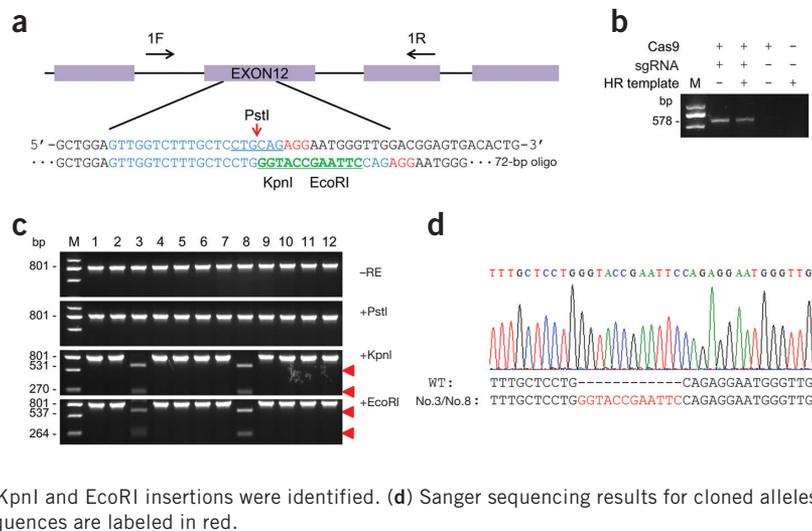


Figure 2 HDR-mediated genome modification in rice protoplasts. (a) Schematic of the oligo targeting site in *OsPDS*. The sgRNA targeting sequence is in blue, and the PAM sequence in red. The 72-bp donor oligo is shown under the target site, with 12-bp insertions (KpnI + EcoRI site) in green. (b) PCR amplification of the protoplast genomic DNA predigested with PstI to enrich for sgRNA:Cas9-induced mutations. Specific 1F and 1R primers were used. HR, homologous recombination. (c) Targeted integration of the KpnI and EcoRI restriction sites. The enrichment PCR product (+Cas9, +sgRNA, +HR template) was cloned into pEASY-Blunt vector (TransGen). Lanes 1–12, representative PCR products of cloned alleles for digesting assay; –RE, PCR amplification of colonies with M13F/R primers; +PstI, +KpnI and +EcoRI, PCR products digested with PstI, KpnI and EcoRI, respectively. Two cloned alleles (no. 3 and no. 8, arrowhead) with KpnI and EcoRI insertions were identified. (d) Sanger sequencing results for cloned alleles no. 3 and no. 8 show HDR-mediated targeting. Inserted sequences are labeled in red.



sequences of the form 5'-A-N_(19–21)-GG-3' identified 32 targets on average per cDNA. The wheat A and D genomes yielded similar results (Supplementary Table 3). Our findings establish that the sgRNA:Cas9 system can be used for rice and wheat genome modification, the first plants shown to be amenable to this gene editing approach.

Note: Supplementary information is available in the online version of the paper (doi:10.1038/nbt.2650).

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (201263, 383601 and 31200273), the Ministry of Agriculture of China (2011ZX08002-004 and 2013ZX08010-002) and Chinese Academy of Sciences (KSCX2-EW-N-06, KSCX2-EW-J-6). We thank X. Wang for bioinformatics analysis.

AUTHOR CONTRIBUTIONS

Q.S., Y.W., J.L., Y.Z., K.C., Z.L., J.J.X., J.-L.Q. and C.G. designed the experiments; Q.S., Y.W., J.L., Y.Z., K.Z. and J.L. performed experiments; Q.S., Y.W., J.L., J.-L.Q. and C.G. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Qiwei Shan^{1,4}, Yanpeng Wang^{1,4}, Jun Li^{1,4}, Yi Zhang¹, Kunling Chen¹, Zhen Liang¹, Kang Zhang¹, Jinxing Liu¹, Jianzhong Jeff Xi², Jin-Long Qiu³ & Caixia Gao¹

¹State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. ²Institute of Molecular Medicine, Peking University, Beijing, China.

³State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. ⁴These authors contributed equally to this work.

e-mail: cxgao@genetics.ac.cn or qiujl@im.ac.cn

- Zhang, F. *et al. Proc. Natl. Acad. Sci. USA* **107**, 12028–12033 (2010).
- Chen, K. & Gao, C. *J. Genet. Genomics* **40**, 271–279

- (2013).
- Mussolino, C. & Cathomen, T. *Nat. Biotechnol.* **31**, 208–209 (2013).
- Barrangou, R. *et al. Science* **315**, 1709–1712 (2007).
- Jinek, M. *et al. Science* **337**, 816–821 (2012).
- Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
- Cong, L. *et al. Science* **339**, 819–823 (2013).

- Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. *Nat. Biotechnol.* **31**, 233–239 (2013).
- Hwang, W.Y. *et al. Nat. Biotechnol.* **31**, 227–229 (2013).
- Chang, N. *et al. Cell Res.* **23**, 465–472 (2013).
- Wang, H. *et al. Cell* **153**, 910–918 (2013).
- Shan, Q. *et al. Mol. Plant* **6**, 1365–1368 (2013).
- Fu, Y. *et al. Nat. Biotechnol.* advance online publication, doi:10.1038/nbt.2623 (23 June 2013).

Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9

To the Editor:

Elucidation and manipulation of human, animal and plant genomes is key to basic biology research, medical advances and crop improvement. The development of targeted genome editing, particularly homologous recombination-based gene replacement, is of great value in all organisms. Recent advances in engineered nucleases with programmable DNA-binding specificities, such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), have provided valuable means to create targeted mutations in metazoan and plant genomes with high specificity^{1–6}. However, these technologies demand elaborate design and assembly of individual DNA-binding proteins for each DNA target site^{1–6}. Recently, a simple, versatile and efficient genome engineering technology has been developed based on the

bacterial clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas) adaptive immune systems⁷. In a type II CRISPR-Cas system from *Streptococcus pyogenes*, a single Cas9 endonuclease guided by a duplex of mature CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA) cleaves trespassing DNA from bacteriophage or plasmids in a sequence-specific manner⁷. By reconstitution of the *S. pyogenes* Cas9 (SpCas9) and an artificial chimera of crRNA and tracrRNA called synthetic-guide RNA (sgRNA) in eukaryotic cells, including yeast, zebrafish, mouse and human cells, targeted genome editing has been achieved through either error-prone nonhomologous end joining (NHEJ) or homology-directed repair (HDR) of the intended cleavage site^{7–14}. Here, we show the feasibility and efficacy of sgRNA:Cas9-based genome editing

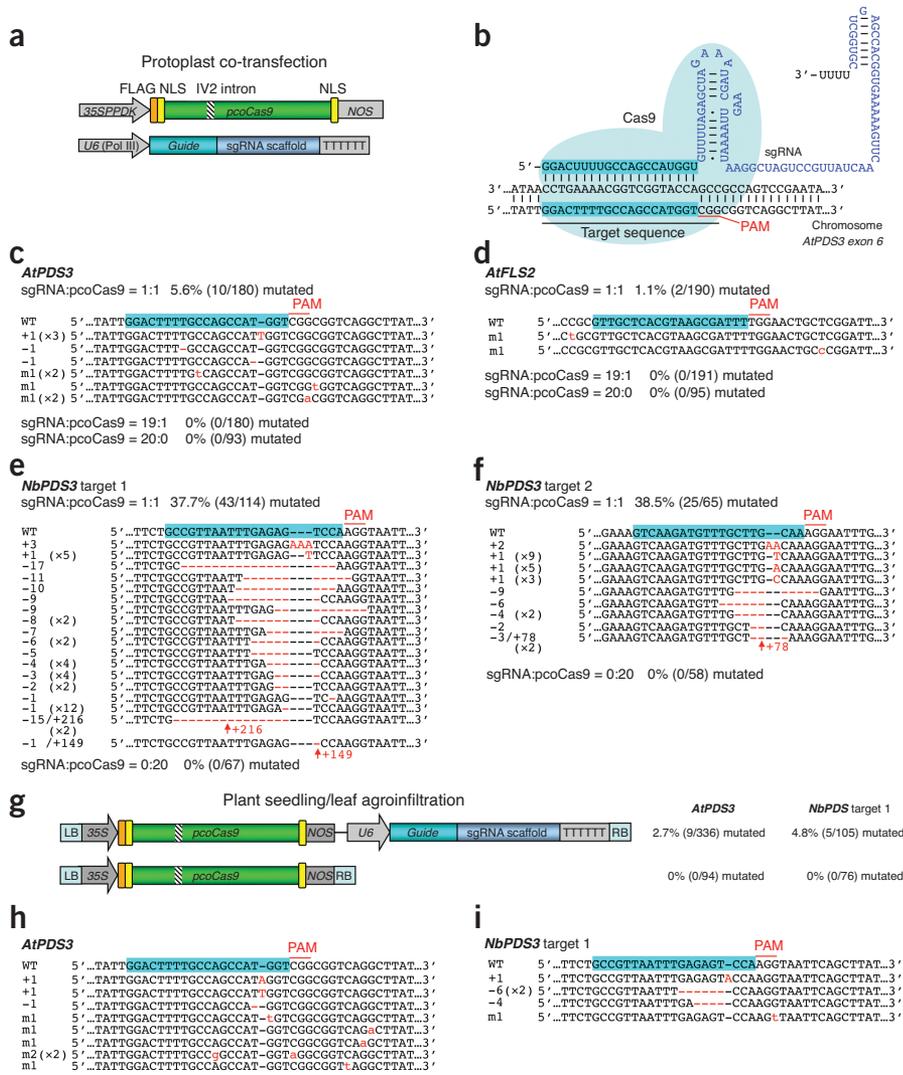


Figure 1 Targeted plant genome editing by sgRNA:pcoCas9. (a) sgRNA:pcoCas9 constructs for protoplast co-transfection. NLS, nuclear localization sequence. (b) Diagram of the sgRNA:pcoCas9 complex targeting the *Arabidopsis AtPDS3* exon 6. (c–f) Targeted genome editing on *AtPDS3* (c) and *AtFLS2* (d) in *Arabidopsis* protoplasts and *NbPDS* (e, f) in *Nicotiana benthamiana* protoplasts. (g) Binary plasmids for genome editing of *AtPDS3* and *NbPDS* in *Arabidopsis* and *N. benthamiana* plants, respectively, by *Agrobacterium*-mediated transient gene expression. (h, i) Targeted genome editing on *AtPDS3* in *Arabidopsis* seedlings (h) and *NbPDS* in *N. benthamiana* leaves (i). The mutation rate in c–g was calculated based on the mutant/total alleles of randomly selected clonal amplicons of the target locus. In c–f, h and i, blue shadow marks the target sequence recognized by cognate sgRNA. PAM, the protospacer adjacent motif. DNA insertions, deletions and point mutations are shown in red as upper case letters, dashes and lower case letters, respectively. The upright arrow and number in red indicate a long insertion.

technology in the model plants *Arabidopsis thaliana* and *Nicotiana benthamiana*.

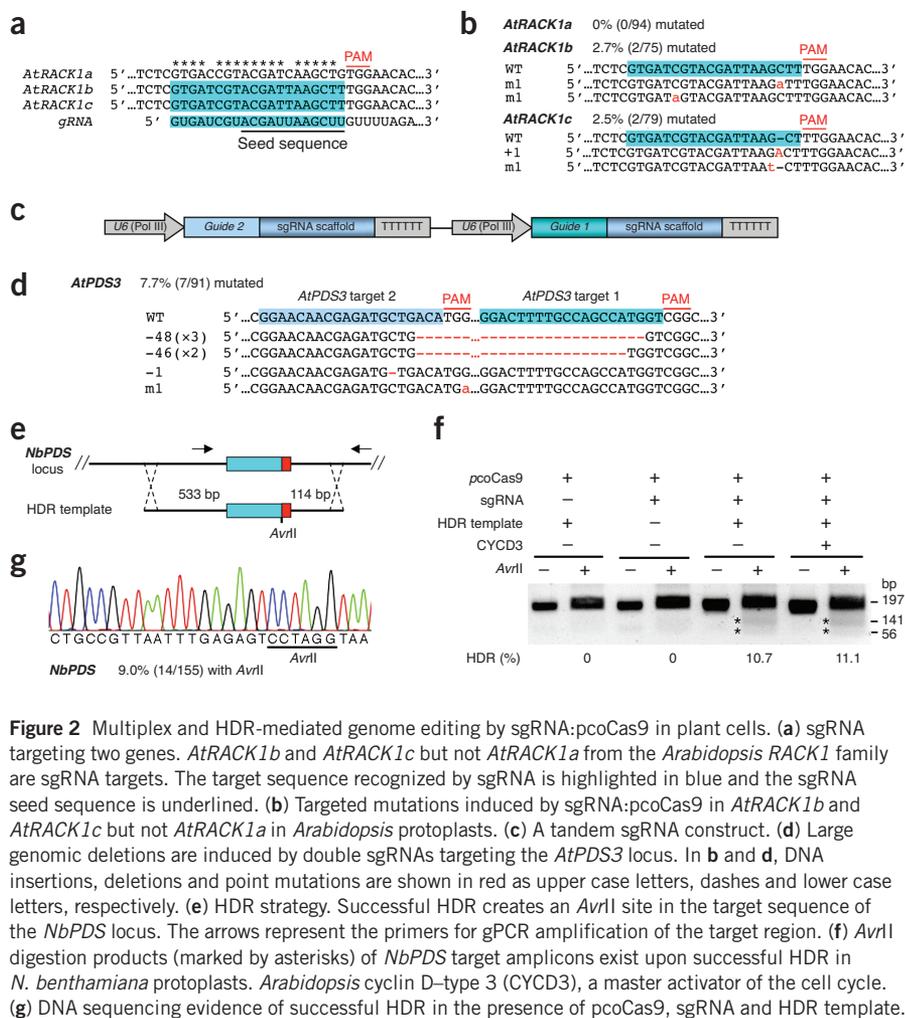
To explore the use of sgRNA:Cas9 technology for plant genome engineering, we first expressed a plant codon-optimized *SpCas9* (*pcoCas9*) and an sgRNA targeting *Arabidopsis thaliana PDS3* (*PHYTOENE DESATURASE*) (Fig. 1a, b and Supplementary Sequences) in *Arabidopsis* mesophyll protoplasts, which are freshly isolated leaf cells without cell walls. The protoplast transient expression system supports highly efficient DNA co-transfection

and protein expression¹⁵. The *pcoCas9* was expressed under the hybrid constitutive 35SPPDK promoter¹⁵, whereas the sgRNA was transcribed from the *Arabidopsis U6* polymerase III promoter (Fig. 1a). Notably, *pcoCas9* was expressed at a substantially higher level than the humanized *SpCas9* (ref. 9) using the same expression vector in *Arabidopsis* protoplasts (Supplementary Fig. 1). In addition, *pcoCas9* encodes nuclear localization sequences at both protein termini (Fig. 1a) for optimal protein nuclear localization⁸. A potato *IV2* intron (Fig. 1a)

was inserted to minimize adverse effects on bacterial growth¹⁶ resulting from potential leaky expression and nuclease activities of *pcoCas9* in *Escherichia coli* during cloning.

To determine the mutagenesis efficiency of the sgRNA:pcoCas9 system in *Arabidopsis* protoplasts, we cloned and Sanger-sequenced genomic PCR (gPCR) amplicons of the target region using total genomic DNA (gDNA) from transfected protoplasts as templates (Supplementary Methods). With a DNA ratio of sgRNA:pcoCas9 at 1:1 during co-transfection, we detected ten mutated *AtPDS3* target alleles among 180 randomly sequenced amplicons, reaching an approximate mutagenesis frequency of 5.6% (Fig. 1c). Of note, a ratio of sgRNA:pcoCas9 at 1:19 failed to induce any mutation in 180 sequenced amplicons, and no mutation was detected among 93 sequenced amplicons when *pcoCas9* was expressed alone (Fig. 1c). For a second gene, *AtFLS2* (*FLAGELLIN SENSITIVE 2*), tested in *Arabidopsis* protoplasts, the sgRNA:pcoCas9-mediated mutagenesis also only occurred with a DNA ratio of sgRNA:pcoCas9 at 1:1 but not at 1:19 (Fig. 1d). In this case, a lower mutagenesis frequency (1.1%) was observed (Fig. 1d). Taken together, these results suggest that sgRNA expression is the limiting factor for optimal targeting and mutagenesis in plant cells, as in human cells¹².

To extend the application of sgRNA:pcoCas9-mediated genome editing to other plant systems, we carried out a parallel study using *N. benthamiana* protoplasts. Notably, we targeted *NbPDS* (the *N. benthamiana* ortholog of *AtPDS3*) at two different sites, and obtained substantially higher mutagenesis frequencies than in *Arabidopsis*, namely 37.7% for the first target site (Fig. 1e) and 38.5% for the second target site (Fig. 1f). The sgRNA:pcoCas9-induced mutagenesis frequently led to considerable DNA deletions or insertions but rare single-nucleotide (nt) substitutions in *N. benthamiana* cells (Fig. 1e, f and Supplementary Fig. 2a, b), as in animal and human cells displaying relatively high mutation rates (e.g., 37.6% in K562 cells and 24.6% in 293T cells)^{7–14}. In contrast, single-nucleotide deletions, insertions or substitutions were most frequently detected in *Arabidopsis* cells with relatively low mutation rates ranging from 1.1% to 5.6% (Fig. 1c, d). The use of high-fidelity DNA polymerase in amplifying these short (~300 bp) target regions (Supplementary Methods) and the absence of mutagenesis in control experiments (Fig. 1c, d) excluded the possibility that the single-nucleotide mutations observed in



Arabidopsis protoplasts were introduced through PCR amplification. Our results demonstrate that the sgRNA:pcoCas9 system is effective in plant cells. Whether the different genome mutagenesis frequencies and patterns in *Arabidopsis* and *N. benthamiana* are due to distinct plant genotypes or physiological states requires future investigation.

To validate the occurrence of the sgRNA:pcoCas9-induced targeted mutagenesis in *PDS* *in planta*, we transiently co-expressed *pcoCas9* and *AtPDS3*- or *NbPDS*-targeting sgRNA on a single binary plasmid (Fig. 1g) in intact leaves of 2-week-old *Arabidopsis* seedlings or 5-week-old *N. benthamiana* plants through *Agrobacterium* leaf infiltration (agroinfiltration). Biallelic disruption of *PDS* in the *Arabidopsis* or *N. benthamiana* genome would be expected to abolish carotenoid biosynthesis and promote chlorophyll oxidation, leading to a photobleached phenotype. We did not observe any visible albino spot on agroinfiltrated leaves from *Arabidopsis* or *N. benthamiana* plants 7 days after

infiltration. This suggests that there were either no cells with biallelic disruption of *PDS* or the population of photobleached cells was too small, as the cell division might have ceased in the infiltrated leaves. Careful screens of single cells after the degradation of existing chlorophyll is necessary for further characterization using fluorescent microscopy. By sequencing target gPCR amplicons, we did detect precise genomic mutations in the *AtPDS3* and *NbPDS* target sequence in cells from agroinfiltrated leaves with a mutagenesis frequency of 2.7% for *AtPDS3* and 4.8% for *NbPDS* (Fig. 1g). Considering that agroinfiltration has lower efficiency and higher variability in gene transfer than the protoplast transfection¹⁵, these mutagenesis frequencies might reflect dilution of the targeted mutations by wild-type gDNA from leaf cells without successful DNA delivery. Notably, the different sgRNA:pcoCas9-induced mutagenesis patterns in *Arabidopsis* and *N. benthamiana* protoplasts were also observed in corresponding whole plants. Although targeted mutations in *Arabidopsis*

seedlings were frequently single-nucleotide substitutions (Fig. 1h), those in *N. benthamiana* plants often involved longer DNA deletions (Fig. 1i). The leaves infiltrated with *Agrobacterium* expressing *pcoCas9* alone did not induce mutations in the target regions (Fig. 1g). These data show that the sgRNA:pcoCas9 system is also effective *in planta*.

To test whether the sgRNA:pcoCas9 system allows multiplex genome editing in *Arabidopsis* protoplasts, we first identified an identical sgRNA target site (target candidate no. 2, Supplementary Fig. 3) for both *AtRACK1b* and *AtRACK1c*, two members of the *Arabidopsis* RECEPTOR FOR ACTIVATED C KINASE 1 (*RACK1*) family (Fig. 2a). By co-expressing *pcoCas9* and the cognate sgRNA, we observed mutations in both target genes with a similar mutagenesis frequency (2.5–2.7%; Fig. 2b). Only single-nucleotide substitutions or insertions were detected in these *Arabidopsis* genes (Fig. 2b). Notably, no mutation was detected in a homologous sequence from *AtRACK1a* (Fig. 2b), which contains a valid protospacer adjacent motif (PAM) but two mismatches to the 12-nt seed sequence governing the sgRNA specificity^{7–9,13,16,17} (Fig. 2a), illustrating the high specificity of the sgRNA:pcoCas9-directed genome editing in plant cells. We further co-expressed *pcoCas9* and tandem sgRNAs aiming for two juxtaposed targets in *AtPDS3* with a 24-bp spacer (Fig. 2c). Interestingly, this simultaneous targeting with two sgRNAs led to deletions of up to 48 bp genomic segments between these two target sites by sgRNA:pcoCas9 with a mutation frequency of 7.7% (Fig. 2d and Supplementary Fig. 2c). Taken together, these results demonstrated that the sgRNA:pcoCas9 system could facilitate multiplex genome editing in plants.

We next addressed whether the presence of a DNA donor upon sgRNA:pcoCas9-mediated generation of a double-strand break would lead to gene replacement by HDR, which could precisely integrate an intended mutation from the DNA donor into the target site. We co-expressed *pcoCas9* and the sgRNA aiming for the *NbPDS* target 1 in *N. benthamiana* protoplasts and concurrently supplied a double-stranded DNA donor that contains a unique *AvrII* site flanked by a 533-bp left homology arm and a 114-bp right homology arm to the *NbPDS* locus (Fig. 2e). *AvrII* digestion of gPCR amplicons spanning the *NbPDS* target site revealed *AvrII* incorporation in the target locus with a frequency of 10.7%, and this incorporation strictly relied on both sgRNA and the DNA

donor (Fig. 2f). Sanger sequencing further verified the anticipated creation of the AvrII site in the target sequence without additional modifications and indicated an HDR-mediated gene replacement at a frequency of 9.0% (Fig. 2g). In addition, we detected NHEJ-mediated targeted mutagenesis at the *NbPDS* locus with a frequency of 14.2% (Supplementary Fig. 4). As mesophyll protoplasts are isolated from differentiated leaves without active cell division, we tested the possibility of enhancing HDR by triggering ectopic cell division. Co-expression of *Arabidopsis* *CYCD3* (*CYCLIN D-TYPE 3*), a master activator of the cell cycle, hardly promoted the HDR in *N. benthamiana* protoplasts (Fig. 2f). Exploration of HDR in *Arabidopsis* protoplasts was unsuccessful, presumably owing to intrinsically low efficiency of HDR in *Arabidopsis*¹⁸.

To facilitate genome-wide application of the sgRNA:pcoCas9 technology in *Arabidopsis*, we generated, using bioinformatics, a database containing a total of 1,466,718 unique sgRNA target sequences in *Arabidopsis* exons (Supplementary Database), which cover >99% (26,942 out of 27,206) of the nuclear protein-encoding genes defined by TAIR10 (The *Arabidopsis* Information Resource 10, http://arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp/). Targeting efficacy and specificity of selected sgRNA target candidates from this database need to be experimentally determined each time during future implementation. We also introduced a facile method to manually design a shared sgRNA target site specific for multiple homologous target genes by aligning their coding sequences and carrying out a BLAST search to evaluate off-target possibilities (Supplementary Fig. 3). The sgRNA:pcoCas9 technology enables an easy reprogramming of DNA targeting specificity by changing the 20-nt guide sequence in the sgRNA without modifying the pcoCas9 protein. We have established a simple and rapid procedure to create a custom sgRNA through overlapping PCR (Supplementary Fig. 5 and Supplementary Table 1). Thus, it is feasible to use single or tandemly expressed sgRNAs (Fig. 2c) to simultaneously target multigene families, which is not easily done with ZFNs and TALENs.

We have tested a total of seven target sequences in five target genes in *Arabidopsis* or *N. benthamiana*, and obtained targeted mutagenesis in all cases. The variation in mutagenesis efficiency among different genes in *Arabidopsis* may stem from distinct sgRNA binding strength to individual target

sequences or distinct chromatin structure and epigenetic state at individual target loci, which requires future investigation. We have demonstrated that plant protoplasts provide a useful system to rapidly evaluate the efficiency of the sgRNA:pcoCas9-mediated genome editing at a specific genomic locus. Our data also suggest that targeting an *Arabidopsis* gene with multiple sgRNAs could improve the success rate of targeted mutagenesis and generate deletions to ensure gene knockout. Notably, sgRNA:pcoCas9 achieved high efficiency of HDR-mediated gene replacement in *N. benthamiana* protoplasts. The simplicity and versatility of the sgRNA:pcoCas9 technology demonstrated in this work promise marker gene-independent and antibiotic selection-free genome engineering with high precision in diverse plant species to advance basic science and biotech.

Note: Supplementary information is available in the online version of the paper (doi:10.1038/nbt.2654).

ACKNOWLEDGMENTS

We thank F. Ausubel for critical reading of the article, D. Voytas for discussion on the HDR strategy, Y. Xiong for the *CYCD3* expression plasmid. J.E.L. is supported by the MGH ECOR Postdoctoral Fellowship for Medical Discovery. The Research is supported by the Department of Energy grant DE-FG02-02ER63445 to G.M.C., the National Science Foundation grant IOS-0843244 and the National Institutes of Health grants R01 GM60493 and R01 GM70567 to J.S.

AUTHOR CONTRIBUTIONS

J.-F.L. and J.S. designed experiments; J.-F.L. and D.Z. performed experiments; J.A., J.E.N., M.M. and G.M.C.

conducted bioinformatics analyses; J.B. supplied plant materials; J.-F.L. and J.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Jian-Feng Li^{1,2}, Julie E Norville^{2,3}, John Aach^{2,3}, Matthew McCormack^{1,2}, Dandan Zhang^{1,2}, Jenifer Bush^{1,2}, George M Church^{2,3} & Jen Sheen^{1,2}

¹Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. ²Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. ³Wyss Institute for Biological Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA. e-mail: sheen@molbio.mgh.harvard.edu

- Zhang, F. *et al.* *Proc. Natl. Acad. Sci. USA* **107**, 12028–12033 (2010).
- Zhang, Y. *et al.* *Plant Physiol.* **161**, 20–27 (2013).
- Li, T. *et al.* *Nat. Biotechnol.* **30**, 390–392 (2012).
- Gaj, T. *et al.* *Trends Biotechnol.* **31**, 397–405 (2013).
- Mussolino, C. & Cathomen, T. *Nat. Biotechnol.* **31**, 208–209 (2013).
- Streubel, J. *et al.* *Nat. Biotechnol.* **30**, 593–595 (2012).
- Jinek, M. *et al.* *Science* **337**, 816–821 (2012).
- Cong, L. *et al.* *Science* **339**, 819–823 (2013).
- Mali, P. *et al.* *Science* **339**, 823–826 (2013).
- Hwang, W.Y. *et al.* *Nat. Biotechnol.* **31**, 227–229 (2013).
- Cho, S.W. *et al.* *Nat. Biotechnol.* **31**, 230–232 (2013).
- Jinek, M. *et al.* *eLIFE* **2**, e00471 (2013).
- DiCarlo, J.E. *et al.* *Nucleic Acids Res.* **41**, 4336–4343 (2013).
- Wang, H. *et al.* *Cell* **153**, 910–918 (2013).
- Yoo, S.D. *et al.* *Nat. Protoc.* **2**, 1565–1572 (2007).
- Qi, L.S. *et al.* *Cell* **152**, 1173–1183 (2013).
- Jiang, W. *et al.* *Nat. Biotechnol.* **31**, 233–239 (2013).
- de Pater, S. *et al.* *Plant Biotechnol. J.* **11**, 510–515 (2013).

Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease

To the Editor:

Sustainable intensification of crop production is essential to ensure food demand is matched by supply as the human population continues to increase¹. This will require high-yielding crop varieties that can be grown sustainably with fewer inputs on less land. Both plant breeding and genetic modification (GM) methods make valuable contributions to varietal improvement, but targeted genome engineering promises to be critical to elevating future yields. Most such methods require targeting DNA breaks to defined locations followed by either nonhomologous end joining (NHEJ) or homologous recombination². Zinc finger nucleases (ZFNs) and transcription activator-like effector

nucleases (TALENs) can be engineered to create such breaks, but these systems require two different DNA binding proteins flanking a sequence of interest, each with a C-terminal FokI nuclease module. We report here that the bacterial clustered, regularly interspaced, short palindromic repeats (CRISPR) system, comprising a CRISPR-associated (Cas)9 protein and an engineered single guide RNA (sgRNA) that specifies a targeted nucleic acid sequence³, is applicable to plants to induce mutations at defined loci.

To test the potential of the Cas9 system to induce gene knockouts in plants, we took advantage of *Agrobacterium tumefaciens*-mediated transient expression assays (agroinfiltration) to co-express a Cas9 variant