Improved plant cytosine base editors with high editing activity, purity, and specificity

Qiurong Ren1,†, Simon Sretenovic2,†, Guanqing Liu3,4,†, Zhaohui Zhong1, Jiaheng Wang1, Lan Huang1, Xu Tang1, Yachong Guo1, Li Liu1, Yuechao Wu3,4, Jie Zhou1, Yuxin Zhao1, Han Yang1, Yao He1, Shishi Liu1, Desuo Yin2,5, Rocio Mayorga2, Xuelian Zheng1, Tao Zhang3,4,6,†, Yiping Qi1,†, and Yong Zhang1,†

1Department of Biotechnology, School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China
2Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Maryland, USA
3Jiangsu Key Laboratory of Crop Genetics and Physiology Key Laboratory of Plant Functional Genomics of the Ministry of Education/ Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Agricultural College of Yangzhou University, Yangzhou, China
4Department of Biotechnology, School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China
5Food Crop Institute, Hubei Academy of Agricultural Sciences, Wuhan, Hubei, China
6Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of Education of China, Yangzhou University, Yangzhou, China

Received 1 March 2021; revised 30 April 2021; accepted 17 May 2021.
*Correspondence (Tel: 1-301-405-8682; email yiping@umd.edu (Y. Q.); Tel 086-028-61830670; email: zhangyong916@uestc.edu.cn (Y. Z.); Tel 086-0514-87977729; email: zhangtao@yzu.edu.cn (T. Z.))
†These authors contributed equally to this work.

Keywords: cytosine base editors, rice, Arabidopsis, multiplexed base editing, off-target effects, whole genome sequencing.

Summary

Cytosine base editors (CBEs) are great additions to the expanding genome editing toolbox. To improve C-to-T base editing in plants, we first compared seven cytidine deaminases in the BE3-like configuration in rice. We found A3A/Y130F-CBE_V01 resulted in the highest C-to-T base editing efficiency in both rice and Arabidopsis. Furthermore, we demonstrated this A3A/Y130F cytidine deaminase could be improved by SpyMacCas9-mediated C-to-T base editing at A-rich PAMs. To showcase its applications, we first applied A3A/Y130F-CBE_V01 for multiplexed editing to generate microRNA-resistant mRNA transcripts as well as pre-mature stop codons in multiple seed trait genes. In addition, we harnessed A3A/Y130F-CBE_V01 for efficient artificial evolution of novel ALS and EPSPS alleles which conferred herbicide resistance in rice. To further improve C-to-T base editing, multiple CBE_V02, CBE_V03 and CBE_V04 systems were developed and tested in rice protoplasts. The CBE_V04 systems were found to have improved editing activity and purity with focal recruitment of more uracil DNA glycosylase inhibitors (UGIs) by the engineered single guide RNA 2.0 scaffold. Finally, we used whole-genome sequencing (WGS) to compare six CBE_V01 systems and four CBE_V04 systems for genome-wide off-target effects in rice. Different levels of cytidine deaminase-dependent and sgRNA-independent off-target effects were indeed revealed by WGS among edited lines by these CBE systems. We also investigated genomewide sgRNA-independent off-target effects by different CBEs in rice. This comprehensive study compared 21 different CBE systems, and benchmarked PmCDA1-CBE_V04 and A3A/Y130F-CBE_V04 as next-generation plant CBEs with high editing efficiency, purity, and specificity.

Introduction

Since the first demonstration in mammalian cells (Komor et al., 2016; Nishida et al., 2016), cytosine base editors (CBEs) have been undergoing rapid evolution for broadened editing scope, refined editing windows, reduced by-products, and enhanced specificity (Anzalone et al., 2020). In recent years, multiple CBEs have been developed and demonstrated in plants with different cytidine deaminases such as rAPOBEC1 (Li et al., 2017; Lu and Zhu, 2017; Zong et al., 2017), hAID (Ren et al., 2018), PmCDA1 (Shimatani et al., 2017; Tang et al., 2019), and human APOBEC3A (A3A) (Cheng et al., 2021; Zong et al., 2018) and APOBEC3B (A3B) (Jin et al., 2020).

Cytosine base editors are powerful tools for plant genetics and breeding. For example, CBEs can introduce non-sense or mis-sense mutations for genetic knockouts (Molla and Yang, 2019; Zhang et al., 2019b). CBEs have also been applied for achieving gain-of-function traits such as herbicide resistance in rice (Kuang et al., 2020) and wheat (Zhang et al., 2019a). Moreover, CBEs are also promising tools for editing cis elements to fine-tune gene expression. For example, increased sugar content in strawberry fruits was recently created with this approach (Xing et al., 2020).

Insertion and deletion (indel) mutations are major byproducts in C-to-T base editing applications. The conversion of C-to-U by cytidine deaminases used in CBEs typically triggers base excision repair, which is likely to cause DNA double-strand breaks and...
indel mutations due to imprecise non-homologous end-joining repair. Upgrading BE2 to BE3 architecture is based on the fusion of an inhibitor of uracil DNA glycosylase (UGI) to Cas9D10A nickase (nCas9), which increases the base editing efficiency and reduces the formation of indel byproducts (Komor et al., 2016). While BE3 only utilizes one copy of UGI, BE4 contains two copies of UGI fused to nCas9 in tandem, which showed higher editing efficiency and product purity (Komor et al., 2017). Similarly, co-expression of multiple copies of UGI with the self-cleaving 2A peptide also helped to improve C-to-T editing activity and purity in human cells (Wang et al., 2018; Wang et al., 2017) as well as in plants (Qin et al., 2019). Hence, it is possible that editing purity and activity may be further improved with the enhanced focal concentration of UGIs at the target site.

Recently, it was reported that rAPOBEC1-based BE3 resulted in genome-wide off-target C-to-T mutations in mice (Zuo et al., 2019) and rice (Jin et al., 2019). Engineered CBEs with minimized off-target effects have been developed and demonstrated in human cells, such as eA3A-BE3 (Gehrke et al., 2018), A3G-BES (Lee et al., 2020), and YE1-BE3-FNLS (Zuo et al., 2020). Despite the fact that a variety of CBEs have been demonstrated in plants (Gurel et al., 2020; Molla and Yang, 2019; Zhang et al., 2019b), the genome-wide off-target effects in plants remain untested for most of them. It is hence imperative to further develop CBE systems of high editing activity and purity, coupled with a genome-wide assessment of off-target effects of these improved CBEs and other widely used CBEs in plants.

Results

Comparison of CBE_V01 base editors in rice and Arabidopsis protoplasts

To improve C-to-T base editing in plants, we systematically assessed a series of CBEs with a primary focus on rice, a major crop. Four CG rich sites (OsCGRSS5, OsCGRSS6, OsCGRSS7 and OsCGRSS8) were chosen for the assessment (Figure 1a). First, we generated seven CBE_V01 editors in the BE3-like configuration (Figure 1b) which included different cytidine deaminases and engineered derivatives, such as rAPOBEC1 (Komor et al., 2016), PmCDA1 (Nishida et al., 2016; Shimatani et al., 2017; Tang et al., 2019), hAID (Ren et al., 2018), A3A (Wang et al., 2018; Zong et al., 2018), A3A/Y130F (Wang et al., 2018), eA3A (Gehrke et al., 2018), and eA3A/Y130F, which was a newly engineered A3A variant by introducing the Y130F mutation into eA3A. In rice protoplasts, A3A/Y130F-CBE_V01 showed the highest C-to-T conversion efficiency (~37.9% on average), followed by PmCDA1-CBE_V01, hAID-CBE_V01 and A3A-CBE_V01, while rAPOBEC1-CBE_V01 and eA3A-CBE_V01 showed poor editing activity and eA3A/Y130F-CBE_V01 showed no editing (Figure 1c). These CBEs showed different base editing windows (Figure 1d). A3A/Y130F-CBE_V01 had a broad editing window spanning from 4th to 15th nucleotide in the 5′ to 3′ direction within the protospacer (Figure 1d). In contrast, PmCDA1-CBE_V01 favours editing from first to sixth nucleotide in the protospacer (Figure 1d). Within the editing windows, all the Cytosines are editable by the top-performing editors at variable efficiencies (Figure S1). Indels and C-to-G editing byproducts were observed at one of the four target sites, OsCGRSS8 (Figure S2), indicating such byproducts are target-specific.

To further test A3A/Y130F-CBE_V01, we targeted the methylated CpG island in the promoter of AtFWA in Arabidopsis (Figure 1e). Among all four CBE_V01 editors compared, A3A/Y130F-CBE_V01 showed the highest editing efficiency in Arabidopsis protoplasts (Figure 1f), suggesting this CBE is efficient at editing methylated DNA, as in human cells (Wang et al., 2018). The editing window for this target site in Arabidopsis (Figures 1g and S3) also appeared to be consistent with those in rice (Figure 1d).

A3A/Y130F outperforms PmCDA1 and hAID using the iSpyMacCas9 based BE3 platform in rice

We further compared the three top cytidine deaminases (A3A/Y130F, PmCDA1 and hAID) by installing them into the iSpyMacCas9 platform which recognizes A-rich PAMs (Chatterjee et al., 2020; Sretenovic et al., 2021). In rice protoplasts, A3A/Y130F-iSpyMacCas9-CBE_V01 consistently outperformed the other two iSpyMacCas9 CBEs based on hAID and PmCDA1 at all four A-rich PAM sites (Figure 2a). We carried stable rice transgenesis with A3A/Y130F-iSpyMacCas9-CBE_V01 for a TAAA PAM site at the OsROC5-gRNA05 target site. It resulted in 68.8% base editing efficiency in transgenic T₀ rice lines (Figure 2b), much higher than the base editing efficiency achieved by PmCDA- and hAID-based CBE_V01 systems at the same target site (Sretenovic et al., 2021). Further analysis showed that C7 and C8 in the protospacer were favoured for C-to-T base conversion, generating monoallelic and biallelic editing events across many T₀ lines (Figure 2c). Occasionally, C-to-G editing and indels were detected in some T₀ lines (Figures 2c and S4).

Multiplexed base editing of two OsMIR156 target sites by A3A/Y130F-CBE_V01

We next sought to apply A3A/Y130F-CBE_V01 for multiplexed editing of important trait genes in rice. A multiplexed T-DNA construct was made for simultaneous base editing of the target sites of OsMI516 in OsSPL14 that controls ideal rice architecture (Jiao et al., 2010) and OsSPL16 that controls rice grain size, shape, and quality (Wang et al., 2012) (Figure 3a). The majority of T₀ rice lines (17/22, 77.3%) contained simultaneous edits at both genes (Figure S5). Interestingly, the majority of edited lines carried deletions at the OsSPL14-sgRNA01 site, while few edited lines had deletions at the OsSPL16-sgRNA01 site (Figure S5). This discrepancy could be explained by the presence of three consecutive Cytosines within the editing window at the OsSPL14-sgRNA01 site, making it prone for generating DSBs during the deamination reaction. We chose two edited T₀ lines for molecular analysis using qRT-PCR. The M₀-1979-4-3 line had monoallelic base editing at the OsSPL14-sgRNA01 site and biallelic base editing at the OsSPL16-sgRNA01 site. These edits had destroyed the OsMIR156 binding sites in the mRNAs of OsSPL14 and OsSPL16 (Figure 5). The M₀-1979-6-1 has a homozygous deletion at the OsSPL14-sgRNA01 site, creating a new resistance element at the OsSPL14-sgRNA01 site. This mutation also abolished the OsMIR156 binding sites in the mRNAs of OsSPL14 and OsSPL16 (Figure S5). Indeed, elevated expression of OsSPL14 and OsSPL16 was detected in these two edited lines (Figure 3b). Notably, there was only a two-fold increase of OsSPL14 expression in the M₀-1979-4-3 line, as opposed to a two-fold increase of OsSPL14 expression in the M₀-1979-6-1 line (Figure 3b). This could be well explained by the fact that only one allele of OsSPL14 was edited in the M₀-1979-4-3 line. Taken together, our data suggest mRNAs of multiple genes can be simultaneously base edited with efficient CBEs such as A3A/Y130F-CBE_V01, rendering them untargetable by certain miRNAs.
We previously used CRISPR-Cas9 to simultaneously edit OsGS3, OsGW2, and OsGN1a to improve grain yield (Zhou et al., 2019). To further demonstrate multiplexed base editing by A3A/Y130F-CBE_V01, we attempted to simultaneously knock out these genes by installing premature stop codons with C-to-T base editing (Figure 3c). Based on a single T-DNA multiplexed editing construct, we achieved very high multiplexed base editing efficiency in stable transgenic plants as more than half of the T0 rice lines (13/25, 52%) are triple mutants (Figure S6). We further characterized four T0 lines (M0-1980-7-1, M0-1980-2-1, M0-1980-14-3 and M0-1980-6-1) that were double and triple null mutants due to the successful installation of premature stop codons (Figure 3d). Indeed, we observed increased grain length (due to OsGS3 knockout) (Figure 3e,f) and grain width (due to OsGW2 knockout) (Figure 3e,g) (Ashikari et al., 2005; Zhou et al., 2019). We analysed the T1 generation of the M0-1980-7-2 T0 plant and found the mutations were inherited according to the Mendelian segregation pattern (Table S1), supporting germline transmission of these mutations. These data suggest A3A/Y130F-CBE_V01 is an efficient CBE for simultaneous knock out of multiple protein-coding genes by the installation of premature stop codons.

Targeted evolution of OsALS and OsEPSPS for herbicide resistance by A3AY130F-CBE_V01 and select sgRNA libraries

Base editors have a great potential for targeted protein evolution for improved traits such as herbicide resistance in plants (Zhang et al., 2019b; Zhang and Qi, 2019), as demonstrated recently (Kuang et al., 2020; Li et al., 2020; Ren et al., 2021b). It is critical to broaden the targeting range of base editing for its application in targeted evolution. To this end, we generated A3A/Y130F-nCas9-NG-CBE_V01 based on Cas9-NG (Nishimasu et al., 2018; Zhong et al., 2019). We first validated this new editor at five NG PAM sites in rice protoplasts. Base editing was detected at all five target sites, albeit with relatively low editing efficiency (Figure 4a). We then applied both A3A/Y130F-CBE_V01 and A3A/Y130F-nCas9-NG-CBE_V01 for the targeted evolution of herbicide resistance in rice. Our first target gene was OsALS. We targeted OsALS at 15 sites, including 10 NGG PAM sites, and 5
NG PAM sites, which were targeted by A3A/Y130F-CBE_V01, and A3A/Y130F-nCas9-NG-CBE_V01, respectively (Figure 4b). Screening of ~120 rice calli transformed with the pooled sgRNA library identified three T0 lines that survived the selection of 0.4 µM bispyribac sodium herbicide treatment (Figure 4c). One plant (M0-180-8) carried a D550N mutation and two plants (M0-180-15-1/2) carried the same D349N mutation (Figure 4d,e). While D349(A)/(E) mutations were previously documented in other plants (Yu and Powles, 2014), we demonstrated that D349N could also result in herbicide resistance in rice. The D550N mutation appears to be an entirely novel herbicide resistance allele in plants. Both mutations appeared to result in protein structural changes that reduce the binding affinity of OsALS to the herbicide (Figure 4d,e).

We also targeted OsEPSPS gene with a library of eight single guide RNAs (sgRNAs), for editing two NGG PAM sites and six NG sites, which were again targeted by A3A/Y130F-CBE_V01 and A3A/Y130F-nCas9-NG-CBE_V01, respectively (Figure 4f). Two T0 lines were identified conferring resistance to the herbicide glyphosate (Figure 4g). Sequence analysis revealed the P173L mutation in both lines (Figure 4h). The P173L mutation was previously reported in maize, but not in rice (Sammons and Gaines, 2014). Taken together, we showed a successful targeted evolution in two independent examples by cytosine base editing of OsALS and OsEPSPS. Notably, all the herbicide resistance alleles of OsALS and OsEPSPS were generated by A3A/Y130F-nCas9-NG-CBE_V01 that recognizes relaxed NG PAM sites.

Development and comparison of multiple CBE_V02 and CBE_V03 systems in rice protoplasts

We sought to further improve C-to-T base editing. First, we explored a CBE_V02 system based on flanking the nCas9 with (a) Comparison of three iSpyMacCas9-derived CBE_V01 systems at four target sites with A-rich PAMs in rice protoplasts. The error bars represent standard deviations of three biological replicates. (b) Base editing frequency of iSpyMacCas9-derived A3A/Y130F-CBE_V01 in stable rice lines. (c) Genotypes of base edited T0 lines. Note the edited positions are indicated by asterisks.
two different cytidine deaminases or putting a tandem fusion of
different cytidine deaminases to nCas9 (Figure S7A). We com-
pared these four CBE_V02 systems with three CBE_V01 systems
in rice protoplasts by editing OsCGRS55 and OsCGRS56 sites. In
general, these CBE_V02 editors showed reduced editing activity
when compared to the CBE_V01 editors (Figure S7B).

Figure 3 Application of A3A/Y130F-CBE_V01 for multiplexed editing of trait genes in rice. (a) Schematics of the target sites in OsSPL14 and OsSPL16. Note the sgRNAs were designed to mutate OsMIR156 target sites. (b) Quantitative RT-PCR analysis of OsSPL14 and OsSPL16 transcripts in edited T0 rice lines. Each target contains three biological replicates. Data are presented as mean values ± SD. (c) Schematics of the target sites in OsGS3, OsGW2 and OsGN1a. (d) Genotypes of four T0 lines containing multiplexed base editing events. (e) Grain size phenotypes of base-edited double and triple mutants. (f, g) Quantitative comparison of rice grain length and width in the wildtype and mutants (n = 10). Error bars represent standard deviations of 10 biological replicates. Different letters indicate significant differences (uppercase letters mean P < 0.01; lowercase letters mean P < 0.05; one-way ANOVA, Turkey test).
As MS2 coat protein (MCP) binds to the MS2 RNA aptamer in dimers, the gRNA2.0 scaffold with two MS2 stem-loops (Konermann et al., 2015) can recruit four additional copies of the effector protein when fused to MCP. Previously, MS2-MCP interaction was used for recruiting more transcriptional activators for enhanced CRISPR activation both in human cells (Konermann et al., 2015) and in plants (Lowder et al., 2018). We reasoned this system could be used for recruiting cytidine deaminase in base editing. Hence, we developed a CBE_03 system that uses the MS2-MCP interaction to recruit additional copies of the same cytidine deaminase (Figure S8A). This time, we tested the three CBE_03 editors in rice protoplasts at the target sites that we previously used for assessing CBE_01 editors, OsCGR555, OsCGR556, OsCGR557, and OsCGR558. These CBE_V03 editors, however, showed compromised editing activity (Figure S8B,C). Indel byproducts were detected at three out of four sites by A3A/Y130F-CBE_V03 (Figure S9), suggesting high-frequency of DNA DSB formation by the CBE_V03 system presumably due to the recruitment of multiple cytidine deaminases. Collectively, our experience of developing CBE_V02 and CBE_V03 systems showed that C-to-T base editing efficiency could not be enhanced by simply recruiting more cytidine deaminases, either by nCas9 fusion or by recruitment through the sgRNA scaffold.

Figure 4 Application of A3A/Y130F-CBE_V01 targeted evolution of herbicide resistance in rice. (a) Editing efficiency and window for A3A/Y130F-nCas9-NG-CBE_V01 at five target sites in rice protoplasts. The error bars represent standard deviations of three biological replicates. Data are presented as mean values ± SEM. (b) Schematics of the target sites in OsALS for targeted evolution of herbicide resistance. Note both Cas9 and Cas9-NG based A3A/Y130F-CBE_V01 systems were used. (c) Herbicide-resistant T0 lines selected by growing the transformed and regenerated rice seedlings on the rooting medium with 0.4 µM bispyribac sodium. Scale bar = 5 cm. (d, e) Genotypes (the upper panel) and OsALS protein structural models (the lower panel) for the herbicide resistance mutations (D550N and D349N) in T0 lines pLR180-8 and pRQ150-15-1/2. (f) Schematics of the target sites in OsEPSPS for targeted evolution of herbicide resistance. (g) Herbicide-resistant T0 lines selected by growing the transformed and regenerated rice seedlings on the regeneration medium with 100 µM glyphosate. (h) Genotypes of pRQ181-1-1 and pRQ181-1-2 that carry OsEPSPS alleles that are resistant to glyphosate.
Development and comparison of four CBE_V04 systems in rice protoplasts

High editing efficiency with reduced indel formation has been previously achieved by fusing more UGIs to nCas9 such as upgrading BE3 to BE4 (which contains two tandem copies of UGI fused to nCas9) (Komor et al., 2017) or co-expressing more UGIs (Wang et al., 2017). Given UGI is a relatively small protein (<100 Aa), we reasoned that the MS2-MCP strategy could be promising for recruiting UGI and hence an improvement of C-to-T base editing. Based on this strategy, one sgRNA could recruit four additional copies of MCP-UGI (Figure 5a), resulting in a high focal concentration of UGIs, unmatched by the previous approaches (Komor et al., 2017; Wang et al., 2017). We tested four such CBE_V04 editors in rice protoplasts. Compared to the CBE_V01 editors, these CBE_V04 editors did not show significant compromise in base editing efficiency (Figure 5b, as compared to Figure 1c). The editing windows of these CBE_V04 editors were also similar to those of the corresponding CBE_V01 editors (Figure 5c, as compared to Figure 1d). Consistent with the observations of the CBE_V01 platform, A3A/Y130F was also the most efficient one in the CBE_V04 platform (Figure 5b). By comparing A3A/Y130F-CBE_V04 and A3A/Y130F-CBE_V01 at 10 additional target sites in rice protoplasts, we found that A3A/Y130F-CBE_V04 showed comparable or higher editing activity than A3A/Y130F-CBE_V01 (Figure 5d). Bulked analysis of these 10 sites showed that A3A/Y130F-CBE_V04 had a significantly smaller dynamic range than A3A/Y130F-CBE_V04 for the base editing frequencies (Figure 5e), indicating A3A/Y130F-CBE_V04 is a more robust CBE. We further analysed indel frequencies across these ten target sites. At six out of ten target sites (OsALS-sgRNA01, OsALS-sgRNA03, OsALS-sgRNA04, OsALS-sgRNA10, OsALS-sgRNA14 and OsALS-sgRNA16), A3A/Y130F-CBE_V04 showed markedly reduced indel frequencies than A3A/Y130F-CBE_V01 (Figure 5f). At the remaining four target sites, the indel frequencies by A3A/Y130F-CBE_V01 were already quite low, which masked the potential improvement by A3A/Y130F-CBE_V04 (Figure 5f). The reduction of indel formation in samples edited by A3A/Y130F-CBE_V04 was very clear in a bulked analysis of all ten target sites (Figure 5g). Further analysis showed A3A/Y130F-CBE_V04 and A3A/Y130F-CBE_V01 had similar editing windows (Figure 5h). Together, these data suggest A3A/Y130F-CBE_V04 represents a more robust CBE with higher editing purity and without altering the base editing windows.

Improved base editing efficiency and purity of CBE_V04 systems over CBE_V01 systems in rice T0 lines

To further confirm that CBE_V04 systems are better editors than CBE_V01 systems, we compared the top three CBE_V01 editors (PmCDA1-CBE_V01, hAID-CBE_V01 and A3A/Y130F-CBE_V01) and the corresponding CBE_V04 editors based on the same cytidine deaminases in transgenic rice lines. These six CBEs were tested at editing the four target sites used in previous protoplast assessments: OsCGRS55, OsCGRS56, OsCGRS57, and OsCGRS58. A total of 24 T-DNA constructs were generated and used for making transgenic rice plants. For each construct, 20 to 33 individual T0 lines were obtained for genotyping. Among the three CBE_V01 systems, A3A/Y130F-CBE_V01 showed the highest average pure C-to-T editing efficiency (23.1%) (Table 1). Unlike the data in protoplasts, indels appeared to be major byproducts at all target sites for all three CBE_V01 systems with average editing efficiency over 50% (Table 1 and Figures S10–S12), likely because these sites are CG rich. For PmCDA1-CBE_V04 and hAID-CBE_V04, the average editing efficiencies for pure C-to-T editing were nearly doubled when compared to their CBE_V01 counterparts: 33.1% and 31.4%, as compared to 18.7% and 16.8%, respectively (Table 1). For A3A/Y130F, the average editing efficiency of pure C-to-T editing increased from 23.1% with the CBE_V01 platform to 34.1% with the CBE_V04 platform (Table 1). Notably, for all three cytidine deaminases, there was a two-fold reduction of indel frequencies with the CBE_V04 platform when compared to the CBE_V01 platform (Table 1 and Figures S13–S15). Collectively, these data from stable transgenic lines demonstrated that the CBE_V04 editors indeed have greatly improved C-to-T editing efficiency and purity, consistent with the observations in rice protoplasts.

Genome-wide assessment of off-target effects for 10 CBE systems in rice using whole-genome sequencing

We previously used whole-genome sequencing (WGS) to investigate the off-target effects of Cas9 and Cas12a in rice (Tang et al., 2018a). Here, we adopted a similar WGS pipeline to comprehensively assess genome-wide off-target effects of 10 CBE editors, including six CBE_V01 editors and four CBE_V04 editors (Figure 6a). A total of 59 plants were chosen for WGS including 31 edited lines by ten CBE editors (Table S2), 18 control lines that expressed base editors without sgRNAs, four wildtype (WT) plants and six other control plants for tissue culture, Agrobacterium transformation and Cas9 expression (Figure 6a). The T-DNA insertion sites for all transgenic plants were mapped to the rice genome (Figure S16). The numbers of indels (~50–100) in all CBE-expressing plants with or without sgRNAs were similar to those in control plants (Figure 6b), consistent with our previous report (Tang et al., 2018a), suggesting these indels are part of the somaclonal variation due to tissue culture.

The discovered single nucleotide variations (SNVs) in all samples appeared to be randomly distributed across the rice genome (Figure S17) and presented in all annotated sequence categories (Figure S18). About 80% or more of SNV alleles are either heterozygous (0.40–0.60 allele frequency) or mosaic (0.00–0.40) (Figure S19), suggesting they will either segregate or not be inherited to the next generation. About 100-130 single nucleotide variations (SNVs) were found in three control plant groups, indicating the baseline level of SNVs caused by somaclonal variations during rice tissue culture (Figure 6c) (Tang et al., 2018a). The CBE_V01 and CBE_V04 systems that expressed PmCDA1 and rAPOBEC1 showed slightly increased SNV counts of up to ~200 (Figure 6c). Slightly more SNVs (~200 to 400) were detected in the A3A/Y130F-CBE_V01 and V04 plants (Figure 6c). Much higher levels of SNVs (~300 to 1200) were detected in the hAID-CBE_V01 and V04 plants (Figure 6c). The various elevated SNV levels for different CBEs appeared to be sgRNA-independent, indicating genome-wide off-target effects.

To further investigate such off-target effects, we analysed the detected SNVs for all six possible nucleotide substitutions. The fraction of C:G>T:A mutations was ~30% for three tissue culture-related control groups. It was elevated to ~50% in the rAPOBEC1-CBE_V01 plants with or without sgRNAs (Figure 6d), suggesting that some of these C:G>T:A mutations were due to off-target effects of rAPOBEC1, as reported recently (Jin et al., 2019; Zuo et al., 2019). The fractions of C:G>T:A mutations rose to ~75% in plants that express hAID-CBE_V01 and V04 (Figure 6e), indicating higher levels of off-targeting. The fractions of C:G>T:A mutations for A3A/Y130F-CBE_V01 and A3A/Y130F-
CBE_V04 were ~60% and ~50%, respectively (Figure 6f). The fractions of C:G>T:A mutations for A3A-CBE_V01 and eA3A-CBE_V01 were both ~40% (Figure 6g), indicating a very low level of off-targeting. The minimal off-target effect of eA3A-CBE was previously demonstrated in human cells (Gehrke et al., 2018), which, however, may be partly explained by its inherent low activity (Figure 1c) (Doman et al., 2020). Notably, both PmCDA1-CBE_V01 and V04 editors displayed a background C:G>T:A
mutation level (Figure 6h), suggesting PmCDA1 has the least tendency for causing off-target C:G→T:A mutations when compared to other cytidine deaminases of high activities.

**Detection of sgRNA-dependent off-target effects by WGS**

In multiple CBE_V04 systems (e.g., rAPOBEC1, hAID and A3A/Y130F), base-edited plants showed higher fractions of C:G→T:A mutations than control plants expressing cytidine deaminases without sgRNAs (Figure 6d–f), which could be partly attributed to sgRNA-dependent off-targeting. Relatively few mutations were identified in the sgRNA-expressing samples at the top off-target sites predicted by Cas-OFFinder (Bae et al., 2014) (Figure S20A). Occasionally, such mutations were present in both biological replicates, suggesting sgRNA-dependent off-targeting (Figure S20B). Most of these mutations were at off-target sites with more than two nucleotide mismatches to the protospacers (Figure S20C), indicating only a fraction of them might be caused by sgRNA-dependent editing.

Further allelic analysis of the mutations at the top off-target sites revealed C-to-T conversions and small indels to some extent, which are hallmarks of CBE editing activity (Figures 7 and S21).

**Table 1** Comparison of top CBE_V01 and CBE_V04 editors in transgenic rice lines

| CBE system       | Targeted sites | Tested T0 lines | Pure C to T (number; ratio) | All C to T (number; ratio) | Other SNVs (number; ratio) | Pure InDels (number; ratio) | All InDels (number; ratio) |
|------------------|----------------|-----------------|-----------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| **PmCDA1-CBE_V01** |                |                 |                             |                           |                           |                           |                           |
| OsCGRS55         | 32             | 6; 18.8%        | 22; 68.8%                   | 3; 9.4%                   | 6; 18.8%                   | 22; 68.8%                 |                           |
| OsCGRS56         | 21             | 2; 9.5%         | 2; 9.5%                     | 0; 0.0%                   | 1; 4.8%                    | 1; 4.8%                   |                           |
| OsCGRS57         | 20             | 6; 30.0%        | 18; 90.0%                   | 0; 0.0%                   | 0; 0.0%                    | 12; 60.0%                 |                           |
| OsCGRS58         | 24             | 4; 16.7%        | 15; 62.5%                   | 0; 0.0%                   | 9; 37.5%                   | 20; 63.3%                 |                           |
| Average          |                | 18.7%           | 57.7%                       | 2.3%                      | 15.3%                      | 54.2%                     |                           |
| **hAID-CBE_V01**  |                |                 |                             |                           |                           |                           |                           |
| OsCGRS55         | 30             | 3; 10.0%        | 14; 46.7%                   | 1; 3.3%                   | 7; 23.3%                   | 18; 60.0%                 |                           |
| OsCGRS56         | 22             | 5; 22.7%        | 9; 40.9%                    | 0; 0.0%                   | 3; 13.6%                   | 7; 31.8%                  |                           |
| OsCGRS57         | 26             | 9; 34.6%        | 21; 80.8%                   | 1; 3.9%                   | 4; 15.4%                   | 16; 61.5%                 |                           |
| OsCGRS58         | 22             | 0; 0.0%         | 1; 4.6%                     | 0; 0.0%                   | 14; 63.6%                  | 15; 68.2%                 |                           |
| Average          |                | 16.8%           | 43.2%                       | 1.8%                      | 29.0%                      | 55.4%                     |                           |
| **A3A/Y130F-CBE_V01** |            |                 |                             |                           |                           |                           |                           |
| OsCGRS55         | 32             | 2; 6.3%         | 14; 43.8%                   | 2; 6.3%                   | 17; 53.1%                  | 29; 90.6%                 |                           |
| OsCGRS56         | 22             | 3; 13.6%        | 11; 50.0%                   | 0; 0.0%                   | 4; 18.2%                   | 12; 54.5%                 |                           |
| OsCGRS57         | 27             | 16; 59.3%       | 20; 74.1%                   | 0; 0.0%                   | 0; 0.0%                    | 4; 14.8%                  |                           |
| OsCGRS58         | 23             | 3; 13.0%        | 11; 47.8%                   | 2; 8.7%                   | 10; 43.5%                  | 20; 87.0%                 |                           |
| Average          |                | 21.3%           | 53.9%                       | 3.7%                      | 28.7%                      | 61.7%                     |                           |
| **PmCDA1-CBE_V04** |                |                 |                             |                           |                           |                           |                           |
| OsCGRS55         | 25             | 8; 32.0%        | 12; 48.0%                   | 2; 8.0%                   | 3; 12.0%                   | 5; 20.0%                  |                           |
| OsCGRS56         | 22             | 0; 0.0%         | 2; 9.1%                     | 0; 0.0%                   | 0; 0.0%                    | 2; 9.1%                   |                           |
| OsCGRS57         | 26             | 17; 65.4%       | 25; 96.2%                   | 4; 15.4%                  | 1; 3.9%                    | 6; 23.1%                  |                           |
| OsCGRS58         | 20             | 7; 35.0%        | 13; 65.0%                   | 0; 0.0%                   | 5; 25.0%                   | 11; 55.0%                 |                           |
| Average          |                | 33.1%           | 54.6%                       | 5.9%                      | 10.2%                      | 26.8%                     |                           |
| **hAID-CBE_V04**  |                |                 |                             |                           |                           |                           |                           |
| OsCGRS55         | 33             | 14; 42.4%       | 22; 66.7%                   | 0; 0.0%                   | 4; 12.1%                   | 12; 36.4%                 |                           |
| OsCGRS56         | 28             | 7; 25.0%        | 10; 35.7%                   | 0; 0.0%                   | 0; 0.0%                    | 3; 10.7%                  |                           |
| OsCGRS57         | 20             | 0; 0.0%         | 0; 0.0%                     | 0; 0.0%                   | 2; 10.0%                   | 2; 10.0%                  |                           |
| OsCGRS58         | 24             | 14; 58.3%       | 14; 58.3%                   | 0; 0.0%                   | 0; 0.0%                    | 0; 0.0%                   |                           |
| Average          |                | 31.4%           | 40.2%                       | 0.0%                      | 5.5%                       | 14.3%                     |                           |
| **A3A/Y130F-CBE_V04** |            |                 |                             |                           |                           |                           |                           |
| OsCGRS55         | 30             | 5; 16.7%        | 19; 63.3%                   | 3; 10.0%                  | 3; 10.0%                   | 14; 46.7%                 |                           |
| OsCGRS56         | 22             | 9; 40.9%        | 13; 59.1%                   | 2; 9.1%                   | 1; 4.5%                    | 3; 13.6%                  |                           |
| OsCGRS57         | 20             | 13; 65.0%       | 16; 80.0%                   | 5; 25.0%                  | 0; 0.0%                    | 0; 0.0%                   |                           |
| OsCGRS58         | 22             | 3; 13.6%        | 18; 81.8%                   | 2; 9.1%                   | 1; 4.6%                    | 14; 63.6%                 |                           |
| Average          |                | 34.1%           | 71.1%                       | 13.3%                     | 4.8%                       | 31.0%                     |                           |
Interestingly, far fewer sgRNA-dependent off-target mutations were identified in A3A/Y130F-CBE_V04 samples than in A3A/Y130F-CBE_V01 samples for either OsCGRS55 (4 vs 7) or OsCGRS57 (2 vs 4) target site (Figure 7a–d). Consistently, off-target mutations were found at two and three off-target sites for OsCGRS55 and OsCGRS57, respectively, in the PmCDA1-CBE_V01 samples, but only one at OsCGRS55 and none at OsCGRS57 in the PmCDA1-CBE_V04 samples (Figure 7e–g).

Furthermore, similar results were observed when comparing CBE_V01 and CBE_V04 systems based on rAPOBEC1 and hAID (Figure S21). We also randomly selected some off-target sites for Sanger sequencing, and the data validated the WGS results (Figure S22). These data suggest CBEs could induce sgRNA-dependent off-target mutations at sites with high sequence homology, which however could be reduced by using the CBE_V04 systems with focal recruitment of more UGIs.

Figure 6 Off-target assessment of ten CBE systems in rice by whole-genome sequencing. (a) A flow chart for the rice whole-genome sequencing (WGS) experiment. (b) All indels in each sample type identified by WGS. (c) All single nucleotide variations (SNVs) in each sample type identified by WGS. (d–h) Frequency analysis of each base change type among each sample type identified by WGS. (b–h) Error bars represent standard deviations. Each dot represents an independent sample. Data are presented as mean values ± SD.
Discussion

In this study, we first compared multiple cytidine deaminases and developed an efficient A3A/Y130F-CBE_V01 system for plant C-to-T base editing. Previously, high editing efficiency was reported for the A3A-based CBEs in human cells (Doman et al., 2020; Gehrke et al., 2018; Wang et al., 2018), plants (Xing et al., 2020; Zong et al., 2018) and yeast (Tan et al., 2020). Remarkably, we found a single Y130F mutation in A3A further promoted base editing efficiency. The Y130F mutation was previously proposed to impact the deaminase-single strand DNA (ssDNA) interaction (Shi et al., 2017). Our data showed A3A/Y130F-CBE_V01 further improved the editing efficiency over A3A-CBE_V01 while the eA3A-CBE_V01 had decreased activity in rice cells, and these results are consistent with recent reports in human cells (Doman et al., 2020; Wang et al., 2018). In separate studies, we have found the superior performance of A3A/Y130F-CBE_V01 in dicot plants such as tomato and poplar, which will be published elsewhere. Thus, A3A/Y130F-CBE_V01 should represent one of the best CBEs for C-to-T base editing in plants. This notion was further proven when we found A3A/Y130F outperforming PmCDA1 and hAID when fused to SpyMacCas9 for editing the A-rich PAMs in rice (Sretenovic et al., 2021).

To further demonstrate the A3A/Y130F-CBE_V01 editor, we showed multiple examples of crop trait engineering in rice. In the first example, we showcased multiplexed editing of OsSPL14 and OsSPL16 which resulted in edited mRNAs resistant to OsMIR156 cleavage, as supported by elevated mRNA levels for both target genes. This however is only a proof-of-concept demonstration. In order to truly engineer improved plants by editing mRNA target sites, one needs to show that edited mRNAs are still fully functional. In the second example, we simultaneously edited OsGS3, OsGW2 and OsGN1a. We found A3A/Y130F-CBE_V01 was very efficient at introducing premature stop codons in these target genes, subsequently generating edited T0 lines with anticipated grain phenotypes (Zhou et al., 2019). Introduction of early stop codons is a more predictable way to make null alleles for a protein-coding gene compared to frame-shift mutations generated by targeted mutagenesis of Cas9 nuclease. Hence, efficient editing by A3A/Y130F-CBE_V01 for premature stop codons could be advantageous to Cas9 nuclease for knocking out protein-coding genes in plants. In the third example, we targeted OsALS for evolving herbicide resistance with two small sets of sgRNA libraries, one coupled with A3A/Y130F-CBE_V01 and the other coupled with A3A/Y130F-nCas9-NG-CBE_V01. We successfully identified herbicide-resistant D349(A/E) mutations that were previously documented in other plants (Yu and Powles, 2014). Furthermore, we found a novel D55ON mutation that also confers herbicide resistance in rice. As this mutation has not been reported in the literature, it will be interesting to see whether transferring this mutation to other plant species will also generate herbicide resistance and whether there will be any additive effects when this mutation is stacked on top of other known herbicide resistance mutations. In the fourth example, we targeted OsEPSPS with an even smaller collection of sgRNAs and obtained the P173L mutation, which was previously reported in maize, but not in rice (Sammons and Gaines, 2014). Interestingly, all the herbicide resistance alleles of OsALS and OsEPSPS were obtained by A3A/Y130F-nCas9-NG-CBE_V01, indicating the importance of using Cas9 variants that can recognize relaxed PAMs. It is thus anticipated that A3A/Y130F may offer high-efficiency C-to-T base editing when coupled with the PAM-less SpRY(D10A) nickase (Li et al., 2021; Ren et al., 2021a,b; Xu et al., 2021).

We further developed four CBE_V04 editors with recruiting more UGIs with the MS2-MCP interaction and a gRNA2.0 scaffold. Based on analyses in both rice protoplasts and stable transgenic lines, A3A/Y130F-CBE_V04 was the best performer, followed by PmCDA1-CBE_V04. All three CBE_V04 editors showed a dramatic reduction of indel frequencies than their CBE_V01 counterparts, presumably due to the enhanced focal concentration of UGI (five copies in CBE_V04 vs a single copy in CBE_V01). Data from 10 additional target sites in rice further demonstrated that A3A/Y130F-CBE_V04 consistently outperformed A3A/Y130F-CBE_V01 with higher editing efficiency and purity, e.g., reduced indel levels. Hence, the CBE_V04 systems, notably A3A/Y130F-CBE_V04, represent further improved CBEs in plants. In the future, we expect these systems to be further validated and applied in many different plant species.

With recruiting additional copies of cytidine deaminases by direct nCas9 fusion (the CBE_V02 systems) or by the MS2-MCP strategy (the CBE_V03 systems), we could not improve C-to-T base editing efficiency. Apparently, recruiting more cytidine deaminases somehow resulted in interference among them, leading to compromised editing activity. This lesson may inspire us to find a better way of engineering base editing systems when multiple deaminases are to be used. For example, numerous recent studies reported simultaneous C-to-T and A-to-G base editing when fusing a cytidine deaminase and adenosine deaminase either at both ends of nCas9 or in tandem to the N-terminus of nCas9 (Grunewald et al., 2020; Li et al., 2020; Sakata et al., 2020; Zhang et al., 2020). Compromised base editing activities were observed in such systems, especially for A-to-G base editing, when compared with single base editors (Li et al., 2020; Zhang et al., 2020). These studies, along with our results, suggest that it is a common phenomenon that deaminases may interfere with one another when targeting the same site. Longer linkers or different fusion configurations should be tried to minimize such interferences.

Recent studies reported sgRNA-independent genome-wide C-to-T mutations by rAPOBEC1-BE3 in rice (Zuo et al., 2019) and in rice (Jin et al., 2019). Improved CBEs with reduced off-targeting were developed and demonstrated in human cells, such as eA3A (Gehrke et al., 2018) and YE1-BE3-FNLS (Zuo et al., 2020). With WGS analysis in rice, we revealed genome-wide and sgRNA-independent C:G>T:A off-target mutations caused by expression of the rAPOBEC1 cytidine deaminase, consistent with the previous report (Zuo et al., 2019). Among all the six CBE_V01 systems, hAID-CBE_V01 generated the most sgRNA-independent off-target SNVs, followed by A3A/Y130F-CBE_V01. The numbers of off-target SNVs for A3A-CBE_V01, eA3A-CBE_V01 and PmCDA1-CBE_V01 were similar or smaller to rAPOBEC1-CBE_V01, averaging around 200, which were approximately twice as many as the spontaneous SNVs (~100) generated during rice tissue culture. With similar levels of off-target effects, A3A-CBE_V01 and PmCDA1-CBE_V01, however, were much more efficient than rAPOBEC1-CBE_V01 for on-target editing. Interestingly, A3A/Y130F-CBE_V04 showed a reduced level of sgRNA-independent off-target mutations than A3A/Y130F-CBE_V01, averaging around 200 SNVs. Remarkably, sgRNA-independent off-target mutations were undetectable for PmCDA1 base editors. Consistently, our further analysis showed that CBE_V04 systems had reduced sgRNA-dependent off-target effects when compared to CBE_V01 systems. While we were conducting this
Figure 7 Sequence analysis of potential sgRNA-dependent off-target sites in A3AY130F-CBE systems and PmcDA1-CBE systems. (a) OsCGR555 sgRNA off-target sites in A3AY130F-CBE-V01. (b) OsCGR557 sgRNA off-target sites in A3AY130F-CBE-V01. (c) OsCGR555 sgRNA off-target sites in A3AY130F-CBE-V04. (d) OsCGR557 sgRNA off-target sites in A3AY130F-CBE-V04. (e) OsCGR555 sgRNA off-target sites in PmcDA1-CBE-V01. (f) OsCGR557 sgRNA off-target sites in PmcDA1-CBE-V01. (g) OsCGR555 sgRNA off-target sites in PmcDA1-CBE-V04. For all panels, identified potential sgRNA-dependent off-target sites were listed on the top, while the genotypes of both alleles for these off-target sites were listed in the bottom.
research, Jin et al. reported rationally designed APOBEC3A and APOBEC3B CBEs with high editing specificity (Jin et al., 2020). In the future, it may be worthwhile to further engineer A3A/Y130F CBE variants for improved performance with rational design or protein evolution.

**Conclusion**

In this study, we compared seven CBE_V01 editors, three iSpyMacCas9 CBE_V01 editors, four CBE_V02 editors, three CBE_V03 editors, and four CBE_V04 for C-to-T base editing in plants. We found A3A/Y130F-CBE_V01 as a top performer for C-to-T editing in both rice and Arabidopsis, and A3A/Y130F-derived iSpyMacCas9 and Cas9-NG CBEs offered high-efficiency base editing at NAAR and NG PAM sites, respectively, broadening targeting ranges in the plant genomes. We demonstrated A3A/Y130F-CBE_V04 and PmCDA1-CBE_V04 as further improved CBEs with robust editing activity and improved editing purity. The WGS analysis revealed an undetectable and a low level of sgRNA-independent off-target effect by PmCDA1-CBE_V04 and A3A/Y130F-CBE_V04, respectively. Both PmCDA1-CBE_V04 and A3A/Y130F-CBE_V04 also showed reduced sgRNA-dependent off-target effects, thus making them the preferred CBEs for plant genome editing due to their high editing efficiency, purity, and specificity. These CBE systems, as newly developed and validated precise plant genome editing tools, are available at Addgene for wide distribution to the plant research community.

**Experimental procedures**

**Vector construction**

All target sites were listed in Table S3. The oligos and gBlocks were summarized in Table S4. Gateway compatible attL1-attR5 entry clones pyPQ265 (Addgene #164712) and pyPQ266 (Addgene #164713) were prepared using homologous recombination of two PCR amicloners in Escherichia coli DH5α of maize codon-optimized (2) Cas9(D10A) nicksense fragment from pyPQ266-D10A vector with primers zCas9-F1-Rec and zCas9-R1-Rec (for pyPQ265) and with primers zCas9-F2-Rec and zCas9-R2-Rec (for pyPQ266). PCR amplified backbone from pyPQ255 (Addgene #124310) with primers UGI-F1-Rec and APB-R1-Rec and PCR amplified backbone from pyPQ256 (Addgene #124312) with primers CDA-F1-Rec and attl1-Rec were used to prepare pyPQ265 and pyPQ266, respectively. pyPQ267A2 was prepared by cloning the PCR amplified sequence with primers MfeI-T2A-F and NLS-NosT-R into the BsaI and BbsI sites of pyPQ255. The hAID-XTEN-nCas9 part fragment was cloned into pTX503 at HpaI site and SpeI site by T4 ligation. To generate pGEL160 for the expression of hAID-CBE_V02 vector, hAID-XTEN and nCas9 part were synthesized. The hAID-XTEN-nCas9 part fragment was cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL161, the hAID-CBE_V03 vector, hAID-XTEN and nCas9 part were synthesized. The hAID-XTEN-nCas9 part fragment was cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL162, the rAPOBEC1-32aa linker-hAID-XTEN and nCas9 fragment was synthesized and cloned into pGEL160 at SbfI and BstEII sites. To generate pGEL163, the hAID-32aa linker- rAPOBEC1-XTEN and nCas9 fragment was synthesized and cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL164, the rAPOBEC1-32aa linker-hAID-XTEN and nCas9 fragment was synthesized and cloned into pGEL160 at SbfI and BstEII sites. To generate pGEL165, the rAPOBEC1-32aa linker-hAID-XTEN-nCas9 fragment was synthesized and cloned into pGEL160 at SbfI and BstEII sites. To generate pGEL166, the rAPOBEC1-32aa linker-hAID-XTEN-nCas9 fragment was synthesized and cloned into pGEL160 at SbfI and BstEII sites. To generate the T-DNA vectors, the plasmid vectors resulted from this assembly start with "pLR" (Table S5).

To generate pGEL158, the synthesized rAPOBEC1-XTEN-nCas9 (D10A)-UGI fragment was cloned into pTX503 at SbfI site and SpeI site by T4 ligation. To generate pGEL159 for the expression of hAID-CBE_V02 vector, hAID-XTEN and nCas9 part were synthesized. The hAID-XTEN-nCas9 part fragment was cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL160, the nCas9-PmCDA1-UGI fragment was synthesized and cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL161, the hAID-XTEN-nCas9 fragment was cloned into pGEL160 at BstEII and SbfI sites. To generate pGEL162, the rAPOBEC1-32aa linker-hAID-XTEN and nCas9 fragment was synthesized and cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL163, the hAID-32aa linker- rAPOBEC1-XTEN and nCas9 fragment was synthesized and cloned into pGEL158 at SbfI and BstEII sites. To generate pGEL164, the rAPOBEC1-32aa linker-hAID-XTEN-nCas9 fragment was synthesized and cloned into pGEL160 at SbfI and BstEII sites. To generate the T-DNA vectors, the oligos were synthesized based on the target sites selected. The oligos were annealed into small DNA fragment and then inserted into backbone vector by Golden Gate Reaction.

**Protoplast transformation and stable transformation**

The Japonica rice cultivar Nipponbare and Arabidopsis Columbia (Col-0) ecotype were used. Polyethylene glycol (PEG) transfection of rice protoplasts was performed at 32 °C according to our previously published protocol (Ren et al., 2019; Tang et al., 2016; Zhong et al., 2020; Zhong et al., 2019). Polyethylene glycol (PEG)
transfection of Arabidopsis protoplasts was performed at 25 °C (Zhang et al., 2013). Rice stable transformation was carried out by adopting the same procedure that we published previously (Tang et al., 2017; Wang et al., 2019; Zhong et al., 2018; Zhou et al., 2019).

**Mutagenesis analysis**

For assessing mutagenesis in protoplasts, protoplasts of rice or Arabidopsis were collected after 48 h post-transfection. DNA was then extracted with the CTAB method (Ren et al., 2019; Zhong et al., 2020). Targeted mutagenesis was quantified by next-generation sequencing (NGS) in which the genomic regions flanking the target sites were PCR-amplified using barcoded primers. The PCR amplicons were sequenced by Novogene using an Illumina HiSeqX platform and analysed by CRISPRMatch (You et al., 2018). For mutation analysis in stable transgenic rice T0 lines, single-strand conformation polymorphism (SSCP) (Zheng et al., 2016) or restriction fraction length polymorphism (RFLP) followed by Sanger sequencing as done in our previous studies (Wang et al., 2019; Zhou et al., 2017; Zhou et al., 2019).

**Quantitative RT-PCR analysis**

Total RNA was extracted from plant leaves using TRIzol Universal RNA reagent (TIANGEN, Beijing, China) and treated with RNase-free DNase I (TIANGEN) according to the manufacturer’s protocol. To generate qRT-PCR template, the extracted RNA was reverse transcribed using the HiScript III RT SuperMix for qPCR kit (Vazyme, Nanjing, China). qRT-PCR was done using ChamQ Universal SYBR qPCR Master Mix (Vazyme). Each qRT-PCR assay was repeated at least three times with three independent RNA preparations, and the rice Actin 1 (OsActin1) gene was used as a reference (Lowder et al., 2018).

**Screen for herbicide-resistant rice lines**

After Agrobacterium-mediated transformation, the rice calli were kept on the selection medium with 50 mg/L hygromycin at 28 °C in dark for 2–3 weeks. Actively grown calli were selected on medium containing 50 mg/L hygromycin and 0.4 μM bispyribac-sodium or 100 uM glyphosate at 28 °C with a 16 h light/8 h dark cycle. Transgenic and herbicide-resistant seedling were identified after ~3 weeks. Regeneration of the green seedlings on the medium containing 50 mg/L hygromycin and 0.4 μM bispyribac-sodium were transferred to the rooting medium with 0.4 μM bispyribac-sodium and grown at 28 °C with a 16 h light/8 h dark cycle.

**WGS and data analysis**

Genomic DNA was extracted using the plant genome DNA kit (Tiangen) as described by the manufacturer. All plant samples were sequenced using the Illumina HiSeq X Ten platform (Bionova, Beijing, China). A total of 775.05Gb WGS data were obtained, with an average sequencing depth of 36.92X. Skewer (version 0.2.2) software was used to remove illumina TruSeq adapter and filter low quality reads with the parameters: `-q 19 -Q 35 -i 50` (Jiang et al., 2014). BWA (v0.7.17) (Li and Durbin, 2009) was used for mapping all cleaned data to rice reference genome TIGR7 (http://rice.plantbiology.msu.edu/) (Kawahara et al., 2013). The average mapping ratio of reads was 99.78%. The average genome coverage was 97.57%. Samtools (v1.7) was used to sort BAM files (Li et al., 2009). Duplicated reads were marked by Picard (v2.23.1). After pre-processing BAM files from each sample, the Genome Analysis Toolkit (GATK, v3.8) was employed to realign reads near indels and recalibrate base quality scores (McKenna et al., 2010). Four programs were used for high confident detection of de novo variants in rice genome. SNVs were detected using LoFreetq* (v2.1.2) (Wilm et al., 2012), MuTect2 (David Benjamin et al., 2019) and VarScan2 (v2.4.3) (Koboldt et al., 2012) software with the somatic methods. Whole-genome indels calling was conducted with MuTect2, VarScan2 and pindel (v0.2.5b9) (Ye et al., 2009) software. Only shared SNVs and indels identified by their callers respectively were kept. BEDTools (v2.29.2) (Quinlan and Hall, 2010) and BCFTools (v1.9) (Li, 2011) were used to filter SNVs/indels and process VCF files. All filtered mutation sites were manually inspected with the IGV software (Robinson et al., 2011). Genome-wide mutations were generated using a 100 kb sliding window and plotted with CIRCOS software (Krzywinski et al., 2009). Genome-wide off-target sites were detected by Cas-OFFinder with up to 10-nt mismatches (Bae et al., 2014). Putative off-target sites were checked in IGV. The sequences of off-target sites were extracted both from the reference genome and WGS data, then aligned with their sgRNAs.

**Statistical analysis**

All statistical analyses were performed with R package version 4.0.2 (http://www.R-project.org/) or Python version 3.7.6.

**Acknowledgements**

This work was supported by the National Transgenic Major Project (award no. 2019ZX08010003-001-002), the National Natural Science Foundation of China (award no. 31771486, 32072045 and 31960423), the Fundamental Research Funds for the Central Universities (award no. ZYGX2019J127) and the State Key Laboratory of Rice Biology (award no. 202000205) to Y. Z. and X. Z. This work was also supported by the National Science Foundation Plant Genome Research Program (award no. IOS-1758745 and IOS-2029889), the U.S. Department of Agriculture Biotechnology Risk Assessment Grant Program (award no. 2018-33522-28789 and 2020-33522-32274) to Y. Q. S. S. is a Foundation for Food and Agriculture Research Fellow.

**Competing interests**

The authors declare no competing interests.

**Author contributions**

Y. Q. and Y. Z. designed the experiments. S. S., Q. R., Z.Z., X. T., D. Y. and R. M. generated all the constructs. Q. R. and Z.Z. did rice protoplast transformation and data analysis. Q. R. and X. T. did the Arabidopsis protoplast transformation and data analysis. Q. R., J. W., L. H., Y. G., L. L., S. S., Y. Z., H. Y., H. K., S. L. and X. Z. conducted rice stable transformation and analysis. Q. R. conducted targeted evolution of herbicide resistance work. Q. R. and J. Z. analysed the protein structure. Q. R., G. L., Y. W. and T. Z. conducted the WGS experiment and analysed the genomic data. Y. Q., Y. Z., Q. R. and S. S. wrote the paper with input from other authors. All authors read and approved the final manuscript.

**Data Availability Statement**

Gateway® compatible attL1-attR5 CBE entry vectors are available from Addgene (Table S6): pYPQ265 (Addgene #164712),
Improved plant cytosine base editors

References

Anzalone, A.V., Koblan, L.W. and Liu, D.R. (2020) Genome editing with CRISPR-Cas nucleases, base editors, transspases and prime editors. Nat. Biotechnol. 38, 824–844.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takahashi, T., Nishimura, A., Angeles, E.R. et al. (2005) Cytokinin oxidase regulates rice grain production. Science 309, 741–745.

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

Chatterjee, P., Lee, J., Nip, L., Koseki, S.R.T., Tsingler, E., Sontheimer, E.J., Jacobson, J.M. et al. (2020) A Cas9 with PAM recognition for adenine dinucleotides. Nat. Commun. 11, 2474.

Cheng, H., Hao, M., Ding, B., Mei, D., Wang, W., Wang, H., Zhou, R. et al. (2021) Base editing with high efficiency in allotetraploid oilseed rape by A3A-PHE system. Plant Biotechnol. J. 19, 87–97.

David Benjamin, T.S., Cibulskis, K., Getz, G., Stewart, C. and Lichtenstein, L. (2019) Calling Somatic SNVs and Indels with Mutect2. bioRxiv. https://doi.org/10.1101/861054

Doman, J.L., Ragarun, A., Nevby, G.A. and Liu, D.R. (2020) Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat. Biotechnol. 38, 620–628.

Gehrke, J.M., Cervantes, O., Clement, M.K., Wu, Y., Zeng, J., Bauer, D.E., Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Cheng, H., Hao, M., Ding, B., Mei, D., Wang, W., Zhou, R. et al. (2021) Base editing with high efficiency in allotetraploid oilseed rape by A3A-PHE system. Plant Biotechnol. J. 19, 87–97.

Gurel, F., Zhang, Y., Sretenovic, S. and Qi, Y. (2020) CRISPR-Cas nucleases and base editors for plant genome engineering. aBIOTECH 1, 74–87.

Jiang, H., Lei, R., Ding, S.W. and Zhu, S. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182.

Jiao, Y., Wang, Y., Yue, D., Wang, J., Yan, M., Liu, G., Dong, X. et al. (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Mol. Cell 37, 493–502.

Jin, S., He, F., Zhu, Z., Luo, Y., Liu, J., Gao, S., Zhang, F. et al. (2020) Rationally designed APOBEC3B cytosine base editors with improved specificity. Mol. Cell 79, 728–740.e6.

Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C. et al. (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science 364, 292–295.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Karaman, H., McCombie, W.R., Qiang, S., Schwartz, D.C. et al. (2013) Improvement of the Orzya sativa Nipponbare reference genome using next generation sequence and optical map data. Rice (N Y) 6, 4.

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

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424.
Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., Voytas, D.F., Tan, J., Zhang, F., Karcher, D. and Bock, R. (2020) Expanding the genome-editing scope in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat. Biotechnol. 38, 856–869.

Sammons, R.D. and Gaines, T.A. (2014) Glyphosate resistance: state of knowledge. Pest Manag. Sci. 70, 1367–1377.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Yamamoto, S., et al (2014) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOLCE3A and APOBEC3B. N. Struct. Mol. Biol. 24, 131–139.

Shen, G., Zheng, X., Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., Xin, X., et al (2018) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol. 161, 20–27.

Zhang, Y. and Qi, Y. (2019) CRISPR enables directed evolution in plants. Genome Biol. 20, 83.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J., et al (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Nat. Biotechnol. 31, 1545–1554.

Zhang, Z., Liu, S., Liu, X., Liu, B., Tang, X., Ren, Q., Zhou, J., et al (2020) Introns-based single guide RNA CRISPR systems for plant genome editing. Rice 7(1), 3.

Zhang, Z., Sretenovic, S., Ren, Q., Yang, L., Bao, Y., Qi, C., Yuan, M., et al (2019) Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. Mol. Plant 12, 1027–1036.

Zhang, Z., Zhang, Y., You, Q., Tang, X., Ren, Q., Liu, S., Yang, L., et al (2018) Plante genome editing using FrnCpf1 and LbcPf1 nucleases at redefined and altered PAM sites. Mol. Plant 11, 999–1002.

Zhou, J., Deng, K., Cheng, Y., Zhang, Z., Tian, L., Tang, X., Yang, B., Wei, J., Wu, J., Wang, R., et al (2018) Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat. Biotechnol. 36, 950–953.

Zong, Y., Zhang, Y., Li, C., Chen, K., Ren, Q., Yang, L., Qi, J.L., et al (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat. Biotechnol. 35, 438–440.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., et al (2019) Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science 364, 289–292.

Zuo, E., Sun, Y., Yuan, T., He, B., Zhou, C., Ying, W., Liu, J., et al (2020) A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. Nat. Methods 17, 600–604.
The editing efficiencies for all editable Cytosines in the editing windows in the ~20nt protospacers at the four target sites. The immediate upstream nucleotide of the Cytosines were examined for potential context effects. The error bars represent standard deviations of three biological replicates. Data are presented as mean values ± SD.

**Figure S4** Genotypes of edited T0 lines with indel byproducts by A3A/Y130F-iSpyMacCas9-CBE_V01 in rice.

**Figure S5** Multiplexed base editing of miRNA binding sites in two target genes in T0 rice lines.

**Figure S6** Multiplexed base editing of three grain trait genes in rice T0 lines.

**Figure S7** Assessment of CBE_V02 that use tandem fusion of cytidine deaminases. (A) Schematic of a novel group of CBEs based on tandem fusion of cytidine deaminases. (B) Comparison of C-to-T base editing efficiency between three CBE_V01 editors (rAPOBEC1, hAID and PmCDA1) and four novel CBEs (CBE_V02) based on tandem fusion of cytidine deaminases in rice protoplasts. The error bars represent standard deviations of two biological replicates. Data are presented as mean values ± SD.

**Figure S8** Assessment of CBE_V03 editors that use MS2-MCP to recruit additional cytidine deaminase. (A) Schematics of the MS2-MCP based recruitment of additional cytidine deaminase. (B, C) Assessment of three CBE_V03 editors in rice protoplasts at OsCGRSS55, 56, 57 and 58 sites. Each dot represents a biological replicate. Each target contains three biological replicates. The maxima, centre and minima of box refer to Upper quartile, median and Lower quartile. The maxima and minima of whiskers refer to maximum value and minimum value. Data are presented as mean values ± SD.

**Figure S9** Frequencies of editing byproducts for the CBE_V03 editors in rice protoplasts. Editing byproducts assessed include indel mutations, C-to-G mutations, and C-to-A mutations. The error bars represent standard deviations of three biological replicates. Data are presented as mean values ± SD.

**Figure S10** Genotype of edited rice T0 lines by PmCDA1-CBE_V01 at two target sites.

**Figure S11** Genotype of edited rice T0 lines by hAID-CBE_V01 at two target sites.

**Figure S12** Genotype of edited rice T0 lines by A3A/Y130F-CBE_V01 at two target sites.

**Figure S13** Genotype of edited rice T0 lines by A3A/Y130F-CBE_V04 at two target sites.

**Figure S14** Genotype of edited rice T0 lines by hAID-CBE_V04.

**Figure S15** Genotype of edited rice T0 lines by A3A/Y130F-CBE_V04 at two target sites.

**Figure S16** Mapped T-DNA insertion sites in all edited rice T0 lines used for WGS. The T-DNA insertion sites for 6 CBE_V01 and 4 CBE_V04 base editors were mapped by WGS and shown in different colours. Note most transgenic lines carried only one T-DNA insertion in the genome.

**Figure S17** Genome-wide distribution of detected SNVs in T0 lines. All SNVs detected in samples edited by 6 CBE_V01 editors and 4 CBE_V04 editors are shown.

**Figure S18** Genomic annotation of detected SNVs in T0 lines. (A) Ratio of C-to-T SNVs in control (without sgRNAs) and positive samples (with sgRNAs). (B) Ratio of C-to-A SNVs in control (without sgRNAs) and positive samples (with sgRNAs). (C) Ratio of C-to-G SNVs in control (without sgRNAs) and positive samples (with sgRNAs). Data are presented as mean values ± SD.

**Figure S19** Allele frequencies of SNVs among CBE-expressing T0 lines. (A) Allele frequency of SNVs in control samples by 6 CBE_V01 editors and 4 CBE_V04 editors without sgRNAs. (B) Allele frequency of SNVs in base edited samples by 6 CBE_V01 editors and 4 CBE_V04 editors with sgRNAs.

**Figure S20** Analysis of potential sgRNA-dependent off-target sites in the rice genome. (A) WGS identified vs in silico predicted putative off-target sites based on different numbers of nucleotide mismatches. (B) Number of shared mutations between replicates with putative off-targets. (C) Putative off-target sites information. Samples from CBE_V04 systems are marked in shade.

**Figure S21** Sequence analysis of potential off-target sites in rAPOBEC1-CBE systems, hAID-CBE systems, and eA3A-CBE systems. (A) Q55 sgRNA off-target sites in rAPOBEC1-CBE-V01. (B) Q57 sgRNA off-target sites in rAPOBEC1-CBE-V01. (C) Q55 sgRNA off-target sites in rAPOBEC1-CBE-V04. (D) Q55 sgRNA off-target sites in hAID-CBE-V01. (E) Q57 sgRNA off-target sites in hAID-CBE-V01. (F) Q57 sgRNA off-target sites in eA3A-CBE-V01.

**Figure S22** Validation of potential sgRNA-dependent off-target sites by Sanger sequencing. Some lines edited by A3A/Y130F-CBE and PmCDA1-CBE systems in Figure 7A–7G were selected for Sanger sequencing, which revealed genotyping data consistent with WGS.

**Table S1** Inheritance of mutations in T1 lines.

**Table S2** Genotype of edited rice T0 lines used for whole-genome sequencing.

**Table S3** Target sites used in this study.

**Table S4** Oligos and gBlocks used in this study.

**Table S5** T-DNA constructs used in this study.

**Table S6** CBE systems used in this study.