their own relatives in the range of those observed in Cas12a-free plants. Therefore, we can conclude that Cas12a can generate transgene-free edited plants (e.g. plant PDS-1-117) whose levels of 'new' mutations not present in the background are negligible and indistinguishable from spontaneous mutations caused during development. A similar conclusion was obtained recently in rice (Tang *et al.*, 2018). Nevertheless, we did observe a relatively large distortion of the alignment of Illumina reads compatible with a duplication event in the vicinity of the target site in at least one plant (PDS-1-118). This structural change would have remained undetected using only PCR-based off-target detection methods. The same line showed an increased number of SVs. These types of events, which are not unusual in traditional plant breeding practices, do not necessarily comprise breeding value to the resulting variety. However, given the strict scrutiny to which genome editing procedures are subjected, we conclude that resequencing of 'elite' lines produced with Cas12a is advisable to discard lines in which rearrangements could occur.

Taken together, our data show that LbCas12a is an effective RNA-guided endonuclease in a broad number of plant species, it is amenable for modular cloning and multiplexing, and it shows efficiencies comparable with classical SpCas9 with similarly low off-target effects and a characteristic tendency to produce larger deletions.

Experimental procedures

GBparts construction

GBparts employed in this study were created through the domestication strategy described in Sarrion-Perdigones *et al.* (2011). Plasmids pY010 (pcDNA3.1-hAsCpf1, Addgene plasmid # 69982) and pY016 (pcDNA3.1-hLbCpf1, Addgene plasmid # 69988), kindly provided by Feng Zhang Laboratory, served as a template for the construction of GB1438 and GB1439 by PCR amplification, using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific). GB1442 and GB1443 were also obtained by PCR amplification of GB1001. An amount of 40 ng of PCR products was subsequently cloned into pUPD2 plasmid to create the above-mentioned GBparts through a BsmBI restriction–ligation reaction. Partially complementary ultramers with sticky ends were used to domesticate GB1444. Separate ultramers were resuspended to a final concentration of 1 μM; then, 5 μL of each was mixed and incubated for 30 min at room temperature to facilitate the hybridization. Finally, 1 μL of this mix was used to set up the domestication reaction following the same protocol as for the PCR products. Once cloned, GBparts were verified by restriction enzyme (RE) analysis and confirmed by sequencing. All the GBparts are listed in Table S1.

Guide RNA assembly on level 1

The design of the guide RNA and the assembly of the expression cassettes on level 1 were performed as described in Vazquez-Vilar *et al.* (2016). On the particular case of Cas12a, the BsaI-mediated restriction–ligation reaction included a complementary pair of oligos of the target (Table S2) flanked by 4 nucleotide overhangs complementary to GB1442 and GB1443. Guide RNA constructs were confirmed by RE analysis and subsequent sequencing. The produced constructs are detailed in Table S1.

Cloning in α- and Ω-level destination vectors

Level 1 assemblies were performed through Golden Gate-like multipartite BsaI restriction–ligation reactions to obtain TUs from basic domesticated level 0 parts. Similarly, several TUs were combined on level >1 with bipartite BsmBI- or BsaI-mediated reactions to create modules. These assembly protocols are detailed in Vazquez-Vilar *et al.* (2017). The generated constructs are included in Table S1. The sequences are available at http://gbcloning.upv.es/search/feature with the GB database ID.

Nicotiana benthamiana transient assays

Transient expression experiments were performed to test the mutagenesis efficiency at several loci as described in Vazquez-Vilar *et al.* (2017, 2016). Each locus was assessed separately for the three RGENs (SpCas9, AsCas12a and LbCas12a). To this end, we mixed equal volumes of *Agrobacterium* cultures of the P19 suppressor of silencing, the RGEN and the crRNA/sgRNA. These cultures were first grown from glycerol stock for 2 days to saturation; then, 10 μL was subcultivated for 16 h. Next, the cultures were pelleted, resuspended in agroinfiltration buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$ and 200 μM acetosyringone) and adjusted to an optical density of 0.2 at 600 nm [estimated 8 active T-DNA copies per cell (Vazquez-Vilar *et al.*, 2017)]. Finally, the three cultures (P19, RGEN and crRNA/sgRNA) were mixed at equal parts to prepare the agroinfiltration mixture. Three independent samples (plants 4–5 weeks old, grown in a stable condition of 24 °C (light)/20 °C (darkness) with a 16-h light/8-h dark photoperiod) were infiltrated to assess the effect of each RGEN at each locus. Three consecutive leaves (second to fifth) were infiltrated in each plant. Five days postinfiltration, one sample per infiltrated plant was collected. Each sample consisted in six pooled leaf discs, two per infiltrated leaf, collected with a 0.5-cm cork-borer (approximately 150 mg of tissue). Immediately, the samples were frozen in liquid nitrogen. Control plants were infiltrated using the same mixture as the targeted locus but with an unrelated crRNA. Samples were ground with a Retsch Mixer Mill MM400 for 1 min at 30 Hz and stored at −80 °C for subsequent genomic DNA extraction (gDNA).

Nicotiana benthamiana stable transformation

Nicotiana benthamiana stable transformation was performed following a modification of transformation and *in vitro* regeneration of the leaf-disc method (Horsch *et al.*, 1985; McCormick *et al.*, 1986). All *in vitro* steps were carried out in a long-day growth chamber (16-h light/8-h dark, 24°C, 60%–70% humidity, 250 μmol/m^2/s). Samples for genotyping were collected once the plants were sufficiently developed to harvest 150 mg of tissue, which was frozen in liquid nitrogen and stored at −80 °C until extraction of gDNA.

Solanum lycopersicum stable transformation

Tomato transformation was performed following an adapted protocol of Ellul *et al.* (2003). Growth conditions and sampling were conducted as described for *N. benthamiana* transformation above.

Arabidopsis thaliana stable transformation

Wild-type Col-0 plants were transformed by the floral-dip method (Clough and Bent, 1998), with a minor modification: 1 min dipping into a solution (sucrose 5% + 0.2 mL Silwet-77/L) containing *Agrobacterium tumefaciens*. Fluorescent seeds, containing the transgene, were identified in a Leica DMS1000 microscope.

Mutagenesis detection of on-target sites

Genomic DNA was extracted from transient and stable expression experiment samples following the CTAB protocol (Murray and Thompson, 1980). The obtained gDNA was used for PCR amplification of the selected locus using MyTaq™ DNA Polymerase (Bioline) and a pair of primers flanking the targeted sites (Table S2). PCR product was confirmed by 1% gel electrophoresis and purified employing QIAquick PCR Purification Kit (QIAGEN), following the manufacturer's protocol. The purified PCR product was utilized to detect mutagenesis by three well-established methods: T7E1 assay, RE analysis and TIDE. For the T7E1 assay (New England Biolabs), 250 ng of the PCR product was submitted to a denaturation–reannealing process using the thermocycler (95 °C for 5 min; 95–85 °C at 2 °C/s; 85–25 °C at 0.1 °C/s; 4 °C hold). The digestion product was visualized in a 2% electrophoresis gel ran for 45 min (Figure S3b–d). The mutagenesis efficiency was estimated measuring the intensity of the nondigested and the digested bands with ImageJ (Table S3), and the data obtained were treated as described in Guschin *et al.* (2010). RE analysis was set up with 500 ng of PCR product and *Eco*RI (Fermentas) to assess mutagenesis at the *XT1* locus. Mutated samples confirmed by either method were sequenced and analysed by TIDE to characterize the indel size and the corresponding efficiency. All

experimental points represent the average of three independent samples.

Mutagenesis detection of off-target sites

Genomic DNA was extracted from 100 mg of *A. thaliana* rosetta leaves following the CTAB protocol (Murray and Thompson, 1980). Paired-end reads were obtained by sequencing the extracted DNA with Illumina NextSeq 550 platform. The resulting raw sequences are available at the NCBI Sequence Read Archive (SRA). Read quality was assessed with FASTQC (Andrews, 2010), and sequences were cleaned using TRIMMOMATIC (Bolger *et al.*, 2014), keeping reads with an average quality value of 29 in Phred 33 scale and a minimum length of 50 nt. Then, these reads were mapped against *A. thaliana* reference genome TAIR10 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001735.4/) using read mapper BWA-MEM (Li, 2013). Mapping statistics can be found in Table S5a. Quality scores were recalibrated using GATK Best Practices (McKenna *et al.*, 2010). For this purpose and for further steps in this pipeline, a collection of SNPs and indels was downloaded from NCBI dbSNP database (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/ archive/arabidopsis_3702/VCF/). SNPs were detected using GATK-HC v4 (McKenna *et al.*, 2010), VARSCAN2 (Koboldt *et al.*, 2012) and FREEBAYES (Garrison and Marth, 2012). For the PDS and the WT samples, a merged set was made using SNPs detected by all three programs using Bcftools (Li, 2011). For the POOL sample, GATK and FREEBAYES were used to produce the set of SNPs. Indels were detected using GATK-HC, VARSCAN2 and PINDEL (Ye *et al.*, 2009). Those indels detected in two out of three programs were used to construct a set of indels using Bcftools and GATK tools. Due to the ploidy of the POOL sample, we omitted the use of VARSCAN2 to the previously described detections. The SVs were detected using Delly (Rausch *et al.*, 2012). Filtering steps were applied in SNP, indel and SV detection using the parameters specified in Table S5b. Using these sets, all the SNPs, indels and SVs were numbered using Bcftools and VCFtools (Danecek *et al.*, 2011). For the PDS samples, variations not found in the WT samples nor in the POOL were taken. The variations found were checked using IGV. Potential off-targets were detected using Cas-OFFinder (Bae *et al.*, 2014b) allowing a maximum of 4 nt mismatches searching variations in these regions using BEDtools (Quinlan and Hall, 2010), extending the off-target region 50 nt in each side, and reviewed with IGV.

Accession numbers

Niben101Scf01519g10008.1 (FT); Niben101Scf01028g01003.1 (TFL1 3.1, TFL1 14.1); Niben101Scf04551g02001.1 (XT2A, XT2B); Niben101Scf04205g03008.1 (XT1); Niben101Scf02459 (CBP); Niben101Scf04196 (RUBISCO); Solyc01g079620.2 (MYB12); and At4g1410 (PDS3).

The raw DNA-sequencing reads have been deposited at Sequence Read Archive (SRA) and can be accessed under the BioProject ID PRJNA497395 (https://www.ncbi.nlm.nih.gov/bioproject).

Acknowledgements

We want to thank Dr. Lynne Yenush for the careful revision of this manuscript and Dr. Feng Zhang Laboratory for kindly providing the plasmids of Cas12a. This work has been funded by Grant BIO201678601-R from Plan Nacional I+D of the Spanish Ministry of Economy and Competitiveness and the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 760331 ('Newcotiana'). Bernabé-Orts, JM is a recipient of a FPI fellowship.

Conflict of interest

The authors declare no conflict of interest.

References

Adikusuma, F., Piltz, S., Corbett, M.A., Turvey, M., McColl, S.R., Helbig, K.J., Beard, M.R. *et al.* (2018) Large deletions induced by Cas9 cleavage. *Nature*, **560**, E8.

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*.

Bae, S., Kweon, J., Kim, H.S. and Kim, J.-S. (2014a) Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods*, **11**, 705–706.

Bae, S., Park, J. and Kim, J.-S. (2014b) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, **30**, 1473–1475.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.

Bentley, K.J., Gewert, R. and Harris, W.J. (1998) Differential efficiency of expression of humanized antibodies in transient transfected mammalian cells. *Hybridoma*, **17**, 559–567.

Biłas, R., Szafran, K., Hnatuszko-Konka, K. and Kononowicz, A.K. (2016) Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tissue Organ Culture*, **127**, 269–287.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

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

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

Chen, J.S., Ma, E., Harrington, L.B., da Costa, M., Tian, X., Palefsky, J.M. and Doudna, J.A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, **360**, 436–439.

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.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, M.A., Handsaker, R.E. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Ellul, P., Garcia-Sogo, B., Pineda, B., Rios, G., Roig, L. and Moreno, V. (2003) The ploidy level of transgenic plants in Agrobacterium-mediated transformation of tomato cotyledons (Lycopersicon esculentum L. Mill.) is genotype and procedure dependent. *Theor. Appl. Genet.* **106**, 231–238.

Endo, A., Masafumi, M., Kaya, H. and Toki, S. (2016) Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. *Sci. Rep.* **6**, 38169.

Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE*, **4**, e5553.

Garrison, E. and Marth, G. (2012) *Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907*.

Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C. and Rebar, E.J. (2010) A rapid and general assay for monitoring endogenous gene modification. *Engineered Zinc Finger Proteins Methods Protocol.* **649**, 247–256.

Horsch, R., Fry, J., Hoffman, N., Eichholtz, D., Rogers, S.A. and Fraley, R. (1985) A simple and general method for transferring genes into plants. *Science*, **227**, 1229–1232.

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

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*, **337**, 816–821.

Khan, M.H.U., Khan, S.U., Muhammad, A., Limin, H., Yang, Y. and Chuchuan, F. (2017) Induced Mutation and Epigenetics Modification in Plants for crop improvement by targeting CRISPR/Cas9 technology. *J. Cell. Physiol.* **233**, 4578–4594.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H. and Kim, J.S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868.

Kim, H., Kim, S.T., Ryu, J., Kang, B.C., Kim, J.S. and Kim, S.G. (2017a) CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* **8**, 14406.

Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.M., Choi, J.W. *et al.* (2017b) *In vivo* high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods*, **14**, 153–159.

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

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A. *et al.* (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576.

Kosicki, M., Tomberg, K. and Bradley, A. (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765.

Kumar, J., Gupta, D.S., Gupta, S., Dubey, S., Gupta, P. and Kumar, S. (2017) Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Rep.* **36**, 1187–1213.

Lee, K., Zhang, Y., Kleinstiver, B.P., Guo, J.A., Aryee, M.J., Miller, J., Malzahn, A. *et al.* (2018) Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol. J.* **17**, 362–372.

Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.

Li, H. (2013) *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997*.

Li, S.-Y., Cheng, Q.-X., Liu, J.-K., Nie, X.-Q., Zhao, G.-P. and Wang, J. (2018) CRISPR-Cas12a has both cis-and trans-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491.

Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F. and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722.

Marraffini, L.A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature*, **526**, 55.

McCormick, S., Niedermeyer, J., Fry, J., Barnason, A., Horsch, R. and Fraley, R. (1986) Leaf disc transformation of cultivated tomato (L. esculentum) using Agrobacterium tumefaciens. *Plant Cell Rep.* **5**, 81–84.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

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

Murray, M.G. and Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**, 4321–4326.

Patron, N.J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O. *et al.* (2015) Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol.* **208**, 13–19.

Peccoud, J., Weber, E., Engler, C., Gruetzner, R., Werner, S. and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, **6**, e16765.

Qin, G., Gu, H., Ma, L., Peng, Y., Deng, X.W., Chen, Z. and Qu, L.-J. (2007) Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res.* **17**, 471.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V. and Korbel, J.O. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, **28**, i333–i339.

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juarez, P., Fernandez-Del-carmen, A., Granell, A. and Orzaez, D. (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS ONE*, **6**, e21622.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J. *et al.* (2013) GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* **162**, 1618–1631.

Sentmanat, M.F., Peters, S.T., Florian, C.P., Connelly, J.P. and Pruett-Miller, S.M. (2018) A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* **8**, 888.

Soda, N., Verma, L. and Giri, J. (2017) CRISPR-Cas9 based plant genome editing: Significance, opportunities and recent advances. *Plant Physiol. Biochem.* **31**, 2–11.

Tabebordbar, M., Zhu, K., Cheng, J.K., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C. *et al.* (2016) *In vivo* gene editing in dystrophic mouse muscle and muscle stem cells. *Science*, **351**, 407–411.

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

Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., Xin, X. *et al.* (2018) A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol.* **19**, 84.

Tóth, E., Weinhardt, N., Bencsura, P., Huszár, K., Kulcsár, P.I., Tálas, A., Fodor, E. *et al.* (2016) Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biol Direct*, **11**, 46.

Vazquez-Vilar, M., Bernabe-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.

Vazquez-Vilar, M., Quijano-Rubio, A., Fernandez-Del-carmen, A., Sarrion-Perdigones, A., Ochoa-Fernandez, R., Ziarsolo, P., Blanca, J. *et al.* (2017) GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Res.* **45**, 2196–2209.

Wang, Y., Liu, X., Ren, C., Zhong, G.-Y., Yang, L., Li, S. and Liang, Z. (2016) Identification of genomic sites for CRISPR/Cas9-based genome editing in the Vitis vinifera genome. *BMC Plant Biol.* **16**, 96.

Wang, M., Mao, Y., Lu, Y., Tao, X. and Zhu, J.-K. (2017) Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Mol. Plant*, **10**, 1011–1013.

Xie, K., Zhang, J. and Yang, Y. (2014) Genome-wide prediction of highly specific guide RNA spacers for CRISPR–Cas9-mediated genome editing in model plants and major crops. *Mol. Plant*, **7**, 923–926.

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.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, **25**, 2865–2871.

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

Yin, X., Biswal, A.K., Dionora, J., Perdigon, K.M., Balahadia, C.P., Mazumdar, S., Chater, C. *et al.* (2017b) CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice. *Plant Cell Rep.* **36**, 745–757.

Yu, Z., Ren, M., Wang, Z., Zhang, B., Rong, Y.S., Jiao, R. and Gao, G. (2013) Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. *Genetics*, **195**, 289–291.

Zaidi, S.S., Mahfouz, M.M. and Mansoor, S. (2017) CRISPR-Cpf1: a new tool for plant genome editing. *Trends Plant Sci.* **22**, 550–553.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., Degennaro, E.M., Winblad, N. *et al.* (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34.

Zsögön, A., Cermak, T., Voytas, D. and Peres, L.E.P. (2016) Genome editing as a tool to achieve the crop ideotype and *de novo* domestication of wild relatives: case study in tomato. *Plant Sci.* **256**, 120–130.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Genotyping results of the *N. benthamiana* transformation with the CGEM targeting XT1 locus.

Figure S2 Summary of the information regarding *A. thaliana* experiments.

Figure S3 Summary of the *N. benthamiana* transient experiments.

Figure S4 Venn diagrams of SNPs and indels detected with different software tools.

Figure S5 Structural variations and large deletions found in a duplication event surrounding the target 1 of PDS-1-118.

Table S1 Constructs generated in this study.

Table S2 Oligonucleotides used in this study.

Table S3 Quantification data extracted from the restriction analysis and T7E1 gels for the efficiency estimation.

Table S4 Stability of the different crRNAs of *TFL1* and *FT* loci.

Table S5 Mapping stats and filtering parameters used in the WGS off-target analysis.