nature biotechnology

Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion

Yuan Zong[1,2,5], Yanpeng Wang[1,2,5], Chao Li[1,2], Rui Zhang[1], Kunling Chen[1], Yidong Ran[3], Jin-Long Qiu[4], Daowen Wang[1] & Caixia Gao[1]

Targeted base editing in plants without the need for a foreign DNA donor or double-stranded DNA cleavage would accelerate genome modification and breeding in a wide array of crops. We used a CRISPR–Cas9 nickase-cytidine deaminase fusion to achieve targeted conversion of cytosine to thymine from position 3 to 9 within the protospacer in both protoplasts and regenerated rice, wheat and maize plants at frequencies of up to 43.48%.

Crop improvement might be accelerated if genetic mutations could be easily incorporated into modern varieties[1]. Genome-wide association studies have shown that single-base changes generate elite trait variation in crop plants[2,3]. Current methods of identifying point mutations (e.g., targeting induced local lesions in genomes, TILLING) are time-consuming and often detect a limited repertoire of point mutations[4,5]. Genome editing in plants via homologous-recombination-mediated DNA repair with CRISPR–Cas9 is inefficient and delivery of DNA repair templates is also challenging[6–8]. Fusion of a Cas9 variant with cytidine deaminase enables the editing of single bases in mammalian genomes and yeasts when directed by single guide RNAs (sgRNAs)[9–12], and targeted single-nucleotide substitutions can be made without either double-strand breaks or a foreign DNA donor. Here we describe the optimization of base-editing in cereal crops.

We used two fusion proteins: nCas9-PBE (plant base editor) and dCas9-PBE. They were composed of either rat cytidine deaminase APOBEC1, a Cas9 variant (Cas9-D10A nickase (nCas9), or catalytically dead Cas9 (dCas9)), and uracil glycosylase inhibitor (UGI), both of which base-edit precisely in human cells[9]. The fusion constructs were codon optimized for cereal plants, and cloned under the maize *Ubiquitin-1* (Ubi-1) gene promoter to generate pnCas9-PBE and pdCas9-PBE (**Fig. 1a** and **Supplementary Fig. 1**).

First, we tested whether pnCas9-PBE and pdCas9-PBE could edit the coding sequence of blue fluorescent protein (BFP) to produce green fluorescent protein (GFP). This requires changing codon 66 from CAC (histidine) to TAC (tyrosine) (**Fig. 1b**). In human cells, this base-editing system created C to T substitutions in a window from position 4 to 8 in the protospacer, counting from the distal end to the protospacer-adjacent motif (PAM)[9]. Therefore, we designed an sgRNA–BFP (**Supplementary Table 1**) with the desired C at position 4 (C_4) of the protospacer (Online Methods). Then we introduced pnCas9-PBE, pUbi-BFPm and pOsU3-BFP-sgRNA (pTaU6-BFP-sgRNA) into rice (and wheat) protoplasts by PEG-mediated transformation. This yielded 5.8% GFP^+ rice cells (6.8% in wheat) whereas substituting pdCas9-PBE yielded only 0.5% GFP^+ rice cells (0.3% in wheat) (**Fig. 1c** and **Supplementary Fig. 2**).

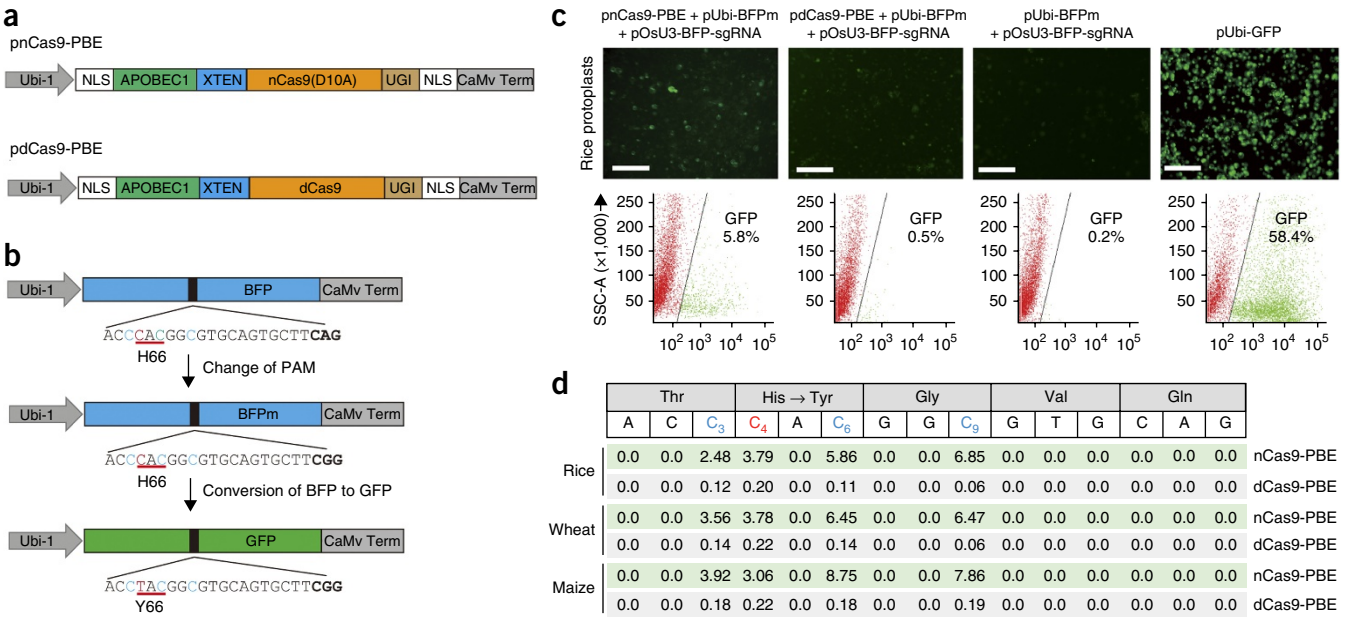
Deep amplicon sequencing revealed that ~4.00% of the 175,000 DNA reads harbored C to T substitutions at C_4 in rice, wheat and maize protoplasts transfected with pnCas9-PBE, pUbi-BFPm and the appropriate sgRNA constructs (Online Methods) (**Fig. 1d**). All the substitutions in rice, wheat and maize were in the C's at positions 3, 6 and 9 of the protospacer, with frequencies of 2.48–3.92%, 5.86–8.75% and 6.47–7.86%, respectively (**Fig. 1d**). The same Cs were mutated when pdCas9-PBE replaced pnCas9-PBE, but frequencies (0.06–0.22%) were much lower (**Fig. 1d**).

We next edited endogenous genes. We designed an sgRNA for each of three rice genes (*OsCDC48*, *OsNRT1.1B* and *OsSPL14*), three different sgRNAs (S1, S2 and S3) for wheat *TaLOX2* (ref. 13) and one sgRNA for maize *ZmCENH3* (ref. 14; **Supplementary Table 1**). As a control, we used wild-type Cas9 (pCas9) to produce deletion and/or insertion mutations (indels). Base-editing in protoplasts of the respective plants was assessed by sequencing 100,000–270,000 reads per locus. pnCas9-PBE induced C to T conversions in all seven target sites, with deamination windows spanning positions 3 to 9 of the protospacers (**Fig. 2a** and **Supplementary Table 2**), which was broader than those in human cells[9]. The frequency of single C to T substitutions was 0.39–7.07%, and it was highest at or near position 7 (**Fig. 2a**). Multiple Cs (two to five) occurred in 0.31–12.48% of the cases; simultaneous editing of two or three Cs was the most common occurrence (**Supplementary Table 2**). pnCas9-PBE produced far fewer indels than pCas9 (0.01–0.22% vs. 6.27–11.68%) (**Fig. 2b**). pnCas9-PBE produced many more editing changes than pdCas9-PBE (0.00–1.29%) (**Fig. 2a** and **Supplementary Table 3**), but a similar numbers of indels (<0.06%) (**Fig. 2b**).

To regenerate base-edited mutant plants, we targeted rice *OsCDC48* (**Fig. 2c** and **Supplementary Table 1**), which regulates senescence and cell death[15], using *Agrobacterium*-mediated transformation (**Supplementary Fig. 3a**). Examination of 92 pH-nCas9-PBE-transformed plants revealed that 40 harbored at least one C to T substitution in the target region (mutation efficiency of 43.48%) (**Fig. 2d** and **Supplementary Fig. 3b**). Substitutions at positions 3, 4, 7 and 8 of the protospacer were found (in seven combination) (**Fig. 2d**).

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]Genovo Biotechnology Co., Ltd, Tianjin, China. [4]State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. [5]These authors contributed equally to this work. Correspondence should be addressed to C.G. (cxgao@genetics.ac.cn).

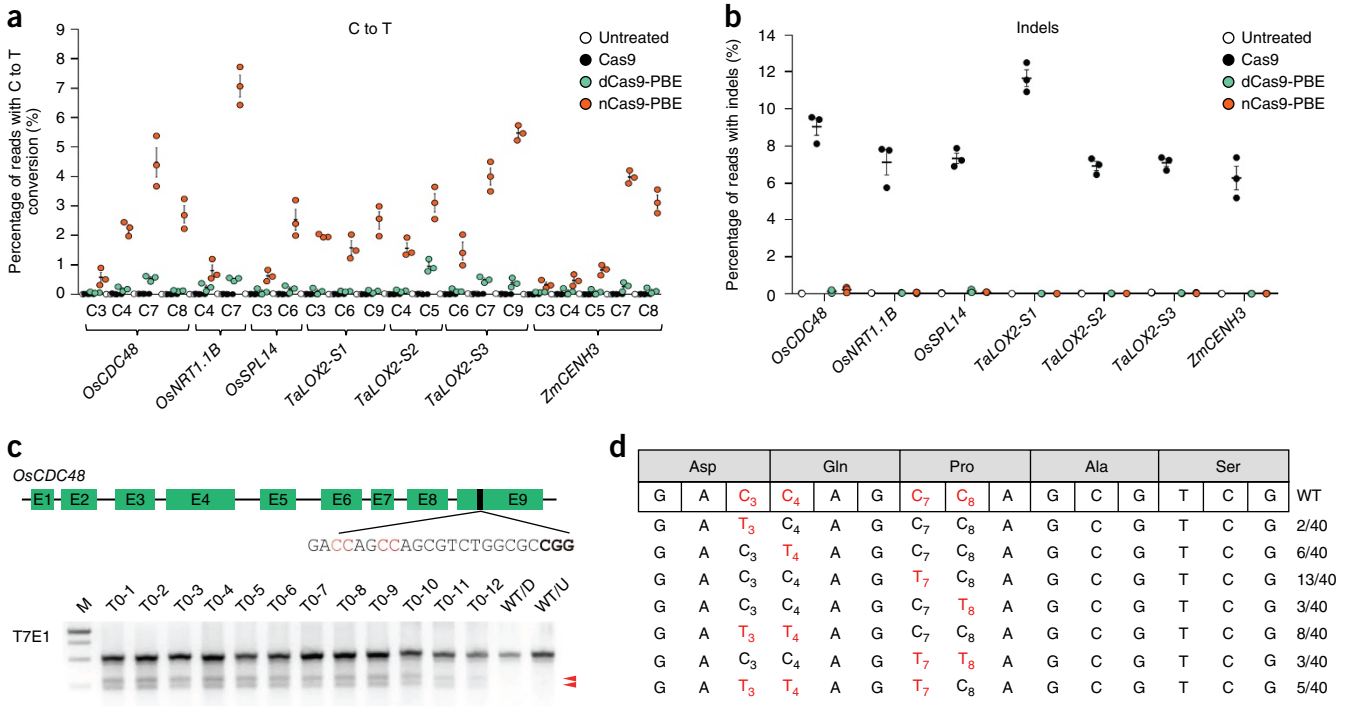
Received 31 October 2016; accepted 5 February 2017; published online 27 February 2017; doi:10.1038/nbt.3811



	Thr			His → Tyr			Gly			Val			Gln			
	A	C	C_3	C_4	A	C_6	G	G	C_9	G	T	G	C	A	G	
Rice	0.0	0.0	2.48	3.79	0.0	5.86	0.0	0.0	6.85	0.0	0.0	0.0	0.0	0.0	0.0	nCas9-PBE
	0.0	0.0	0.12	0.20	0.0	0.11	0.0	0.0	0.06	0.0	0.0	0.0	0.0	0.0	0.0	dCas9-PBE
Wheat	0.0	0.0	3.56	3.78	0.0	6.45	0.0	0.0	6.47	0.0	0.0	0.0	0.0	0.0	0.0	nCas9-PBE
	0.0	0.0	0.14	0.22	0.0	0.14	0.0	0.0	0.06	0.0	0.0	0.0	0.0	0.0	0.0	dCas9-PBE
Maize	0.0	0.0	3.92	3.06	0.0	8.75	0.0	0.0	7.86	0.0	0.0	0.0	0.0	0.0	0.0	nCas9-PBE
	0.0	0.0	0.18	0.22	0.0	0.18	0.0	0.0	0.19	0.0	0.0	0.0	0.0	0.0	0.0	dCas9-PBE

Figure 1 Changing BFP to GFP in protoplasts using n/dCas9-PBE. (**a**) The plant base-editing (PBE) vectors, pnCas9-PBE and pdCas9-PBE. (**b**) The BFP-to-GFP reporter system for comparing nCas9-PBE and dCas9-PBE editing in rice, wheat and maize protoplasts. The four Cs, in the presumed deamination window, are shown in red (conversion to T leads to the H66Y substitution) and blue, respectively. (**c**) Flow cytometry of rice BFP to GFP conversion using nCas9-PBE. Scale bars, 800 μm. (**d**) Frequency (%) of C to T conversion in reads of the BFPm target by nCas9-PBE and dCas9-PBE in rice, wheat and maize protoplasts.

Two of the mutants were homozygous (one C_7 to T_7, one C_8 to T_8) (**Supplementary Fig. 3c**). Among the heterozygotes were four single-base substitutions, two double-base substitutions and one triple-base substitution (**Fig. 2d**). The deamination window (3 to 8) was consistent with the protoplast results. The mutation frequency of individual Cs ranged from 5.00% (C_3, 2/40) to 32.50% (C_7, 13/40) (**Fig. 2d**), and no indels were observed in the target region. We did not detect mutations in the potential off-target regions (Online Methods) (**Supplementary Tables 4** and **5**). By contrast, neither point mutations nor indels were detected among the 87 plants transformed with pH-dCas9-PBE.



	Asp			Gln			Pro			Ala			Ser			
	G	A	C_3	C_4	A	G	C_7	C_8	A	G	C	G	T	C	G	WT
	G	A	T_3	C_4	A	G	C_7	C_8	A	G	C	G	T	C	G	2/40
	G	A	C_3	T_4	A	G	C_7	C_8	A	G	C	G	T	C	G	6/40
	G	A	C_3	C_4	A	G	T_7	C_8	A	G	C	G	T	C	G	13/40
	G	A	C_3	C_4	A	G	C_7	T_8	A	G	C	G	T	C	G	3/40
	G	A	T_3	T_4	A	G	C_7	C_8	A	G	C	G	T	C	G	8/40
	G	A	C_3	C_4	A	G	T_7	T_8	A	G	C	G	T	C	G	3/40
	G	A	T_3	T_4	A	G	T_7	C_8	A	G	C	G	T	C	G	5/40

Figure 2 n/dCas9-PBE editing of endogenous genes in rice, wheat and maize. (**a**,**b**) Frequencies of single C to T conversion (**a**) and indels (**b**) in the seven target sites. Frequencies were calculated from $n = 3$ replicates. Error bars, mean ± s.e.m. (**c**) sgRNA sequence used to target a site in *OsCDC48*. Lower panel: T7E1 assays of 12 representative C-to-T conversions. Arrowheads indicate the bands anticipated from T7E1. (**d**) The seven different mutations (single and multiple C-to-T conversions) and their frequencies among the 40 mutant plants regenerated for *OsCDC48* by nCas9-PBE.

We also delivered pnCas9-PBE and pTaU6-LOX2-S1-sgRNA constructs (**Fig. 1a** and **Supplementary Fig. 4a**) into immature wheat embryos by particle bombardment, and regenerated plants without herbicide selection, as described previously[16]. We obtained two independent heterozygous mutant plants, T0-3 and T0-7, from 160 embryos (**Supplementary Fig. 4b** and **Supplementary Table 5**). Sanger sequencing revealed that T0-3 had C to T substitutions at positions 3, 6 and 9, while T0-7 had one C to T substitution at position 3 (**Supplementary Fig. 4b**). Again, no indels were observed in the target region of either plant. Furthermore, PCR screening with six primer sets, specific for pnCas9-PBE and pTaU6-LOX2-S1-sgRNA (**Supplementary Fig. 5** and **Supplementary Table 5**), confirmed that these base-edited plants did not carry the transgene vectors.

Finally, we analyzed mutations introduced into the CENP-A targeting domain (CATD) of ZmCENH3 (**Fig. 2a** and **Supplementary Fig. 6**). *Agrobacterium*-mediated transformation generated 208 independent T0 transgenic maize lines using the binary vector pBnCas9-PBE, expressing the same sgRNA as used in the protoplast test (**Supplementary Fig. 7a**,**b**). Sanger sequencing identified 21 independent T0 lines carrying C to T mutations. All 21 plants had C_7 to T_7 synonymous substitutions (**Supplementary Fig. 7c**). An experiment still ongoing with different maize genotypes may explain why only this C to T substitution was recovered in transgenic plants when several other substitutions were found in base-edited maize protoplasts.

In summary, nCas9-PBE can accomplish efficient and site-specific C to T base editing in rice, wheat and maize. In these plants, it has a deamination window covering 7 bases of the protospacer and produces virtually no indel mutations. Our findings, together with base-editing reported in rice[17–19], suggest that base-editing is more efficient than TILLING and homologous-recombination-mediated generation of point mutations in crops.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We acknowledge P. Liu for deep sequencing analysis and T. Li for technical support in flow cytometry. This work was supported by grants from the National Key Research and Development Program of China (2016YFD0101804), the National Transgenic Science and Technology Program (2016ZX08010-002), the Chinese Academy of Sciences KFZD-SW-107 and GJHZ1602 and the National Natural Science Foundation of China (31570369).

AUTHOR CONTRIBUTIONS

Y.Z., Y.W. and C.G. designed the experiments; Y.Z., Y.W., C.L., R.Z. and K.C. performed the experiments; Y.R. advised on wheat tissue culture work; D.W., J.-L.Q., Y.Z. and C.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Voytas, D.F. & Gao, C. *PLoS Biol.* **12**, e1001877 (2014).
2. Zhao, K. *et al. Nat. Commun.* **2**, 467 (2011).
3. Henikoff, S. & Comai, L. *Annu. Rev. Plant Biol.* **54**, 375–401 (2003).
4. Henikoff, S., Till, B.J. & Comai, L. *Plant Physiol.* **135**, 630–636 (2004).
5. Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. & Facciotti, D. *Nat. Biotechnol.* **23**, 75–81 (2005).
6. Hsu, P.D., Lander, E.S. & Zhang, F. *Cell* **157**, 1262–1278 (2014).
7. Svitashev, S. *et al. Plant Physiol.* **169**, 931–945 (2015).
8. Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y. & Voytas, D.F. *Genome Biol.* **16**, 232 (2015).
9. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. *Nature* **533**, 420–424 (2016).
10. Nishida, K. *et al. Science* http://dx.doi.org/10.1126/science.aaf8729 (2016).
11. Ma, Y. *et al. Nat. Methods* **13**, 1029–1035 (2016).
12. Hess, G.T. *et al. Nat. Methods* **13**, 1036–1042 (2016).
13. Feng, B. *et al. J. Cereal Sci.* **52**, 387–394 (2010).
14. Zhong, C.X. *et al. Plant Cell* **14**, 2825–2836 (2002).
15. Huang, Q.N. *et al. J. Integr. Plant Biol.* **58**, 12–28 (2016).
16. Zhang, Y. *et al. Nat. Commun.* **7**, 12617 (2016).
17. Lu, Y. & Zhu, J.K. *Mol. Plant* http://dx.doi.org/10.1016/j.molp.2016.11.013 (2016).
18. Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. *et al. Mol. Plant* http://dx.doi.org/10.1016/j.molp.2016.12.001 (2016).
19. Ren, B. *et al. Sci. China Life Sci.* http://engine.scichina.com/doi/10.1007/s11427-016-0406-x (2016).

ONLINE METHODS

Construction of pn/dCas9-PBE and sgRNA expression vectors for protoplast assay. APOBEC1, XTEN, nCas9 (D10A), dCas9 (D10A and H840A) and UGI sequences were codon-optimized for cereal plants, and synthesized commercially (GenScript, Nanjing, China). The resultant fragment, APOBEC1-XTEN-UGI, was cloned into the pUC57 vector, yielding the plasmid pUC57-APOBEC1-XTEN-UGI. The full-length coding sequence of n/dCas9 was amplified using the primer set AflII-F and MluI-R to introduce AflII and MluI restriction sites. The PCR products were digested with AflII and MluI, then were subcloned into pUC57-APOBEC1-XTEN-UGI to generate pUC57-APOBEC1-XTEN-n/dCas9-UGI. Subsequently, the fusion cistrons APOBEC1-XTEN-n/dCas9-UGI were amplified using the primer set BamHI-F and Bsp1047I-R, and the PCR products were digested with BamHI and Bsp1047I and placed downstream of the Ubi-1 promoter in the plasmid pJIT163-Ubi-GFP by replacing the GFP coding sequence, which generated the plasmids pnCas9-PBE and pdCas9-PBE. The construction of sgRNA expression vectors under the control of the promoters *OsU3*, *TaU6* and *ZmU3* for rice, wheat and maize, respectively, was performed as previously described[20–22]. All primer sets used in this work are listed in **Supplementary Table 5**.

Construction of sgRNA-BFP and pUbi-BFPm for protoplast assay. We designed an sgRNA-BFP (**Supplementary Table 1**) with desired C at position 4 (C_4) of the protospacer, with CAG as the downstream protospacer-adjacent motif (PAM). Because CAG is not an optimal PAM for CRISPR–Cas9, we mutated it to CGG, which did not alter fluorescence emission of BFP or GFP. The resultant BFP sequence (BFPm) was cloned downstream of the Ubi-1 promoter to generate the expression construct pUbi-BFPm (**Fig. 1b**). The sgRNA-BFP was transcribed under the control of the promoters *OsU3*, *TaU6* and *ZmU3*, as described in our earlier work[20–22].

Protoplast assay. We used the spring wheat variety Bobwhite, the *Japonica* rice variety Nipponbare, and the maize inbred line Zong31 for preparing the protoplasts used in this study. Protoplast isolation and transformation were performed as previously described[20–22]. Plasmid DNA (10 μg per construct) was introduced into the desired protoplasts by PEG-mediated transfection, with the mean transformation efficiency being 55–70%. At 48 h after transfection, the protoplasts were collected, and genomic DNAs were extracted for deep amplicon sequencing and T7E1 and PCR restriction enzyme digestion assay (PCR-RE assay) (see below).

Deep amplicon sequencing. The genomic DNA samples extracted from the desired protoplast samples at 48 h after transfection were used as the templates. In the first-round PCR, the target region was amplified using site-specific primers (**Supplementary Table 5**). In the second-round PCR, both forward and reverse barcodes were added to the end of the PCR products for library construction (**Supplementary Table 5**). Equal amounts of the final PCR products were pooled. Then the samples were sequenced commercially (Beijing Genomics Institute, Beijing, China) using the Illumina High-Seq 4000 platform. The sgRNA target sites in the sequenced reads were examined for C to T substitutions and indels. For each target site, the amplicon sequencing was repeated three times using the genomic DNA samples extracted from three independent protoplast samples.

Construction of pH/B-n/dCas9-PBE vectors for rice and maize transformation. The APOBEC1-XTEN-n/dCas9-UGI fragment was fused to StuI and SacI-digested pHUE411 (for rice)[23] or pBUE411 (for maize)[23] with the primer set Gibson-F and Gibson-R using the Gibson cloning method[24], yielding the vectors pHUE411/BUE411-APOBEC1-XTEN-n/dCas9-UGI. Pairs of oligonucleotides including *OsCDC48* or *ZmCENH3* targeting sequence were synthesized, annealed and cloned into BsaI-digested pHUE411 or pBUE411 to generate pHUE411-sgRNA-CDC48 and pBUE411-sgRNA-CENH3. Then the sgRNA-CDC48 and sgRNA-CENH3 expression cassettes were excised using PmeI and AvrII, and were subcloned into pHUE411/BUE411-APOBEC1-XTEN-n/dCas9-UGI that had been digested using the same two enzymes. The resultant constructs (pH/B-n/dCas9-PBE) were used to transform the rice cultivar Nipponbare or the maize inbred Zong31 through *Agrobacterium*-mediated transformation (see below).

***Agrobacterium*-mediated transformation of rice callus cells.** *Agrobacterium tumefaciens* strain AGL1 was transformed by the pH-n/dCas9-PBE binary vectors by electroporation. *Agrobacterium*-mediated transformation of the callus cells of Nipponbare was conducted as reported before[20]. Hygromycin (50 μg/ml) was used to select the transgenic plants.

Biolistic delivery of DNA constructs into wheat immature embryo cells. The plasmid DNAs of pnCas9-PBE and pTaU6-LOX2-S1-sgRNA were simultaneously delivered into the immature embryos of Bobwhite via particle bombardment as previously described[16]. After the bombardment, the embryos were cultured for plantlet regeneration on the media without selective agent[16].

***Agrobacterium*-mediated transformation of maize immature embryo cells.** Maize Zong31 plants were grown in the field. Immature embryos, isolated at 12–14 d after anthesis, were used for *Agrobacterium*-mediated transformation, as previously described[22]. Bialaphos-resistant plantlets were regenerated after 9–12 weeks of culture on the selection medium.

Mutant identification by T7E1, PCR-RE assays and Sanger sequencing. T7E1, PCR-RE assays and Sanger sequencing were conducted as described previously[20–22] to identify the rice, wheat or maize mutants harboring C to T conversions in the target regions. For rice and maize, the T0 transgenic plants were examined individually. For wheat, plantlets (usually 3–4) derived from each bombarded immature embryo were pooled for the assays, and the positive pools were examined further to identify individual mutant plantlets[16].

Off-target detection. We examined potential off-target effects of the base editing method among the 40 pH-nCas9-PBE-induced rice mutants of *OsCDC48*. Five potential off-target sites were identified in the Nipponbare genome with Cas-OFFinder[25], each of which had three nucleotide mismatches to the sgRNA target region of *OsCDC48*. Amplicons from the five sites were carefully examined using the T7E1 assay and Sanger sequencing, but neither point mutations nor indels were detected (**Supplementary Tables 4** and **5**).

Data availability. The authors declare that all data supporting the findings of this study are available in the article and its supplementary figures and tables or are available from the corresponding author on request. For sequence data, Rice LOC_Os ID (http://rice.plantbiology.msu.edu/): LOC_Os03g05730 (*OsCDC48*), LOC_Os10g40600 (*OsNRT1.1B*), LOC_Os08g39890 (*OsSPL14*); NCBI GenBank: GU167921 (*TaLOX2*), AF519807 (*ZmCENH3*). NCBI Sequence Read Archive: SRR5215214.

20. Shan, Q. *et al. Nat. Biotechnol.* **31**, 686–688 (2013).
21. Wang, Y. *et al. Nat. Biotechnol.* **32**, 947–951 (2014).
22. Liang, Z., Zhang, K., Chen, K. & Gao, C. *J. Genet. Genomics* **41**, 63–68 (2014).
23. Xing, H.L. *et al. BMC Plant Biol.* **14**, 327 (2014).
24. Kurowska, M. *et al. J. Appl. Genet.* **52**, 371–390 (2011).
25. Bae, S., Park, J. & Kim, J.S. *Bioinformatics* **30**, 1473–1475 (2014).