nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-020-0566-4

Check for updates

Precise, predictable multi-nucleotide deletions in rice and wheat using APOBEC–Cas9

Shengxing Wang[1,4], Yuan Zong[1,4], Qiupeng Lin[1,2,4], Huawei Zhang[1,4], Zhuangzhuang Chai[1,2], Dandan Zhang[2,3], Kunling Chen[1], Jin-Long Qiu[2,3] and Caixia Gao[1,2]✉

Short insertions and deletions can be produced in plant genomes using CRISPR–Cas editors, but reliable production of larger deletions in specific target sites has proven difficult to achieve. We report the development of a series of APOBEC–Cas9 fusion-induced deletion systems (AFIDs) that combine Cas9 with human APOBEC3A (A3A), uracil DNA-glucosidase and apurinic or apyrimidinic site lyase. In rice and wheat, AFID-3 generated deletions from 5′-deaminated C bases to the Cas9-cleavage site. Approximately one-third of deletions produced using AFID-3 in rice and wheat protoplasts (30.2%) and regenerated plants (34.8%) were predictable. We show that eAFID-3, in which the A3A in AFID-3 is replaced with truncated APOBEC3B (A3Bctd), produced more uniform deletions from the preferred TC motif to the double-strand break. AFIDs could be applied to study regulatory regions and protein domains to improve crop plants.

The CRISPR–Cas9 system has been widely applied for genome engineering[1,2]. In brief, a single guide (sg)RNA-guided Cas9 nuclease generates chromosomal double-strand breaks (DSBs), which can have both blunt and staggered ends due to flexible cleavage by the RuvC domain[3]. These DSBs are mainly repaired by nonhomologous end joining (NHEJ), which results in frequent short insertions and deletions (indels)[4–7]. However, the heterogeneity of these small indels makes it technically challenging to target small functional regulatory elements and domains, such as *cis*-acting elements, micro (mi)RNAs and their binding sites, and genomic regions encoding protein domains, which are promising targets for gene function studies, gene therapy and crop improvement[8–12]. Several strategies are available for generating targeted larger deletions, but all have drawbacks, such as limited scope and unpredictability of the deletions produced[13–18]. In particular, the paired guide RNA strategy requires two appropriately spaced and active sgRNAs, which limits its scope, and often generates long deletions outside the desired target[13–15]. The I-TevI:Cas9 fusion cleaves the target DNA strands and generates ~33- to 36-bp deletions; these are thus relatively constant in length but they are hard to produce because they require two different recognition sites (NGG protospacer adjacent motif (PAM) and CNNNG motif) at a specific distance from each other[16]. Exonucleases such as Trex2 and T5 tend to excise one or more nucleotides randomly on either side of the Cas9-mediated DSB, leading to a series of unpredictable deletions[17,18]. The recently developed prime editor uses an engineered Cas9 nickase–reverse transcriptase fusion protein paired with an engineered prime editing guide (peg)RNA to produce desired deletions, insertions and nucleotide substitutions[19], but its efficiency in plants needs to be improved, especially in the case of low-efficiency targets[20]. Thus, new strategies are required to generate predictable multi-nucleotide deletions over the whole genome for precision editing.

Cytosine base editors, consisting of a cytidine deaminase fused with a Cas9 nickase (nCas9) and the uracil glycosylase inhibitor, have been used to achieve C-to-T substitutions in many organisms[2,21]. When an sgRNA pairs with the target DNA strand, the nontarget DNA strand is left as a single-stranded bubble, and this helps the cytidine deaminase to catalyze C-to-U base substitutions[21,22]. In the base excision repair (BER) pathway, uracil DNA glucosidase (UDG) recognizes U•G mismatches, excises uracil generated in genomic DNA and creates an abasic site that leads to nicking of the deaminated strand by apurinic or apyrimidinic site lyase (AP lyase)[21,23].

Based on the mechanism of cytidine deamination and BER, we developed AFIDs that produce predictable multi-nucleotide-targeted deletions within the protospacer (Fig. 1a). In these systems, APOBEC, UDG and AP lyase excise the deaminated C in the active window and produce a single-nucleotide gap in the deaminated strand, and Cas9 cuts both strands, resulting in a predictable single-strand deletion extending from the deaminated C to the Cas9 cleavage site (Fig. 1a). After that, unwinding of the sgRNA and the target strand exposes incompatible 5′-overhanging ends; these are then resected to blunt ends by the Artemis-like nuclease and joined by the DNA ligase complex[4,24], generating a double-strand deletion extending from the deaminated C to the DSB (Fig. 1a).

Results

Design and development of AFID systems. In a previous study of base editing in plants, human APOBEC3A (A3A) exhibited higher deaminase activity than APOBEC1 (ref. [25]), which was initially considered for this work. With the endogenous BER system in mind, we first fused A3A to the N-terminus of *Streptococcus pyogenes* Cas9 (Cas9) without and with *Escherichia coli* UDG fused to the C-terminus of Cas9, generating AFID-1 and AFID-2, respectively (Fig. 1b). To further increase the efficiency of AP site removal, *E. coli* AP lyase was fused with AFID-2 using the ribosomal skipping peptide (P2A), to produce AFID-3 (Fig. 1b).

We first measured the indel and base-editing activities of the three AFID systems in rice and wheat protoplasts using amplicon deep sequencing, with Cas9 as a control. Eight target sites were

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China. [2]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China. [3]State Key Laboratory of Plant Genomics, Institute of Microbiology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China. [4]These authors contributed equally: Shengxing Wang, Yuan Zong, Qiupeng Lin, Huawei Zhang. ✉e-mail: cxgao@genetics.ac.cn

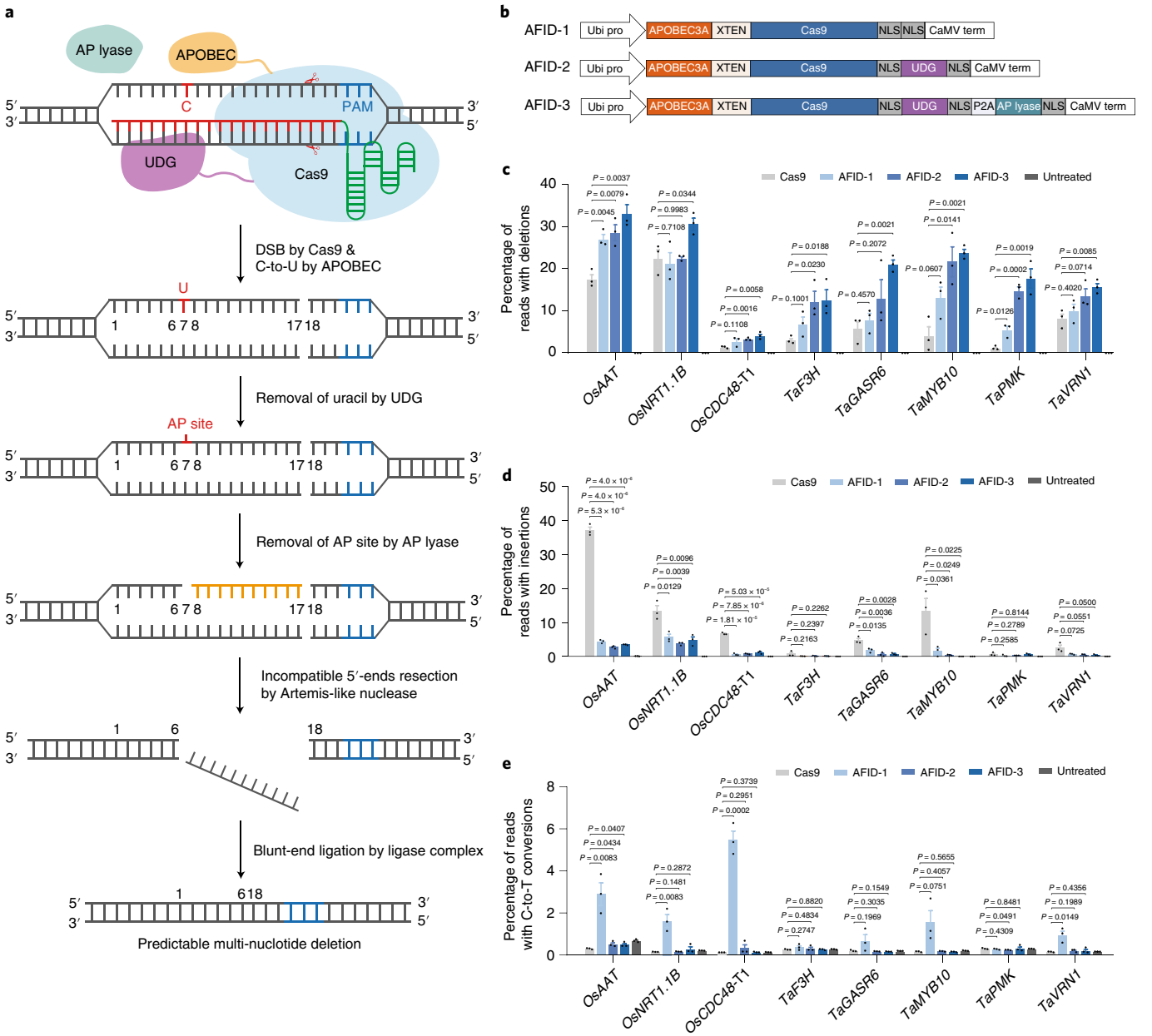
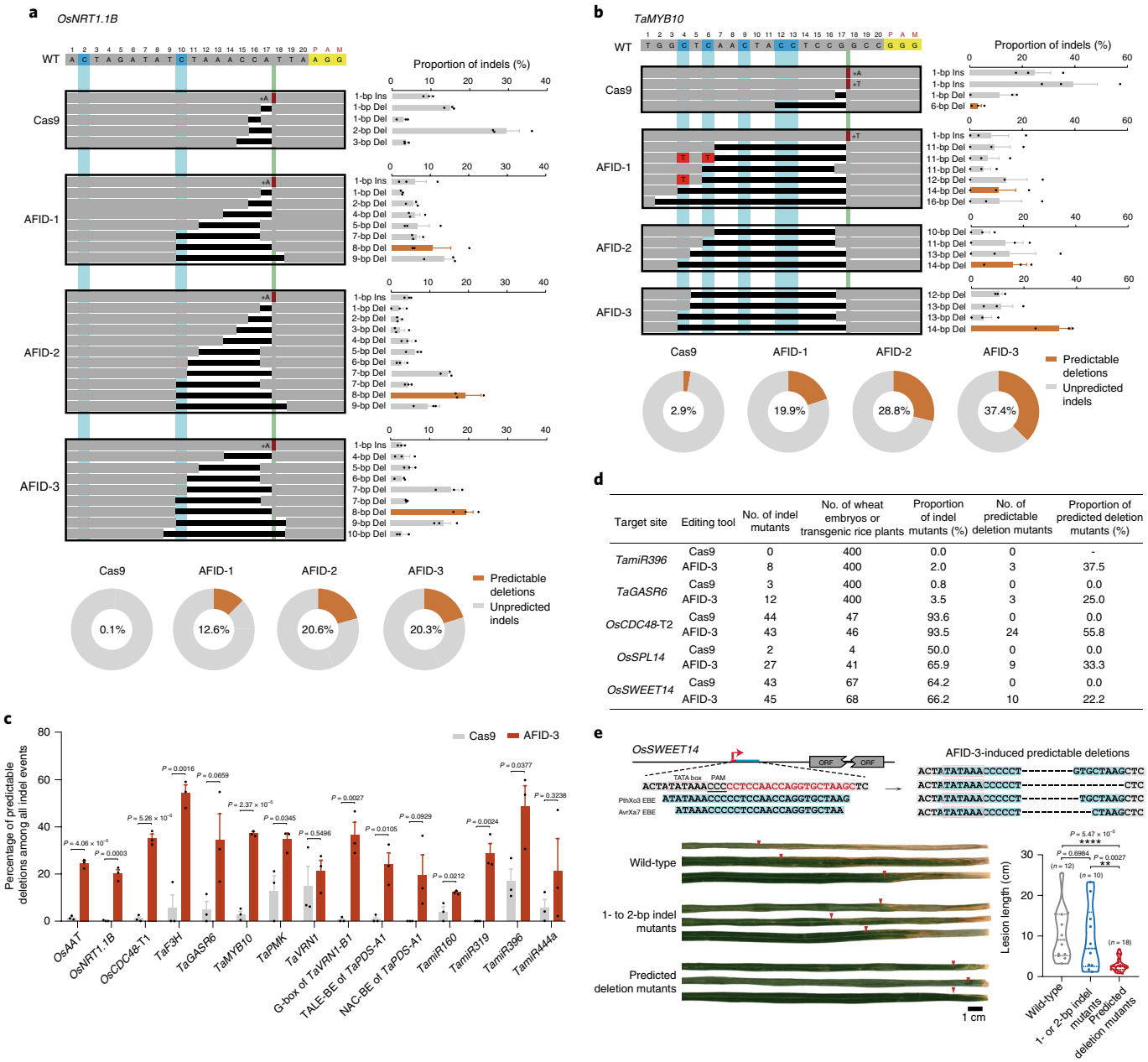


Fig. 1 | Characterization of the AFID system. a, Schematic representation of the AFID system. APOBEC deaminase converts cytidine to uridine on the nontarget strand and UDG then excises uracil from the uridine to generate an AP site, which is removed by AP lyase; Cas9 cuts both strands to form a DSB, leading via the NHEJ repair pathway to 'predictable' deletions extending from the deaminated C to the DSB. **b**, Structures of AFIDs 1–3. **c**,**d**, Comparison of deletion rates (**c**) and insertion rates (**d**) produced by Cas9 and AFIDs 1–3 at targets in rice and wheat protoplasts. Significant differences between Cas9 and AFIDs were tested using two-tailed Student's *t*-tests. **e**, Comparison of base-editing activities produced by Cas9 and AFIDs 1–3 at targets in rice and wheat protoplasts. Frequencies (mean ± s.e.m.) are based on three biologically independent experiments ($n = 3$) in **c**–**e**.

selected—three in rice genes (*OsAAT*, *OsNRT1.1B* and *OsCDC48*) and five in wheat genes (*TaF3H*, *TaGASR6*, *TaMYB10*, *TaPMK* and *TaVRN1*)—and sgRNAs were designed for each (see Supplementary Table 1). The AFID systems, especially AFID-3, had significantly higher deletion efficiencies at these targets than Cas9 ($P = 0.0019$–0.0344) and produced far fewer insertions (Fig. 1c,d). AFID-3 gave the highest deletion rates (4.1–33.1%), 1.4- to 14.6-fold higher than those created by Cas9, 1.2- to 3.3-fold higher than AFID-1 and up to 1.6-fold higher than AFID-2 (Fig. 1c). It is interesting that AFID-3 greatly increased the rates of those deletions that were inefficiently generated by Cas9, such as those at the *TaF3H* and *TaPMK* targets (Fig. 1c,d). We also found that AFID-1 produced a small number of C-to-T conversions at all eight targets with frequencies of 0.4–5.5%, whereas such base substitutions were at the background level for AFID-2 and AFID-3 (Fig. 1e). Evidently, the presence of UDG is essential for obtaining an efficient AFID system, and AP lyase also tends to bias editing events toward deletions.

AFIDs induce predictable deletions in protoplasts. We then compared the genetic changes induced by Cas9 and AFIDs 1–3 at the tested targets. Consistent with previous reports[6,12], the Cas9-induced changes were mainly 1-bp insertions or ~1- to 3-bp deletions, together with a few random deletions of >3 bp (Fig. 2a,b and Supplementary Figs. 1 and 2). In contrast, the deletions induced by



Target site	Editing tool	No. of indel mutants	No. of wheat embryos or transgenic rice plants	Proportion of indel mutants (%)	No. of predictable deletion mutants	Proportion of predicted deletion mutants (%)
TamiR396	Cas9	0	400	0.0	0	-
	AFID-3	8	400	2.0	3	37.5
TaGASR6	Cas9	3	400	0.8	0	0.0
	AFID-3	12	400	3.5	3	25.0
OsCDC48-T2	Cas9	44	47	93.6	0	0.0
	AFID-3	43	46	93.5	24	55.8
OsSPL14	Cas9	2	4	50.0	0	0.0
	AFID-3	27	41	65.9	9	33.3
OsSWEET14	Cas9	43	67	64.2	0	0.0
	AFID-3	45	68	66.2	10	22.2

Fig. 2 | AFID-3 induces predictable, multi-nucleotide-targeted deletions in protoplasts and regenerated plants. a,**b**, Comparison of insertion/deletion mutation types created by Cas9 and AFIDs 1–3 at *OsNRT1.1B* in rice protoplasts (**a**) and *TaMYB10* in wheat protoplasts (**b**). Each mutation type shown was filtered by >2% of indels with two or more repeat experiments. The brown bars represent the predictable deletions extending from the deaminated C bases to the DSB. Del, deletion; Ins, insertion; WT, wild-type. **c**, The proportions of predictable deletions among all indel events generated by Cas9 and AFID-3 at 15 target sites in rice and wheat protoplasts. Significant differences between Cas9 and AFID-3 were tested using two-tailed Student's *t*-tests, and proportions (mean ± s.e.m.) were calculated from three biologically independent experiments ($n = 3$) in **a**–**c**. **d**, Mutation rates produced by Cas9 and AFID-3 in T0-regenerated wheat and rice plants. **e**, Differences in bacterial blight resistance between 1- and 2-bp indel mutants and predicted deletion mutants induced by AFID-3 in the effector-binding elements of *OsSWEET14*. The red arrowheads indicate the endpoint of leaf lesion caused by *Xoo*. Significant differences of lesion length among wild-type, 1- to 2-bp indel mutants and predicted deletion mutants were tested using two-tailed Student's *t*-tests; *n* represents the number of measured leaves, and the violin plot elements show phenotypic data distribution with the median and quartiles. $^{**}P \leq 0.01$; $^{****}P \leq 0.0001$.

AFID-1 were larger (mostly >3 bp), but it occasionally introduced some undesired C-to-T substitutions (Fig. 2a,b and Supplementary Figs. 1 and 2). As expected, the proportion of predictable deletions generated by the various editors at the eight target sites increased from Cas9 (average 6.1%) to AFID-1 (average 17.8%) to AFID-2 (average 31.9%), and finally to AFID-3 (average 32.7%) (Fig. 2a,b and Supplementary Figs. 1 and 2). Evidently, deletion length depends strongly on the position of the 5′-deaminated C and its susceptibility to deamination, for instance, the C-to-T conversion activity of A3A–PBE at C10 (61.5%) was higher than at C2 (15.2%) at the *OsNRT1.1B* target; correspondingly, the most frequent deletions induced by AFID-3 extended from C10 to the DSB (41.5%)

rather than from C2 to the DSB (Fig. 2a,b and Supplementary Figs. 1–3). In addition, we used AFID-3 to target four miRNA genes (*TamiR160*, *TamiR319*, *TamiR396* and *TamiR444a*) and three *cis*-acting elements (the G-box of *TaVRN1-B1* and the TALE- and NAC-binding elements of *TaPDS-A1*), with Cas9 as the control. AFID-3 generated a larger number of longer deletions (~3–17 bp), including predictable deletions, than Cas9, and these deletions were more effective in preventing the formation of pre-miRNAs and disrupting *cis*-acting elements[12] (see Supplementary Fig. 4). In summary, AFID-3 yielded many more predicted deletions of >3 bp in length at all 15 target sites including 4 miRNA genes and 3 *cis*-acting elements (average 30.8%, highest 54.5% at the *TaF3H* target) than Cas9, average 4.8% (Fig. 2c and Supplementary Figs. 1, 2 and 4).

It is interesting that we found that, compared with the AFID-3-generated, predictable deletion types, a considerable percentage of the other deletion types induced by AFID-3 harbored 1-bp insertions that were identical to the nucleotide adjacent to either the 3′-Cas9-cleavage site (average 8.4%) or the 5′-deaminated site (average 15.7%), or at both sites (average 2.6%) (see Supplementary Fig. 5a). We reasoned that the former occurred mainly because of flexible cleavage by the RuvC domain and its sliding one base upstream (distal to the PAM), followed by filling in of the resulting sticky end by DNA polymerase[6,15,26,27], whereas the 1-bp templated cytidylate insertions at 5′-deaminated sites were possibly due to competitive filling in by DNA polymerase, using the incompletely resected overhanging strand as a template[4,15] (see Supplementary Fig. 5b).

AFID-3-induced deletions recovered in regenerated plants. To see whether AFID-3 could induce predictable multi-nucleotide-targeted deletions in regenerated wheat and rice plants, we selected two sites (*TamiR396* and *TaGASR6*) in wheat, using particle bombardment, and two (*OsCDC48*-T2 and *OsSPL14*) in rice, using *Agrobacterium*-mediated transformation. In regenerated wheat plants, we obtained three predictable deletion mutants of *TamiR396* that prevented the formation of pre-miRNA and mature miRNA[12], at a frequency of 37.5% (Fig. 2d and Supplementary Fig. 6). Similarly, in rice plants AFID-3 generated predictable deletion mutants at frequencies of up to 55.8%, far surpassing the frequencies for Cas9 (Fig. 2d and Supplementary Fig. 7). Thus, AFID-3 can indeed induce predictable, targeted deletions in plants.

Multi-nucleotide deletions for blight resistance in rice. Rice *OsSWEET14* is induced by the transcription activator-like (TAL) effectors AvrXa7 and PthXo3 of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) to facilitate infection[28]. AvrXa7 and PthXo3 have similar effector-binding elements, which overlap with the TATA box in the promoter of *OsSWEET14* (Fig. 2e). We attempted to create predictable multi-nucleotide-targeted deletions using AFID-3, with Cas9 as a control. As expected, most of the Cas9-induced mutants were ~1- to 3-bp deletions, whereas >80% of the AFID-3-generated mutants were long deletions (>3 bp), with the proportion of predicted mutants reaching 22.2% (Fig. 2d). Subsequently, we assessed the bacterial blight resistance of these mutants, and found that the predictable deletion mutants had significantly smaller blight lesions than the mutants with ~1- to 2-bp indels ($P = 0.0027$) (Fig. 2e). Thus, the predictable multi-nucleotide-targeted deletions generated by AFID-3 outside the TATA box in the effector-binding element conferred enhanced resistance to bacterial blight without affecting plant growth[29,30].

Truncated APOBEC3B for uniform AFID deletions. Consistent with the wide deamination window of A3A[25], the AFID-3-induced predictable deletions varied in length (Fig. 2a,b and Supplementary Figs. 1–3). We found that A3Bctd displayed not only a higher base-editing efficiency but also a narrower window than other APOBEC deaminases, including APOBEC1 (refs. [22,31]), A3A[25], eA3A[32] and A3B[33,34] (Fig. 3a). To engineer AFIDs producing more uniform products, we therefore replaced A3A in AFID-3 with A3Bctd, generating the enhanced AFID-3 system (eAFID-3) (Fig. 3b). We found that eAFID-3 had higher overall deletion activity (average 17.2%) than Cas9 (average 6.4%) and AFID-3 (average 10.9%) (Fig. 3c), and generated a higher proportion of predictable deletions (average 32.6%) than AFID-3 (average 24.6%) and Cas9 (average 4.0%) over all tested target sites (Fig. 3d,e and Supplementary Figs. 8 and 9).

More importantly, when we compared the proportions of predictable deletion types extending from AC, TC, CC and GC motifs, the proportion of eAFID-3-generated, predictable deletions initiated from TC was on average 2.6-fold (maximum 5.4-fold for *OsCDC48*-T2) higher than in the case of AFID-3 (Fig. 3f), indicating that A3Bctd endowed eAFID-3 with an enhanced preference for TC motifs over AC, CC and GC motifs. On average, 71.3% of the predictable deletions generated by eAFID-3 over the eight tested target sites extended from the preferred TC motif to the DSB, compared with 46.8% for AFID-3 (Fig. 3g).

Given the promising characteristics of eAFID-3, we attempted to use it to obtain predictable in-frame, multi-nucleotide target deletions at two miRNA-binding sites in rice protoplasts. The eAFID-3-generated substantial numbers of predicable in-frame deletions at the miR156-binding site of *OsIPA1* (*OsIPA1*-miRT, 15.7%) and the miR396h-binding site of *OsGRF1* (*OsGRF1*-miRT, 30.2%), far higher than AFID-3 (7.0% at *OsIPA1*-miRT and 19.2% at *OsGRF1*-miRT) and Cas9 (none at either site). Such deletions might prevent the corresponding miRNAs from binding to the coding regions of positive regulator genes, and might thus improve plant growth and development by reducing transcription inhibition[10,35] (Fig. 3h,i and Supplementary Fig. 9). These findings underscore the usefulness of the AFID systems for manipulating regulatory DNAs.

Discussion

In the present study, we systematically combined Cas9, A3A/A3Bctd, UDG and AP lyase to create deletion systems (AFIDs) that generate predictable, multi-nucleotide-targeted deletions within the protospacer. In most cases, Cas9 nuclease creates blunt-end cuts, which can be efficiently ligated without any change by Ku-XRCC4-ligase IV in the NHEJ pathway, which limits its editing efficiency[4,16]. However, the presence of the additional deaminase, UDG and AP lyase can induce the AFID to generate noncompatible DNA ends at target sites; these help to avoid unproductive DNA repair and bias more of the editing events toward deletion mutagenesis (Fig. 1a,c,d). We have also shown that endogenous UDG activity is insufficient to remove all the generated uracils, so that it is very necessary to introduce an exogenous UDG from *E. coli*, humans, and so on, and overexpress it with Cas9. In contrast, overexpression of AP lyase increases deletion efficiency only to a limited extent (Fig. 1c), indicating that the endogenous AP lyase activity in rice and wheat cells is more or less adequate, and the more limited combination of APOBECs, Cas9 and UDG is also a good alternative when limited by carrier capacity.

Differing from predicting Cas9-induced outcomes using the machine-learning models including inDelphi[36], FORECasT[37] and SPROUT[38], AFID-induced outcomes can be predicted from the composition of the protospacer sequences in a more intuitive way, because they are strongly determined by the deamination activity and window of the loading deaminase. In this situation, the high deamination efficiency is the primary factor generating the 5′-deaminated C bases and subsequent single-nucleotide gaps for the final products, and the narrow deamination window dependent on base preference ensures the uniformity of the predictable products.

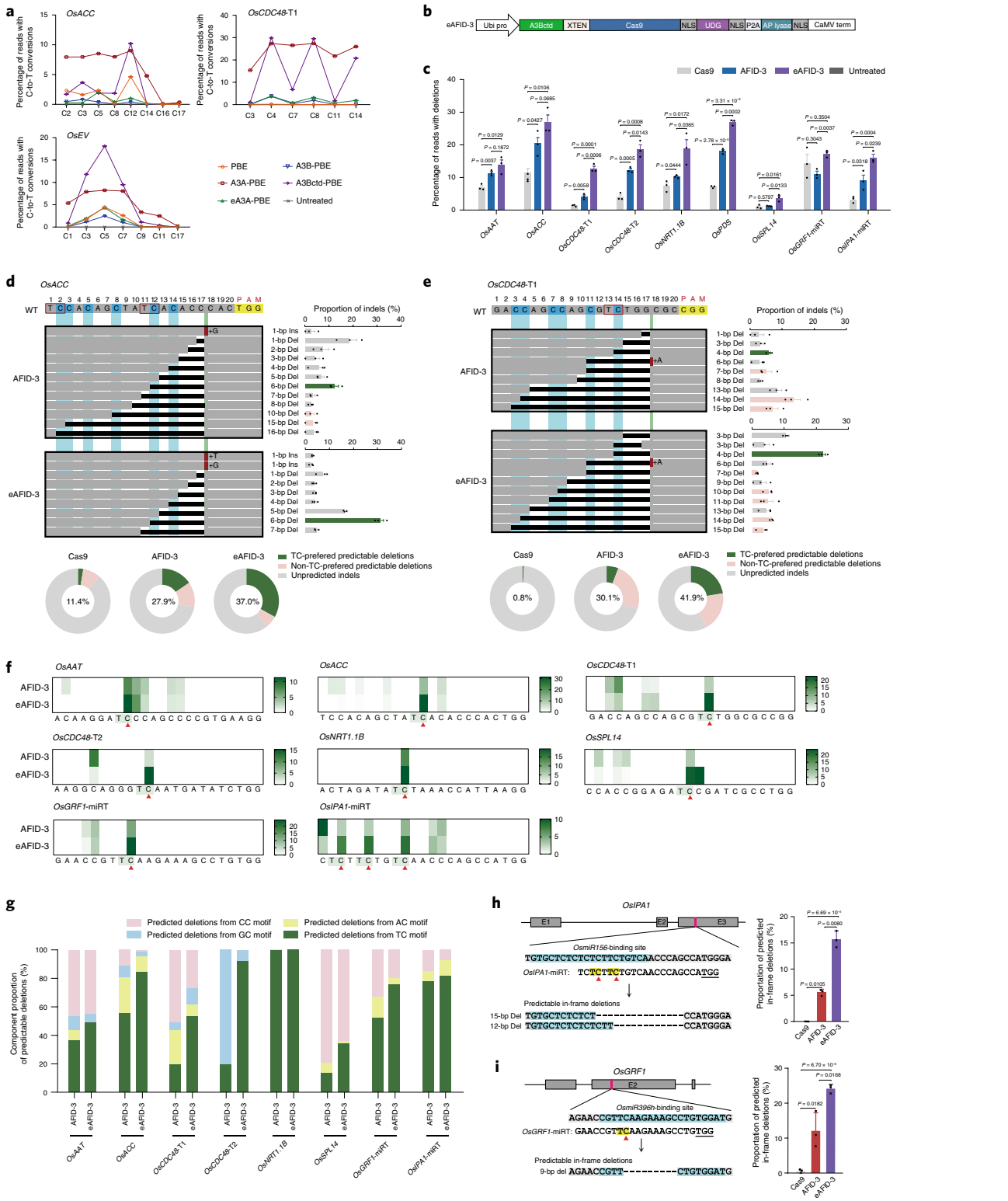


Fig. 3 | A3Bctd allows AFIDs to generate more uniform precise deletions within the protospacer. a, Comparison of base-editing efficiency and deamination window of five base editors in rice protoplasts. **b**, Structure of the eAFID-3 construct. **c**, Comparison of deletion rates induced by Cas9, AFID-3 and eAFID-3 at nine targets in rice protoplasts. **d**,**e**, Comparison of mutation types generated by AFID-3 and eAFID-3 at the target sites of *OsACC* (**d**) and *OsCDC48*-T1 (**e**). The green and light-red bars represented the deletion extending from the TC motif and other motifs (AC, CC and GC) to the DSB, respectively. Each mutation type shown was filtered by >2% of indels with two or more repeat experiments. **f**, Efficiencies of predictable deletions extending from AC, CC, GC and TC motifs to the DSB at nine targets in rice protoplasts sites using heatmaps. **g**, Proportions of predictable deletions from the TC motifs to the DSB at eight target sites. **h**,**i**, Comparison of predictable in-frame deletion types generated by Cas9, AFID-3 and eAFID-3 at the miR156-binding site of *OsIPA1* (**h**) and the miR396h-binding site of *OsGRF1* (**i**). The red arrowheads indicate the deaminated C bases for predictable in-frame deletions that can prevent the binding of miRNA. Frequencies or proportions (mean ± s.e.m.) were calculated from three biologically independent experiments ($n = 3$) in **a** and **c**–**i**, and significant differences among Cas9, AFID-3 and eAFID-3 were tested using two-tailed Student's *t*-tests in **c**, **h** and **i**.

After screening the deamination activity of different cytosine deaminases in protoplasts, we finally selected A3A and A3Bctd as the most promising deaminases for AFID constructions in the present study. As the wide deamination window of A3A, AFID-3 produced a variety of predictable deletions extending from the 5′-deaminated C to the Cas9 cleavage site; when using A3Bctd with enhanced TC preference to replace A3A, eAFID-3 generated larger numbers of predictable and uniform products from the preferred TC motif to the Cas9-generated DSB. Therefore, given the excellent performance of eAFID-3, we should be able to further modify A3Bctd in eAFID-3 using structure-guided, directed evolution approaches[34,39], to obtain variants with adequate activities and unique preferences for AC, CC, GC or TC motifs.

Many small regulatory elements, including miRNAs, miRNA-binding sites and *cis*-acting elements, comprise only about ~5–24 nucleotides; hence AFID-induced predictable deletions are superior to the mutants induced by current tools such as Cas9 (short indels), dual-Cas9 (>30-bp deletions), TevCas9 (33- to 36-bp deletions), Trex2-Cas9 and T5exo-Cas9 (random and unpredictable deletions)[13–18], for disrupting these regulatory elements and removing domains of a given gene in-frame, as well as for high-throughput screening of regulatory elements and domains; they might even be used to facilitate protein evolution[40]. Thus, AFIDs promise to provide robust deletion tools for basic research and genetic improvement.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0566-4.

Received: 20 January 2020; Accepted: 19 May 2020;
Published online: 29 June 2020

References

1. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
2. Chen, K., Wang, Y., Zhang, R., Zhang, H. & Gao, C. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* **70**, 667–697 (2019).
3. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
4. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
5. Stinson, B. M., Moreno, A. T., Walter, J. C. & Loparo, J. J. A mechanism to minimize errors during non-homologous end joining. *Mol. Cell* **77**, 1080–1091 (2020).
6. Taheri-Ghahfarokhi, A. et al. Decoding non-random mutational signatures at Cas9 targeted sites. *Nucleic Acids Res.* **46**, 8417–8434 (2018).
7. van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646 (2016).
8. Burgess, D. G., Xu, J. & Freeling, M. Advances in understanding *cis* regulation of the plant gene with an emphasis on comparative genomics. *Curr. Opin. Plant Biol.* **27**, 141–147 (2015).
9. Li, C. et al. A new rice breeding method: CRISPR/Cas9 system editing of the *Xa13* promoter to cultivate transgene-free bacterial blight-resistant rice. *Plant Biotech. J.* **18**, 313–315 (2020).
10. Li, M. et al. Reassessment of the four yield-related genes *Gn1a*, *DEP1*, *GS3*, and *IPA1* in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* **7**, 377 (2016).
11. Zhang, X., Wang, L., Liu, M. & Li, D. CRISPR/Cas9 system: a powerful technology for *in vivo* and *ex vivo* gene therapy. *Sci. China Life Sci.* **60**, 468–475 (2017).
12. Zhou, J. et al. CRISPR-Cas9 based genome editing reveals new insights into microRNA function and regulation in rice. *Front. Plant Sci.* **8**, 1598 (2017).
13. Bolukbasi, M. F. et al. Orthogonal Cas9–Cas9 chimeras provide a versatile platform for genome editing. *Nat. Commun.* **9**, 4856 (2018).
14. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
15. Shou, J., Li, J., Liu, Y. & Wu, Q. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell* **71**, 498–509 (2018).
16. Wolfs, J. M. et al. Biasing genome-editing events toward precise length deletions with an RNA-guided TevCas9 dual nuclease. *Proc. Natl Acad. Sci. USA* **113**, 14988–14993 (2016).
17. Čermák, T. et al. A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* **29**, 1196–1217 (2017).
18. Zhang, Q., Yin, K., Liu, G., Li, S. & Qiu, J. L. Fusing T5 exonuclease with Cas9 and Cas12a increases the frequency and size of deletion at target sites. *Sci. China Life Sci.* https://doi.org/10.1007/s11427-020-1671-6 (2020).
19. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
20. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
21. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
22. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
23. Lei, L. et al. APOBEC3 induces mutations during repair of CRISPR-Cas9-generated DNA breaks. *Nat. Struct. Mol. Biol.* **25**, 45–52 (2018).
24. Que, Q. et al. Plant DNA repair pathways and their applications in genome engineering. *Methods Protoc.* **1917**, 3–24 (2019).
25. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).
26. Gisler, S. et al. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nat. Commun.* **10**, 1598 (2019).
27. Zuo, Z. & Liu, J. Cas9-catalyzed DNA cleavage generates staggered ends: evidence from molecular dynamics simulations. *Sci. Rep.* **5**, 37584 (2016).
28. Li, T., Liu, B., Spalding, M. H., Weeks, D. P. & Yang, B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* **30**, 390–392 (2012).
29. Oliva, R. et al. Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nat. Biotechnol.* **37**, 1344–1350 (2019).
30. Xu, Z. et al. Engineering broad-spectrum bacterial blight resistance by simultaneously disrupting variable TALE-binding elements of multiple susceptibility genes in rice. *Mol. Plant* **12**, 1434–1446 (2019).
31. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
32. Gehrke, J. M. et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
33. St. Martin, A. et al. A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC-Cas9 or cleavage by Cas9 in living cells. *Nucleic Acids Res.* **46**, e84 (2018).
34. Siriwardena, S. U., Guruge, T. A. & Bhagwat, A. S. Characterization of the catalytic domain of human APOBEC3B and the critical structural role for a conserved methionine. *J. Mol. Biol.* **427**, 3042–3055 (2015).
35. Omidbakhshfard, M., Proost, S., Fujikura, U. & Mueller-Roeber, B. Growth-regulating factors (GRFs): a small transcription factor family with important functions in plant biology. *Mol. Plant* **8**, 998–1010 (2015).
36. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
37. Allen, F. et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–72 (2018).
38. Leenay, R. T. et al. Large dataset enables prediction of repair after CRISPR-Cas9 editing in primary T cells. *Nat. Biotechnol.* **37**, 1034–1037 (2019).
39. Turonyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
40. He, W. et al. De novo identification of essential protein domains from CRISPR-Cas9 tiling-sgRNA knockout screens. *Nat. Commun.* **46**, 505–529 (2019).

Methods

Plasmid construction. To construct vectors for the AFID systems, the *E. coli* UDG, AP lyase and A3Bctd deaminase sequences were obtained from the National Center for Biotechnology Information (NCBI) (GenBank ID: AMB53293.1, WP_115209270.1 and NM_004900.5, respectively), codon optimized for cereal plants and synthesized commercially (Genewiz). Cereal codon-optimized A3A deaminase sequence[25] was fused to the N-terminus of Cas9 with the XTEN linker in AFID-1, and codon-optimized UDG was fused to the C-terminus of AFID-1 by a 21-bp linker in AFID-2. The codon-optimized AP lyase was co-expressing with AFID-2 using P2A in AFID-3. In addition, A3A of AFID-3 was replaced by the A3Bctd deaminase sequence for constructing eAFID-3. All fusion protein sequences were cloned into the pJIT163 backbone for PEG-mediated protoplast transformation, and particle bombardment of immature wheat embryos. To construct the binary vector for *Agrobacterium*-mediated rice transformation, AFID-3 and the sgRNA expression cassettes were integrated into the pHUE411 backbone[41] using a ClonExpressII One Step Cloning Kit (Vazyme). The coding sequences of all AFID constructs are listed in Supplementary sequences. The constructs pOsU3-sgRNA and pTaU6-sgRNA were made as previously described[42,43], and all the sgRNA target sites and oligonucleotide sequences used in this work are listed in Supplementary Table 1 and were synthesized by Beijing Genomics Institute.

PEG-mediated protoplast transformation. We used the winter wheat variety Kenong199 and the *Japonica* rice variety Zhonghua11 to prepare protoplasts. Protoplast isolation and transformation were performed as described[42,43]. Cas9, PBE, A3A–PBE, eA3A–PBE, A3B–PBE, A3Bctd–PBE, AFID-1, AFID-2, AFID-3 and eAFID-3 were co-transformed into protoplasts with sgRNA vectors (1:1), and three biologically independent experiments were performed for each target site. Transformed protoplasts were incubated at 26 °C for 48 h, and then collected for extraction of genomic DNA.

DNA extraction and PCR amplification. Genomic DNA was extracted from protoplasts using the cetrimonium bromide method. The genomic region flanking the sgRNA target site was amplified with TransStart FastPfu DNA polymerase (TransGen Biotech) using site-specific primers (see Supplementary Table 2) in the first-round PCR. In the second round, the amplicons were amplified using nested PCR primers with different barcodes (see Supplementary Table 2). The amplicons were then fractionated by 2% agarose gel electrophoresis and purified with an AxyPrep DNA Gel Extraction kit (Axygen Biosciences) for library construction.

DNA library construction and amplicon deep sequencing. The amplicons were quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), and equal amounts (50 ng per sample) were pooled as separate libraries. The DNA libraries were constructed using a Second Generation Sequencing Rapid DNA Library kit (Illumina), and sequenced using an Illumina NovaSeq 6000 platform (Novogene). The sgRNA target sites in the sequenced reads were examined to analyze mutations type and rates. Analyses of base-editing processivity and indels were performed as previously described[22].

Biolistic transformation of immature wheat embryo cells. Plasmids AFID-3 and pTaU6-sgRNA were simultaneously delivered into immature embryos of Kenong199 via particle bombardment, as previously described[43]. After bombardment, the embryos were cultured on medium without a selective agent to regenerate plantlets.

***Agrobacterium*-mediated transformation of rice callus cells.** The binary vectors pH-Cas9 and pH-AFID-3 were transformed into *A. tumefaciens* strain AGL1 by electroporation. Callus cells of Zhonghua11 were used for *Agrobacterium*-mediated transformation, with hygromycin selection of transgenic plants[42,44].

Mutant identification by PCR-RE and Sanger sequencing. PCR-RE assays and Sanger sequencing were used to identify rice and wheat mutants with indels in target regions, as described previously[42,43].

Inoculation of rice and evaluation of bacterial blight resistance. All rice mutants were grown under 13-h light (28 °C):11-h dark (26 °C) conditions in a plant growth chamber. The TAL-free PH strains harboring the TAL effector AvaXa7 were grown in PS medium at 28 °C overnight, washed and resuspended to in sterile distilled water with an absorbance at 600 nm of 0.5 for inoculation. Leaves of 8-week-old, rice-regenerated plants were cut with scissors, and dipped in the bacterial suspensions. Lesion lengths were recorded at 13-d post-inoculation.

Statistical analysis. All numerical values are presented as mean ± s.e.m. Significant differences between controls and treatments were evaluated using two-tailed Student's *t*-tests.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

NGS data have been deposited in the NCBI Sequence Read Archive database (accession no. PRJNA630559). Two plasmids encoding AFID-3 and eAFID-3 in the present study will be available through Addgene. All data supporting the findings of the present study are available in the article and its supplementary figures and tables, or from the corresponding author on request. For sequence data, *OsAAT* (LOC_Os01g55540), *OsACC* (LOC_Os05g22940), *OsCDC48* (LOC_Os03g05730), *OsEV* (LOC_Os02g11010), *OsNRT1.1B* (LOC_Os10g40600), *OsSPL14*/*OsIPA1* (LOC_Os08g39890) and *OsSWEET14* (LOC_Os11g31190) are from Rice Genome Annotation Project (http://rice.plantbiology.msu.edu); *TaMYB10* (AB191458.1, AB191459.1, AB191460.1) and *TaVRN1* (AY747603.1, AY747604.1, AY747605.1) are from the NCBI (https://www.ncbi.nlm.nih.gov); *TaF3H* (TraesCS2A02G493500, TraesCS2B02G521500, TraesCS2D02G493400), *TaGASR6* (TraesCS1A02G270100, TraesCS1D02G270100), *TaPDS* (TraesCS4A02G004900, TraesCS4B02G300100, TraesCS4D02G299000) and *TaPMK* (TraesCS5A02G449000, TraesCS5B02G453800, TraesCS5D02G455500) are from the International Wheat Genome Sequencing Consortium (https://urgi.versailles.inra.fr); and tae-Pri-miR160, tae-Pri-miR319, tae-Pri-miR396 and tae-Pri-miR444a are from miRBase (http://www.mirbase.org).

References

41. Xing, H. L. et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327 (2014).
42. Shan, Q. et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
43. Wang, Y. et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951 (2014).
44. Zhang, Y., Zhang, Q. & Chen, Q. J. *Agrobacterium*-mediated delivery of CRISPR/Cas reagents for genome editing in plants enters an era of ternary vector systems. *Sci. China Life Sci.* https://doi.org/10.1007/s11427-020-1685-9 (2020).

Acknowledgements

We thank Professor G. Chen (Shanghai Jiao Tong University) for providing the PH strain containing the TAL effector AvaXa7 for evaluation of bacterial blight resistance. This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (Precision Seed Design and Breeding, grant no. XDA24020100), the National Transgenic Science and Technology Program (grant no. 2016ZX08010002), the National Natural Science Foundation of China (grant no. 31788103, 31971370), the National Key Research and Development Program of China (grant no. 2016YFD0101804), the Chinese Academy of Sciences (grant no. QYZDY-SSW-SMC030) and the Postdoctoral Innovative Talent Support Program of China (grant no. BX20190365).

Author contributions

S.W., H.Z. and C.G. designed the project. S.W., Y.Z., Q.L., H.Z., Z.C., K.C. and D.Z. performed the experiments. C.G. supervised the project. S.W., Y.Z., Q.L., H.Z, J.-L.Q. and C.G. wrote the manuscript.

Competing interests

The authors have submitted a patent application based on the results reported in this article.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0566-4.

Correspondence and requests for materials should be addressed to C.G.

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

nature research

Corresponding author(s): Caixia Gao

Last updated by author(s): May 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina NovaSeq platform was used to collect the amplicon deep sequencing data.
Data analysis	Graphpad prism 7 was used to analyze the data. Amplicon sequencing data of base-editing processivity and indels was analyzed using the published code as previously described in reference 22.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available in the article and its Supplementary Information files or are available from the corresponding author on request. Datasets of high-throughput sequencing experiments have been deposited with the National Center for Biotechnology Information (NCBI) Sequence Read Archive database (PRJNA630559) .

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The experiments of protoplasts were performed with three biological repeats. About 500,000 protoplasts were used for each transfection. The number of protoplasts in each transfection was measured by thrombocytometry. The experiment in rice and wheat regenerated plants was performed once, all the regenerated seedlings were sampled, the number of mutants were confirmed by PCR-RE and Sanger sequencing.
Data exclusions	No data exclusion.
Replication	All attempts for replication were successful. For the experiments in rice and wheat protolasts, a minimum of three independent experiments were included.
Randomization	Rice and wheat protoplasts were isolated and randomly separated to each transformation.
Blinding	Not applicable. As samples were processed identically through standard and in some cases automated procedures (DNA sequencing, transfection, DNA isolation) that should not bias outcomes.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and*

	any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*

Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.