Genome engineering in rice using Cas9 variants that recognize NG PAM sequences

Kai Hua, Xiaoping Tao, Peijin Han, Rui Wang, Jian-Kang Zhu

PII: S1674-2052(19)30122-4
DOI: https://doi.org/10.1016/j.molp.2019.03.009
Reference: MOLP 760

To appear in: MOLECULAR PLANT
Accepted Date: 23 March 2019


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All studies published in MOLECULAR PLANT are embargoed until 3PM ET of the day they are published as corrected proofs on-line. Studies cannot be publicized as accepted manuscripts or uncorrected proofs.
Genome engineering in rice using Cas9 variants that recognize NG PAM sequences

Kai Hua¹,³, Xiaoping Tao¹, Peijin Han⁴, Rui Wang¹,³, Jian-Kang Zhu¹,²*

¹Shanghai Center for Plant Stress Biology, CAS Center of Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China
²Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907, USA
³University of Chinese Academy of Sciences, Beijing 100049, China
⁴College of Life Sciences, Shandong University, Jinan 250100, China

*Correspondence: Jian-Kang Zhu (e-mail: jkzhu@sibs.ac.cn)

Short summary:
Both xCas9 and SpCas9-NG can efficiently edit rice endogenous target sites with NG PAMs. The base editors containing xCas9 variants do not work efficiently in rice, but the different forms of adenine and cytosine base editors containing SpCas9-NG can efficiently edit rice genes with a broadened PAM compatibility.
Abstract

CRISPR-Cas9 genome editing relies on sgRNA-target DNA base pairing and a short downstream PAM sequence to recognize target DNA. The strict PAM requirement hinders the applications of CRISPR-Cas9 system since it restricts the targetable sites in the genomes. xCas9 and SpCas9-NG are two recently engineered SpCas9 variants that can recognize more relaxed NG PAMs, which have great potential in addressing the issue of PAM constraint. Here, we use stable transgenic lines to evaluate the efficacies of xCas9 and SpCas9-NG in performing gene editing and base editing in rice. We found that xCas9 can efficiently induce mutations at target sites with NG and GAT PAM sequences in rice. However, base editors containing xCas9 failed to edit most of the tested target sites. SpCas9-NG exhibited a robust editing activity at sites with various NG PAMs without showing any preference for the third nucleotide after NG. In addition, we show that xCas9 and SpCas9-NG have higher specificity than SpCas9 at the CGG PAM site. We demonstrate that different forms of cytosine or adenine base editors containing SpCas9-NG can work efficiently in rice with broadened PAM compatibility. Thus, our work has yielded versatile genome engineering tools that will significantly expand the target scope in rice and other crops.

Key words: gene knockout, base editing, xCas9, SpCas9-NG, rice
Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated protein 9 (Cas9) system has been widely used for genome engineering in both prokaryotes and eukaryotes (Hsu et al., 2014; Komor et al., 2017a). Cas9 target recognition and cleavage rely on the extensive base pairing between single guide RNA (sgRNA) and target DNA and also requires a short protospacer adjacent motif (PAM) sequence flanking the target site (Deveau et al., 2008; Mojica et al., 2009). In bacteria and archaea, the CRISPR-Cas adaptive immune systems utilize PAM sequences to discriminate self and non-self targets, protecting the host genome from self-cleavage (Marraffini and Sontheimer, 2010). However, when the CRISPR-Cas9 system is repurposed for genome engineering, this PAM requirement greatly reduces the targetable sites in the genome. Some applications of the CRISPR-Cas9 system, such as base editing (Gaudelli et al., 2017; Komor et al., 2016) or precision editing through homology-directed repair (HDR) (Paquet et al., 2016), are even more severely affected by the PAM restriction. In theory, the more complex the PAM sequence is, the fewer target sites can be accessed by the Cas9 protein.

To address this limitation, many researchers sought to exploit the natural diversity of the CRISPR-Cas9 system as each Cas9 ortholog recognizes a different PAM sequence. Several Cas9 orthologs with different properties derived from Neisseria meningitides (NmeCas9) (Esvelt et al., 2013), Streptococcus thermophilus (StCas9) (Xu et al., 2014), Staphylococcus aureus (SaCas9) (Ran et al., 2015), Campylobacter jejuni (CjeCas9) (Kim et al., 2017) and Geobacillus thermodenitrificans (ThermoCas9 and GeoCas9) (Harrington et al., 2017; Mougiakos et al., 2017) have been adopted for genome editing. Nevertheless, the PAMs recognized by these Cas9 proteins are relatively complex, restricting their wide use in genome engineering. Among these Cas9 proteins, only SaCas9 has been widely used in plants (Hua et al., 2019; Kaya et al., 2016; Qin et al., 2018). More recently, another Cas9 ortholog from Neisseria meningitides, Nme2Cas9, was identified that recognizes a simple N4CC PAM.
sequence, which has a target site density comparable to the widely used SpCas9 (Edraki et al., 2018). Furthermore, a Cas9 from *Streptococcus canis* (ScCas9) showing a high sequence homology (89.2%) to SpCas9 was reported as capable of recognizing the minimal NNG PAM (Chatterjee et al., 2018). The future applications of these Cas9 variants in genome editing will help alleviate PAM constraint.

In addition to the Cas9 proteins, Cas12a (or Cpf1) and Cas12b (or C2c1) derived from class 2 type V CRISPR-Cas system have also been repurposed for genome engineering (Teng et al., 2018; Zetsche et al., 2015). They possess many features that are different from those of Cas9 proteins. One important difference is that Cas12a and Cas12b recognize AT-rich PAM sequences, complementary to GC-rich PAMs recognized by Cas9 (Teng et al., 2018; Zetsche et al., 2015). It has been demonstrated that FnCpf1 and LbCpf1 can work efficiently in many plant species (Endo et al., 2016; Tang et al., 2017; Wang et al., 2017b).

Although many Cas9 proteins have been applied for genome editing, SpCas9 is still the most commonly used one, in part because it shows robust editing activity in a wide range of organisms and recognizes the simple NGG PAM sequence. In order to further expand the PAM-binding repertoire of SpCas9, the PAM-interaction (PI) domain of SpCas9 was randomly mutated to screen for SpCas9 variants with different PAM binding preferences (Kleinstiver et al., 2015b). The VQR-Cas9 and VRER-Cas9 variants were thus generated, which recognize NGA and NGCG PAMs, respectively (Kleinstiver et al., 2015b). Using a similar approach, an SaCas9 variant, SaKKH-Cas9, was obtained that recognizes a more relaxed PAM sequence, NNNRRT (Kleinstiver et al., 2015a). To develop SpCas9 variants with a broadened PAM compatibility, Hu et al. (2018) recently performed an unbiased, phage-assisted continuous evolution (PACE) experiment to evolve new SpCas9 variants that recognize more relaxed PAMs. Two xCas9 variants, xCas9 3.6 and xCas9 3.7, were generated that can recognize a broad range of PAM sequences including NG, GAA and GAT in *E.coli* and mammalian cells (Hu et al., 2018). Although Wang et al (2018) recently showed that xCas9 could be
used to edit genes in rice calli, no genome-edited transgenic plants have been reported. Moreover, it remains unknown whether xCas9 can be applied to perform base editing in plants. Interestingly, guided by the structure of SpCas9-sgRNA-DNA complex, Nishimasu et al. (2018) rationally designed a SpCas9-NG variant that can also recognize relaxed NG PAMs. SpCas9-NG has been used for gene disruption and cytosine base editing in rice, but most of the gene editing experiments and all of the base editing work were carried out only in rice calli (Endo et al., 2018). More stable transgenic plants are needed to assess the performance of SpCas9-NG in rice.

Although the xCas9 variants and SpCas9-NG can recognize the NG PAM sequences, the mutation sites introduced into them are totally different, implying that different mechanisms may account for their relaxed PAM recognition (Hu et al., 2018; Nishimasu et al., 2018). It was reported that SpCas9-NG recognizes NGH (H=A, T, C) PAMs more efficiently than xCas9 does in human cells (Nishimasu et al., 2018). However, the editing efficiencies of xCas9 and SpCas9-NG in plants have not been directly compared. In this study, we used stable transgenic plants to evaluate the efficacies of xCas9 variants and SpCas9-NG in performing gene disruption and base editing in rice. Both xCas9 and SpCas9-NG could edit target sites with NG PAM sequences in rice. Moreover, we show that xCas9 and SpCas9-NG have higher specificity than that of SpCas9 at a selected CGG PAM site. Surprisingly, we found that cytosine and adenine base editors containing the xCas9 variants failed to edit most of the target sites we tested. In contrast, base editors containing SpCas9-NG could work efficiently in rice with a broadened PAM compatibility.

Results

Gene editing by xCas9 in rice

We introduced the E108G/S217A/A262T/S409I/E480K/E543D/M694I/E1219V and A262T/R324L/S409I/E480K/E543D/M694I/E1219V mutations (Hu et al., 2018) into human codon-optimized SpCas9 (Cong et al., 2013) by PCR to generate xCas9 3.6 and xCas9 3.7 variants, respectively. xCas9 3.6 and xCas9 3.7 were used to replace
the wild type SpCas9 in the pCas9 (OsU6) vector, generating vectors pCas9 3.6 (OsU6) and pCas9 3.7 (OsU6), respectively (Figure 1A). Because xCas9 3.6 and xCas9 3.7 can recognize NG, GAA and GAT PAM sequences in *E. coli* and mammalian cells (Hu et al., 2018), we selected six target sites harboring CGG, AGC, TGA, CGT, GAT and GAA PAM sequences to test the editing activity of xCas9 in rice (Table 1). Stable transgenic plants were generated by using *Agrobacterium*-mediated transformation of the rice variety Nipponbare. At the canonical CGG PAM site tested, both xCas9 3.6 and xCas9 3.7 exhibited very high editing activities, with mutation rates up to 85.3% and 89.3%, which were comparable to that achieved by SpCas9 (87.3%) (Table 1). At the TGA and CGT PAM sites, xCas9 variants also showed robust editing activity, with indel mutation frequencies over 80%, comparable to the editing efficiencies at the canonical CGG PAM site (Table 1). However, at the AGC and GAT PAM sites, xCas9 3.6 and xCas9 3.7 displayed significantly lower editing activity, with mutation rates ranging from 3.2% to 23.8% (Table 1). Neither xCas9 3.6 nor xCas9 3.7 introduced any mutation at the GAA PAM site (Table 1). Collectively, the above results suggest that xCas9 variants can efficiently induce gene mutations at NGG, NGT, NGA PAM sites in rice, but less efficiently at NGC, GAA and GAT PAM sites, which are consistent with the results reported in mammalian cells (Hu et al., 2018).

The editing efficiency of xCas9 3.7 in general was higher or comparable to that of xCas9 3.6 at various PAM sites tested in mammalian cells (Hu et al., 2018). At the six target sites tested here, xCas9 3.6 showed comparable editing efficiencies to xCas9 3.7 at the CGG, TGA and CGT PAM sites, but was 7.4 and 3.4 times as efficient as that of xCas9 3.7 at the AGC and GAT PAM sites, respectively (Table 1).

**Mutation patterns generated by xCas9 at different PAM sites**

SpCas9 exhibits robust editing activity in rice, generating high percentages of possible homozygous and biallelic mutants in the T0 generation at most target sites (Ma et al., 2015). To characterize the mutation patterns generated by SpCas9 and xCas9 variants at different target sites, we examined the Sanger sequencing chromatograms of all
targets using TIDE or DSDecode (Brinkman et al., 2014; Liu et al., 2015). Consistent with previous work (Ma et al., 2015), we found that SpCas9 generated high percentages of potential homozygous (48.3%) and biallelic (34.5%) mutants in rice in the T0 generation at the canonical CGG PAM site (Figure 1B and Table 1). Of note, xCas9 variants also induced high proportions of potential homozygous and biallelic mutants in the T0 generation at the CGG PAM site. xCas9 3.7 generated 40% potential homozygous and 48% biallelic mutants at this PAM site, whereas xCas9 3.6 was slightly less efficient, but still generated 31% potential homozygous and 27.6% potential biallelic mutants (Figure 1B and Table 1). However, at the non-NGG PAM sites, the ratios of possible homozygous and biallelic mutants generated by xCas9 3.6 and xCas9 3.7 were much decreased (Table 1 and Figure 1C-1E). In contrast, the ratio of chimeric mutants was markedly increased at these non-NGG PAM sites (Table 1 and Figure 1C-1E). At the TGA PAM site with a high overall mutation rate, xCas9 3.6 generated twenty four (82.8%) chimeric mutants but only one potential biallelic mutant (3.4%) (Figure 1C and Table 1). xCas9 3.7 was more efficient at the TGA PAM site, but still generated a much higher ratio of chimeras (69.2%) than possible homozygotes (3.8%) and bialeles (23.1%) (Figure 1C and Table 1). xCas9 3.6 induced comparable ratios of potential homozygous and biallelic mutants (in total 47.6%) to chimeric mutants (42.9%) at the CGT PAM site, whereas xCas9 3.7 generated more chimeras (58.3%) than potential homozygotes (0) and bialleles (33.3%) (Table 1 and Figure 1D). At the PAM sites with low overall mutation rates, such as AGC and GAT, the xCas9 variants did not generate any possible homozygous or biallelic mutants (Table 1 and Figure 1E). Taken together, our results indicate that xCas9 variants are more efficient in inducing homozygous or biallelic mutants at the canonical NGG PAM sites than at the non-NGG PAM sites.

**Inheritance of mutations induced by xCas9 variants**

To examine whether the mutations in T0 transgenic lines created by xCas9 variants could be inherited to the next generation, we first examined the mutation transmission at the canonical CGG PAM site, where xCas9 induced a high ratio of potential
homozygous or biallelic mutants. Line 69-10 was a potential homozygous mutant generated by xCas9 3.6, harboring 1-bp T insertion 3 bp upstream of the CGG PAM sequence. All eight T1 seedlings from Line 69-10 were homozygous and had the same 1-bp T insertion in the protospacer region of OsSPL14 (Supplemental Figure 1A), suggesting that Line 69-10 was indeed a homozygous mutant in the T0 generation. For another line, Line 69-5, a chimera in the T0 generation, the eight sequenced T1 seedlings showed six different genotypes (Supplemental Figure 1A). We also selected two lines edited by xCas9 3.7 for inheritance analysis. Line 30-3 and Line 30-6 were potential biallelic mutants in the T0 generation, and the mutated alleles in both lines were inherited to the T1 generation (Supplemental Figure 1B). As xCas9 variants generated high proportions of chimeras in the T0 generation at the non-NGG PAM sites (Table 1 and Figure 1C-1E), we wondered whether these chimeric mutations could be transmitted to the next generation. Here we took the TGA PAM site as an example. Line 70-8 and Line 70-16 edited by xCas9 3.6 were chimeric in the T0 generation and T1 progenies derived from them had diverse genotypes (Supplemental Figure 2A). Eight T1 plants from Line 70-8 showed four different genotypes, including 1 wild type, 1 homozygote, 3 heterozygotes and 3 chimeras with different mutation forms (Supplemental Figure 2A). One biallele, 2 heterozygotes and 4 chimeras with different mutation types were found in the seven T1 seedlings from Line 70-16 (Supplemental Figure 2A). Interestingly, two selected T0 chimeric mutants generated by xCas9 3.7 showed much simpler mutation patterns in the T1 generation. Two homozygotes, two heterozygotes and one chimera were found in the five genotyped T1 progenies from Line 32-13 (Supplemental Figure 2B), whereas two homozygotes and two heterozygotes were obtained from the four genotyped T1 seedlings from Line 32-21 (Supplemental Figure 2B). We conclude that the mutations induced by xCas9 variants can be faithfully transmitted to the next generation.

**Base editing by xCas9 in rice**

The narrow base editing window and PAM requirement substantially limit the target scope of base editors (Gaudelli et al., 2017; Hua et al., 2018; Komor et al., 2016 Li et
al., 2017; Lu et al., 2017; Yan et al., 2018). Base editors containing Cas9 or Cas9 variants that recognize different PAM sequences remarkably expand the targetable sites in the rice genome (Hua et al., 2019; Qin et al., 2018). The xCas9 variants recognize more relaxed PAM sequences (Table 1) (Hu et al., 2018), and thus have the potential to further expand the target scope of base editors in the rice genome. Therefore, we sought to apply the xCas9 variants to perform base editing in rice. We first converted xCas9 3.6 and xCas9 3.7 to nxCas9 3.6 (D10A) and nxCas9 3.7 (D10A) nickases and then used them to replace the nSpCas9 (D10A) nickase in ABE-P1 and CBE-P1 (Hua et al., 2018; Hua et al., 2019), generating the adenine base editors ABE-P6 and ABE-P7, and the cytosine base editors CBE-P6 and CBE-P7, respectively (Figure 2A). The cytidine deaminase rAPOBEC1 and adenine deaminase ecTadA used here are human-codon optimized as described (Gaudelli et al, 2017; Komar et al., 2016). Next, we selected sgRNA1~6 to test the base editing activity of ABE-P6 and ABE-P7 in rice because these sgRNAs are functional, except sgRNA6, in gene disruption assays (Table 1), and there are editable adenines in the canonical base editing window (positions 4-7 in the protospacer region, scoring the PAM sequence as 21-23) (Gaudelli et al., 2017) in their target sites (Table 2). Surprisingly, we found that ABE-P7 did not induce any mutation at these six target sites (Table 2). ABE-P6 could only edit the sgRNA4-targeted site (OsSPL7, containing a CGT PAM) with low efficiency (4.8%, 2/42) (Table 2 and Figure 2B). The thymidine (T) at position 6 in the protospacer region of OsSPL7 was edited by ABE-P6 with a modest efficiency (Figure 2B). Interestingly, even at the OsSPL14 target site with a canonical CGG PAM, ABE-P6 and ABE-P7 failed to edit the multiple adenines in the protospacer region (Table 2). However, this target site can be efficiently edited by ABE-P1 (containing SpCas9 (D10A) nickase) in our previous study (Hua et al., 2018).

We designed sgRNA7~10 to test the base editing activity of CBE-P6 and CBE-P7 in rice (Table 3). The four sgRNA-targeted sites harbor AGG, AGA, AGT and AGC PAM sequence, respectively (Table 3). At the AGG PAM site, CBE-P6 and CBE-P7 could induce C-T substitutions at position (s) 4 and/or 5 in the protospacer region of
$SNB$ (Figure 2C and Table 3). All CBE-P6 edited lines had expected C-T substitutions in the protospacer without indel formation (Table 3). However, two lines edited by CBE-P7 only had an unexpected 5 bp indel in the target region (Table 3 and Figure 2D). Moreover, all CBE-P6 and CBE-P7 edited lines were heterozygous or chimeras, whereas homozygous substitution lines at the $SNB$ target site could be easily obtained using CBE-P1 in our previous work (Hua et al., 2019). In addition, the editing efficiencies of CBE-P6 (17.3%, 12/69) and CBE-P7 (35.7%, 5/14) at the $SNB$ target site were much lower than that achieved by CBE-P1 (80%) (Hua et al., 2019). At the other non-NGG PAM sites, neither CBE-P6 nor CBE-P7 induced any mutation (Table 3). Collectively, the above results suggest that the adenine and cytosine base editors containing the xCas9 variants do not work efficiently in rice, which is different from the observations in mammalian cells (Hu et al., 2018).

**Gene editing by SpCas9-NG in rice**

Recently, another SpCas9 variant, SpCas9-NG, that was rationally designed based on the structure of the SpCas9-sgRNA-DNA complex, can also recognize relaxed NG PAM sequences (Nishimasu et al., 2018). We introduced the R1335A/L1111R/ D1135V/G1218R/E1219F/A1322R/T1337R mutations into SpCas9 to generate the SpCas9-NG variant and used it to replace the SpCas9 in the pCas9 (OsU6) vector, leading to the gene editing vector pCas9-NG (OsU6) (Figure 1A). *In vitro* cleavage assay and *in vivo* genome editing in mammalian cells indicated that SpCas9-NG outperforms xCas9 at NGH PAM sites (H=A, T, C) (Nishimasu et al., 2018). To directly compare the editing efficiencies of xCas9 variants and SpCas9-NG at various NG PAM sites in rice, we selected sgRNA1~4 to test the editing activity of SpCas9-NG (Table 1). It was reported that the editing efficiency of SpCas9-NG at the NGG PAM sites were lower than that of wild type SpCas9 in mammalian cells (Nishimasu et al., 2018). Here we found that SpCas9-NG (65.7%) was less efficient compared to SpCas9 (87.9%) at the CGG PAM site in rice (Table 1). It seems that xCas9 variants were also more efficient than SpCas9-NG at the CGG PAM site, as indel frequencies of xCas9 variants were 1.3-fold higher than that of SpCas9-NG.
Furthermore, we noted that the ratios of potential homozygous (17.4%) and biallelic (26.1%) mutants generated by SpCas9-NG at the CGG PAM site were much decreased compared to SpCas9 and xCas9 variants (Table 1 and Figure 1B). Nearly half of the mutants (47.8%) generated by SpCas9-NG were heterozygous (Table 1 and Figure 1B). At the TGA PAM site, SpCas9-NG (80%) showed a comparable editing efficiency as xCas9 variants (Table 1). However, compared with xCas9 variants, which generated a high ratio of chimeric mutants at the TGA PAM site, the proportion of potential homozygous (15%) and biallelic (55%) mutants generated by SpCas9-NG was much increased (Table 1 and Figure 1C). The editing efficiency of SpCas9-NG at the CGT PAM site (74.2%) was slightly lower than those of xCas9 variants and SpCas9-NG generated 4.3% homozygous and 26.1% biallelic mutants in the T0 generation (Table 1 and Figure 1D). The overall ratio of possible homozygous and biallelic mutants (30.4%) generated by SpCas9-NG was comparable to that of xCas9 3.7 (33.3%), but was lower than that of xCas9 3.6 (47.6%) (Table 1 and Figure 1D). Although SpCas9-NG showed low editing activity at NGC PAM sites in mammalian cells (Nishimasu et al., 2018), we found that the AGC PAM site we selected here could be efficiently edited by SpCas9-NG. The mutation frequency of SpCas9-NG (81.8%) was 3.4-fold and 25.6-fold as high as that of xCas9 3.6 and xCas9 3.7, respectively (Table 1). Moreover, when we used SpCas9-NG, a high proportion of putative homozygous (3.7%) and biallelic (51.9%) mutants could be obtained at the AGC PAM site in the T0 generation (Table 1 and Figure 1E). In contrast, no homozygous or biallelic mutants could be generated by xCas9 variants at this site in the T0 generation (Table 1 and Figure 1E). Collectively, our results showed that SpCas9-NG has robust editing activity at several selected PAM sites (CGG, AGC, TGA, and CGT) in rice.

**xCas9 and SpCas9-NG are highly specific in rice**

Compared with SpCas9, both xCas9 and SpCas9-NG recognize more relaxed PAM sequences (Hu et al., 2018; Nishimasu et al., 2018). Therefore, these Cas9 variants may have much higher risks of off-targeting. To evaluate the specificity of xCas9
variants and SpCas9-NG in rice, we examined their off-target editing possibility at the sgRNA1 target site. Potential off-target sites of sgRNA1 were predicted by the website tool Cas-OFFinder (Bae et al., 2014). Rice genomic sequences with up to 3 mismatches to sgRNA1 and harboring an NG PAM were listed as potential off-target sites (Supplemental Table 1). For each Cas9 variant, we selected four transgenic lines with on-target mutations at the sgRNA1 target site for off-target analysis. Nine potential off-target sites were each PCR amplified from these transgenic lines and purified for Sanger sequencing. We found that xCas9 variants and SpCas9-NG did not induce any mutation at these sites (Supplemental Table 1). As a control, we also tested whether SpCas9 could induce off-target editing at these predicted off-target sites. Surprisingly, we found that SpCas9 induced extensive off-target editing at off-target site 1, which has one mismatch to sgRNA1 at the fifth position upstream of the CGG PAM sequence (Supplemental Table 1 and Supplemental Figure 3). No mutation was detected at other off-target sites (Supplemental Table 1). These results indicate that xCas9 variants and SpCas9-NG are more specific than SpCas9 at the sgRNA1 target site, even though they recognize broadened PAM sequences.

**Base editing by SpCas9-NG in rice**

Since SpCas9-NG could generate targeted mutations at various NG PAM sites with high efficiency (Table 1), we applied it to perform base editing. The new base editors ABE-NG and CBE-NG were constructed using the nSpCas9-NG (D10A) nickase to replace the nSpCas9 (D10A) in ABE-P1 and CBE-P1 (Figure 3A and Figure 4A). We first used the sgRNA1~4 to test the base editing activity of ABE-NG in rice (Table 2). The overall adenine base editing efficiencies of ABE-NG at these four target sites were relative low compared to the indel formation frequencies of pCas9-NG (OsU6) at these sites, even though they used the same sgRNAs (Table 1 and Table 2). At the sgRNA1 target site (CGG PAM), we obtained one base-edited line from 38 transgenic lines (Table 1). Therefore, the editing efficiency of ABE-NG at this target site (2.6%) was 10% of that of ABE-P1 (26%) (Hua et al., 2018). The edited line harbored a T-C substitution at position 7 in the protospacer of *OsSPL14* (Figure 3B). Moreover, the
base editing efficiencies of ABE-NG at the sgRNA2 (2%, AGC PAM) and sgRNA4 (2.9%, TGA PAM) target sites were also very low, as we could only get one base-substituted line from 51 and 35 transgenic lines, respectively (Table 2, Figure 3C and 3D). However, the base editing efficiency of ABE-NG at the sgRNA3 target site (CGT PAM) was 11.9% (5/42) (Table 2). All the five edited lines had a T-C substitution at position 8 in the protospacer of OsIAA13 (Table 2 and Figure 3E). Of note, combining all adenine base edited events together, only one adenine (or thymidine) in the protospacer region of four target genes could be edited (Table 2) and no homozygous substitution line was generated. The above results suggest that the adenine base editor containing SpCas9-NG can work in rice at endogenous NG PAM sites, and thus can be used to expand the target scope of ABES in rice.

We recently demonstrated that the TadA*7.10-nCas9 fusion can be used to improve the base editing efficiency of ABES containing different Cas9 proteins or variants in rice (Hua et al., submitted). To test whether this strategy could improve the base editing efficiency of ABE-NG, we generated a simplified adenine base editor, ABE-NG-S, by fusing TadA*7.10 to the N-terminus of nSpCas9-NG (D10A) (Figure 3A). We selected sgRNA1 and sgRNA2 to test the editing activity of ABE-NG-S. The base editing efficiency of ABE-NG-S at the sgRNA1 target site (2.8%) was comparable to that of ABE-NG, as we obtained only one base substitution line from 35 transgenic lines (Table 2 and Figure 3B). At the sgRNA2 target site, however, the base editing efficiency of ABE-NG-S (7.7%, 3/39) was 3.9-fold as high as that of ABE-NG (Table 2 and Figure 3C). Therefore, the simplified base editor ABE-NG-S shows a great potential for improving the base editing efficiency at some NG PAM sites.

Next, we used sgRNA7~10 to test the editing activity of CBE-NG (Table 3). At the sgRNA7 target site (AGG PAM), 26 out of 36 transgenic lines had mutations in the target region (Table 3). However, we found that only 15 lines had our expected C-T substitutions at position(s) 4 and/or 5 in the protospacer region of SNB and several lines were homozygous mutants. Nine lines had additional indel mutations or other forms of base conversions in protospacer region besides the desired C-T substitutions,
whereas the rest 2 edited lines had only undesired indels at the target site (Table 3 and Figure 4B). The cytosine base editing efficiencies at the other non-NGG PAM sites were much lower than that at the AGG PAM site (Table 3). Nine out of 33 transgenic lines had mutations at the sgRNA8 target site (AGA PAM). The 9 edited lines could be classified into four different genotypes: 4 lines only had the expected C-T substitutions at positions 4 and 6 in the protospacer region of OsSPL7; 3 lines had both indel mutations and C-T substitutions at the target site; 1 line had a C-A substitution in addition to a C-T substitution and 1 line only had an indel at the target site (Table 3 and Figure 4C). The base editing efficiency of CBE-NG at the AGT PAM site was 4.5% (1/22), with only one base substitution line was found, which harbored a C-T substitution at position 6 in the protospacer region of PMS3 (Table 3 and Figure 4D). We did not detect any mutation at the sgRNA10 target site (AGC PAM) from 35 transgenic lines (Table 3).

Recently, Ren et al. (2018) showed that a CBE containing a hyperactive hAID cytidine deaminase mutant had higher editing activity in rice. We wondered whether the hyperactive hAID mutant could improve the editing efficiency of CBE harboring an nSpCas9-NG (D10A) nickase in rice. To test this, we changed the rAPOBEC1 cytidine deaminase in CBE-NG to hAID and generated a new CBE, CBE-NG-hAID (Figure 4A). We used the CBE-NG-hAID to edit the sgRNA10 target site which failed to be edited by CBE-NG. Unfortunately, we found that this site was also resistant to editing by CBE-NG-hAID. No mutation was detected at the target site from 25 transgenic lines (Table 3). However, at the sgRNA7 target site, CBE-NG-hAID showed a high base editing activity. Eight out of 24 transgenic lines harbored the expected C-T substitutions in the target region and another transgenic line had an undesired indel mutation (Table 3). Compared with CBE-NG, the base editing window of CBE-NG-hAID was expanded remarkably at the sgRNA7 target site because all cytosine in the protospacer region of SNB (at positions 4, 5, 14, 17) could be efficiently edited by CBE-NG-hAID (Figure 4E). Moreover, we noted that even the cytosine upstream of the protospacer of SNB could be edited by CBE-NG-hAID and these editing events were further confirmed by TA cloning and sequencing (Figure 4E).
and 4F). Our results suggest that CBE-NG-hAID may have different editing properties compared to CBE-NG, as it shows potential to edit both strands outside of the protospacer region at the \textit{SNB} locus.

**Discussion**

In this study, we successfully applied the xCas9 and SpCas9-NG variants to perform gene editing at various NG PAM sites in rice. Although the xCas9 variants were inefficient in rice when they were used in the base editors, different forms of adenine and cytosine base editors containing SpCas9-NG could be used to edit rice endogenous genes with a broadened PAM compatibility.

We showed that xCas9 variants could generate targeted mutations at all tested PAM sites except the GAA PAM site in rice (Table 1). Wang et al (2018) recently showed that xCas9 variants could induce mutations at the GAA PAM sites in rice calli, but the editing efficiencies were very low for xCas9 3.6 (0-4.17%) and xCas9 3.7 (2.08-12.5%). Therefore, more target sites will need to be tested to determine whether xCas9 variants can efficiently edit sites in rice with the GAA PAM. We noted that the editing activity of xCas9 variants in rice in our work here is different from that reported by Wang et al (2018). They showed that xCas9 3.6 and xCas9 3.7 edited various NG PAM sites in rice calli with low efficiency (all sites are below 20%), even at the canonical NGG PAM sites. No gene-edited transgenic lines were regenerated from the rice calli (Wang et al., 2018). Using stable rice transgenic lines, however, we showed here that xCas9 3.6 and xCas9 3.7 can work efficiently in rice with mutation rates up to 80% at several PAM sites (Table 1). We should mention that a human-codon optimized SpCas9 was used in our study (Cong et al., 2013), whereas Wang et al (2018) used a rice codon-optimized SpCas9 (Wang et al., 2015). It is unclear whether the different codon usages of SpCas9 or differences in some other components in the binary vector may have caused the discrepancies in the editing efficiencies of xCas9 variants.

In mammalian cells, xCas9 3.7 in general has comparable or higher editing activity than xCas9 3.6 at various PAM sites tested (Hu et al., 2018). Recently, it was reported
that xCas9 3.7 outperformed xCas9 3.6 for inducing gene mutations in rice calli (Wang et al., 2018). However, we found that xCas9 3.6 showed comparable or even higher editing activity than xCas9 3.7 at the six target sites we tested (Table 1). Considering that we have tested only a subset of NG PAMs and a limited number of sites, more targets need to be tested in order to draw firm conclusions on the efficiencies of xCas9 variants in performing gene mutagenesis in rice.

An interesting finding in our study is that the base editors containing xCas9 variants failed to edit the most sites we tested (Table 2 and Table 3), which is different from the results from mammalian cells (Hu et al., 2018). In our study, the same set of sgRNAs was used to perform gene mutagenesis and adenine base editing in rice. We found that the sgRNA activities for causing indel mutations and for inducing adenine base conversion were not correlated for sgRNA1~5 (Table 1 and Table 2). xCas9 variants could induce indels (some with high efficiency) at the five target sites, whereas ABE-P7 did not induce adenine base conversion at these sites and ABE-P6 only modestly edited the sgRNA4 target site. Importantly, even at a canonical CGG PAM site (sgRNA1) that could be edited by ABE-P1 in our previous work (Hua et al., 2018), ABE-P6 and ABE-P7 failed to induce any mutation. We hypothesized that fusing adenine deaminase to the N-terminus of xCas9 (D10A) nickases may compromise their binding to the target sites. However, we found that xCas9 3.6 fused with the adenine deaminase could induce indel mutations at the CGG and CGT PAM sites (Supplemental Figure 4), suggesting that the deaminase fusion did not affect binding affinity of xCas9 3.6 to the target site. Therefore, the reasons for the inefficiencies of the base editors containing xCas9 variants in rice remain unknown. Future efforts are needed to optimize the architecture of base editors containing xCas9 variants to improve their editing activity in rice.

Off-target editing is an important issue in genome engineering. xCas9 variants were evolved to recognize broadened PAM sequences with even higher editing specificity (Hu et al., 2018). SpCas9-NG and SpCas9 had similar cleavage specificity in mammalian cells (Nishimasu et al., 2018) but was recently reported to induce fewer off-target mutations in rice at NGG PAMs than SpCas9 owing to its lower cleavage
activity (Endo et al., 2018). We observed that xCas9 and SpCas9-NG have higher specificity than SpCas9 in rice (Supplemental Table 1). SpCas9 induced extensive editing at the predicted off-target site 1 of sgRNA1 (Supplemental Figure 3), whereas xCas9 and SpCas9-NG variants did not. Considering that xCas9 variants show comparable editing activity to that of SpCas9 at NGG PAMs, we suggest that they can substitute for SpCas9 in some applications to minimize the off-target effects. Unlike xCas9 variants, the SpCas9-NG-containing base editors could edit various NG PAM sites. ABE-NG had relatively low editing activity at the four different PAM sites we tested, but the edited products were very clean. Only A-G (or T-C) substitutions were found in the target regions (Table 2). Moreover, we found that a simplified ABE-NG-S base editor could slightly improve the editing efficiency at an TGA PAM site (Table 2). Whether it can improve adenine base editing efficiency at the other NG PAM sites need further work. Although SpCas9-NG was recently used to perform cytosine base editing in rice calli, no stable base-edited transgenic lines were generated (Endo et al., 2018). In this study, we showed that CBE-NG can be used to generate base substitution lines at AGG, AGA and AGT PAM sites with various efficiencies (Table 3). Strikingly, an AGC PAM site failed to be edited by both CBE-NG and CBE-NG-hAID (Table 3), possibly because the sgRNA we selected was suboptimal. More target sites are needed to test whether the CBE-NG and CBE-NG-hAID can edit NGC PAM sites in rice. In addition, we found that a subset of CBE-NG and CBE-NG-hAID edited lines had indels or other forms of base conversions in the target regions besides the desired C-T substitutions, perhaps due to the high uracil glycosylase activity that cleaves the uracil in the U-G mismatch generated by cytidine deaminase in rice. We expect that expressing uracil glycosylase inhibitor (UGI) at higher levels or fusion more UGI to the CBE-NG or CBE-NG-hAID may improve the product purity of the base edited plants (Komor et al., 2017b; Wang et al., 2017a).

In conclusion, our results show that both xCas9 and SpCas9-NG variants recognizing the relaxed NG PAMs can work efficiently in rice and SpCas9-NG is more suitable for performing base editing in rice. The versatile genome engineering tools developed
here significantly expand the target scope in rice and other crops and thus benefit basic plant research and genetic improvement in crops.

**Methods**

**Vector construction**

In this study, xCas9 and SpCas9-NG variants were mutated from a human-codon optimized SpCas9 (Cong et al., 2013) by PCR. The corresponding mutation sites in xCas9 3.6, xCas9 3.7 and SpCas9-NG have been previously reported (Hu et al., 2018). The binary vector pCas9 (OsU6) which contains a human codon-optimized SpCas9 gene driven by a maize ubiquitin promoter and a sgRNA scaffold driven by the OsU6 promoter in the pCAMBIA1300 backbone, was modified to construct the vectors for gene disruption assay. xCas9 3.6, xCas9 3.7 and SpCas9-NG were used to replace the SpCas9 in the pCas9 (OsU6) vector, generating the vectors pCas9 3.6 (OsU6), pCas9 3.7 (OsU6) and pCas9-NG (OsU6), respectively. The vectors pRABEsp-OsU6 (also known as ABE-P1) (Hua et al., 2018) and pRCBEsp-OsU6 (also known as CBE-P1) (Hua et al., 2019) were used to construct the new adenine and cytosine base editing vectors. The nSpCas9 (D10A) nickase in ABE-P1 and CBE-P1 were replaced by the nxCas9 3.6 (D10A), nxCas9 3.7 (D10A) and nSpCas9-NG (D10A) nickases, leading to the base editors ABE-P6, ABE-P7, ABE-NG, CBE-P6, CBE-P7 and CBE-NG, respectively.

To construct a simplified adenine base editor, ABE-NG-S, the nSpCas9 (D10A) nickase in the ABE-P1S (submitted) was changed to the nSpCas9-NG (D10A) nickase. A hypersensitive form of human AID cytidine deaminase (hAID) was synthesized from Sangon Biotech (Shanghai, China) as described by Ren et al (2018). The hAID was used to substitute the rAPOBEC1 cytidine deaminase in CBE-NG, leading to the base editor CBE-NG-hAID. ABE-P6-WT was constructed by replacing nxCas9 3.6 (D10A) nickase in ABE-P6 with xCas9 3.6.

Pairs of sgRNA primers were annealed on a PCR machine and then ligated into the BsaI digested binary vectors. The accuracy of sgRNA sequences was confirmed by sequencing before rice transformation. Primers for sgRNAs are listed in Supplemental
Table 2.

**Generation of rice transgenic plants**

All binary vectors were introduced into *Agrobacterium tumefaciens* strain EHA105 through the standard freeze/thaw method. The *Agrobacterium* then infected the rice embryogenic calli induced from the mature seeds of the rice variety Nipponbare. The whole rice transformation processes were performed as described elsewhere (Nishimura et al., 2007).

**Detection of on-target mutations in rice transgenic plants**

Rice genomic DNA was extracted from the three randomly sampled leaves from T0 or T1 transgenic lines. The on-target site of each sgRNA was amplified from the genomic DNA of independent T0 transgenic lines. The PCR amplicons were purified and subjected to Sanger sequencing. The Sanger sequencing chromatograms were decoded by the TIDE or DSDecode algorithm to analyze the mutation patterns at each target site (Brinkman et al., 2014; Liu et al., 2015). The genotypes of some transgenic lines were needed to be confirmed by TA cloning. Ten clones were randomly selected for sequencing. To examine whether the mutations induced by xCas9 variants can transmit to next generation, two independent T1 transgenic lines of sgRNA1 and sgRNA3 were selected for analysis. We randomly selected 5-8 seedlings in each line for on-target site genotyping. Primers used for on-target sites amplification and sequencing are listed in Supplemental Table 3.

**Off-target analysis**

Potential off-target sites for sgRNA1 in the rice genome were predicted by the online tool Cas-OFFinder (Bae et al., 2014). The potential off-target sites of SpCas9 were a subset of xCas9 and SpCas9-NG variants. Homologous sequences in the rice genome with up to 3 bp mismatches to sgRNA1 and harboring an NG PAM were listed as potential off-target sites. For SpCas9 or SpCas9 variants, we each selected four independent transgenic lines with on-target mutation at sgRNA1 target site for off-target analysis. All potential off-target sites were PCR amplified from these edited lines and were subjected to Sanger sequencing to detect off-target editing events. Primers used for off-target sites amplification and sequencing are listed in
Supplemental Table 4.

Reference


Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M.


Author Contributions

K.H. designed and conducted the experiments, analyzed the data and wrote the manuscript. X.T. conducted rice transformation. P.H. and R.W. performed some genotyping experiments. J.-K. Z. supervised the project and edited the manuscript.

Acknowledgements

This work was supported by the Chinese Academy of Sciences.

Conflict of interest statement
The authors declare no conflict of interests.
Table 1. Summary of mutation efficiencies and mutation patterns at different target sites for SpCas9, xCas9 3.6, xCas9 3.7 and SpCas9-NG

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>Target gene</th>
<th>PAM sequence</th>
<th>Cas9 variant</th>
<th>Mutation frequency</th>
<th>Homo</th>
<th>Bi</th>
<th>Het</th>
<th>Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgRNA1</td>
<td>OsSPL14</td>
<td>CGG</td>
<td>SpCas9</td>
<td>87.9% (29/33)</td>
<td>10/29 (34.5%)</td>
<td>0/29 (0)</td>
<td>5/29 (17.2%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.6</td>
<td>85.3% (29/34)</td>
<td>9/29 (31.0%)</td>
<td>8/29 (27.6%)</td>
<td>0/29 (0)</td>
<td>12/29 (41.4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.7</td>
<td>89.3% (25/28)</td>
<td>10/25 (40%)</td>
<td>12/25 (48%)</td>
<td>1/25 (4%)</td>
<td>2/25 (8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SpCas9-NG</td>
<td>65.7% (23/35)</td>
<td>4/23 (17.4%)</td>
<td>6/23 (26.1%)</td>
<td>11/23 (47.8%)</td>
<td>2/23 (8.7%)</td>
</tr>
<tr>
<td>sgRNA2</td>
<td>LF</td>
<td>AGC</td>
<td>xCas9 3.6</td>
<td>23.8% (5/21)</td>
<td>0/5 (0)</td>
<td>0/5 (0)</td>
<td>3/5 (60%)</td>
<td>2/5 (40%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.7</td>
<td>3.2% (1/31)</td>
<td>0/1 (0)</td>
<td>0/1 (0)</td>
<td>1/1 (100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SpCas9-NG</td>
<td>81.8% (27/33)</td>
<td>1/27 (3.7%)</td>
<td>14/27 (51.9%)</td>
<td>4/27 (14.8%)</td>
<td>8/27 (29.6%)</td>
</tr>
<tr>
<td>sgRNA3</td>
<td>OsIAA13</td>
<td>TGA</td>
<td>xCas9 3.6</td>
<td>82.9% (29/35)</td>
<td>1/29 (3.4%)</td>
<td>6/29 (23.1%)</td>
<td>1/26 (3.8%)</td>
<td>18/26 (69.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.7</td>
<td>89.7% (26/29)</td>
<td>1/26 (3.8%)</td>
<td>6/26 (23.1%)</td>
<td>1/26 (3.8%)</td>
<td>18/26 (69.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SpCas9-NG</td>
<td>80% (20/25)</td>
<td>3/20 (15%)</td>
<td>11/20 (55%)</td>
<td>0/20 (0)</td>
<td>6/20 (30%)</td>
</tr>
<tr>
<td>sgRNA4</td>
<td>OsSPL7</td>
<td>CGT</td>
<td>xCas9 3.6</td>
<td>84% (21/25)</td>
<td>4/21 (19%)</td>
<td>6/21 (28.6%)</td>
<td>2/21 (9.5%)</td>
<td>9/21 (42.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.7</td>
<td>85.7% (12/14)</td>
<td>0/12 (0)</td>
<td>4/12 (33.3%)</td>
<td>1/12 (8.3%)</td>
<td>7/12 (58.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SpCas9-NG</td>
<td>74.2% (23/31)</td>
<td>1/23 (4.3%)</td>
<td>6/23 (26.1%)</td>
<td>8/23 (34.8%)</td>
<td>8 (34.8%)</td>
</tr>
<tr>
<td>sgRNA5</td>
<td>OsSPL4</td>
<td>GAT</td>
<td>xCas9 3.6</td>
<td>11.1% (5/45)</td>
<td>0/5 (0)</td>
<td>0/5 (0)</td>
<td>1/5 (20%)</td>
<td>4/5 (80%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.7</td>
<td>3.3% (1/30)</td>
<td>0/1 (0)</td>
<td>0/1 (0)</td>
<td>1/1 (100%)</td>
<td></td>
</tr>
<tr>
<td>sgRNA6</td>
<td>OsMADS57</td>
<td>GAA</td>
<td>xCas9 3.6</td>
<td>0 (0/3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>xCas9 3.7</td>
<td>0 (0/3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Note: Homo, Bi, Het and Chi represent homozygote, biallele, heterozygote and chimera, respectively.
Table 2. Summary of adenine base editing efficiencies at different target sites for ABE-P6, ABE-P7, ABE-NG and ABE-NG-S

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>Target gene</th>
<th>Protospacer-PAM sequence</th>
<th>Base editor</th>
<th>Total transgenic lines</th>
<th>Number of edited lines</th>
<th>Editing efficiency</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgRNA1</td>
<td>OsSPL14</td>
<td>AGAGAGAGAGCACAGCTCGAGT</td>
<td>ABE-P6</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>40 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-P7</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>22 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG</td>
<td>38</td>
<td>1</td>
<td>2.6%</td>
<td>37 WT+1 A6-G6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG-S</td>
<td>35</td>
<td>1</td>
<td>2.9%</td>
<td>24 WT+1 A6-G6</td>
</tr>
<tr>
<td>sgRNA2</td>
<td>LF1</td>
<td>TGAGATGAGGGTCCATGCTAGC</td>
<td>ABE-P6</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>44 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-P7</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>22 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG</td>
<td>51</td>
<td>1</td>
<td>2%</td>
<td>50 WT+1 A6-G6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG-S</td>
<td>39</td>
<td>3</td>
<td>7.7%</td>
<td>36 WT+3 A6-G6</td>
</tr>
<tr>
<td>sgRNA3</td>
<td>OsIAA13</td>
<td>CCAAACACTGCTGCGCGCTGA</td>
<td>ABE-P6</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>36 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-P7</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>14 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG</td>
<td>42</td>
<td>5</td>
<td>11.9%</td>
<td>37 WT+5 A6-G6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG-S</td>
<td>35</td>
<td>1</td>
<td>2.9%</td>
<td>34 WT+1 A6-G6</td>
</tr>
<tr>
<td>sgRNA4</td>
<td>OsSPL7</td>
<td>GAGAGAGAGCACCAGCGACGA</td>
<td>ABE-P6</td>
<td>42</td>
<td>2</td>
<td>4.8%</td>
<td>40 WT+2 A6-G6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-P7</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>35 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-NG</td>
<td>35</td>
<td>1</td>
<td>2.9%</td>
<td>34 WT+1 A6-G6</td>
</tr>
<tr>
<td>sgRNA5</td>
<td>OsSPL4</td>
<td>AGAGAGAGAGAGAGAAGCTGAAG</td>
<td>ABE-P6</td>
<td>68</td>
<td>0</td>
<td>0</td>
<td>68 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-P7</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>35 WT</td>
</tr>
<tr>
<td>sgRNA6</td>
<td>OsMADS57</td>
<td>AGCAACTGCGACAATTGCGAAGA</td>
<td>ABE-P6</td>
<td>48</td>
<td>0</td>
<td>0</td>
<td>48 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ABE-P7</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>35 WT</td>
</tr>
</tbody>
</table>

Note: The PAM sequences are written in bold and the adenines in the canonical base editing window defined by Gaudelli et al. (2017) (positions 4-7 in the protospacer region, scoring the PAM sequence as 21-23) are underlined.
Table 3. Summary of cytosine base editing efficiencies at different target sites for CBE-P6, CBE-P7, CBE-NG and CBE-NG-hAID

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>Target gene</th>
<th>Protospacer-PAM sequence</th>
<th>Base editor</th>
<th>Total genotyped lines</th>
<th>Number of edited lines</th>
<th>Mutation rate</th>
<th>Mutation pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgRNA7</td>
<td>SNB</td>
<td>AAATTTCAATTAAGG</td>
<td>CBE-P6</td>
<td>69</td>
<td>12</td>
<td>17.3%</td>
<td>57 WT+12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-P7</td>
<td>14</td>
<td>5</td>
<td>35.7%</td>
<td>9 WT+3 Sub+2 Indel</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG</td>
<td>36</td>
<td>26</td>
<td>72.2%</td>
<td>10 WT+15 Sub+9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG-hAID</td>
<td>24</td>
<td>9</td>
<td>37.5%</td>
<td>15 WT+8 Sub+1 Indel</td>
</tr>
<tr>
<td>sgRNA8</td>
<td>OsSPL7</td>
<td>TCTCTCTCTGTCATGCTAGA</td>
<td>CBE-P6</td>
<td>48</td>
<td>0</td>
<td>0</td>
<td>48 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-P7</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>15 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG</td>
<td>33</td>
<td>9</td>
<td>27.2%</td>
<td>24 WT+5 Sub+3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG-hAID</td>
<td>24</td>
<td>9</td>
<td>37.5%</td>
<td>15 WT+8 Sub+1 Indel</td>
</tr>
<tr>
<td>sgRNA9</td>
<td>PMS3</td>
<td>TTTGTCATCAACCACCAACAGGT</td>
<td>CBE-P6</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>29 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-P7</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>27 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG</td>
<td>22</td>
<td>1</td>
<td>4.5%</td>
<td>21 WT+1 Sub</td>
</tr>
<tr>
<td>sgRNA10</td>
<td>OsSPL14</td>
<td>TCTCTCTCTGTCATGCGAGCC</td>
<td>CBE-P6</td>
<td>47</td>
<td>0</td>
<td>0</td>
<td>17 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-P7</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>35 WT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CBE-NG-hAID</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>25 WT</td>
</tr>
</tbody>
</table>

Note: The PAM sequences are written in bold and the cytosine in the base editing window defined by Komor et al. (2016) (positions 4-8 in the protospacer region, scoring the PAM sequence as 21-23) are underlined. Sub means that the target region only has expected C-T substitutions, whereas Sub-indel represents that the target region has both C-T substitutions and unexpected indels.
Figure legends

Figure 1. Gene mutagenesis by xCas9 and SpCas9-NG variants in rice. A, Diagrams of the pCas9 (OsU6), pCas9 3.6 (OsU6), pCas9 3.7 (OsU6) and pCas9-NG (OsU6) vectors used for gene mutagenesis in rice. NLS, nuclear localization signal. B, Frequencies of different mutation types at a CGG PAM site mutated by SpCas9, xCas9 3.6, xCas9 3.7 and SpCas9-NG. C, Frequencies of different mutation types at a TGA PAM site mutated by xCas9 3.6, xCas9 3.7 or SpCas9-NG. D, Frequencies of different mutation types at a CGT PAM site mutated by xCas9 3.6, xCas9 3.7 or SpCas9-NG. E, Frequencies of different mutation types at an AGC PAM site mutated by xCas9 3.6, xCas9 3.7 or SpCas9-NG. In B-E, the frequencies of different mutation types were calculated by scoring the number of plants with specific mutation type divided by the total number of mutated transgenic lines.

Figure 2. Base editing by xCas9 in rice. A, Schematic view of ABE-P6, ABE-P7, CBE-P6 and CBE-P7 base editors. The D10A mutation inactivates the nuclease activity of RuvC domain in xCas9 3.6 and xCas9 3.7. B, Sequencing chromatograms of two base edited lines, Line 76-9 and Line 76-18, at the OsSPL7 target site edited by ABE-P6. Arrow points to the position with edited base. C, Sequencing chromatograms of two base edited lines, Line 48-17 and Line 54-11, at the SNB target site. Line 48-17 was generated using CBE-P6 and Line 54-11 was generated using CBE-P7. Arrows point to the positions with edited bases. D, A representative line, Line 54-10, has a 5-bp deletion at the SNB target site edited by CBE-P7. Sequencing chromatogram of Line 54-10 at the SNB target site is shown. Arrow points to the mutation site.

Figure 3. Adenine base editing by SpCas9-NG in rice. A, Schematic view of ABE-NG and ABE-NG-S adenine base editors. The D10A mutation inactivates the nuclease activity of RuvC domain in SpCas9-NG. B, Both ABE-NG and ABE-NG-S can edit a CGG PAM site in OsSPL14. Only one transgenic line was edited by ABE-NG and ABE-NG-S, respectively. Their sequencing chromatograms at the


*OsSPL14* target site are shown. C, Both ABE-NG and ABE-NG-S can edit *LF1* with an AGC PAM. The sequencing chromatograms of representative edited lines at the *LF1* target site are shown. D, *OsSPL7* with a CGT PAM can be edited by ABE-NG. Only one base-edited transgenic line was obtained and its sequencing chromatogram at *OsSPL7* target site is shown. E, ABE-NG can edit a TGA PAM site in *OsIAA13*. Sequencing chromatograms of two base edited lines, Line 112-11 and Line 112-15, at the *OsIAA13* target site are shown. Arrows in B-E point to the edited bases.

**Figure 4. Cytosine base editing by SpCas9-NG in rice.** A, Diagrams of the CBE-NG and CBE-NG-hAID cytosine base editors used in this study. B, CBE-NG edits an AGG PAM site in *SNB*. Line 117-2 is a homozygous substitution line, whereas Line 117-21 has undesired indel at the target site besides the expected C-T base conversions. Their sequencing chromatograms at the *SNB* target site are shown. C, *OsSPL7* with an AGA PAM site can be edited by ABE-NG. Line 118-9 has only the expected C-T base substitutions, whereas Line 118-15 has both C-T and C-A base conversions. Their sequencing chromatograms at the *OsSPL7* target site are shown. The C-A conversion is marked by a blue arrow. D, *PMS3* with an AGT PAM site can be edited by CBE-NG. Only one base-edited transgenic line was obtained and its sequencing chromatogram at the *PMS3* target site is shown. E, CBE-NG-hAID edits an AGG PAM site in *SNB* with an expanded base editing window. In Line 125-17, cytosine in the protospacer and upstream of the protospacer could be edited. In Line 125-25, cytosine at positions 4, 5 and 14 in the protospacer of *SNB* were edited. Sequencing chromatograms of these two lines at the *SNB* target site are shown. F, TA cloning results for Line 125-17 at the *SNB* target site. Ten clones were randomly selected for sequencing. The different genotypes and their corresponding sequencing chromatograms are shown. Arrows in B-F point to the edited bases or the mutation sites. The substituted bases are highlighted in blue in B, C and F.
A

B

C

D

E

A

B

C

D

E
A

OsU6 Sp sgRNA → ZmUbi

D(10A) NLS

TadA wt+7-10 nxCas9 3.6 D(10A) NOS

ABE-P6

OsU6 Sp sgRNA → ZmUbi

D(10A) NLS

TadA wt+7-10 nxCas9 3.7 D(10A) NOS

ABE-P7

OsU6 Sp sgRNA → ZmUbi

rAPOBEC1 nxCas9 3.6 D(10A) UGI NOS

CBE-P6

OsU6 Sp sgRNA → ZmUbi

rAPOBEC1 nxCas9 3.7 D(10A) UGI NOS

CBE-P7

B

PAM ACG

TCGTCGGCTGGTGCTCTCTC OsSPL7

C

AATCCTGATGATGCTGCAGT AGG SNB

SNB edited by CBE-P6

Line 48-17

SNB edited by CBE-P7

Line 54-11

D

Reference AATCCTGATGATGCTGCAGT AGG SNB

Allele 1 AATCCTGATGATGCTGCAGT AGG WT

Allele 2 —— TGATGATGCTGCAGT AGG 5 bp deletion

Line 54-10
A

ABE-NG

OsU6  Sp sgRNA → ZmUbi → TadA wt  TadA 7-10  nSpCas9-NG (D10A)  NLS

ABE-NG-S

OsU6  Sp sgRNA → ZmUbi → TadA 7-10  nSpCas9-NG (D10A)  NLS

B

PAM

CCG  ACTCGAGCTGTGCTCTCTCTCTC  OsSPL14

Edited by ABE-NG

PAM

Line113-17

Edited by ABE-NG-S

PAM

Line109-9

D

PAM

ACG  TCGTCGGCTGTGCTCTCTCTCTCTCTCTCT

OsSPL7

PAM

Line114-13

C

TGCGATGAAGGTCTCATGCT

PAM

AGC  LF1

Edited by ABE-NG

Line111-17

Edited by ABE-NG-S

Line107-19

Line107-25

E

PAM

TCA  GGGCGCAGGCAGTGTTGG  OsIAA13

Line112-11

Line112-15