nature plants

BRIEF COMMUNICATION

PUBLISHED: 17 FEBRUARY 2017 | VOLUME: 3 | ARTICLE NUMBER: 17018

A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants

Xu Tang[1†], Levi G. Lowder[2†], Tao Zhang[3], Aimee A. Malzahn[4], Xuelian Zheng[1], Daniel F. Voytas[5], Zhaohui Zhong[1], Yiyi Chen[1], Qiurong Ren[1], Qian Li[1], Elida R. Kirkland[2], Yong Zhang[1]* and Yiping Qi[4,6]*

Clustered regularly interspaced short palindromic repeats (CRISPR)–Cpf1 has emerged as an effective genome editing tool in animals. Here we compare the activity of Cpf1 from *Acidaminococcus* sp. *BV3L6* (As) and *Lachnospiraceae bacterium ND2006* (Lb) in plants, using a dual RNA polymerase II promoter expression system. LbCpf1 generated biallelic mutations at nearly 100% efficiency at four independent sites in rice T0 transgenic plants. Moreover, we repurposed AsCpf1 and LbCpf1 for efficient transcriptional repression in *Arabidopsis*, and demonstrated a more than tenfold reduction in *miR159b* transcription. Our data suggest promising applications of CRISPR–Cpf1 for editing plant genomes and modulating the plant transcriptome.

Sequence-specific nucleases (SSNs) create DNA double strand breaks at predefined genomic loci and hence provide opportunities for precise genome editing based on DNA repair. SSNs, such as zinc finger nucleases (ZFNs) and TAL effector nucleases (TALENs), are time consuming to construct, as protein engineering of their DNA binding domains is required to achieve the requisite target specificity[1,2]. Consequently, CRISPR–Cas9, an RNA-guided endonuclease that targets DNA sites through nucleotide base pairing, has become the preferred SSN for genome editing in plants[3].

CRISPR–Cpf1, a new class 2 CRISPR–Cas system, was recently used to edit genomes in human cells[4], mice[5,6] and *Drosophila*[7]. The Cpf1 system differs from Cas9 by at least five aspects: (1) the protospacer adjacent motif (PAM) is 'TTTN', which helps target AT-rich regions and complements the popular SpCas9 system ('NGG' PAM); (2) Cpf1 creates 5′ staggered ends, which potentially can facilitate precise gene replacement using non-homologous end joining (NHEJ); (3) Cpf1 cleaves DNA at sites distal to the PAM. Such distal cleavage allows previously mutated sequences to be severed repeatedly, promoting homology-dependent repair (HDR); (4) repetitive cleavage, coupled with extensive processing of staggered 5′ DNA ends, may also promote large chromosomal deletions; (5) the Cpf1 crRNA length (~43 nt) is less than half that of Cas9, making it more suitable for multiplexed genome editing and packaging into viral vectors. For all these reasons, therefore, CRISPR–Cpf1 is an attractive tool for plant genome editing.

To develop and test Cpf1 in plants, we focused on AsCpf1 and LbCpf1, both of which showed DNA cleavage activity in human cells[4]. In rice protoplasts, we expressed the plant codon optimized AsCpf1 and LbCpf1 under the maize ubiquitin (ZmUbi) promoter and the CRISPR RNA (crRNA) under the rice U6 (OsU6) promoter. However, at the targets tested, we detected little cleavage activity (data not shown). The first nucleotide of As-crRNA and Lb-crRNA is 'U' (Fig. 1a,b), which cannot be accommodated by RNA polymerase III (Pol III) promoters such as U6 (transcription starts with 'G') and U3 (transcription starts at 'A'). Also, Pol III-transcribed crRNAs contain 3′ terminal poly U sequences. These 3′ poly U sequences are immediately adjacent to protospacer sequences involved in DNA recognition. We reasoned, therefore, that sequences of the crRNA might be highly specific and critical for Cpf1 activity. To accommodate this sequence specificity, we used a double ribozyme system that precisely processes the crRNAs[8] (Supplementary Fig. 1). Both the crRNA ribozyme cassette and Cpf1 were expressed by a ZmUbi promoter to achieve coordinate expression (Fig. 1a,b). We tested this dual Pol II promoter system for targeting six sites in three rice genes (*OsPDS*, *OsDEP1* and *OsROC5*), and we assayed mutation frequencies in protoplasts. With AsCpf1, three out of six target sites showed significant mutagenesis resulting from error-prone NHEJ, as revealed by restriction fragment length polymorphism (RFLP) analysis (Supplementary Fig. 2). For LbCpf1, all six target sites showed high frequencies of mutagenesis by NHEJ (Supplementary Fig. 2). Mutation frequencies were then quantified using high-throughput DNA sequencing. Mutation frequencies induced by AsCpf1 ranged from 0.6 to 10% whereas mutation frequencies induced by LbCpf1 ranged from 15 to 25% across the six targets (Fig. 1c). More than 90% of mutations produced by both AsCpf1 and LbCpf1 were deletions (Fig. 1d), and the majority ranged from 6 to 13 bp in size (Fig. 1e). These deletions are considerably larger than those induced by Cas9 (1–3 bp)[3], and might prove advantageous for Cpf1 use as a mutagen. Deletions mostly occurred 13–23 nucleotides distal to the PAM site for LbCpf1 (Fig. 1f) and AsCpf1 (Supplementary Fig. 3); more rarely, insertions were also recovered at these sites (Supplementary Figs 3 and 4).

Initial testing of LbCpf1 proved superior to AsCpf1 in generating mutations. Based on these findings, we accessed target specificity of LbCpf1 by testing whether the enzyme is capable of tolerating mismatches between on-target DNA and the crRNA protospacer sequence. Six double mismatch mutations (at positions 1–2, 5–6, 9–10, 13–14, 17–18 and 21–22) were introduced into the 23 nt protospacer of Lb-crRNA, which targets the first exon of *OsPDS* (Fig. 1g). These six constructs, along with the on-target control, were transformed into rice protoplasts, and mutagenesis was assessed by high-throughput DNA sequencing. Mismatches in the first 18 nt of the protospacer completely destroyed cleavage activity, whereas mismatches at positions 21–22 reduced activity by a half, and generated a

[1]Department of Biotechnology, School of Life Science and Technology, Centre for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China. [2]Department of Biology, East Carolina University, Greenville, North Carolina 27834, USA. [3]Jiangsu Key Laboratory of Crop Genetics and Physiology, Co-Innovation Centre for Modern Production Technology of Grain Crops, Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou University, Yangzhou 225009, China. [4]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Maryland 20742, USA. [5]Department of Genetics, Cell Biology & Development and Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA. [6]Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, Maryland 20850, USA. †These authors contributed equally to this work. *e-mail: yiping@umd.edu; zhangyong916@uestc.edu.cn

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

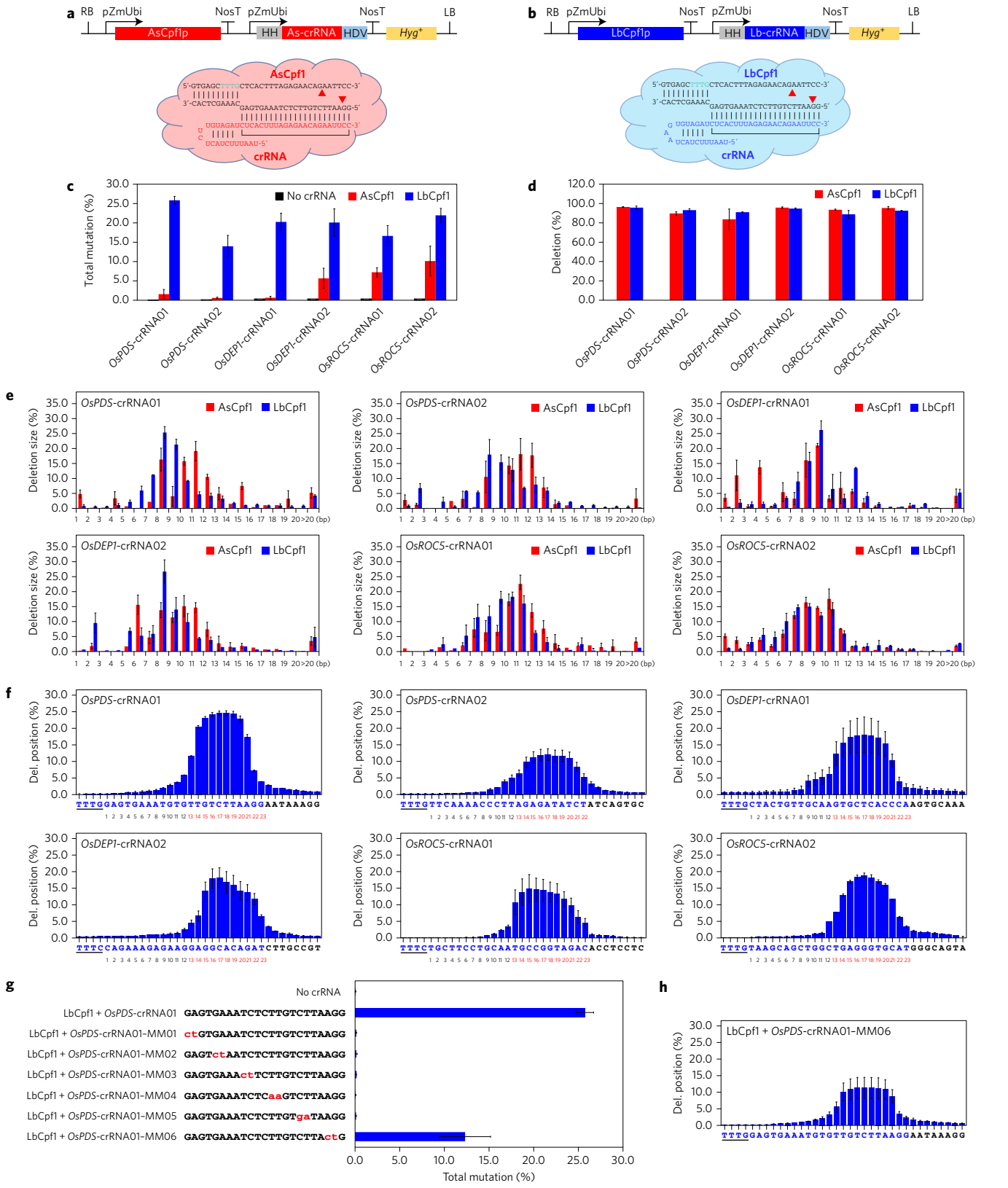
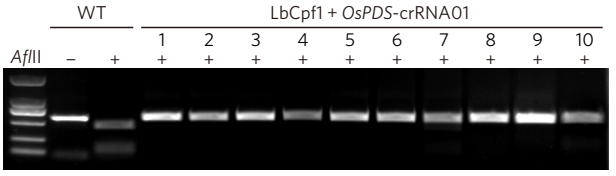


Figure 1 | Comprehensive analysis of AsCpf1 and LbCpf1 activities in rice protoplasts. a, The AsCpf1 expression system and the AsCpf1-crRNA complex. RB, right border; LB, left border; pZmUbi, *Zea mays* Ubiquitin promoter; NosT, nopaline synthase terminator; HH, hammerhead ribozyme; HDV, hepatitis delta virus ribozyme; *Hyg*$^{+}$, hygromycin resistance gene. **b**, The LbCpf1 expression system and the LbCpf1-crRNA complex. Approximate DNA double-strand break positions are illustrated by red triangles. **c**, Comparison of mutation frequencies by AsCpf1 and LbCpf1. **d**, Comparison of deletion frequencies by AsCpf1 and LbCpf1. **e**, Comparison of deletion sizes by AsCpf1 and LbCpf1; bp, base pairs. **f**, Positions of deleted bases by LbCpf1. PAM sequences are underlined. **g**, Off-targeting analysis with mismatch (MM) crRNAs at the *OsPDS* locus. **h**, Positions of deleted bases by LbCpf1 and *OsPDS*-crRNA01–MM06. Error bars represent standard deviations of two biological replicates.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

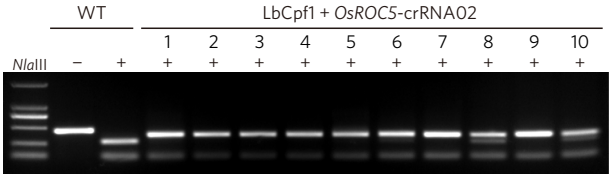
a

Targeted rice gene	LbCpf1 reagents	Tested T0 lines	Mutated T0 lines: number, ratio	Biallelic mutation lines: number, ratio
OsPDS	*OsPDS*-crRNA01	15	15, 100%	15, 100%
OsPDS	*OsPDS*-crRNA02	13	13, 100%	13, 100%
OsDEP1	*OsDEP1*-crRNA02	16	16, 100%	15, 93.8%
OsROC5	*OsROC5*-crRNA02	19	19, 100%	19, 100%



WT (*OsPDS*): GTGAGCTTTGGAGTGAAATCTCTTGTCTTAAGGAATAAAGGAA
OsPDS-crRNA01-01
Locus-1: GTGAGCTTTGGAGTGAAATCT----------GAATAAAGGAA -11 bp
Locus-2: GTGAGCTTTGGAGTGAAATCTCTT---------AATAAAGGAA -9 bp
OsPDS-crRNA01-02
Locus-1: GTGAGCTTTGGAGTGA----------GGAA -23 bp
Locus-2: GTGAGCTTTGGAGTGAAATCTCTT---------AATAAAGGAA -9 bp
OsPDS-crRNA01-03
Locus-1: GTGAGCTTTGGAGTGAAATC----------AAGGAA -17 bp
Locus-2: GTGAGCTTTGGAGTGAAATCTCTTGTC-----GAATAAAGGAA -5 bp
OsPDS-crRNA01-04
Locus-1: GTGAGCTTTGGAGTGA----------GGAATAAAGGAA -15 bp
Locus-2: GTGAGCTTTGGAGTGAAATCTCTTGTtTTAAGGAATAAAGGAA -1 bp/+1 bp
OsPDS-crRNA01-05
Locus-1: GTGAGCTTTGGAGTGAAATCTCT---------GAATAAAGGAA -9 bp
Locus-2: GTGAGCTTTGGAGTGAAATCTCTTG--------AATAAAGGAA -8 bp
OsPDS-crRNA01-06
Locus-1: GTGAGCTTT----------AAGGAATAAAGGAA -20 bp
Locus-2: GTGAGCTTTGGAGTGAAATCT----------GAATAAAGGAA -11 bp





WT (*OsROC5*): TCCGGTTTTGTAAGCAGCTGGCTGAGGGTGCATGGGCAGTAGT
OsROC5-crRNA02-01
Locus-1: TCCGGTTTTGTAAGCAGCTGGCTGAG------TGGGCAGTAGT -6 bp
Locus-2: TCCGGTTTTGTAAGCAGCTGGCTGAG------TGGGCAGTAGT -6 bp
OsROC5-crRNA02-02
Locus-1: TCCGGTTTTGTAAGCAGCTGGCT--------ATGGGCAGTAGT -8 bp
Locus-2: TCCGGTTTTGTAAGCAGCTGGCT---------GGGCAGTAGT -10 bp
OsROC5-crRNA02-03
Locus-1: TCCGGTTTTGTA----------GT -29 bp
Locus-2: TCCGGTTTTGTAAGCAGCTGGCT---------TGGGCAGTAGT -9 bp
OsROC5-crRNA02-04
Locus-1: TCCGGTTTTGTAAGCAGCTGGCTG----------GGCAGTAGT -10 bp
Locus-2: TCCGGTTTTGTAAGCAGCTGGCTGAGGG---ATGGGCAGTAGT -3 bp
OsROC5-crRNA02-05
Locus-1: TCCGGTTTTGTAAGCAGCTGGCTGA--------AGT -15 bp
Locus-2: TCCGGTTTTGTAAGCAGCTGGCTGA---------GGCAGTAGT -9 bp
OsROC5-crRNA02-06
Locus-1: TCCGGTTT--------AGT -32 bp
Locus-2: TCCGGTTTTGTAAGCAGCTGGCTGAG----CATGGGCAGTAGT -4 bp



Figure 2 | Highly efficient genome editing in rice T0 lines by LbCpf1. a, A summary of genotyping results on stable transgenic lines at four target sites. **b**, RFLP analysis of independent T0 lines (as indicated in a) and biallelic mutations with *OsPDS*-crRNA01 revealed by Sanger sequencing. The PAM sequence is in red and the target sequence is in blue. **c**, The phenotypes of the wild type (WT; left) and a *OsPDS* mutant (right, *OsPDS*-crRNA01-03). **d**, RFLP analysis of independent T0 lines (as indicated in a) and biallelic mutations with *OsROC5*-crRNA02 revealed by Sanger sequencing. The PAM sequence is in red and the target sequence is in blue. Owing to two additional *Nla*III sites in the PCR amplicons, the mutated bands are slightly smaller than the uncut WT product. Note the cleaved bands for lines no. 8 and no. 10 resulted from *Nla*III sites (CATG) created by mutagenesis. **e**, The phenotypes of the WT (left) and a *OsROC5* mutant with curly leaves (right, *OsROC5*-crRNA02-02). #, number of different transgenic lines.

similar deletion pattern to the on-target crRNA (Fig. 1g,h). These data suggest CRISPR–Cpf1 has high targeting specificity in plant cells, consistent with the observations in human cells[9,10].

We next tested whether we could generate rice plants with Cpf1-induced mutations. Four LbCpf1 constructs, targeting three rice genes, were transformed into rice calli by *Agrobacterium*. Analysis of individual T0 transgenic plants from independent calli revealed 100% had mutations at the target sites (Fig. 2a). The vast majority of mutations were found to be biallelic. All 15 T0 lines transformed with *OsPDS*-crRNA01 contained biallelic mutations and one line (no. 14) was homozygous (Fig. 2b; Supplementary Fig. 5); photobleaching was the phenotypic consequence of these loss of function mutations (Fig. 2c). Among 19 T0 lines transformed with reagents targeting *OsROC5*-crRNA02, all had biallelic mutations and one line (no. 1) was homozygous (Fig. 2d; Supplementary Fig. 6); biallelic knockout conferred a curly leaf phenotype (Fig. 2e). For *OsPDS*-crRNA02, all 13 T0 lines had biallelic mutations and two of them were homozygous (Supplementary Fig. 7). For *OsDEP1*-crRNA02,

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

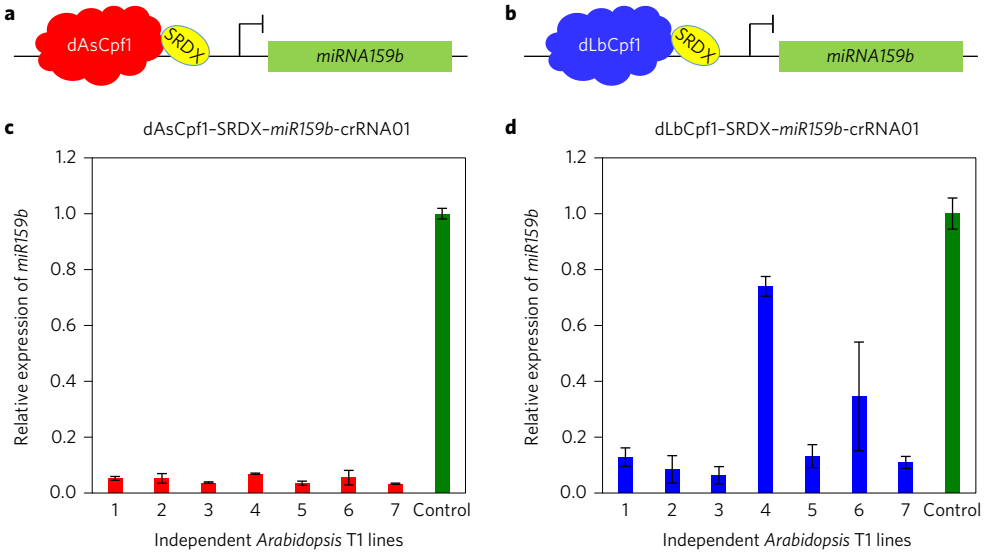


Figure 3 | Effective transcriptional repression in *Arabidopsis* by dAsCpf1-SRDX and dLbCpf1-SRDX. a,**b**, An illustration of targeted repression of *miR159b* by dAsCpf1-SRDX (**a**) or dLbCpf1-SRDX (**b**). **c**,**d**, Quantitative real-time (qRT)-PCR data showing targeted repression of *miR159b* in independent transgenic T1 lines by dAsCpf1-SRDX and dLbCpf1-SRDX, respectively. For each line, the data is in technical triplicate. Seven individual GUS (β-glucuronidase)-expressing lines with the same marker gene (*Hyg*$^{+}$) were bulked as control. The data are normalized to the *ACTIN 2* gene. Error bars represent standard deviations.

15 out of 16 regenerated T0 lines had biallelic mutations and the remaining one (no. 8) was heterozygous (Supplementary Fig. 8). To confirm the observed mutations were homogenous throughout individual plants, we genotyped three T0 lines each for *OsPDS*-crRNA02 and *OsDEP1*-crRNA02. For each line, DNA was extracted from three independent leaves and polymerase chain reaction (PCR) products encompassing the target site were subjected to Sanger sequencing. All sampled leaves from individual plants carried the same mutations (Supplementary Fig. 9). Thus, the LbCpf1-induced mutations in rice plants were largely non-mosaic, suggesting the nuclease acts early and effectively at a single-cell stage. Such mutations should be readily transmitted to the next generation.

Previously, we demonstrated that CRISPR–Cas9 can be engineered to regulate gene expression in plants[11]. The thymine rich PAM makes CRISPR–Cpf1 very suitable for targeting AT-rich promoter regions, which motivated us to repurpose Cpf1 as a transcriptional repressor. The nuclease domains of AsCpf1 and LbCpf1 were deactivated by mutations, generating dAsCpf1 (D908A) and dLbCpf1 (D832A)[4]. These dCpf1 reagents were then fused to three copies of the SRDX transcriptional repressor[11] (Supplementary Fig. 10). We developed a dual ubiquitin promoter system to express CRISPR–dCpf1–SRDX repressors (Supplementary Fig. 10) and then tested these in *Arabidopsis* by targeting the promoter of a non-coding RNA, *miR159b*. With the dAsCpf1-SRDX repressor, expression of *miR159b* was less than 10% of the wild type in seven randomly chosen T1 transgenic lines (Fig. 3a,c). We also targeted *miR159b* with dLbCpf1–SRDX guided by the same crRNA and found similar repression activity across T1 transgenic lines, albeit with more variation (Fig. 3b,d). Thus, we demonstrated effective transcriptional repression with Cpf1-based regulators in *Arabidopsis*. The results suggest that although AsCpf1 is less potent as a nuclease, it effectively binds DNA, perhaps even more tightly than LbCpf1. Further investigation of this property is warranted.

While we were preparing this manuscript, two other groups reported that CRISPR–Cpf1 could edit plant genomes, albeit at a lower efficiency[12,13]. These studies used Pol III-expressed crRNAs that contain additional nucleotides on the 5′ and 3′ ends, which may be suboptimal in plants. In this study, we demonstrated that LbCpf1, when coupled with Pol II-expressed and ribozyme-processed crRNAs, is a highly effective mutagen in rice. For the first time in eukaryotic cells, we repurposed AsCpf1 and LbCpf1 as transcriptional repressors and demonstrated their high efficacy for targeted gene repression *in vivo*. CRISPR–Cpf1 holds great promise for editing plant genomes, modulating the transcriptome and other applications.

Methods

Construction of gateway compatible CRISPR–Cpf1 vectors. Details about construction of three sets of Gateway modular vectors for Cpf1, crRNA and PolII promoters of choice are available in the Supplementary Information. All the oligos and gBlocks used in this study are summarized in Supplementary Table 1.

Assembly of T-DNA expression vectors. Each individual T-DNA expression vector (Supplementary Table 2) was assembled from a single Multi-site Pro LR reaction (1-5-2) with the attR1–attR2 destination vector pYPQ203 or pYPQ202, an attL1–attR5 Cpf1 entry clone (pYPQ220, pYPQ230, pYPQ223 or pYPQ233) and an attL5–attL2 crRNA expression entry clone using Gateway LR clonase II (Invitrogen). Refer to Supplementary Methods for more details.

Rice protoplast transformation and stable transformation. The *Japonica* cultivar Nipponbare was used in this study. Polyethylene glycol transformation of rice protoplasts with T-DNA vectors was carried out according to our previously published protocol[14]. Rice stable transformation was conducted as published previously[14].

Mutagenesis analysis at target sites. Genomic DNA was extracted from transformed rice protoplasts or transgenic lines using the CTAB method[15]. Mutagenesis at target sites was analysed in two methods. For RFLP analysis, each target site was amplified by PCR followed by restriction digestion with corresponding enzymes (see Supplementary Methods for details).

High-throughput sequencing analysis. The genomic region flanking Cpf1 target sites were PCR-amplified using barcoded primers for all 42 samples containing two biological replicates (Supplementary Table 3). Purified DNA samples were quantified by Qubit 2.0 Fluorometer (Life Technologies) and were sequenced using Illumina Hiseq 2,500 planform. More than 50,000 reads were generated for each sample. To generate a full-length target sequence with 250 bp paired-end Illumina reads, the paired-ends were joined by single reads using flash software[16]. We then used BWA[17] to map all joined reads to *OsPDS* (Os03g0184000), *OsDEP1* (Os09g0441900) and *OsROC5* (Os02g0674800). Data processing and counting analysis were carried out using python and R.

***Arabidopsis* transformation, screen, RNA extraction and qRT–PCR analysis.** *Arabidopsis* ecotype Columbia (Col-0) was used in this study. The procedures all followed our previous described protocols[11].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Data availability. *Accession codes.* The eight Gateway-compatible vectors for this CRISPR–Cpf1 system are available from Addgene (https://www.addgene.org/): pYPQ141-ZmUbi-RZ-As (no. 86196) and pYPQ141-ZmUbi-RZ-Lb (no. 86197), pYPQ202 (no. 86198), pYPQ203 (no. 86207), pYPQ220 (no. 86208), pYPQ223 (no. 86209), pYPQ230 (no. 86210), pYPQ233 (no. 86211). The high-throughput sequencing data sets have been submitted to the National Center for Biotechnology information (NCBI) database under Sequence Read Archive (SRA) BioSample ID SRS1840609.

Received 16 November 2016; accepted 27 January 2017; published 17 February 2017

References

1. Carroll, D. *Genetics* **188,** 773–782 (2011).
2. Bogdanove, A. J. & Voytas, D. F. *Science* **333,** 1843–1846 (2011).
3. Paul, J. W. III & Qi, Y. *Plant Cell Rep.* **35,** 1417–1427 (2016).
4. Zetsche, B. *et al. Cell* **163,** 759–771 (2015).
5. Kim, Y. *et al. Nat. Biotechnol.* **34,** 808–810 (2016).
6. Hur, J. K. *et al. Nat. Biotechnol.* **34,** 807–808 (2016).
7. Port, F. & Bullock, S. L. *Nat. Methods* **13,** 852–854 (2016).
8. Gao, Y. & Zhao, Y. *J. Integr. Plant Biol.* **56,** 343–349 (2014).
9. Kleinstiver, B. P. *et al. Nat. Biotechnol.* **34,** 869–874 (2016).
10. Kim, D. *et al. Nat. Biotechnol.* **34,** 863–868 (2016).
11. Lowder, L. G. *et al. Plant Physiol.* **169,** 971–985 (2015).
12. Xu, R. *et al. Plant Biotechnol. J.* http://doi.org/bzh2 (2016).
13. Endo, A., Masafumi, M., Kaya, H. & Toki, S. *Sci. Rep.* **6,** 38169 (2016).
14. Tang, X. *et al. Mol. Plant* **9,** 1088–1091 (2016).
15. Murray, M. G. & Thompson, W. F. *Nucleic Acids Res.* **8,** 4321–4325 (1980).
16. Magoc, T. & Salzberg, S. L. *Bioinformatics* **27,** 2957–2963 (2011).
17. Li, H. & Durbin, R. *Bioinformatics* **26,** 589–595 (2010).

Acknowledgements

This work was supported by grants including the National Science Foundation of China (31330017, 31271420 and 31371682) and the National Transgenic Major Project (2014ZX0801003B-002) to X.Z. and Y.Z., a Collaborative Funding Grant from North Carolina Biotechnology Center and Syngenta Biotechnology (2016-CFG-8003) and a start-up fund from University of Maryland-College Park to Y.Q., and National Science Foundation (MCB 0209818, DBI 0923827 and 105-1339209) to D.F.V.

Author contributions

Y.Q., Y.Z. and D.F.V. designed the experiments. Y.Q., L.G.L. and A.A.M. generated all the constructs. X.T. and Y.Z. performed the transient assays in protoplasts and prepared samples for deep sequencing. T.Z., Y.Z. and X.T. analysed the deep sequencing data. X.T., X.Z., Z.Z., Y.C., Q.R. and Q.L. generated stable transgenic rice and analysed the plants. L.G.L. and E.R.K. produced *Arabidopsis* transcriptional repression data. Y.Q., Y.Z. and D.F.V. wrote the paper with input from other authors. All authors read and approved the final manuscript.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.Z. and Y.Q.

How to cite this article: Tang, X. *et al.* A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* **3,** 17018 (2017).

Competing interests

The authors declare no competing financial interests.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.