ABSTRACT

Title of Dissertation: DEVELOPING AND IMPROVING CRISPR-BASED GENOME EDITING TECHNOLOGIES IN PLANTS

Simon Sretenovic, Doctor of Philosophy, 2022

Dissertation directed by: Dr. Yiping Qi, Associate Professor, Department of Plant Science and Landscape Architecture

Earlier genome editing technologies were developed based on programmable nucleases including zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN), both requiring tedious protein engineering. By contrast, clustered regularly interspaced short palindromic repeats (CRISPR) systems, such as CRISPR-Cas9, has revolutionized the genome editing field in the past decade due to ease of guiding Cas9 endonuclease to the target site by programmable guide RNAs. However, not every target site can be edited due to Cas9 endonucleases' recognition of so-called protospacer adjacent motif (PAM) when binding to the target site. For example, the PAM for the widely used SpCas9 is NGG (N=A, C, T or G). This drastically limits targeting scope in the genomes. Thus, researchers have developed engineered Cas9 variants recognizing more relaxed PAMs and tested them in mammalian cell lines. Repair of Cas9 endonuclease-induced double strand breaks through non-homologous end joining (NHEJ) DNA repair pathway typically generates unspecified insertions and deletions, which is a missed

opportunity for introducing precise edits. To confer precise genome editing, CRISPR-Cas9 derived base editors and prime editors have been developed. In this work, expanding the plant genome editing scope with engineered Cas9 variants, improving precise cytosine and adenine base editing in plants as well as demonstrating prime editing in plant cells were pursued. The technologies were tested in the model crop, rice, in transiently transformed protoplasts and stably transformed T0 lines. Findings suggest that engineered Cas9 variants can drastically expand the targeting scope for targeted mutagenesis and base editing in plants. Additionally, newer genome editing technologies such as base editors and prime editors can be applied in plants to achieve precise genome editing with varying efficiencies. These validated and useful CRISPR-Cas9 genome editing toolkits have been deposited to the public vector depository, Addgene. Adoption of these genome editing technologies by plant scientists and breeders will enable basic research discoveries and accelerate breeding of next generation crops, ensuring global food security amidst climate change and increasing global population.

DEVELOPING AND IMPROVING CRISPR-BASED GENOME EDITING TECHNOLOGIES IN PLANTS

by

Simon Sretenovic

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Yiping Qi, Chair
Professor Gary Coleman
Professor James Culver
Professor José Feijó
Professor Shunyuan Xiao

Acknowledgements

I would like to express my gratitude to my mentor and chair of my committee, Prof. Yiping Qi for his patience, support and feedback. My gratitude also extends to my defense committee for their guidance.

This work would not have been possible without the help from my lab mates as well as our collaborators Prof. Yong Zhang from the University of electronic science and technology of China, Prof. Tao Zhang from Yangzhou University in China and their lab members. Their efforts are tremendously appreciated and listed in introduction sections of chapters 3, 4, and 5. Thank you all for a fruitful collaboration!

For financial support, I would like to thank Rockey FFAR Fellow program as well as Syngenta and Inari Agriculture. Furthermore, Dr. Qiudeng Que from Syngenta and Dr. Neena Pyzocha, Dr. Hélène Bergès and Dr. Greg Peel from Inari Agriculture, as well as my mentor, Prof. Yiping Qi, provided valuable guidance in my professional development through Rockey FFAR Fellowship. Prof. Rebecca Dunning, the director of Rockey FFAR Fellows program, contributed tremendous efforts to provide a carefully curated experience with plethora of opportunities to hone the non-technical skills via virtual as well as in person sessions. Thank you all for your contributions towards my professional development!

I am thankful to the Department of plant science and landscape architecture and the UMD Research greenhouse complex staff for providing the logistic and administrative support during our research.

Last but surely not least, I am also grateful to my parents for their love and support. I also appreciate all my friends stateside for spending joyful time together or an ocean away for cheering me on.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A3A, APOBEC3A	apolipoprotein B mRNA editing enzyme catalytic subunit 3A
ABE	adenine base editor
ADAR	adenosine deaminase acting on RNA
AP site	apyrimidinic/apurinic site
BE	base editor
BH domain	bridge helix domain
Cas9	CRISPR-associated protein 9
Cas9n	CRISPR-associated protein 9 nickase
Cas9-WT	CRISPR-associated protein 9 wild type
CBE	cytosine base editor
CGBE	cytosine to guanine base editor
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	crispr RNA
dCas9	dead or deactivated CRISPR-associated protein 9
DNA	deoxyribonucleic acid
DSB	double stranded break
dsDNA	double stranded DNA (deoxyribonucleic acid)
eA3A, eAPOBEC3A	engineered apolipoprotein B mRNA editing enzyme catalytic subunit 3A
EMS	ethyl methanesulfonate
ExoI	exonuclease I
GATK	genome analysis toolkit
hAID	human activation-induced deaminase
HDR	homology dependent repair
Indel	insertion and/or deletion
MCP	MS2 bacteriophage coat protein
MLH1dn	dominant negative human MutL homolog 1
MMEJ	microhomology-mediated end joining
M-MLV-RT	Moloney Murine leukemia virus reverse transcriptase
MMR	mismatch DNA repair
NHEJ	non-homologous end joining
nsgRNA	nicking single guide RNA
NUC lobe	nuclease lobe
PAM	protospacer adjacent motif

PBS	primer binding site
pco	plant codon optimization
pco-Cas9	plant codon optimized CRISPR-associated protein 9
PE	prime editor
pegRNA	prime editing guide RNA
PI domain	PAM interacting domain, protospacer adjacent motif interacting domain
PmCDA1	Petromyzon marinus cytidine deaminase1
pre-crRNA	precursor crisprRNA
rAPOBEC1	rat apolipoprotein B mRNA editing enzyme catalytic subunit 1
RBE	RNA base editors, ribonucleic acid base editor
REC lobe	recognition lobe
REPAIR	RNA editing for programmable A to I replacement
RESCUE	RNA editing for specific C to U exchange
RNA	ribonucleic acid
RNP	ribonucleoprotein complex
RT	reverse transcriptase
RT template	reverse transcriptase template
scFv	single-chain variable fragment
sgRNA/gRNA	single guide RNA/guide RNA
SpCas9	*Streptococcus pyogenes* Cas9
ssDNA	single stranded DNA
TadA	tRNA-specific adenosine deaminase
TALEN	transcription activator-like effector nucleases
T-DNA	transfer DNA
tracrRNA	trans-activating crispr RNA
tRNA	transfer RNA
UNG	uracil N-glycosylase
WGS	whole genome sequencing
XRCC1	X-ray repair cross complementing protein 1
z-Cas9	maize codon optimized CRISPR-associated protein 9
ZFN	zinc finger nucleases
ZmUBI promoter	maize ubiquitin 1 promoter

Chapter 1: Literature review

1.1 Genome editing

1.1.1 Genome editing platforms

Throughout history, crops have been conventionally bred by harnessing naturally occurring mutations to produce progeny with beneficial traits. Modern breeding technologies include germline mutation breeding with physical or chemical mutagens, RNA interference gene silencing, transgenes based biotech crop development, and genome editing with nucleases. In broadest of terms, genome editing can be described as a manipulation of a living organism's genomic DNA sequence by replacing, deleting or inserting another DNA sequence. The concept of genome editing materialized especially with the discovery that endonucleases could be engineered to create site-specific double stranded breaks (DSBs) in the DNA molecules that could increase the introduction of desired DNA sequence into genomic DNA via homologous recombination by more than one thousand fold (1–3). These sequence specific endonucleases include meganucleases (also called homing nucleases) (4,5), zinc finger nucleases (ZFNs) (6), transcription activator-like effector nucleases (TALENs) (7,8), and clustered regularly interspaced short palindromic repeats (CRISPR) associated nucleases (9–12). Meganucleases, ZFNs and TALENs depend on physical protein-DNA interaction to confer target site specificity, which is difficult to engineer and multiplex. CRISPR technology has become the obvious choice for genome editing

applications due to its programmable RNA-guided property and high genome editing efficiency (10).

1.1.2 Classification of CRISPR-Cas systems

CRISPR-Cas systems are adaptive defense systems of bacteria and archaea that protect the prokaryotic cells from invading viral DNA as well as mobile genetic elements (13). A typical CRISPR-Cas locus consists of a CRISPR array, comprised of several short direct repeats (of approximately 25 to 36 nucleotides in size) separated by unique spacer sequences, and adjacent *Cas* genes that form one or more operons (14). CRISPR-Cas mediated defense occurs in three phases: adaptive, expression and interference phase. During the adaptive phase, bacteria and archaea carrying one or more CRISPR loci respond to the invading DNA by integrating short fragments of the invading DNA (protospacers) into the host chromosome, proximally to the CRISPR array (15–17). During the expression and interference phases, transcription of the repeat-spacer sequence into precursor crispr RNA (pre-crRNA) followed by the enzymatic cleavage produces crispr RNA (crRNA) that can pair with the complementary protospacer of the invading DNA (18). Target recognition by crRNAs directs the silencing of the invading DNA by Cas proteins that function as a crRNA-protein complex (19,20).

Classification of diverse CRISPR systems is based on Cas proteins functional, structural and phylogenetic properties into two classes. Class 2 CRISPR systems process pre-crRNA and mediate invading DNA cleavage in the expression and interference phases, respectively, with a single multifunctional protein compared to

Class 1 systems (21). Each class is further divided into several types according to their signature proteins. Type I with Cas3, type III with Cas10 and type IV with Cas8 signature protein belong to class 1 while type II with Cas9, type V with Cas12a-i, Cas12j (22), Cas12k (23), and type VI with Cas13a-d signature proteins belong to class 2 as reviewed by Makarova et al. (24). Each type of CRISPR systems is organized in subtypes based on operon organization and Cas proteins found in CRISPR loci. It is important to note that classification of CRISPR systems is constantly under development and revision due to identification of novel CRISPR systems.

1.1.3 CRISPR-SpCas9 genome editor description and mechanism of introducing DSB

CRISPR-Cas9 genome editor consists of RNA-guided Cas9 endonuclease and synthetic single guide RNA (sgRNA/gRNA), which is comprised of crispr RNA (crRNA) and trans-activating crispr RNA (tracrRNA). The crRNA part contains an approximately 20 nucleotides long spacer that complements a sequence of the target site, also called a protospacer, while the tracrRNA part of sgRNA plays a pivotal role in Cas9 recruitment (10). Successful targeted DNA cleavage also requires a recognition of protospacer adjacent motif (PAM), located downstream of protospacer, directly by Cas9. After the introduction of DSB in the targeted region of a genome, either non-homologous end joining (NHEJ) (25), microhomology-mediated end joining (MMEJ) (26) or homology-directed repair (HDR) (27) mechanism of DNA repair is activated to repair the double strand gap in DNA. Error prone NHEJ and MMEJ DNA repair mechanisms may disrupt targeted genes by introducing insertions, deletions, substitutions or other DNA rearrangements at the DSB site (28,29), while error free

HDR confers insertion of template DNA through homologous recombination (27) (Figure 1).

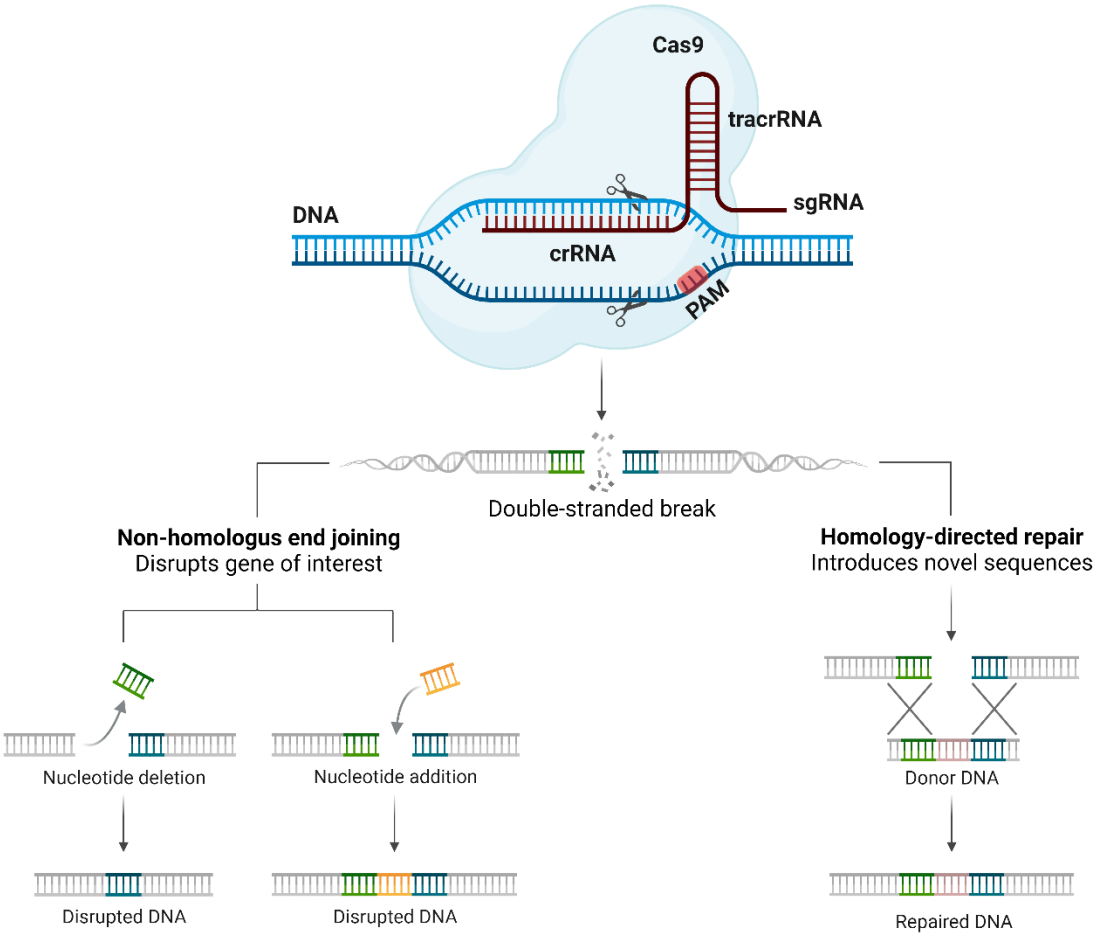


Figure 1: The mechanism of introducing double stranded DNA break by CRISPR-Cas9. sgRNA consists of crRNA that contains approximately 20 nt recognition sequence and tracrRNA that recruits Cas9. Successful targeted dsDNA cleavage also requires a recognition of protospacer adjacent motif (PAM), located downstream of protospacer (highlighted in red), directly by Cas9. After the introduction of blunt DSB at the target site three base pairs upstream of PAM (depicted by scissors), either non-homologous end joining (NHEJ) or homology-directed repair (HDR) mechanism of DNA repair is activated to repair the double strand gap in DNA. Error prone NHEJ repair mechanisms may disrupt targeted genes by introducing insertions, deletions, substitutions or other DNA rearrangements at the DSB site, while error free HDR confers insertion of template DNA through homologous recombination.

Streptococcus pyogenes Cas9 (SpCas9, Cas9-WT) recognizes a simple, G rich PAM (5'-NGG-3') (10), making it the commonly utilized CRISPR-Cas9 system in the vast majority of organisms, including plants. SpCas9 is a large (1386 amino acid, 4.1 kbp, 160 kDa) multidomain and multifunctional DNA endonuclease. The crystal structure of CRISPR-Cas9 endonuclease revealed that Cas9 consists of two lobes, a recognition (REC) lobe and a nuclease (NUC) lobe (30). The REC lobe contains REC1 domain responsible for binding sgRNA, REC2, and bridge helix (BH) domain crucial for initiating DNA cleavage upon binding of target DNA. The NUC lobe contains RuvC and HNH domains as well as PAM interacting (PI) domain (30). Cas9 is capable of cleaving dsDNA within the protospacer, three base pairs upstream of the PAM using two distinct nuclease domains. An HNH-like nuclease domain cleaves the DNA strand complementary to the sgRNA sequence (target strand), while an RuvC-like nuclease domain cleaves the DNA strand opposite the complementary strand (nontarget strand) (31,32). Introduction of a D10A mutation in the RuvC domain or H840A in the HNH domain generates a Cas9 nickase (Cas9n) that only cleaves the target or nontarget strand, respectively (10,11,33,34). When both point mutations are introduced, the Cas9 becomes catalytically inactive (dCas9). Cas9-WT, Cas9n or dCas9 can be utilized to develop many versatile genome engineering toolkits like base editors (35–39), prime editors (40) and transcriptional regulators (41). The CRISPR-Cas9 editor with sgRNA can be delivered as DNA using *Agrobacterium*-mediated transfer DNA (T-DNA) or biolistic-mediated transformation, whereas RNA reagents encoding the CRISPR-Cas9 editor and sgRNA or ribonucleoprotein (RNP) complex can be transformed or delivered using biolistic delivery to protoplasts or somatic and germline tissues (42).

A drawback of CRISPR-Cas9 genome editing is introducing genome edits at unintended genomic loci, which is called off-target editing. The off-target editing of Cas9 is usually a major concern of CRISPR technology, especially at sites that are similar, but not exact to the intended target site (43).

1.1.4 Expanding targeting scope with engineered Cas9 variants and orthologues

A key disadvantage of CRISPR-SpCas9 genome editor is its narrow targeting scope conferring editing of targets associated with 5'-NGG-3' canonical PAMs and with lower activity at 5'-NAG-3' PAM sites (10). Researchers have opted for two approaches of obtaining Cas9 with altered PAM requirements, the first relying on screening SpCas9 orthologs and the second relying on Cas9 protein engineering. Several SpCas9 orthologs have been explored for genome editing including *Staphylococcus aureus* Cas9 (SaCas9) recognizing NNGRRT PAMs (44), *Streptococcus thermophilus* Cas9s (St1Cas9 recognizing NNAGAAW PAMs and St3Cas9 NGGNG PAMs) (45), *Brevibacillus laterosporus* Cas9 (BlatCas9) recognizing NNNNCND PAMs (46) and many others mainly demonstrated in mammalian cell lines (N=A/C/T/G, R=A/G, W=A/T, D=A/G/T). In general, Cas9 orthologs expand the targeting scope beyond target sites associated with canonical NGG PAMs of SpCas9 and can be used for orthogonal genome engineering.

Within the Cas9 protein engineering approach of expanding the targeting scope, there are three ways. The first way is based on protein directed evolution. For example, xCas9-3.7 was developed using phage-assisted continuous evolution to target relaxed

NG, GAA, GAT and CAA PAMs (47). The second way relies on structure-directed mutagenesis. For example, Cas9-NG and Cas9-NGv1 (48), also claimed to target NG PAMs in mammalian cell lines, were engineered in this manner harboring seven amino acid substitutions. A near PAM-less SpCas9 variant, called SpRY Cas9, has also been developed in this way carrying eleven amino acid substitutions (49). Additionally, SpCas9-VQR, -EQR, and -VRER variants can recognize NGA, NGAG and NGCG PAMs, respectively (50). The third way includes swapping PI domains between Cas9 orthologs. The SpyMacCas9 variant has been established in this way by grafting the PI domain from *Streptococcus macacae* Cas9 (SmacCas9) recognizing NAAN PAMs to its well established ortholog from *Streptococcus pyogenes* (SpCas9) (51). Compared to canonical SpCas9, all the engineered variants with relaxed PAMs generally performed worse than SpCas9 at sites associated with canonical PAMs.

1.1.5 CRISPR-Cas9 as a recruitment platform

The sgRNA programmable DNA binding and cleaving properties of CRISPR-Cas9 system provide a powerful recruiting platform of many functional domains through protein fusion or sgRNA-protein interactions. Those functional domains include FokI nuclease domains (50), several transcriptional activators and repressors, deaminases (35,36), reverse transcriptase (40), epigenome modifiers, modifiers of DNA repair pathways and others. Functional domains can be fused directly to N or C terminus of Cas9 protein through flexible linker (like XTEN (52)) or self-cleaving peptides (like T2A (53)). The unhindered genome editing of Cas9 with modified or extended at the 3’ end tracrRNA part of the sgRNA allows for the recruitment of functional domains through sgRNA-protein interaction. Several RNA-binding protein–RNA aptamer

systems have been used for recruitment of functional domains, including MCP-MS2 (33,54–56). In the MCP-MS2 system, as MS2 bacteriophage coat protein (MCP) binds to MS2 RNA aptamer in dimers, the gRNA2.0 scaffold with two MS2 stem loops can recruit four additional copies of the functional domain fused to MCP protein via a linker (56,57). To increase the focal concentration of the functional domains at the target site, the supernova tagging system (SunTag) utilizes a repeat of GCN4 epitopes fused to dCas9, which are bound by scFv antibodies fused to functional domains (58). By coupling the SunTag system with the MS2-MCP interaction to drastically increase the focal concentration of activators, the CRISPR-Act3.0 was developed for multiplexed gene activation in plants (59) as well as CRISPR-Combo facilitating simultaneous genome editing (targeted mutagenesis or base editing) and gene activation (60). The aforementioned protein recruiting strategies can be used individually or in combination to drastically increase the focal concentration of the same or various functional domains, expanding genome engineering outcomes in plants.

1.2 Base editing

1.2.1 Base editing platforms

Base editing is a precise, CRISPR-based genome editing technology enabling direct, irreversible conversion of one target nucleotide into another in a programmable manner, without requiring the introduction of DSB at the target site or a donor template (35). In general, base editors (BEs) are comprised of catalytically impaired Cas endonuclease that is fused via a linker to nucleotide deaminase and in certain cases to proteins modifying inherent DNA repair mechanisms. BEs can introduce single

nucleotide conversions at the target sites in RNA as well as in nuclear or organellar DNA. There are two types of RNA base editors (RBEs): RBE for A to I (inosine) editing and RBE for C to U (uracil) editing. The RNA editing for programmable A to I replacement (REPAIR) system has been developed based on Cas13b (that has no PAM requirement) fused to adenosine deaminase acting on RNA (ADAR) (61). ADAR converts A to I, which is during the cellular splicing and translation read as guanosine. REPAIRv1 and REPAIRv2 systems have also been prepared based on engineered ADAR that confers reduced off-target editing (61). Using directed evolution approach, researchers have further evolved ADAR to act like cytidine deaminase acting on double stranded RNA. Fusing evolved ADAR to dCas13 resulted in RNA editing for specific C to U exchange (RESCUE) system, which is capable of both adenosine and cytidine deamination (62).

DNA base editors can be grouped into cytosine BEs (CBEs), adenine BEs (ABEs), C to G BEs (CGBEs), dual BEs as well as organellar BEs. Organellar BEs heavily rely on DddA cytidine deaminase acting on dsDNA conferring unwinding dsDNA and C to T transition (63). DddA is usually coupled with TALEN platform (8) for easier translocation into the plant mitochondria as well as plastids. Dual base editors consist of cytidine and adenosine deaminases as well as uracil-glycosylase inhibitor (UGI) that increases C to T base editing efficiency. These effectors are fused to a catalytically impaired Cas9 nuclease (Cas9n) and can simultaneously introduce C to T and A to G substitutions at the target site (64–68). CGBEs are the only base editors capable of introducing transversions opposed to transitions introduced with other BEs. CGBEs are

comprised of uracil N-glycosylase (UNG) fused to Cas9n and rAPOBEC1 cytidine deaminase (39) or engineered rAPOBEC1 with (R33A) (38). During C to G base editing, C is deaminated by cytidine deaminase to U and UNG generates an apyrimidinic/apurinic (AP) site by removing U. Subsequently, error-prone polymerase incorporates G at the AP site, resulting in base transversion (38,39). Another CGBE was developed by fusing X-ray repair cross complementing protein 1 (XRCC1) to APOBEC1-Cas9n (37). In this case, a cellular UNG removes the U generated by C deamination with APOBEC1 deaminase and an AP site is created. Afterwards, XRCC1, a base excision repair protein helps with the preferred incorporation of a G at the AP site resulting in a C to G conversion (37).

1.2.2 C to T base editing

Base editing technology was first introduced in 2016 with invention of the CBEs (35). First generation of CBE (BE1) was engineered by fusing a cytidine deaminase with catalytically inactivated Cas9 (dCas9) to abolish dCas9 cleaving activity while retaining its DNA binding activity. In human cell lines, BE1 demonstrated C to T base editing efficiency spanning between a few percent to about ten percent. Later generations of CBE (BE3 and BE4) employed Cas9n with the D10A mutation to inactivate catalytic activity of RuvC-like domain while HNH domain can still cleave the complementary, non-edited strand (35). BE3 (Figure 2A) and BE4 generations also use Cas9n fused UGI with BE4 containing two copies of UGI. Within the nucleoprotein complex, sgRNA guides CBE to the intended target site where the dsDNA is nicked. N terminally fused cytidine deaminase to Cas9n confers hydrolytic deamination of C within the target site to U. U is recognized by cell replication or repair machinery as T

thus facilitating a permanent C to T transition. UGI inhibits the action of UNG that confers reversion of U back to C thus increasing the base editing efficiency (69) (Figure 2B). In human cell lines, BE3 achieved base editing efficiency of up to 40% (35). BE3 configuration of CBEs have become the preferred choice of many researchers including plant biologists who have demonstrated CBEs in many plant species not limited to rice (70–75), *Arabidopsis* (76–79), wheat (73,80), maize (73), tomato (75,81), potato (71,81–83), watermelon (84), cotton (85), soybean (86), apple (87), pear (87), strawberry (88), rapeseed (89,90), moss (91), and poplar(92). CBEs displayed various base editing efficiencies in protoplasts or stable transgenic T0 lines, with on average lower base editing activities than in mammalian cell lines.

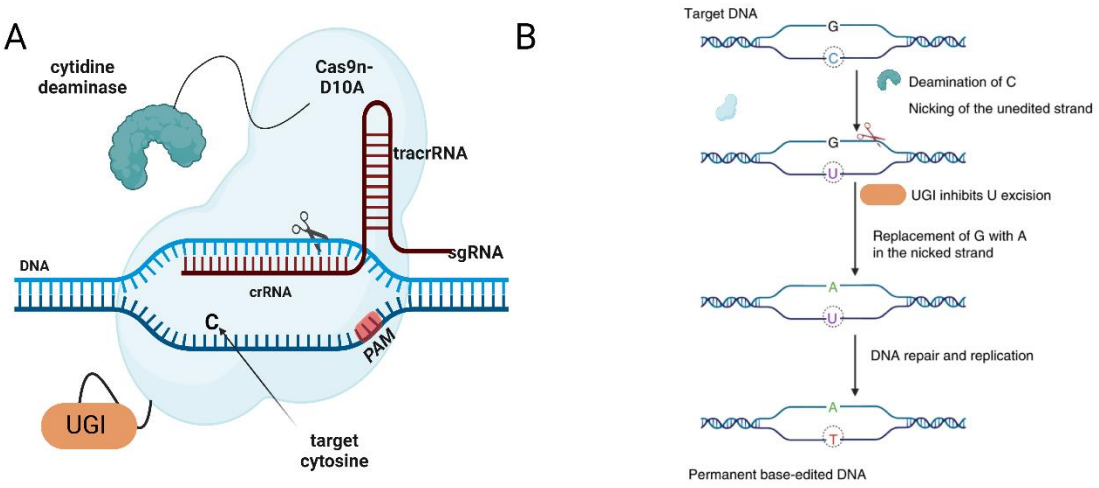


Figure 2: C to T base editing of genomic DNA. **A)** BE3 architecture of a C to T base editor and **B)** schematics of the mechanism of introducing permanent C to T transition at the target site. The cytidine deaminase catalyzes hydrolytic deamination of a target C to U. The generated U•G mismatch can be resolved by cellular mismatch repair, base excision repair or replication machinery. The Cas9n makes a nick in the G containing unedited DNA strand which prompts the removal of G by the cellular mismatch repair pathway. The use of U as a template for repair leads to the intended T•A base pair formation. Cellular uracil DNA N-glycosylase (UNG) removes U from the DNA, leading to a reversion to the original C•G base pair. Uracil glycosylase inhibitor (UGI) increases the rate of T•A introduction by inhibiting the excision of U.

Several cytidine deaminases have been repurposed for C to T base editing, including rAPOBEC1 (35), PmCDA1 (75,93,94), hAID (95), A3A (71,96), as well as engineered A3A/Y130F (96) to confer base editing at methylated regions, and eA3A to reduce the bystander effect when multiple Cs are present in the activity window (97). The activity window can be defined as the scope of nucleotides within the protospacer that are optimally edited by a BE. The activity window can vary among CBEs and can on average span between four and twelve nucleotides.

One of the base editing applications constraints is the targeting scope of SpCas9 being limited to targets associated with canonical NGG PAMs. To overcome this constraint, base editors have been coupled to several orthologous Cas9 proteins as well as engineered Cas9 variants with alternative PAM requirements (98). Another constraint of base editing is off-target editing occurring at genomic sequences with high homology to the intended target site, also referred to as gRNA-dependent off-target editing. CBEs seem to have fewer gRNA-depended off-target effects compared to Cas9 genome editors targeting the same target sites (99–101). Base editors can also introduce sgRNA-independent off-target editing at genomic sites that do not bear any sequence similarities with the intendent target site. Whole genome sequencing studies have revealed that CBEs could generate genome wide gRNA-independent off-target mutations, being attributed to cytidine deaminases, in mouse (102) and rice (103). CBEs can also cause transcriptome-wide off-target effects (104). While CBEs are a potent and precise genome editing platforms, they could be further improved in

activity, specificity and purity by adopting novel cytidine deaminases for base editing as well as modifying CBEs' architecture.

1.2.3 A to G base editing

Since none of the reported naturally occurring adenosine deaminases work on DNA, Liu and coworkers evolved a transfer RNA (tRNA) adenosine deaminase (TadA) variant that works on ssDNA (36). The deamination of adenine by evolved adenosine deaminase produces inosine, which is treated by polymerases as a guanine, thus during the replication A to G transition in installed (36) (Figure 3). ABEs have been demonstrated in various plants, including rice (105–110), wheat (107), *Arabidopsis* (111), *Brassica napus* (111), *Nicotiana benthamiana* (112), poplar (92), and moss (91).

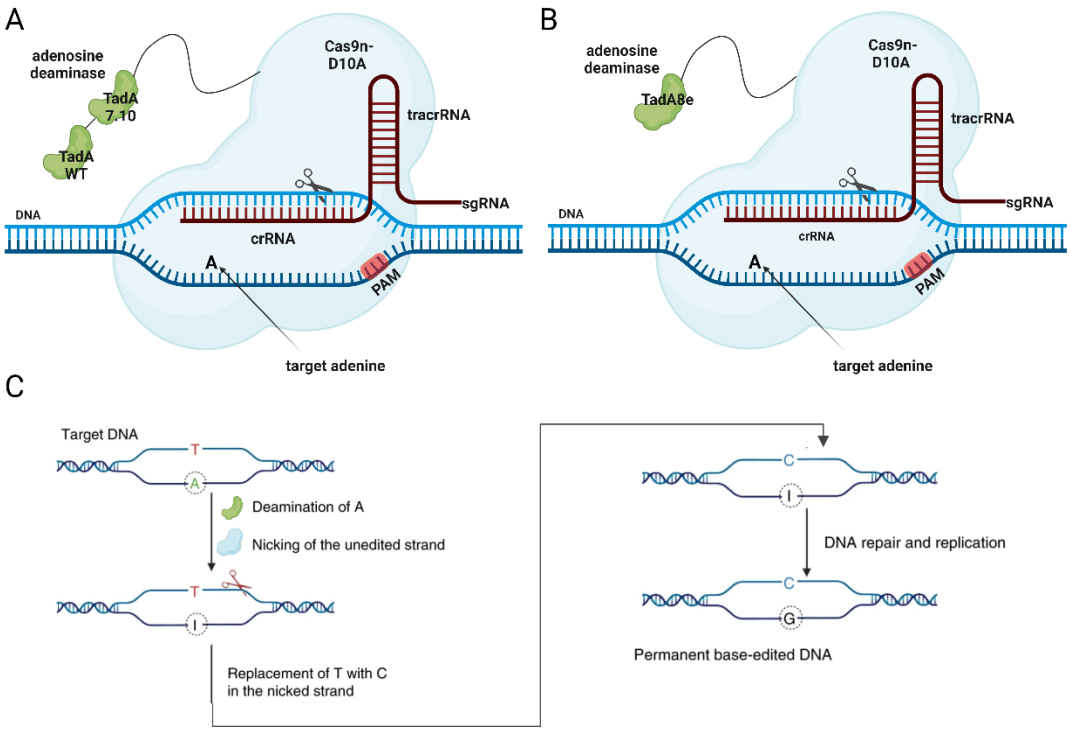


Figure 3: A to G base editing of genomic DNA. **A**) A to G base editor with evolved TadA-7.10 and TadA-WT adenosine deaminases. **B**) A to G base editor with evolved TadA8e adenosine deaminase. **C**) Schematics of the mechanism of introducing permanent A to G transition at the target site. Once the ABE is recruited to the targeted site, the target A is

deaminated to inosine (I), thus an A•T pair becomes I•T pair. I is recognized as G and pairs with C during the replication. The I•T pair is resolved by cellular mismatch repair or DNA replication. The Cas9n nicks the T containing strand to prompt the removal of T by the mismatch repair machinery and introduce C opposite to I form an I•C pair that eventually becomes a G•C pair.

Several versions of TadA adenosine deaminase were tested with TadA-7.10 demonstrating the highest A to G base editing activities amongst the early evolved variants. TadA-7.10-Cas9n ABE favorably edited the target As located at the protospacer 4-9 positions (counting the NGG PAM of SpCas9 as 21-23) (36). Although earlier reports on ABEs in mammalian cell lines and plant protoplasts described the use of wtTadA-TadA-7.10 heterodimer (99,101), the latest studies demonstrated that the TadA-7.10 monomer is sufficient for successful ABE activity in mammalian cell lines (113–115) and rice (116–118). TadA-7.10 adenosine deaminase was further evolved to generate TadA8e, with substantially increased deamination kinetics (113). Fusing TadA8e evolved deaminase with Cas9n yielded ABE8e, that has demonstrated high efficiency base editing in rice (116,118–120), *Nicotiana benthamiana* (112) and mammalian cell lines (113). In mammalian cell lines, the activity window of ABE8e was between 4 and 8 positions within the protospacer (counting the NGG PAM of SpCas9 as 21-23) and differ quite dramatically when TadA8e was fused with orthogonal Cas9ns that expanded base editing scope to target sites associated with non-canonical PAMs (113). By introducing two additional mutations (V82S and Q154R) into TadA8e, ABE9 was prepared for improved base editing in rice (116).

Compared with CBEs, ABEs produce cleaner products with less unintended edits. ABEs in rice and wheat plants did not exhibit unintended edits at both on- and off-

target genomic sites (106,107). However, a study reported on unintended proximal base editing to the target sites in rice plants expressing ABEs (105). Furthermore, a recent study of genome-wide analyses of TadA8e-SpRY Cas9 ABE revealed substantial off-target effects in rice plants (121,122). ABEs can also cause transcriptome-wide off-target effects (122).

1.3 Prime editing

1.3.1 Prime editing platform

Prime editing is a versatile and precise CRISPR-Cas9 based genome editing technology that directly writes small-scale desired edits to a DNA target site using an engineered reverse transcriptase (RT) fused to a C terminal of a Cas9n-H840A. Prime editors (PEs) can introduce all 12 possible substitutions, small indels as well as combinations thereof with quite favorable intended editing to unintended byproduct ratios (40). RT is programmed with a prime editing guide RNA (pegRNA) that specifies the target site and encodes desired edit. PegRNA is a modified sgRNA with 3' extension of RT template and primer binding site (PBS) sequences (Figure 4A). The Cas9n-H840A is used to cleave the edited strand of dsDNA at the target site. Afterwards, the nicked strand is used for priming the reverse transcription of an edit-encoding extension (RT template) on the pegRNA directly into the target site (40) (Figure 4B). This results in a branched intermediate comprised of two competing ssDNA flaps. The 3' flap contains the edited sequence while the 5' flap contains the unedited sequence and is preferentially cleaved by structure specific endonucleases like FEN1 (123) or 5' exonuclease such as ExoI (124) in mammalian cells. Ligation of the 3' flap incorporates

the edited DNA strand into the heteroduplex DNA containing an edited and unedited strand. Finally, to resolve the heteroduplex, DNA repair machinery permanently installs the intended edit by copying the information from the edited strand to the complementary strand (Figure 4C).

In the seminal paper about prime editing (40), three versions of prime editing systems were introduced. Prime editor 1 (PE1) consists of Cas9n-H840A with a C terminal fusion of a WT Moloney Murine leukemia virus RT (M-MLV-RT). PE2 incorporates engineered M-MLV-RT pentamutant (D200N/L603W/T330P/T306K/W313F) with increased thermostability, processivity, DNA-RNA substrate affinity and inactivated RNaseH activity. Compared to PE1, PE2 demonstrated about threefold increase in editing efficiency in human cell lines. PE3 involves nicking the non-edited strand to stimulate DNA repair machinery and therefore further increases the editing efficiency by two to four fold compared to PE2. Since the PE3 approach includes double nicking of two complementary DNA strand, a potential DSB formation could lead to indel formation due to NHEJ repair. To reduce the unintended editing outcomes, PE3b was developed. PE3b uses a nicking gRNA with a protospacer that matches the edited strand but not original allele thus temporally separating nicks introduction. In human cell lines, PE3b resulted in thirteen fold decrease in the average number of indels compared to PE3 in human cell lines without compromising editing efficiencies ranging from approximately 30% to 80% (40). Prime editing has been also demonstrated in plant species, including rice (125–132), wheat (126), maize (133), potato and moss (134), *Arabidopsis* and *N. benthamiana* (131), tomato (135) and others. Surprisingly, in plants PE3 or PE3b generally do not yield higher editing

efficiencies than PE2 (125–128,136). Overall, PE2 and PE3 mediated editing efficiencies reported in plants were up to a few percent, drastically lower than the efficiencies reported in the seminal paper in mammalian cell lines (40). PEs have a relatively narrow editing window focused around the PAM site and are capable of introducing insertions and deletions up to 100 bp. As PEs need three tiers of specific hybridization to DNA (pegRNA spacer-target site for binding and nicking the PAM strand at the target site, pegRNA PBS-target DNA for initiating reverse transcription, and RT product-target DNA for flap resolution), the off-target effects of PEs are low. In rice, pegRNA-dependent off-target editing was detected below 0.3% at target sites similar but not exact to the intended target site (137). Furthermore, whole genome sequencing of 29 PE-edited rice plants confirmed that PEs do not induce genome-wide pegRNA-independent off-target edits (137).

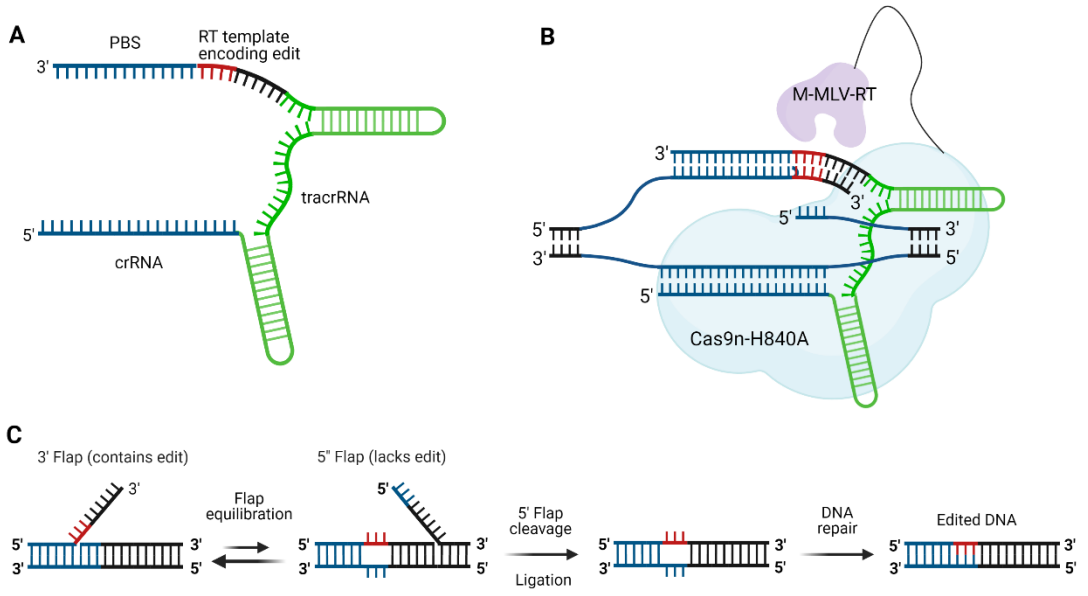


Figure 4: Prime editing of genomic DNA. **A**) pegRNA consists of crRNA, tracrRNA, RT template and primer binding site (PBS). crRNA guides the prime editor (PE) to the target site, tracrRNA recruits and renders Cas9n-H840A nickase active, the RT template encodes the desired edit and the PBS primes the reverse transcription. **B**) PE, a ribonucleoprotein consisting

of pegRNA, Cas9n-H840A nickase and C-terminally fused M-MLV-RT located at the target site within the genomic DNA. Cas9n-mediated R-loop enables the nicked exposed single DNA strand to be primed and extended with M-MLV-RT. **C**) The mechanism that leads to the permanent incorporation of the desired edit into targeted region of the double-stranded DNA. Reverse transcription generates a branched intermediate consisting of two competing single-stranded DNA flaps. The 5′ flap containing the non-edited sequence is preferentially cleaved and the 3′ flap containing the desired edit is ligated and forms heteroduplex DNA. DNA mismatch repair mechanism enable the permanent installation of the edit into double-stranded DNA.

1.3.2 Improving prime editing

While prime editing technology holds huge potential for plant genome editing, its efficiency needs further improvement. In order to achieve optimal editing efficiency at specific target site, a pegRNA should be tailored for each target independently. The RT template and PBS within a pegRNA present several possible design for introducing the same intended edit. In general, RT template should be between 10-20 nucleotides long, while PBS should be approximately 13 nucleotides long, but can be between 7 to 17 nucleotides long. The melting temperature (Tm) is important for the stability of DNA-RNA hybrids and can dramatically influence prime editing efficiency. The recommended Tm of a PBS is approximately 30°C to obtain maximal efficiency (138). Several strategies have been developed to boost PE efficiency, including the use of paired pegRNAs encoding the same edits in trans and designing a PBS with a Tm of 30°C (138). Extending pegRNA at 3’ end with a sequence that forms a secondary structure can also increase prime editing efficiency by preventing pegRNA degradation by cellular exonucleases (139). Using pooled CRISPR interference screen, researchers discovered that mismatch DNA repair (MMR) activity suppressed the efficiency and outcome purity of prime editing. These insights led to the development of PE4

(PE2+MLH1dn) and PE5 (PE3+MLHdn) systems that coexpress dominant negative MLH1 (MLH1dn) to transiently inhibit MMR, enhance prime editing efficiency and purity (140). PE4 and PE5 can be coupled with extended pegRNAs (139) to prevent their degradation by exonucleases to further improve prime editing.

Chapter 2: Materials and methods

2.1 Vector construction

T-DNA vectors for genome editing, base editing and prime editing were prepared using Golden Gate and Gateway LR assembly reactions based on protocols described previously (41). Briefly, forward and reverse primers were phosphorylated with T4 polynucleotide kinase (NEB, catalog #M0201*), annealed, and ligated with T4 DNA ligase (NEB, catalog #M0202*) into BsmBI (ThermoFisher, Catalog #ER045*) restriction digested pYPQ141C (Addgene #69292, with the OsU6 promoter) or pYPQ141D (Addgene #69293, with OsU3 promoter) sgRNA entry clones in single step Golden Gate reactions. Individual three way Gateway LR reactions were prepared using attL5-attL2 sgRNA entry clone, attL1-attR5 genome/base/prime editor entry clone, and attR1-attR2 destination vector pPYQ203 (Addgene #86207) containing the ZmUBI promoter for genome/base/prime editor expression. Both sgRNA and genome/base/prime editor entry clone recombination regions were confirmed with Sanger sequencing. Final T-DNA vectors were confirmed by restriction digestion with EcoRI-HF (NEB, catalog #R3101*). Genome/base/prime editor entry clones demonstrating high performance in rice plants were deposited in Addgene vector database. List of DNA oligonucleotides used to prepare the sgRNAs (that define the target sites), list of target sites and list of final T-DNA vectors can be found in Supplementary materials of publication (141) for xCas9-3.7, Cas9-NGv1 and Cas9-NG genome and base editors used in chapter 3, (142,143) for (i)SpyMacCas9 genome and base editors used in chapter 3, (144) for SpRY Cas9 genome and base editors used in chapter 3, (143) for Cas9-WT base editors used in chapter 4. List of DNA

oligonucleotides used to prepare the sgRNAs, list of target sites and list of final T-DNA vectors can be found in Supplementary materials of publication (125) for prime editors used in chapter 5. List of cytidine deaminases adopted for base editing with NCBI GenBank accession numbers and list of final T-DNA vectors used for screening novel base editors used in chapter 4 can be found in Supplementary Table 1 and 2.

Molecular cloning procedures used to prepare genome/base/prime editors used in this dissertation can be found in Supplementary materials of publication (141) for xCas9-3.7, Cas9-NGv1 and Cas9-NG genome and base editors used in chapter 3, (142,143) for (i)SpyMacCas9 genome and base editors used in chapter 3, (144) for SpRY Cas9 genome and base editors used in chapter 3, (143) for Cas9-WT base editors used in chapter 4. Entry clones of novel base editors used in chapter 4 were prepared by ligation with T4 DNA ligase (NEB, catalog #M0202*) of BsrGI-HF (NEB, catalog #R3575*) and BsaI-HFv2 (NEB, catalog #R3733*) digested pYPQ265E2 (Addgene #164719) backbone and digested synthetic dsDNA inserts (encoding cytidine deaminases) purchased from Twist Bioscience.

2.2 Targeted genes and target sites

For testing PAM preference of engineered Cas9 variants in chapter 3, two target sites were chosen for each tested PAM to reduce the sequence specific effects that could prevent Cas9 variants from introducing DSBs. Target sites for base editing were chosen based on PAM preferences of Cas9 variants as well as preferred position of Cs or As in the activity windows of base editors. Target sites for prime editing were chosen based on 5'-NGG-3' PAM preference of canonical Cas9-H840A nickase.

The information about pursued targeted genes and target sites can be found in Supplementary materials of publication (141) for xCas9-3.7, Cas9-NGv1 and Cas9-NG genome and base editors used in chapter 3, (142,143) for (i)SpyMacCas9 genome and base editors used in chapter 3, (144) for SpRY Cas9 genome and base editors used in chapter 3, (143) for Cas9-WT base editors used in chapter 4, and (125) for prime editors used in chapter 5.

2.3 Rice protoplast isolation and transformation

For rice protoplast assay, either *Oryza sativa* Japonica Nipponbare or Kitaake cultivars were used. All rice protoplasts assays in this dissertation were transformed with T-DNA encoded genome/base/prime editors. The rice seedlings were grown on ½ MS solid medium for 10-15 days in the dark at 28°C. Rice protoplast extraction and transformation were perfomred according to previously published protocols in research articles (41,145) and book chapters (146,147). Briefly, stamens of etiolated rice seedlings or healthy leaves were cut into 0.5-1.0 mm strips and transfered into enzyme solution followed by vacuum-infiltration for 30 minutes. Afterwards, the strips were incubated at 60-80 rpm for 8 hours at 25°C in dark. Each digestion suspension was filtered through a 40 µm strainer. After washing the suspension twice with W5 washing buffer, the optional sucrose gradient centrifugation step was performed. Protoplasts were then examined and their titer was adjusted to $2x10^6$ mL^{-1}. For protoplast transformation, 30 ul of midipreped (Qiagen Midiprep kit) plasmid DNA with 1 µg/µL concentration was used to transform 200 µL of protoplasts with $2x10^6$ mL^{-1} concentration by gently adding 230 µL of 40% polyethylene glycol (PEG) transformation buffer. After half an hour incubation at room temperature in dark, the

transformation was concluded by adding 900 µL of W5 washing buffer. The protoplasts were concentrated using centrifugation, transferred into 12 wells culture plates and incubated at 32°C in dark for two days. The protoplast transformation efficiency was determined by GFP screening of at least 50 protoplasts after 12-20 hours using fluorescence microscopy and was above 90% as determined by the number of protoplasts emitting fluorescence divided by a total number of protoplasts. Control samples in rice protoplast assay were protoplasts transformed with deionized water instead of T-DNA plasmids.

2.4 Agrobacterium tumefaciens-mediated stable transformation of rice

The Japonica cultivars Kitaake or Nipponbare was used for rice stable transformation. Since Kitaake cultivar plants are light insensitive, it is easier to obtain the seeds after plant regeneration. *Agrobacterium* mediated transformation was performed according to a previously published protocol (41). Briefly, a successful transformation of *Agrobacterium tumefaciens* EHA105 with T-DNA plasmids was performed using a freeze thaw method (148). For rice transformation, dehusked seeds were disinfected in 70% ethanol for 1 min and washed five times with sterilized deionized water. The seeds were further disinfected in 2.5% sodium hypochlorite containing a drop of Tween 20 for 15 min. After washing the seeds for five times, the seeds were disinfected again 2.5% sodium hypochlorite without Tween 20 for another 15 min. The seeds were washed again for five times, dried and cultured on solid medium at 28°C in dark for 2-3 weeks. Actively growing calli were collected for subculturing at 28°C in dark for 1-2 weeks. Transformed *Agrobacterium* cultures were collected and resuspended in liquid LB medium with 100 µM acetosyringone to OD_{600} of 0.06-0.1. Rice calli were

submerged in the *Agrobacterium* suspension for half an hour, dried afterwards and cocultured on solid medium, for 3 days at 25°C in dark. The infected calli were washed five times with sterile deionsed water, dried, transfered to selection medium and incubated for 5 weeks in dark at 28°C. During the five-week selection step, infected calli were transferred to fresh selection medium every two weeks. After the selection step, actively growing calli were transferred to regeneration media and incubated at 28°C for 3-4 weeks at 16h light – 8 h dark cycle. Afterwards, seedlings were transferred to sterile large test tubes containing fresh solid regeneration medium and incubated for 2-3 weeks at same environmental conditions before being transferred to soil for genotyping and seed collection.

2.5 Validation of genome/base/prime editing mutagenesis

2.5.1 Validation of mutagenesis in rice protoplasts

For assessment of genome/base/prime editing in rice protoplasts, protoplasts were collected after 48 h after transformation. Genomic DNA was isolated from transformed protoplasts using CTAB method (149) or with elution buffer from Phire plant direct PCR kit (ThermoFisher, catalog #F130WH). Genome editing mutagenesis of engineered Cas9 variants was determined either with next generation sequencing (NGS) of PCR amplicons or with restriction fragment length polymorphism assay (RFLP). Base editing and prime editing mutagenesis was determined with NGS of PCR amplicons. In case of RFLP, target sites were chosen based on the presence of a type II restriction enzyme site superimposed at the cleavage sites, allowing for the validation of target site mutations with RFLP assay. Target sites were PCR amplified and

subjected to restriction digestion. ImageJ software was utilized to obtain the electrophoresis run fragment band intensities of bands representing uncut and cut DNA fragments. Genome-editing efficiencies were calculated as the ratios of uncut bands against all (cut and uncut) bands. Positive (cut wild type target site) and negative controls (uncut WT target site) were carried out for system calibration. For NGS of PCR amplicons validation, the flanking genomic regions of the target sites were PCR amplified using barcoded primes. The PCR amplicons were sequenced with Illumina HiSeq X platform and analyzed by either CRISPRMatch (150), CAS-Analyzer (151), or CRISPResso2 (152). Top off-target sites were predicted with Cas-OFFinder (153). Further information can be found in Supplementary materials of publication (141) for xCas9-3.7, Cas9-NGv1 and Cas9-NG genome and base editors used in chapter 3, (142,143) for (i)SpyMacCas9 genome and base editors used in chapter 3, (144) for SpRY Cas9 genome and base editors used in chapter 3, (143) for Cas9-WT base editors used in chapter 4, and (125) for prime editors used in chapter 5.

2.5.2 Validation of mutagenesis in rice T0 lines

Sanger sequencing was used to determine genetic modifications of genome and base editing at target sites in T0 plants. Using T0 genomic DNA as a template (isolated from plant tissue with CTAB method (149) or using elution buffer from Phire plant direct PCR kit (ThermoFisher, catalog #F130WH)), the genomic regions containing target sites were PCR amplified. The specificity of PCR reactions was verified by gel electrophoresis. PCR reactions were enzymatically cleaned with the Exo-CIP Rapid PCR cleanup kit (NEB, catalog #E1050*) and Sanger sequenced by Genewiz. The raw data were analyzed with SnapGene software. Multiple superimposed sequencing

chromatograms (resulting from indel-mediated frameshifts) were analyzed using DSDecodeM web tool (154). Further information can be found in Supplementary materials of publication (141) for xCas9-3.7, Cas9-NGv1 and Cas9-NG genome and base editors used in chapter 3, (142,143) for (i)SpyMacCas9 genome and base editors used in chapter 3, (144) for SpRY Cas9 genome and base editors used in chapter 3, (143) for Cas9-WT base editors used in chapter 4, and (125) for prime editors used in chapter 5.

2.6 Whole genome sequencing (WGS) and data analysis

Whole genome sequencing (WGS) of T0 rice lines was conducted to determine genome-wide assessment of off-target effects by base editors in chapter 4. A WGS pipeline is described comprehensively in (143). Succinctly, genomic DNA was isolated using plant genome DNA kit (Tiangen) and sequenced using Illumina HiSeq X platform providing an average sequencing depth of 36.92x. Skewer software was used to remove Illumina TruSeq adapters and filter low quality reads (155). BWA software (156) was used for mapping all cleaned data to rice reference genome TIGR7 (157). The average mapping ration of reads was 99.78% and average genome coverage was 97.57%. Samtools was used for sorting BAM files (158), duplicated reads were marked by Picard. After pre-processing BAM files from each sample, the Genome Analysis Toolkit (GATK) was utilized to realign reads near indels and recalibrate base quality scores (159). Four programs were used for high confidence detection of *de novo* variants in rice genome. SNVs were detected using LoFreq (160), MuTect2 (161) and VarScan2 (162) software with somatic methods. Whole genome indels calling was conducted with MuTect2, varScan2 and Pindel (163) software. Only shared SNVs and

indels identified by their callers were kept. BEDtools (164) and BCFtools (165) were used to filter SNVs/indels and process VCF files. All filtered mutation sites were manually inspected with the IGV software (166). Genome wide mutations were generated using a 100 kb sliding window and plotted with CIRCOS software (167).

Chapter 3: Engineered Cas9 variants can expand the targeting scope of Cas9-based genome editing

3.1 Introduction

A major drawback of CRISPR-Cas9 technology is its dependence on protospacer associated motif (PAM), which in case of SpCas9 is 5'-NGG-3' (10), thus limiting the CRISPR-SpCas9 targeting scope to G-rich regions of the genome. To overcome this limitation, several research groups have engineered Cas9 variants that recognize a more relaxed PAMs or are claimed to be PAM-less. In this chapter, the use of engineered Cas9 variants xCas9-3.7, Cas9-NG, Cas9-NGv1, (i)SpyMacCas9 and SpRY Cas9 was explored for expanding the targeting scope of genome editing and base editing in rice protoplasts and T0 lines. The engineered Cas9 variants were previously demonstrated in either mammalian cell lines or in vitro. The work presented in this chapter has been published (141–144). Members of Yong Zhang lab from the University of electronic science and technology of China performed the rice protoplasts assay as well as generated T0 rice lines for testing of xCas9-3.7, Cas9-NG, Cas9-NGv1 and SpRY-Cas9 engineered variants and analyzed the data. Dr. Desuo Yin, a visiting scholar in Qi lab, helped with the generation of T0 rice lines for testing (i)SpyMacCas9 engineered variants.

3.2 High-fidelity xCas9-3.7 and non-canonical PAM-targeting Cas9-NG

3.2.1 xCas9-3.7 demonstrated high editing efficiency at canonical NGG PAM sites and low efficiency at non-canonical NG PAM sites

To investigate the genome editing potential of xCas9-3.7 (47), a comparison was made with wild type SpCas9 (Cas9-WT) at two nearly complete sets of NGN PAMs (31 sites) in rice protoplasts. Amongst the eight tested target sites associated with the canonical PAMs, xCas9 demonstrated similar mutation frequency at AGG02, GGG02, and TGG02 sites, lower mutation frequency at AGG01, CGG01, and GGG01 sites and higher mutation frequency at CGG01 and TGG01 sites compared to Cas9-WT (Figure 5). xCas9 and Cas9-WT retained mutation frequencies between 5% and 80% at canonical NGG PAM sites, except at AGG01 and CGG02 sites where xCas9 showed low activity. The data suggest that xCas9 predominantly demonstrated high editing activity at the canonical NGG PAM sites, which is consistent with the original report in human cell lines (47). However, xCas9 demonstrated low mutation frequencies at twenty-four tested target sites associated with non-canonical NG PAMs, exceeding 5% mutation frequencies only at CGA02, CGC02, and TGC02 sites (Figure 5). Cas9-WT performed similarly poor at non-canonical NG PAM sites, which is consistent with the seminal Cas9-WT report (10). Stable rice transformations targeting CGA01, CGT02, and TGC02 sites was also pursued. In T0 generation, 2 out of 18 (11.1%) lines were mutated at CGA01 site, 1 out of 15 (6.7%) lines at CGT02 site, and 4 out of 19 (21.1%) lines at TGC02 target site.

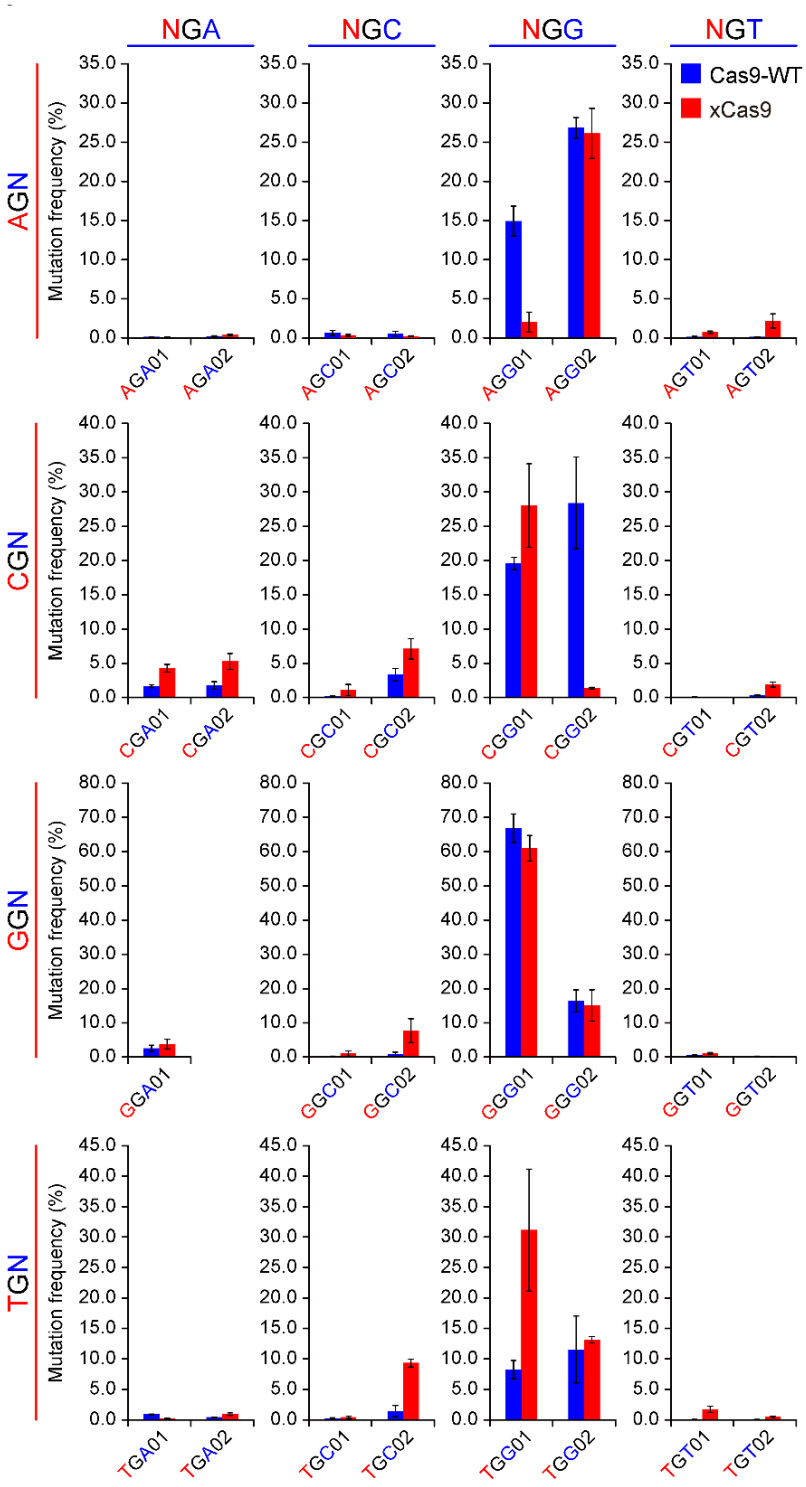


Figure 5: Genome editing with Cas9-WT and xCas9 at 32 NGN PAM sites in rice protoplasts. The error bars represent standard deviations of three biological replicates.

Table 1: Genome editing with xCas9 at CGA01, CGT02, and TGC02 sites in stable T0 rice lines

PAM index	Targeted rice genes	Cas9 variants	Tested T0 lines	Mutated T0 lines (number; ratio)	Biallelic mutation lines (number; ratio)	T-DNA
CGA01	*OsGS3*	xCas9	18	2; 11.1%	0; 0.0%	pLR940
CGT02	*OsDEP1*	xCas9	15	1; 6.7%	0; 0.0%	pLR948
TGC02	*OsDEP1*	xCas9	19	4; 21.1%	0; 0.0%	pLR952
CGC02	*OsDEP1*	xCas9	28	0; 0.0%	0; 0.0%	pLR956

3.2.2 xCas9 showcased superior targeting specificity compared to Cas9-WT

One of the xCas9's advantages is its high fidelity, which was first noted in the seminal report in mammalian cell lines (47) and later validated by protein structure analysis (168). For xCas9 and Cas9-WT targeting specificity comparison, AGG02 PAM site was chosen at which both xCas9 and Cas9-WT demonstrated similar mutation frequency of approximately 25% (Figure 5). Every base in the AGG02 protospacer was systematically mutated from 1st to 19th position relative to the PAM and tested in rice protoplasts (Figure 6). Only the non-mutated protospacers resulted in the same mutation frequency of approximately 25% with both xCas9 and Cas9-WT, which was consistent with the previous observation (Figure 5). Remarkably, xCas9 showed lower activity compared to Cas9-WT with 14 out of 19 mismatched protospacers (Figure 6). For protospacers mismatched at 12th to 19th positions relative to the PAM, xCas9 demonstrated lower activity. While the mismatched protospacers at the 5' end of the target site were tolerated by Cas9-WT, these mismatched drastically reduced xCas9 mutation frequency to about 10% to 20% of the Cas9-WT mutation frequency (Figure 6). The 4th to 10th position of the target site, where xCas9 with mismatched protospacers demonstrated higher mutation frequency compared to Cas9-WT, are a part of the seed sequence known for its increased intolerance to mismatches (10–12). Overall, the

mutation frequencies of xCas9 and Cas9-WT with mismatched protospacers were low in the seed sequence. Core seed sequence spans from the 1st to the 3rd position of the target site relative to the PAM, where xCas9 also displayed much lower mutation frequency compared to Cas9-WT with mismatched protospacers. Altogether, the data showcased xCas9 as more intolerant to single nucleotide mutations and having higher targeting fidelity compared to Cas9-WT at most positions within the protospacer.

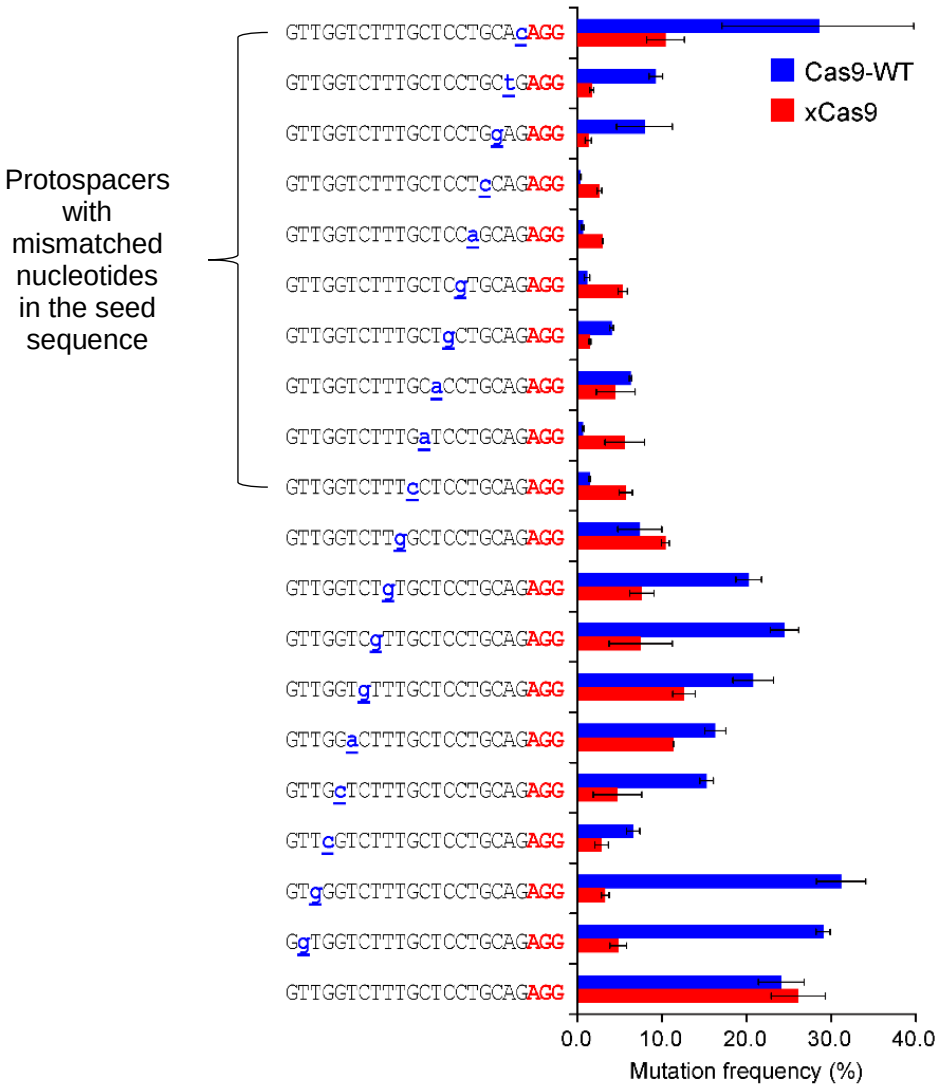


Figure 6: Targeting specificity of Cas9-WT and xCas9 at the AGG02 target site with the mismatched protospacers. The position of mismatched nucleotides within the protospacer is counted upstream relative to the 5' nucleotide of the PAM. The error bars represent standard deviations of three biological replicates.

3.2.3 Cas9-NG variants extended high-efficiency genome editing to non-canonical NG PAM sites

To evaluate the genome editing potential of Cas9-NGv1 and Cas9-NG (48), the two engineered Cas9 variants were compared with wild type SpCas9 (Cas9-WT) at two complete sets of NGN PAMs (32 sites) in rice protoplasts. The results demonstrated both Cas9-NG variants had drastically reduced mutation frequencies at target sites associated with canonical PAMs, which was also observed in human cell lines (48). The mutation frequency of either Cas9-NG variant was approximately a third or less compared to Cas9-WT at the eight canonical NGG PAM sites (Figure 7). At 18 out of the 24 non-canonical NG PAM sites, Cas9-NG variants showed detectable mutation frequencies if 3% was the cutoff threshold (Figure 7). Furthermore, amongst half of these 18 sites, at least one of the Cas9-NG variants demonstrated mutation frequency of 10% or higher. Conversely, Cas9-WT edited only two non-canonical NG PAM sites (CGC02 and TGA01) with low mutation frequencies. A comparison between Cas9-NG and Cas9-NGv1 displayed Cas9-NG having higher activity at AGT01, GGA02, TGT01, and TGT02 target sites and lower activity at AGC01 and GGT01 target sites (Figure 7). In 4 out of 8 NGT PAM sites, at least one Cas-NG variant demonstrated mutation frequency of at least approximately 30%. Altogether, the results feature that Cas9-NG variants lost the strong preference for the third nucleotide of the PAM, though still favoring the third nucleotide to be "T". Cas9-NGv1 demonstrated the highest

mutation frequency in plants at an NGT PAM target site (169), further supporting this observation.

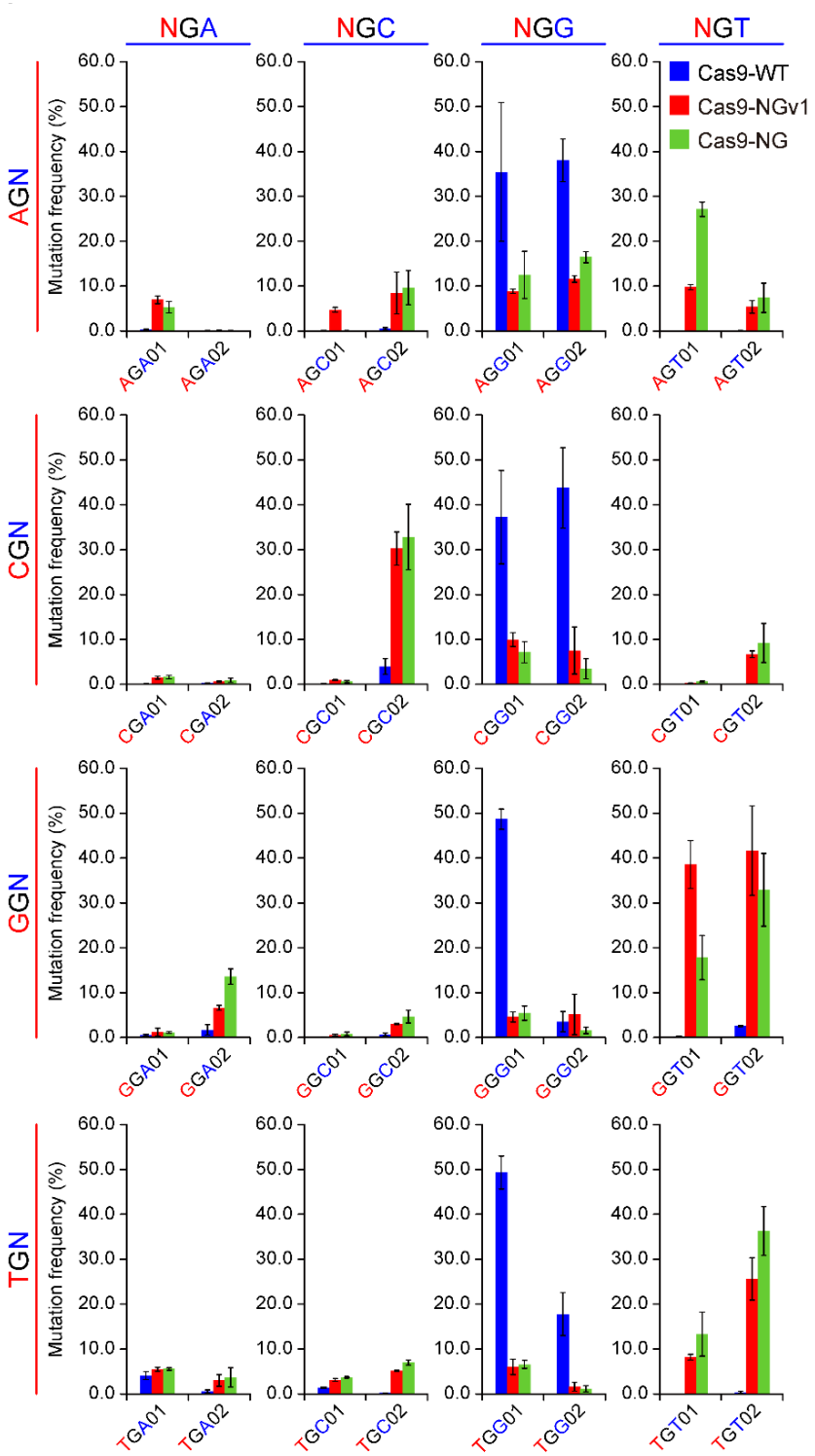


Figure 7: Genome editing with Cas9-WT, Cas9-NGv1, and Cas9-NG at 32 NGN PAM sites in rice protoplasts. The error bars represent standard deviations of three biological replicates.

Being interested in how Cas9-NG variants perform in stable T0 rice lines, Cas9-NG and Cas9-NGv1 were tested at the CGC02 target site, where xCas9 failed to produce mutated stable T0 lines (Table 1). 18 out of 32 (56.3%) T0 lines were mutated in case of Cas9-NG and 6 out of 33 (18.2%) T0 lines were mutated in case of Cas9-NGv1 (Table 2). Comparing the two variants at the three additional non-canonical NG PAM sites (AGG02, CGT02, and GGT02) further solidified Cas9-NG superior performance over Cas9-NGv1 in generating edited T0 lines (Table 2).

Table 2: Genome editing with Cas9-NG and Cas9-NGv1 at CGC02, AGG02, CGT02, and GGT02 sites in stable T0 rice lines

PAM index	Targeted rice genes	Cas9 variants	Tested T0 lines	Mutated T0 lines (number; ratio)	Biallelic mutation lines (number; ratio)	T-DNA
CGC02	*OsDEP1*	Cas9-NG	32	18; 56.3%	5; 15.6%	pZHZ393
		Cas9-NGv1	33	6; 18.2%	2; 6.0%	pZHZ281
AGG02	*OsPDS*	Cas9-NG	15	5; 33.3%	5; 33.3%	pZHZ387
		Cas9-NGv1	25	2; 8.0%	0; 0.0%	pZHZ253
CGT02	*OsPDS*	Cas9-NG	18	4; 22.2%	2; 11.1%	pZHZ397
		Cas9-NGv1	28	1; 3.5%	0; 0.0%	pZHZ285
GGT02	*OsDEP1*	Cas9-NG	52	20; 38.5%	3; 5.8%	pZHZ405
		Cas9-NGv1	22	3; 13.6%	0; 0.0%	pZHZ293

3.2.4 Cas9-NG variants surpassed xCas9 in editing GAT, GAA, and CAA non-canonical PAM target sites

In the seminal report (47), xCas9 mediated targeting of sites in mammalian cell lines associated with non-canonical PAMs like GAT, GAA, and CAA was pursued.

However, application of xCas9 in rice for targeting these PAMs mostly failed (170). Considering the improved performance of the Cas9-NG variants at editing non-canonical NG PAM sites, Cas9-NG and Cas9-NGv1 were evaluated at targeting additional non-canonical PAMs. Specifically, targeted were two GAA PAM sites, two GAT PAM sites and a CAA PAM site with Cas9-WT, Cas9-NGv1, Cas9-NG, and xCas9 in rice protoplasts (Figure 8). Amongst the five tested target sites, only Cas9-NG variants demonstrated detectable mutation frequencies, and at GAT01 target site, the mutation frequency exceeded 20%. On the other hand, mutation frequencies at other target sites were low at approximately a few percent. In conclusion, the data indicate Cas9-NG is superior than xCas9 and Cas9-WT at targeting sites associated with GAT, GAA, and CAA PAMs, though with variable efficiency.

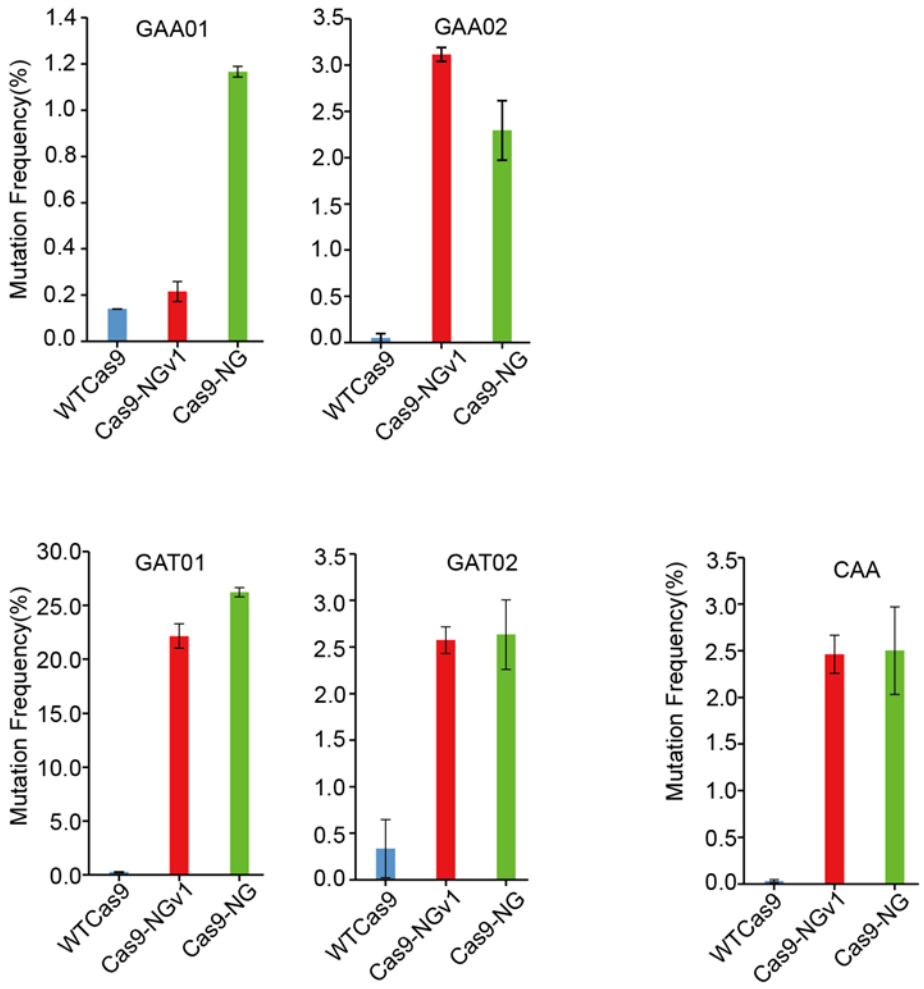


Figure 8: Genome editing with Cas9-WT, Cas9-NGv1, Cas9-NG, and xCas9 GAA, GAT, and CAA PAM sites in rice protoplasts. The error bars represent standard deviations of three biological replicates.

3.2.5 Cas9-NG derived C to T base editors expanded the base editing scope

Since the data indicate that Cas9-NG demonstrated superior mutation frequencies compared to Cas9-NGv1 (Figure 7, 8, Table 2), Cas9n-NG, xCas9n, and Cas9n-WT based C to T base editors with PmCDA1 cytidine deaminase were prepared. Targeting CGC02 target site using rice protoplast assay revealed Cas9n-WT based CBE

performed rather poorly, whereas xCas9n based CBE demonstrated C to T transition efficiency of approximately 2%, and Cas9n-NG based CBE outperforming both aforementioned CBEs with approximately 20 % C to T editing efficiency (Figure 9A). Analysis of C to T transition rate relative to the position of cytosine within the target site revealed that Cas9n-NG-PmCDA1 demonstrated a large activity window spanning over ten nucleotides and activity window was focused towards the 5' end of the target site (Figure 9B). Amongst the three GC sites within the activity window, only C4 was converted efficiently (Figure 9B). Furthermore, targeted were also additional four target sites associated with CGG, AGC, AGT, and TGT PAMs with Cas9n-NG-PmCDA1. This CBE performed rather poorly at CGG canonical PAM (Figure 9C), which was expected taking into consideration the previous genome editing results of Cas9-NG (Figure 7). However, C to T base editing activity increased at the three non-canonical NG PAM sites up to 6 % (Figure 9C).

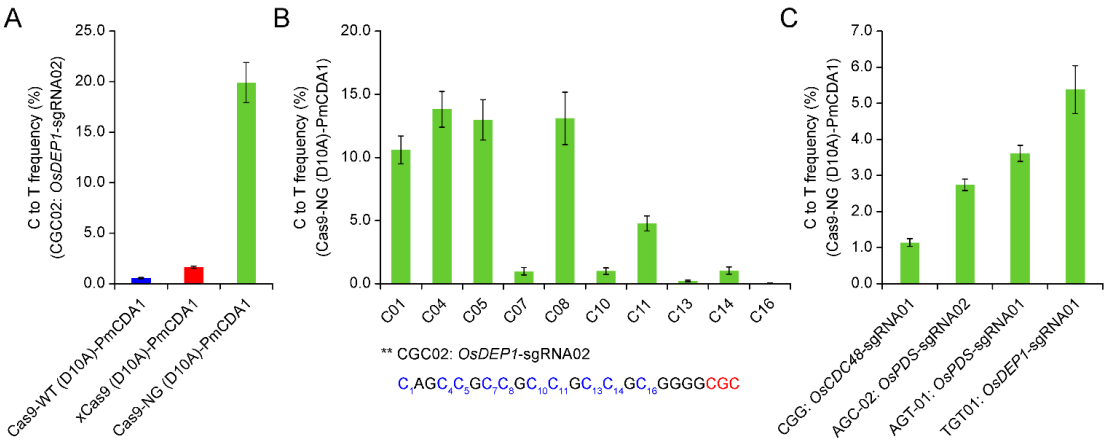


Figure 9: Expanding the base editing technology with Cas9-NG derived C to T editor. **A)** C to T base editing at CGC02 target site with Cas9-WT, xCas9, and Cas9-NG derived CBEs based on PmCDA1 cytidine deaminase. **B)** C to T base transition frequencies within the CGC02 target site achieved by Cas9n-NG-PmCDA1. The entire target sequence is shown below the bar chart with PAM in red and cytosines in blue. **C)** C to T base editing frequencies at four target sites

(CGG, AGC-02, AGT-01, and TGT01) achieved with Cas9n-NG-PmCDA1. **A-C)** All the data were obtained using rice protoplast assay. The error bars represent standard deviations of three biological replicates.

To evaluate how well Cas9n-NG-PmCDA1 performs C to T base editing in stable T0 rice lines, CGC02 and TGT01 sites were targeted. The vast majority of mutation events were base editing events detected in 9 out of 20 lines at the CGC02 target site and in 7 out of 23 lines at the TGT01 target site (Table 3). Base conversion profiles varied amongst the independent T0 lines suggesting that a diverse collection of base edited plants can be cultivated using Cas9n-NG based CBE with PmCDA1 (Figure 10A, B).

Table 3: C to T base editing with Cas9n-NG-PmCDA1 at CGC02 and TGT01sites in stable T0 rice lines

PAM index	Targeted rice sites	Tested T0 lines	Base editing lines (number; ratio)	InDel lines (number; ratio)	T-DNA
CGC02	*OsDEP1*-sgRNA02	20	9; 45.0%	3; 15.0%	pZHZ365
TGT01	*OsDEP1*-sgRNA01	23	7; 30.4%	1; 4.3%	pZHZ366

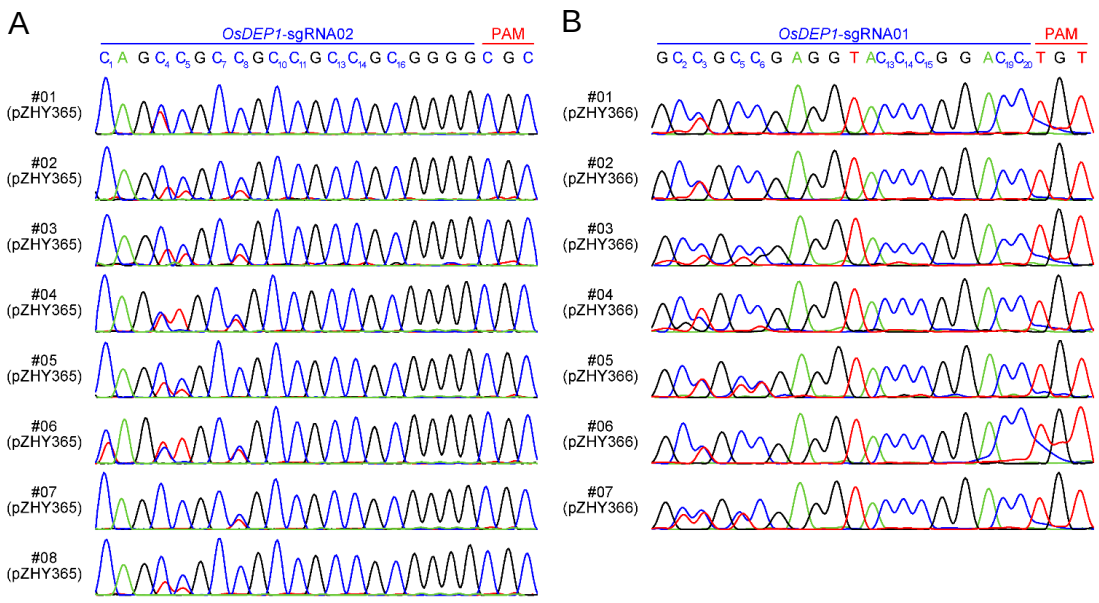


Figure 10: Sanger sequencing profiles of T0 rice lines with Cas9n-NG-PmCDA1 at CGC02 **(A)** and TGT01 **(B)** sites.

3.3 Hybrid (i)SpyMacCas9 targeting A-rich PAM target sites

3.3.1 iSpyMacCas9 demonstrated high editing efficiency at NAAR PAM target sites

Four hybrid Cas9 variants were initially assessed at four independent target sites associated with NAA PAMs using rice protoplast assay. These four hybrid Cas9s include plant codon optimized (pco) and maize optimized (z) SpyMacCas9s and their improved versions pco-iSpyMacCas9 and z-iSpyMacCas9. The results showed that the two NAA PAM sites were edited and z-iSpyMacCas9 appeared to be the best performing one, resulting in approximately 30% mutation frequency at SM17 target site (Figure 11A). After narrowing in on z-iSpyMacCas9, this Cas9 variant was further assessed for its PAM requirement by targeting two complete sets of NAAN PAMs (32

sites) in the rice genome using rice protoplast assay. The data suggested z-iSpyMacCas9 could edit the vast majority of target sites associated with NAAA and NAAG PAMs and demonstrated low mutation frequencies at sites associated with NAAT and especially NAAC PAMs (Figure 11B). The data, along with reports in human cell lines (51) and rabbit pronuclear-stage embryos (171), suggest that z-iSpyMacCas9 activity is greatly affected by the fourth nucleotide within the PAMs. Based on the results, z-iSpyMacCas9 prefers NAAR (R=A, G) PAMs in rice.

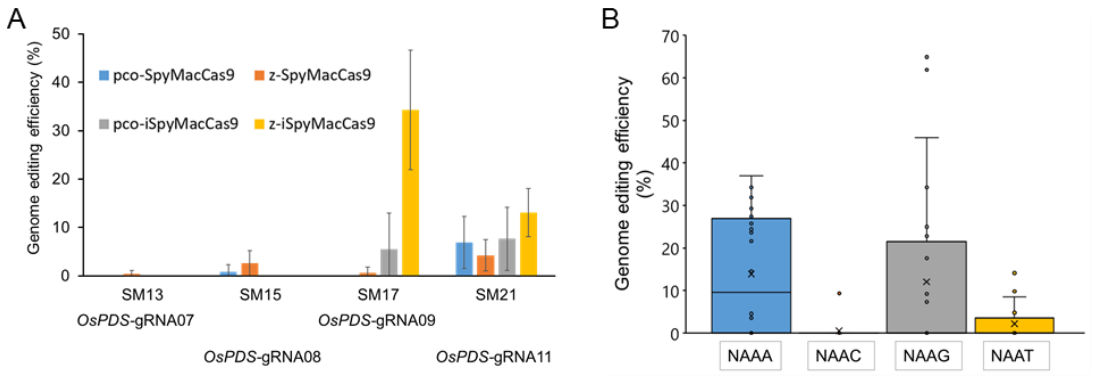


Figure 11: Assessing genome editing efficiency of hybrid Cas9s in rice protoplasts. **A)** Genome editing efficiencies at SM13, SM15, SM17, and SM21 target sites by four SpyMacCas9s. **B)** Genome editing efficiencies at target sites associated with NAAA, NAAC, NAAG, and NAAT PAMs by z-iSpyMacCas9. **A, B)** The error bars represent standard deviations of three biological replicates.

Furthermore, z-SpyMacCas9 and z-iSpyMacCas9 variants' performance was investigated in stable T0 rice lines. At SM21 site associated with NAAG PAM, 20% of T0 lines were mutated (1 out of 5) in case of z-SpyMacCas9 and 57.1% (4 out of 7) in case of z-iSpyMacCas9 (Table 4). Only z-iSpyMacCas9 was able to generate biallelic mutation lines (Table 4). The results indicating z-iSpyMacCas9's superior performance were consistent with previous observation in rice protoplasts (Figure

11A). Focusing on z-iSpyMacCas9, mutation frequencies in T0 lines at SM6 and SM17 target sites associated with NAAG and NAAA, respectively, were evaluated. 8 out of 9 (88.9%) edited T0 lines at SM6 target site were observed and 6 out of 16 (37.5%) edited T0 lines at SM17 target site (Table 4). Taken together, the data demonstrated that z-iSpyMacCas9 can generate mutations at target sites associated with NAAR PAMs in stable T0 rice lines.

Table 4: Genome editing with z-(i)SpyMacCas9 variants at SM6, SM17, and SM21 sites associated with NAAR PAMs in stable T0 rice lines

Constructs	PAM index	Targeted rice sites	Cas9 variants	Tested T0 lines	Mutated T0 lines (number; ratio)	Biallelic mutation lines (number; ratio)
pLR1820	GAAG	OsPDS-sgRNA11-SM21	z-SpyMacCas9	5	1; 20%	0; 0.0%
pLR1829	GAAA	OsPDS-sgRNA09-SM17	z-iSpyMacCas9	16	6; 37.5%	1; 6.3%
pLR1830	GAAG	OsPDS-sgRNA11-SM21	z-iSpyMacCas9	7	4; 57.1%	1;14.2%
pLR2223	AAAG	OsROC5-sgRNA3-SM6	z-iSpyMacCas9	9	8; 88.9%	8; 100%

3.3.2 z-iSpyMacCas9 derived C to T base editors expanded the base editing scope to A-rich genomic regions

To evaluate the C to T base editing potential of z-iSpyMacCas9 derived CBEs with hAID, PmCDA1 and A3A/Y130F cytidine deaminases, their activities were assessed at four target sites associated with NAAR PAMs in rice protoplasts. At SM6 target site, the PmCDA1-based z-iSpyMacCas9 CBE outperformed hAID-based z-iSpyMacCas9 CBE, while at the remaining target sites both CBEs demonstrated similar C to T editing

activity (Figure 12). At all four target sites, hA3A/Y130F-based z-iSpyMacCas9 CBE performed at least slightly better than the other two CBEs (Figure 12).

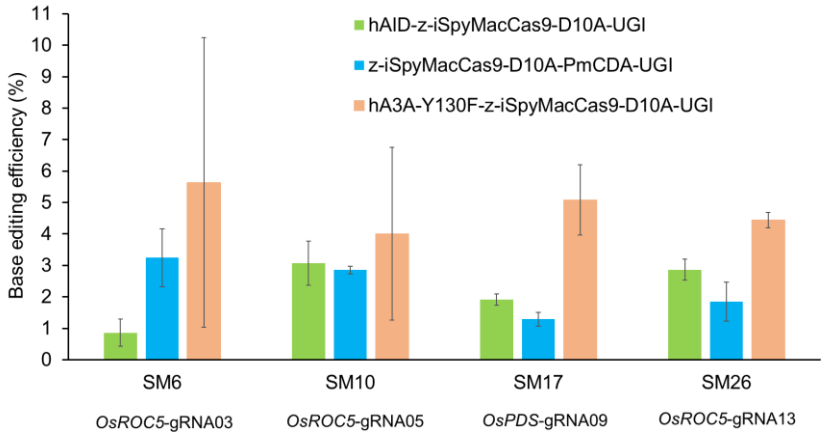


Figure 12: C to T base editing at SM6, SM10, SM17, and SM26 target sites with hAID z-iSpyMacCas9 CBE, PmCDA1 z-iSpyMacCas9 CBE and hA3A/Y130F z-iSpyMacCas9 CBE in rice protoplasts. The error bars represent standard deviations of three biological replicates.

The comparison of PmCDA1 and hA3A/Y130F based z-iSpyMacCas9 CBEs in stable T0 rice lines at SM10 target site showed hA3A/Y130F-based z-iSpyMacCas9 CBE generated higher editing frequency with 11 out of 16 (68.8%) of edited T0 lines, compared to 3 out of 16 (18.8%) edited T0 lines with PmCDA1-based z-iSpyMacCas9 CBE (Table 5). At SM26 site, PmCDA1-based z-iSpyMacCas9 CBE generated 7 out of 18 (38.9%) edited T0 lines (Table 5).

Table 5: Base editing with z-iSpyMacCas9 CBE with PmCDA1 and hA3A/Y130F cytidine deaminases at SM10 and SM26 sites in stable T0 rice lines

Constructs	PAM index	Targeted rice sites	Base editor	Tested T0 lines	Mutated T0 lines (number; ratio)	Base edited T0 lines (number; ratio)	Indels in T0 lines (number; ratio)
pLR2769	TAAA	OsROC5-sgRNA05-SM10	z-iSpyMacCas9-PmCDA1-UGI	16	5; 31.3%	3; 18.8%	2; 12.5%
pLR2773	CAAA	OsROC5-sgRNA13-SM26	z-iSpyMacCas9-PmCDA1-UGI	18	9; 50%	7; 38.9%	5; 27.8%
pLR2770	TAAA	OsROC5-sgRNA05-SM10	hA3A-Y130F-z-iSpyMacCas9-D10A-UGI	16	13; 81.3%	11; 68.8%	5; 31.3%

3.3.3 z-iSpyMacCas9 derived A to G base editor expanded the base editing scope to A-rich genomic regions

After a successful demonstration of C to T base editing with z-iSpyMacCas9 derived CBEs, expanding the A to G base editing scope was pursued to sites associated with NAAR PAMs. Towards this end, the A to G base editing potential of z-iSpyMacCas9 derived ABE with the evolved TadA7.10 adenosine deaminase was evaluated at four target sites in rice protoplasts. The overall base editing efficiencies were rather low, reaching up to 0.5%, suggesting A to G base editing in general is not very efficient in non-dividing cells (Figure 13). Therefore, iSpyMacCas9 derived ABE was also tested in stable T0 rice lines at SM10 target site that showed an average activity in rice protoplasts. iSpyMacCas9 derived ABE demonstrated A to G editing in 3 out of 27 (12.5%) T0 rice lines and none of the edited lines carried byproduct indels (Table 6) which is a stark contrast compared to iSpyMacCas9 derived CBEs (Table 5). Thus, iSpyMacCas9 derived ABE for high purity A to G base editing at NAAR PAM sites in rice was developed.

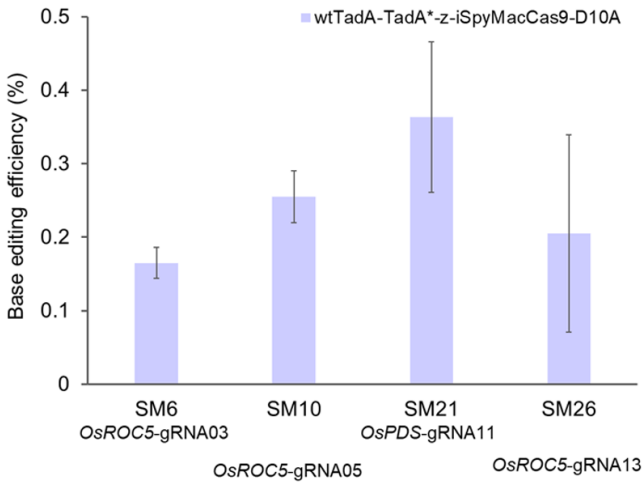


Figure 13: A to G base editing at SM6, SM10, SM21 and SM26 target sites with z-iSpyMacCas9 ABE with TadA7.10 evolved deaminase in rice protoplasts. The error bars represent standard deviations of three biological replicates.

Table 6: Base editing with z-iSpyMacCas9 ABE at SM10 site in stable T0 rice lines

Constructs	PAM index	Targeted rice sites	Base editor	Tested T0 lines	Mutated T0 lines (number; ratio)	Base edited T0 lines (number; ratio)	Indels in T0 lines (number; ratio)
pLR2771	TAAA	OsROC5-sgRNA05-SM10	wtTadA-TadA*-z-iSpyCacCas9	24	3; 12.5%	3; 12.5%	0; 0.0%

3.4 PAM-less genome editing with SpRY Cas9 variant

3.4.1 SpRY Cas9 demonstrated robust genome editing at most of NNN PAM sites

I wanted to investigate whether the near PAM-less nature of SpRY Cas9 demonstrated in human cell lines (49) extends to rice. Towards this end, 59 sites associated with NNN PAMs in rice protoplasts were targeted. The data demonstrated SpRY Cas9 managed to edit most target sites, albeit with different efficiencies (Figure 14). Conversely, Cas9-WT only showed editing at sites associated with canonical NGG PAMs and to a lesser extent NAG PAM sites as previously demonstrated (10). SpRY Cas9 exhibited

approximately four times higher mutation frequencies on average at NAG and NAT PAM sites compared to NAA and NAC PAM sites (Figure 14A). While SpRY Cas9 outperformed Cas9-WT at editing NAG PAM sites (Figure 14A), Cas9-WT was superior only at editing canonical NGG PAM sites (Figure 14B). At NGN PAM sites, SpRY Cas9 demonstrated between 5 and 20% median mutation frequency and showcased higher editing at NGG PAM sites compared to NGA, NGC, and NGT PAM sites (Figure 14B). At NYN PAM sites (Y = C or T), SpRY Cas9 generated approximately 5 to 10% median editing efficiency (Figure 14C, D). Compared to xCas9 (Figure 4), Cas9-NGv1 (Figure 7), and Cas9-NG (Figure 7), SpRY Cas9 could edit all NNN PAM groups and showed higher mutation frequencies at NRN PAM sites (R = A, G) than NYN PAM sites (Figure 14), which is consistent with the results obtained in the human cell lines (49). Furthermore, a strong bias at the third nucleotide position of the PAM was not noticed, which was evident in case of Cas9-NG (Figure 7). Compared to xCas9, Cas9-NGv1, Cas9-NG, z-iSpyMacCas9, and especially SpRY Cas9 expanded the genome editing scope to sites associated with non-canonical NGG PAM sites.

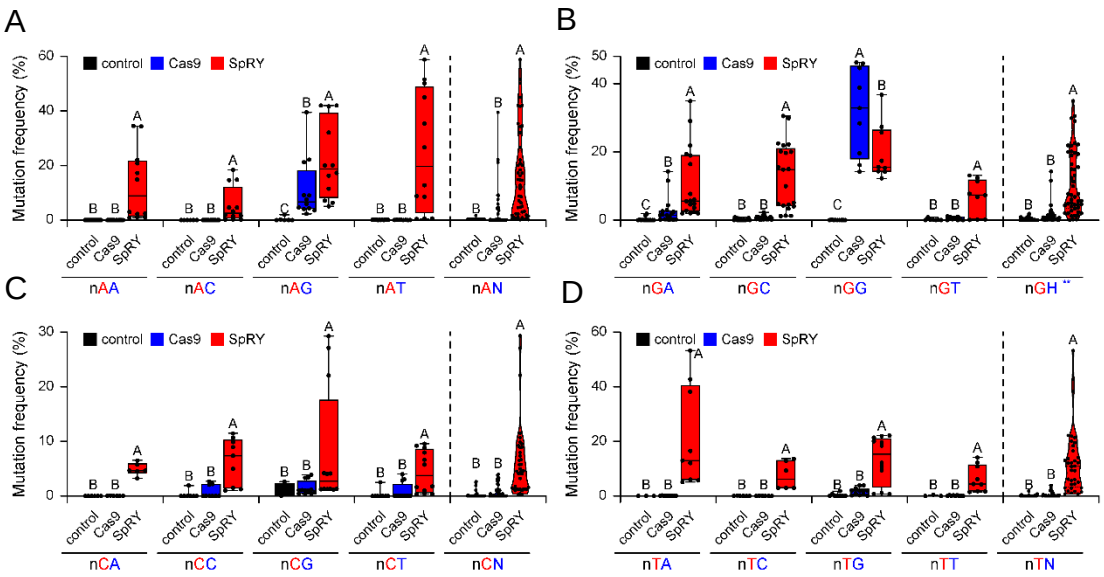


Figure 14: Genome editing with Cas9-WT (Cas9) and SpRY Cas9 (SpRY) at **A)** 16 NAN PAM sites, **B)** 19 NGN PAM sites, **C)** 12 NCN PAM sites, and **D)** 12 NTN PAM sites in rice protoplasts. **A-D)** Left panel (four box and whiskers plots left of the dashed vertical lines) depicts results of editing at PAM subgroups and right panel editing at the PAM group. Each dot represents a biological replicate and each target site was tested with three biological replicates. Different capital letters indicate significant difference ($P<0.05$; one way variance analysis, Bonferroni post hoc test). Samples with the same capital letters did not demonstrate significant difference. The median, interquartile range (IQR) and 1.5x IQR are shown. The maxima, center and minima of each box refer to the upper quartile (Q3), median (Q2) and lower quartile (Q1). The maxima and minima of whiskers refer to Q3+1.5 IQR and Q1-1.5 IQR, respectively.

Furthermore, the characteristics of SpRY Cas9 editing were examined by comparing the non-homologous end joining (NHEJ) editing profiles of SpRY Cas9 and Cas9-WT at NAG PAM sites and canonical NGG PAM sites obtained in rice protoplasts. The analysis of deletion positions within the target sites showed a large overlap between SpRY Cas9 and Cas9-WT (Figure 15A); however, SpRY Cas9 introduced a higher frequency of larger deletions (Figure 15B). For example, SpRY Cas9 introduced a higher frequency of 5 bp deletions at both target sites in comparison to Cas9-WT (Figure 15B) suggesting SpRY Cas9 could be more suitable for knocking out

microRNA genes (172) or engineering quantitative trait variation through promoter editing (173).

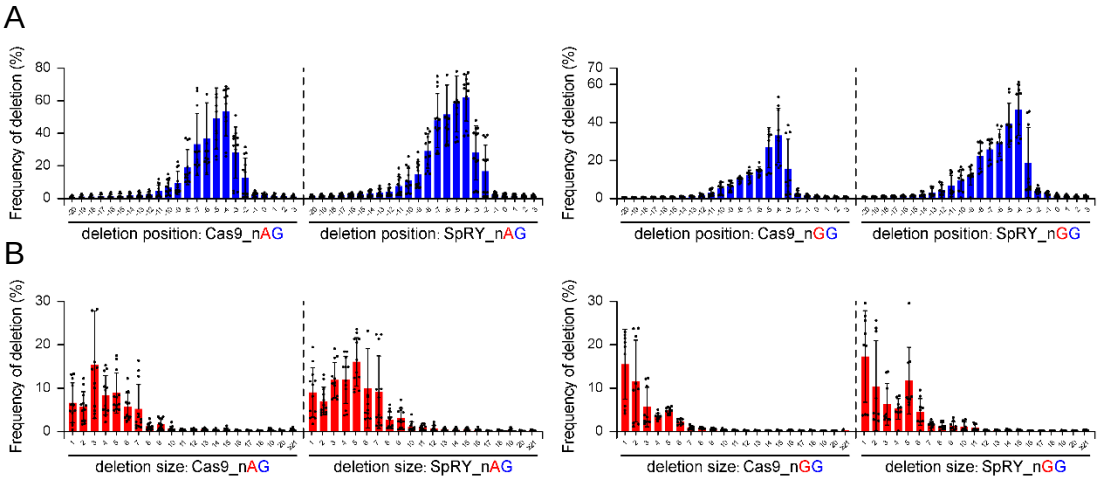


Figure 15: Comparison of **A)** deletion positions and **B)** deletion sizes introduced by Cas9-WT (Cas9) and SpRY Cas9 (SpRY) at three NGG and four NAG PAM sites in rice protoplasts. **A, B)** Each dot represents a biological replicate and each target site was tested with three biological replicates. Data are presented as mean values ± standard deviation. Twelve target sites were tested for NAG PAM group and nine target sites for canonical NGG PAM group.

To assess how well SpRY Cas9 performs genome editing in T0 rice lines, Cas9-WT and SpRY Cas9 were compared at two target sites. SpRY Cas9 generated 10 out of 16 (62.5%) edited T0 lines at OsPDS-AGG-02 target site with canonical NGG PAM compared to Cas9-WT with 15 out of 19 (79%) edited T0 lines (Table 7). At an NAG PAM site (OsPDS-CAG-01), SpRY Cas9 generated 4 out of 17 (23.5%) mutated T0 rice lines compared to Cas9-WT with 3 out of 22 (13.6%) mutated T0 rice lines. At both target sites, biallelic T0 lines were also generated. Additionally, SpRY Cas9 at 10 more target sites with relaxed PAM was also tested with editing efficiencies reaching between 5 and 100% of screened T0 lines and several lines showed high frequency of biallelic editing (Table 7). The data suggest SpRY Cas9 is robust for targeted

mutagenesis at relaxed PAM sites in rice T0 lines, which is consistent with the results obtained in protoplasts (Figure 14).

Table 7: Genome editing with SpRYCas9 at twelve target sites in stable T0 rice lines

Target gene	Reagents	sgRNA+PAM	Tested T0 lines	Mutated T0 lines (number; ratio)	Biallelic T0 lines (number; ratio)	Self cleavage T0 lines (number; ratio)
OsPDS	Cas9_*OsPDS*-AGG-02	GTTGGTCTTTGCTCCTGCAGAGG	19	15; 79.0%	15; 79.0%	0; 0.0%
OsPDS	SpRY_*OsPDS*-AGG-02	GTTGGTCTTTGCTCCTGCAGAGG	16	10; 62.5%	7; 43.8%	4; 25.0%
OsPDS	Cas9_*OsPDS*-CAG-01	TTATGTTCATAGATGACAGGCAG	22	3; 13.6%	1; 4.6%	0; 0.0%
OsPDS	SpRY_*OsPDS*-CAG-01	TTATGTTCATAGATGACAGGCAG	17	4; 23.5%	2; 11.8%	5; 29.4%
OsPDS	SpRY_*OsPDS*-TAA-03	ACCTCCACTAGAAAACACAATAA	16	15; 93.8%	13; 81.3%	12; 75.0%
OsPDS	SpRY_*OsPDS*-GAA-01	GGCATTTCTACCTTATCGATGAA	15	2; 13.3%	0; 0.0%	3; 20.0%
OsDEP1	SpRY_*OsDEP1*-GAT-02	GACCTGTGCGGCCGGCGGCGGAT	20	20; 100.0%	20; 100.0%	9; 45.0%
OsPDS	SpRY_*OsPDS*-CAC-01	CGCAAGTAGCAGCATCCAAGCAC	19	12; 63.2%	8; 42.1%	9; 47.4%
OsPDS	SpRY_*OsPDS*-TGC-02	ACATGCTGACTACTTTTCAGTGC	23	22; 95.7%	13; 56.5%	19; 82.6%
OsDEP1	SpRY_*OsDEP1*-GGT-01	CCGCCGGCCGCACAGGTCCGGGT	22	9; 40.9%	8; 36.4%	0; 0.0%
OsPDS	SpRY_*OsPDS*-TTG-01	GCAATAATAGAGTACAAAGGTTG	15	7; 46.7%	1; 6.7%	12; 80.0%
OsPDS	SpRY_*OsPDS*-CTG-02	GGAAAGTCCTGGCAAACAACCTG	19	1; 5.3%	1; 5.3%	7; 36.8%
OsPDS	SpRY_*OsPDS*-ACT-01	TCTGCAGGTTGTTTGCCAGGACT	23	2; 8.7%	0; 0.0%	12; 52.2%
OsPDS	SpRY_*OsPDS*-TCA-01	ACAACCTGCAGAAAAGCCAATCA	24	3; 12.5%	0; 0.0%	19; 79.2%

3.4.2 PAM-less nature of SpRY Cas9 genome editing made the SpRY Cas9 amenable to vector self-editing

The canonical sgRNA scaffold contains the GTT trinucleotide sequence downstream of the protospacer (10). SpRY Cas9 showed approximately 5% median mutation frequency at targets associated with NTT PAMs in rice protoplasts (Figure 14D), indicating that the GTT PAM is targetable by SpRY Cas9, which is further supported by SpRY Cas9 mediated larger deletion introduction at the target sites (Figure 14B). Analyzing genotyping results of SpRY Cas9 mediated genome editing of twelve target sites in T0 lines (Table 7), evidence of T-DNA self-editing for eleven out of twelve constructs was indeed found, including for the one targeting an AGG canonical PAM site. Conversely, Cas9-WT did not show T-DNA self-editing at the two tested target sites. SpRY Cas9 mediated on-target editing efficiencies at NRN PAM sites (R = A,

G) were in most cases higher than T-DNA self-editing efficiencies at those sites, which is not the case with on-target editing efficiencies at NYN PAM sites (Y = C, T) (Table 7). These data supported the idea that GTT PAM on T-DNA was favored by SpRY Cas9 over many NYN PAM sites, which could have contributed to lower on-target genome editing at NYN PAM sites. Nevertheless, since SpRY-Cas9 generated mutated T0 lines at all twelve tested target sites, SpRY Cas9 is a robust genome editor for editing relaxed PAM sites, despite self-editing.

The PAM-less nature of SpRY Cas9 can increase the probability of off-targeting in two ways. Firstly, by editing target sites with high sequence similarity to the intended target site regardless of PAM sequences. To assess this type of off-targeting, genotyping results of T0 rice lines with editing at sites associated with GAA, TTG, GGT, AGG, and ACT PAMs were analyzed. Top off-target sites, containing one to three mismatches between the protospacer and off-target site, were identified in rice genome by Cas-OFFinder (153). No off-target mutations were detected at these sites (Table 8).

Table 8: Off-target editing at ten putative off-target sites with SpRY Cas9 in stable T0 rice lines

Targeted locus	Putative off target site	Putative off-target locus	Off-target site (sgRNA+PAM)	Mismatched bases	Tested T0	Mutation rate
OsPDS-GAA-01	OsPDS-GAA-01-OT1	Chr1:41499760..41499782	GGCATgTCTACCTgATCGAT **AAT**	2	15	0%
	OsPDS-GAA-01-OT2	Chr7:26606814..26606836	GGCATTTCcACCTTgTCGAT **GCC**	2	15	0%
OsPDS-TTG-01	OsPDS-TTG-01-OT1	Chr1:34915892..34915914	GtAATAATAGAGTgCAAAGG **CAT**	2	15	0%
	OsPDS-TTG-01-OT2	Chr10:3511356..3511378	GCAAaAATAGAGTACAAAtG **AGC**	2	15	0%
OsDEP1-GGT-01	OsDEP1-GGT-01-OT1	Chr1:25248134..25248156	CCGCCcGCCGCACAGGTCgc **CGC**	3	22	0%
	OsDEP1-GGT-01-OT2	Chr1:30488346..30488368	gCGCCGGCCGCgCgGGTCCG **TGT**	3	22	0%
OsPDS-AGG-02	OsPDS-AGG-02-OT1	Chr11:3595300..3595322	GTTGGTCTTgcCTCCTGCAG **ATC**	2	16	0%
	OsPDS-AGG-02-OT2	Chr11:8386155..8386177	GTTGGTCTTTcCTCCTaCAG **GTC**	2	16	0%
OsPDS-ACT-01	OsPDS-ACT-01-OT1	Chr2:15634854..15634876	TCTGCAaGTTcTTTGCCAGG **AAA**	2	23	0%
	OsPDS-ACT-01-OT2	Chr5:22181041..22181063	TCTGCAaGTTcTTTGCCAGG **AAA**	2	23	0%

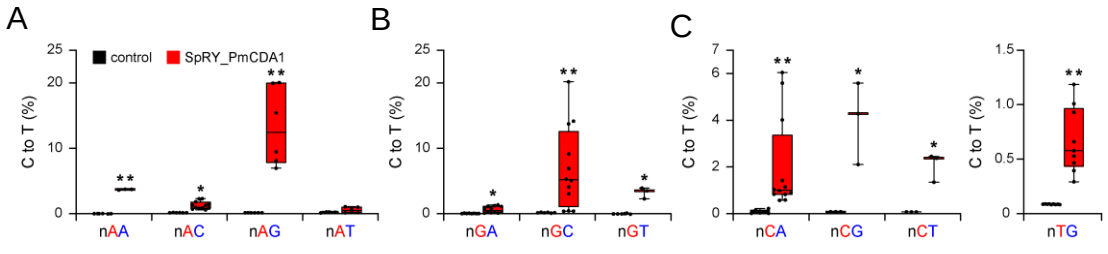
The second way that could contribute to off-targeting due to PAM-less nature of SpRY Cas9 results from the *de novo* generated sgRNAs due to T-DNA self-editing. To assess this, eight edited T0 lines were chosen that exhibited small self-editing deletions resulting from SpRY Cas9 mediated genome editing at four target sites (SpRY_OsPDS-TTG-01, SpRY_OsDEP1-GAT-02, SpRY_OsPDS-TAA-03, and SpRY_OsPDS-GAA-01) (Table 7). These lines carried newly generated sgRNAs with altered protospacers due to T-DNA self-editing. Top off-target sites preferred by these new sgRNAs with altered protospacers were assessed and one T0 line (# 3139-3-1) was found containing an off-target mutation caused by a *de novo* generated sgRNA with 17 nucleotides perfect match to this off-target site (Table 9). No mutations were found at the putative off-target sites in the remaining seven edited T0 lines. Thus, despite self-editing, high off-targeting effects associated with SpRY Cas9 mediated genome editing were not found.

Table 9: Off-target editing at eight putative off-target sites in OsDEP1-GAT-02 mutated T0 rice line with SpRY Cas9

Sample ID	*de novo* generated sgRNA	Nucleotide number (bp)	Backbone PAM	Putative off-target locus	Off-target sites for new sgRNA	Mutation
#3139-3-1	GACCTGTGCGGCCGGCG	17	GTT	Chr1:25248139..25248159	GACCTGTGCGGCgGGCG **GCG**	No
				Chr1:31733946..31733966	cACCTGTGCGGCCGGCt **GTG**	No
				Chr11:10842666..10842686	GgaCTGTGCGGCCGGCG **TCG**	No
				Chr3:6091713..6091733	GttCTGTGCGGCCGGCG **GCG**	Yes
				Chr10:17622414..17622434	GcgCTGTGCGGCCGGCG **GTG**	No
				Chr10:19125988..19126008	GAgCTGTGCGGCgGGCG **AGG**	No
				Chr3:32397023..32397043	GACCTGcGCGGCCGcCG **CGG**	No
				Chr2:12109106..12109126	GAaCTGTGCcGCCGGCG **GGG**	No

3.4.3 PAM-less nature of SpRY Cas9 conferred PAM-less C to T base editing

To further utilize the PAM-less nature of SpRY Cas9, C to T base editing in rice protoplasts was investigated. Results of testing SpRY Cas9n-PmCDA1 at 26 target sites demonstrated PAM-less base editing as it resulted in C to T base editing at many NAN (Figure 16A), NGN (Figure 16B), NCN (Figure 16C), and NTN (Figure 16D) PAM sites with varying efficiencies. Consistent with the genome editing data (Figure 14), SpRY Cas9n-PmCDA1 demonstrated higher editing efficiencies at NRN PAM sites compared to NTN PAM sites (Figures 16A-D). Analysis of base editing data from 26 target sites revealed a base editing window was spanning from first to eighth nucleotide in the protospacer from the 5' end (Figure 16E), consistent with Cas9n-NG-PmCDA1 (Figure 9B) and other studies that used PmCDA1-based CBEs (75,93).



D

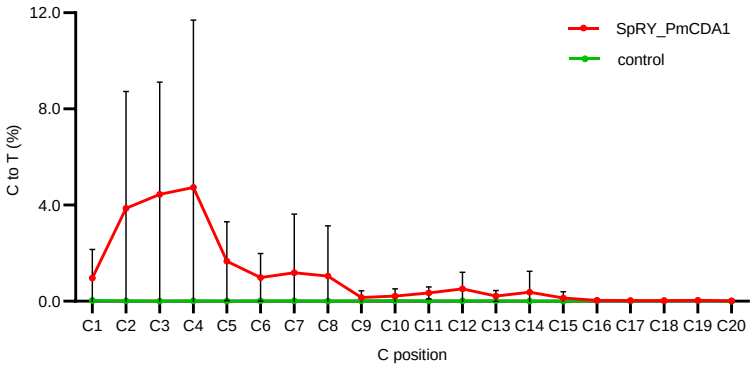


Figure 16: C to T base editing with SpRY Cas9n-PmCDA1 CBE at 26 target sites associated with **A)** NAN PAMs, **B)** NGN PAMs, **C)** NCN PAMs, and **D)** NTN PAMs in rice protoplasts. **A-D)** Each dot represents a biological replicate and each target site was tested with three biological replicates. Asterisks indicate significant difference assessed by independent-sample two-tailed t-tests (* $P<0.05$, ** $P<0.01$) The median, interquartile range (IQR) and 1.5x IQR are shown. The maxima, center and minima of each box refer to the upper quartile (Q3), median (Q2) and lower quartile (Q1). The maxima and minima of whiskers refer to Q3+1.5 IQR and Q1-1.5 IQR, respectively. **E)** Base editing window of PmCDA1-SpRY Cas9n, calculated from editing 26 target sites in rice protoplasts. Data are presented as mean values ± standard deviation of three biological replicates.

Additionally, SpRY Cas9n-PmCDA1 CBE was tested in T0 rice lines. At three tested target sites associated with relaxed PAMs, between 10% and 73% of the lines were base edited amongst approximately 20 screened T0 lines per target site (Table 10). At the OsDEP1-GGT-01 site, 3 out of 19 (15.8%) T0 lines contained byproduct editing and 2 out of 19 (10.5%) T0 lines displayed T-DNA self-editing. At the OsALS-sgRNA22 site, 7 out of 18 (38.9%) T0 lines contained byproduct editing and 8 out of 18 (11.1%) T0 lines displayed T-DNA self-editing (Table 10). At all three target sites, on target base editing was favored over T-DNA self-editing. Based on the data from rice protoplasts and stable T0 lines, it was concluded that SpRY Cas9 derived CBE conferred precise C to T base editing in a PAM-less fashion.

Table 10: PmCDA1-SpRY Cas9n mediated C to T base editing at three target sites in stable T0 rice lines

Reagents	sgRNA+PAM	Tested T0 lines	Mutated T0 lines (number; ratio)	C to T T0 lines (number; ratio)	Other SNP T0 lines (number; ratio)	Self cleavage T0 lines (number; ratio)
SpRY-PmCDA1_*OsDEP1*-GAT-02	GACCTGTGCGGCCGGCGGCGGAT	20	2; 10.0%	2; 10.0%	0; 0.0%	0; 0.0%
SpRY-PmCDA1_*OsDEP1*-GGT-01	GCCGCCGGCCGCACAGGTCCGGGT	19	8; 42.1%	5; 26.3%	3; 15.8%	2; 10.5%
SpRY-PmCDA1_*OsALS*-sgRNA22	CCCCACTTGGGATCATAGGCAGC	18	13; 72.2%	6; 33.3%	7; 38.9%	2; 11.1%

3.4.4 SpRY Cas9 ABE conferred robust A to G base editing at relaxed PAM target sites in rice stable lines

Finally yet importantly, the SpRY Cas9 based ABE with an evolved TadA8e adenosine deaminase (SpRY-ABE8e) was tested at several target sites with relaxed and canonical NGG PAMs in rice protoplasts. At TAA-04, CGG-03, and TTG-01 target sites, base editing efficiency was observed at approximately 1% while editing efficiencies at other tested target sites were below 0.5% (Figure 17A, B). Analysis of base editing data from seven tested target sites revealed a base editing window spanning from fourth to eighth nucleotide in the protospacer from the 5' end (Figure 17C), consistent with ABE8e editing data in human cell lines (113). A much lower A to G base editing activities were observed (Figure 13; 17A, B) compared to C to T editing activities (Figure 9A, C; 12; 16A-D; 17A, B) in general in rice protoplasts, which is consistent with published literature (64) and could be due to low activity of A to G base editing pathway in non-replicating protoplasts.

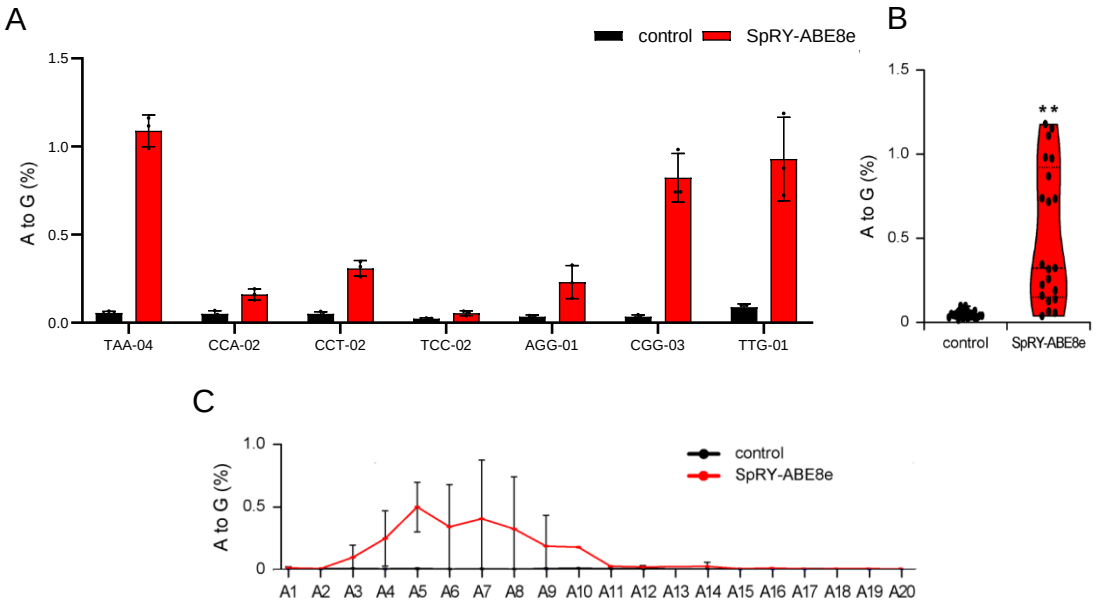


Figure 17: A to G base editing at seven target sites depicted **A)** individually or **B)** collectively with SpRY Cas9n ABE with TadA8e evolved deaminase (SpRY-ABE8e) in rice protoplasts. **C)** A to G base editing window, calculated from editing 7 target sites using data derived from **A)**. **A, C)** Data are presented as mean values ± standard deviation of three biological replicates. **B)** Depicted are three biological replicates of 7 target sites. Two asterisks indicate significant difference assessed by independent-sample two-tailed t-tests ($P<0.01$).

SpRY-ABE8e was also assessed in T0 rice lines. At the two tested target sites associated with relaxed PAMs, SpRY-ABE8e generated 15 out of 19 (79.0%) edited T0 rice lines and 9 out of 20 (45.0%) edited T0 rice lines (Table 11), while at one target site no edited T0 rice line could be obtained. Notably, no byproduct editing was observed amongst the edited lines. It is worth noting that testing base editing at three target sites exhibited high frequency of T-DNA self-editing even in case of the OsPDS-CCA-02 target site, where mutated T0 lines at the target site could not be obtained (Table 11).

Table 11: SpRY-ABE8e mediated A to G base editing at three target sites in stable T0 rice lines

Reagents	sgRNA+PAM	Tested T0 lines	Mutated T0 lines (number; ratio)	A to G T0 lines (number; ratio)	Other SNP T0 lines (number; ratio)	Self cleavage T0 lines (number; ratio)
SpRY-ABE8e_*OsPDS*-TAA-04	AGAAACAGTGAACAACCCACTAA	19	15; 79.0%	15; 79.0%	0; 0.0%	8; 42.1%
SpRY-ABE8e_*OsPDS*-TTG-01	GCAATAATAGAGTACAAAGGTTG	20	9; 45.0%	9; 45.0%	0; 0.0%	14; 70.0%
SpRY-ABE8e_*OsPDS*-CCA-02	GCAAACAACCTGCAGAAAAGCCA	18	0; 0.0%	0; 0.0%	0; 0.0%	5; 27.8%

Such a high frequency of T-DNA self-editing could contribute towards secondary off-targeting due to *de novo* generated sgRNAs. For assessment of this, 13 edited T0 lines were chosen exhibiting self-editing resulting from SpRY-ABE8e mediated base editing at three target sites (OsPDS-TAA-04, OsPDS-TTG-01, and OsPDS-CCA-02) (Table 11). These lines carried newly generated sgRNAs with altered protospacers due to T-DNA self-editing. Top off-target sites preferred by these new sgRNAs with altered protospacers were assessed and two secondary off-target events were found in two independent T0 lines (# 3185-1-1 and #3185-14-1). These lines were carrying off-target mutations caused by the same *de novo* generated sgRNA with only one mismatched nucleotide to the off-target site (Table 12). No secondary off-target mutations were found at the putative off-target sites in the remaining eleven edited T0 lines. Thus, SpRY-ABE8e offers robust A to G base editing in PAM-less fashion in stable rice lines, albeit with frequent self-editing.

Table 12: Off-target editing at putative off-target sites in OsPDS-TTG-01 mutated T0 rice lines with SpRY-ABE8e

Sample ID	*de novo* generated sgRNA	Nucleotide number (bp)	Backbone PAM	Putative off-target locus	Off-target sites for new sgRNA	Mutation
#3185-1-1	GCAATGGTAGAGTACAAAGG	20	GTT	Chr6:27724531..27724554	aCAATGGTAGAaTACAAAGG **GTG**	Yes
				Chr11:18995798..18995821	GCAAatGTAGAGaACAAAGG **GCA**	No
#3185-14-1	GCAATGGTAGAGTACAAAGG	20	GTT	Chr6:27724531..27724554	aCAATGGTAGAaTACAAAGG **GTG**	Yes
				Chr11:18995798..18995821	GCAAatGTAGAGaACAAAGG **GCA**	No

Chapter 4: Improved and novel C to T base editors can confer higher base editing activity, purity and refined editing windows

4.1 Introduction

During the green revolution occurring between 1950s and late 1960s, C to T mutations were randomly introduced into plant genomes using the chemical mutagen ethyl methanesulfonate (EMS). In this chapter, it was explored how targeted C to T transitions could be introduced using base editing, a CRISPR based technology. Building on the seminal report of C to T base editing in mammalian cell lines (35), C to T base editors' architecture was modified by fusing or recruiting additional DNA or base excision repair modifying enzymes to increase base editing purity and activity in rice protoplasts and T0 lines. Additionally, base editing off-target activity was explored by whole genome sequencing (WGS) in rice T0 lines. Furthermore, novel C to T base editors were prepared by adopting cytidine deaminases from various organisms for base editing. Some of the data presented in this chapter have been published (143). Members of Yong Zhang lab from the University of electronic science and technology of China performed the rice protoplasts assays as well as generated T0 rice lines for testing the improved base editors with modified architecture and analyzed the data. Members of Tao Zhang lab from Yangzhou University in China performed the WGS experiment and analyzed the data.

4.2 Improving C to T base editing activity, specificity, and purity by modifying the CBE architecture

4.2.1 A3A/Y130F and PmCDA1 derived CBEs demonstrated the highest base editing

To improve C to T base editing in plants, a comparison of seven CBEs was carried out at four CG rich target sites (OsCGRS55, OsCGRS56, OsCGRS57, and OsCGRS58) (Figure 18A) in rice protoplasts. The seven base editors were in a BE3 like configuration and were based on rAPOBEC1 (35), PmCDA1 (75,93,94), hAID (95), A3A (71,96), as well as engineered A3A/Y130F (96), eA3A (97) and eA3A/Y130F, a newly engineered A3A variant introducing Y130F mutation into eA3A (Figure 18B). The data demonstrated A3A/Y130F-CBE_V01 displayed the highest C to T transition efficiency at approximately 40% on average followed by PmCDA1-CBE_V01, hAID-CBE_V01 and A3A-CBE_V01 (Figure 18C). While rAPOBEC1-CBE_V01 and eA3A-CBE_V01 showed poor editing activity, eA3A/Y130F-CBE_V01 showed no editing. Tested CBEs demonstrated different base editing windows at four target sites (Figure 18D). These results are consistent with reports in human cell lines (96,174). Amongst CBE_V01, A3A/Y130F-CBE_V01 should represent one of the best CBEs for C to T base editing in plants. This notion was supported also when A3A/Y130F-z-iSpyMacCas9 CBE outperformed PmCDA1- and hAID-based z-iSpyMacCas9 CBEs (Figure 12). PmCDA1-CBE_V01 favored editing from first to sixth nucleotide at the 5’ end of the protospacer while A3A/Y130F-CBE_V01 had a broad activity window stretching from the 4^{th} to 15^{th} nucleotide in the 5’ to 3’ direction within the protospacer.

At OsCGRS57 target site, decrease in editing activity at the middle of the protospacer was observed which was a target specific event (Figure 18D).

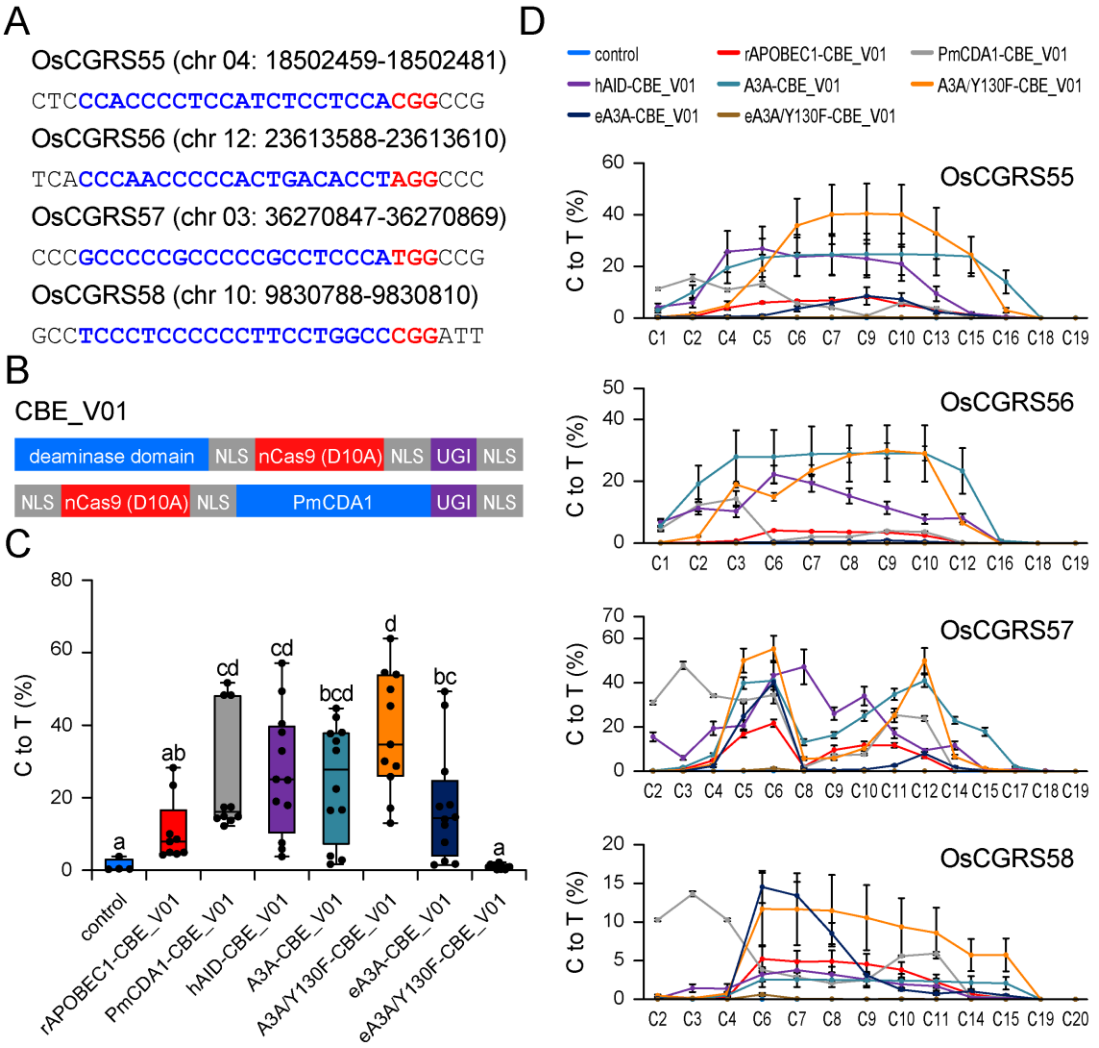


Figure 18: Assessment of seven C to T base editors at four C rich target sites in rice protoplasts. **A)** Four CG rich target sites, at which C to T base editing was pursued. **B)** Base editing was pursued with seven base editors of first version (CBE_V01). Cytidine deaminases were fused to N terminal of Cas9n-WT except PmCDA1 was fused to C terminal end of Cas9n-WT. UGI=uracil glycosylase inhibitor and NLS=nuclear localization signal. **C)** C to T base editing at four target sites with seven CBE_V01s. Each dot represents a biological replicate and each target site was tested with at least two biological replicates. Different lowercase letters indicate significant difference ($P<0.05$; one way ANOVA, Duncan test). Samples with the same capital letters did not show a significant difference. The maxima, center and minima of each box refer to the upper quartile (Q3), median (Q2) and lower quartile (Q1). The maxima and minima of whiskers refer to the maximum and minimum values. **D)** Editing windows of seven CBE_V01

editors at four target sites in rice protoplasts. Each target site was tested with at least two biological replicates. Data are presented as mean value ± standard error of the mean (SEM).

To further improve C to T base editing, CBE_V02 architecture was tested based on flanking the Cas9n-WT with two different cytidine deaminases or fusing in tandem different cytidine deaminases to Cas9n-WT (Figure 19A). A comparison was carried out including four CBE_V02 with three CBE_V01 at editing OsCGRS55 and OSCGRS56 target sites (Figure 19B). On average, CBE_V02 editors demonstrated reduced editing activity comparing to CBE_V01 editors.

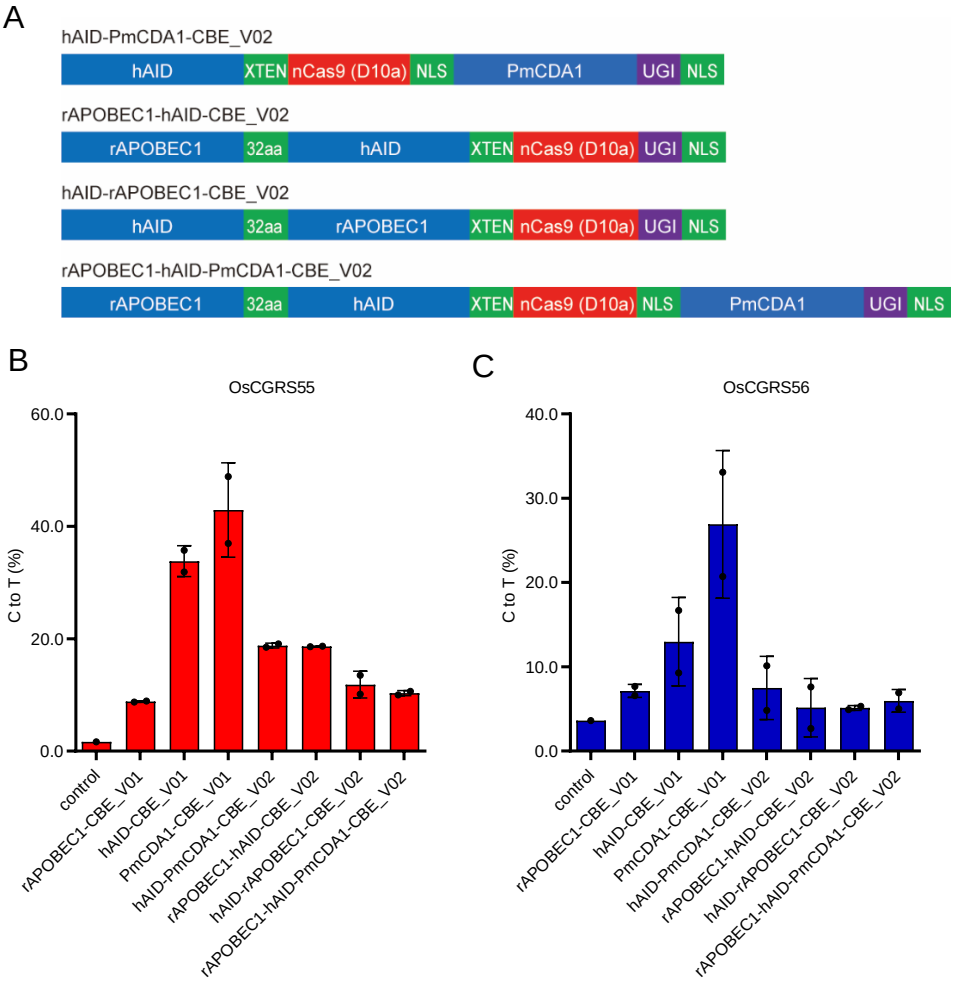


Figure 19: C to T base editing with CBEs_V01 and CBEs_V02 at two C rich target sites in rice protoplasts. **A)** CBE_V02 architecture. Cas9 nickase is fused to the tandem of different cytidine deaminases or with two different deaminases. 32aa = 32 amino acid flexible linker, XTEN = 16-residue flexible linker, UGI=uracil glycosylase inhibitor and NLS=nuclear localization signal. **B)** Base editing with selected CBE_V01 and CBE_V02 at OsCGRS55 and OsCGRS56 target sites in rice protoplasts. Each dot represents a biological replicate. Error bars represent standard deviation of two biological replicates.

To overcome reduced C to T base editing activity observed by CBEs_V02, CBE_V03 editors were developed (Figure 20A). CBEs_V03 used MS2-MCP interaction (56) to recruit four copies of the same cytidine deaminase through an engineered sgRNA scaffold in addition to directly fusing deaminases using flexible linkers as in CBE_V01. Testing CBE_V03s at the same four target sites as in case of CBE_V01 revealed impaired base editing activity (Figure 20B) compared to CBE_V01 editing activity. Several studies have reported on simultaneous C to T and A to G editing when fusing cytidine and adenosine deaminase either to both ends of Cas9n or in tandem to the N terminal end of Cas9n (64–67). Compromised base editing activities were demonstrated in such dual base editor platforms, especially in case of A to G editing, in comparison to single base editors (64,66). These studies, along with presented results (Figures 19B, 20B), demonstrated that C to T base editing efficiency could not be improved by simply increasing the focal concentration of the cytidine deaminases at the target sites, probably due to steric interference amongst deaminases.

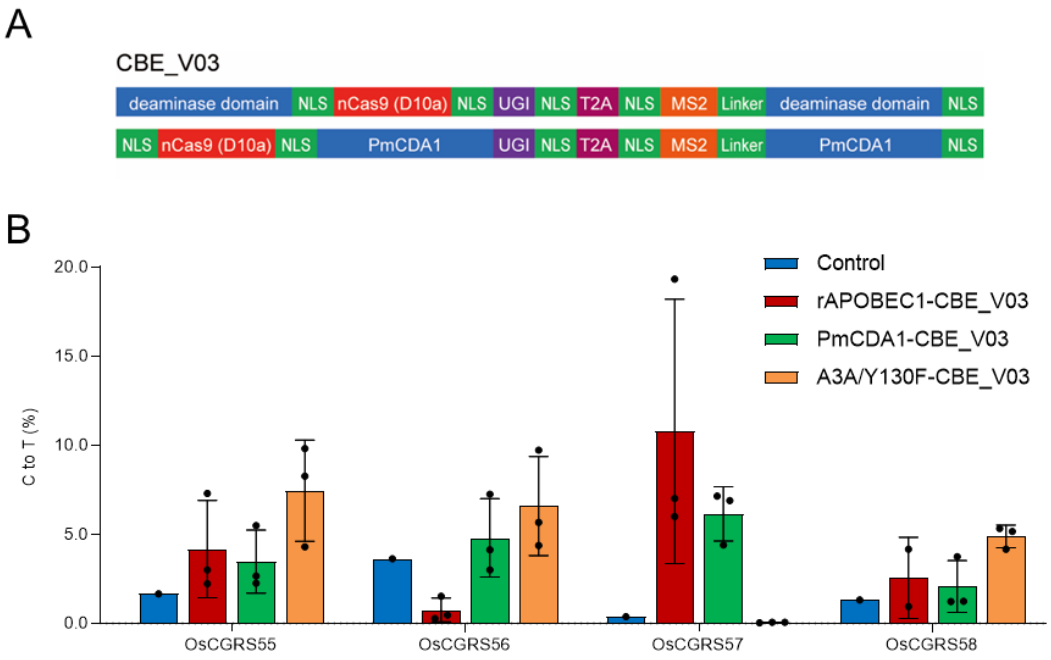


Figure 20: C to T base editing with CBEs_V03 at four C rich target sites in rice protoplasts. **A)** CBE_V03 architecture. Deaminase domain is fused to Cas9n through a flexible linker and four deaminase domains are recruited through MS2-MCP interaction. UGI=uracil glycosylase inhibitor, NLS=nuclear localization signal and T2A= self cleaving peptide. **B)** Base editing with three CBEs_V03 at OsCGRS55, OsCGRS56, OsCGRS57, and OsCGRS58 target sites in rice protoplasts. Each dot represents a biological replicate. Error bars represent standard deviation of three biological replicates.

Since high editing efficiency with reduced indel formation has already been developed by fusing more uracil glycosylase inhibitors (UGI) to Cas9n (175) or co-expressing more UGIs (176), the utilization of MS2-MCP recruitment method was promising to increase the focal concentration of UGIs at the target site (Figure 21A). Four such CBEs_V04 were tested in rice protoplasts and compared to CBEs_V01, CBE_V04 editors did not show reduced editing activity (Figure 21B compared to 18C) and had a similar base editing windows (Figure 21C compared to 18D). The A3A/Y130F-CBE_V4 displayed the highest average activity amongst CBE_V04 editors (Figure 21B) like A3A/Y130F-CBE_V1 amongst CBE_V01 editors (Figure 18C). Comparison

of A3A/Y130F-CBE_V01 and A3A/Y130F-CBE_V04 at ten additional target sites showed A3A/Y130F-CBE_V04 had comparable or higher editing activity than A3A/Y130F-CBE_V01 (Figure 21D). Grouped analysis of these ten target sites showed A3A/Y130F-CBE_V04 had a drastically smaller dynamic range compared to A3A/Y130F-CBE_V01 for the base editing frequencies (Figure 21E), indicating A3A/Y130F-CBE_V04 was a more robust base editor. Furthermore, indel frequencies were analyzed at these ten target sites. At six out of ten tested target sites (OsALS-sgRNA01, OsALS-sgRNA03, OsALS-sgRNA04, OsALS-sgRNA10, OsALS-sgRNA14, and OsALS-sgRNA16), A3A/Y130F-CBE_V04 showed significantly lower indel frequencies than A3A/Y130F-CBE_V01 (Figure 21F). Grouped indel analysis of the ten target sites further showed reduced indel formation when A3A/Y130F-CBE_V04 was used (Figure 21G). In conclusion, the data suggest A3A/Y130F-CBE_V04 is a robust CBE with higher editing purity compared to A3A/Y130F-CBE_V01, presumably due to increased focal concentration of UGIs at the target site.

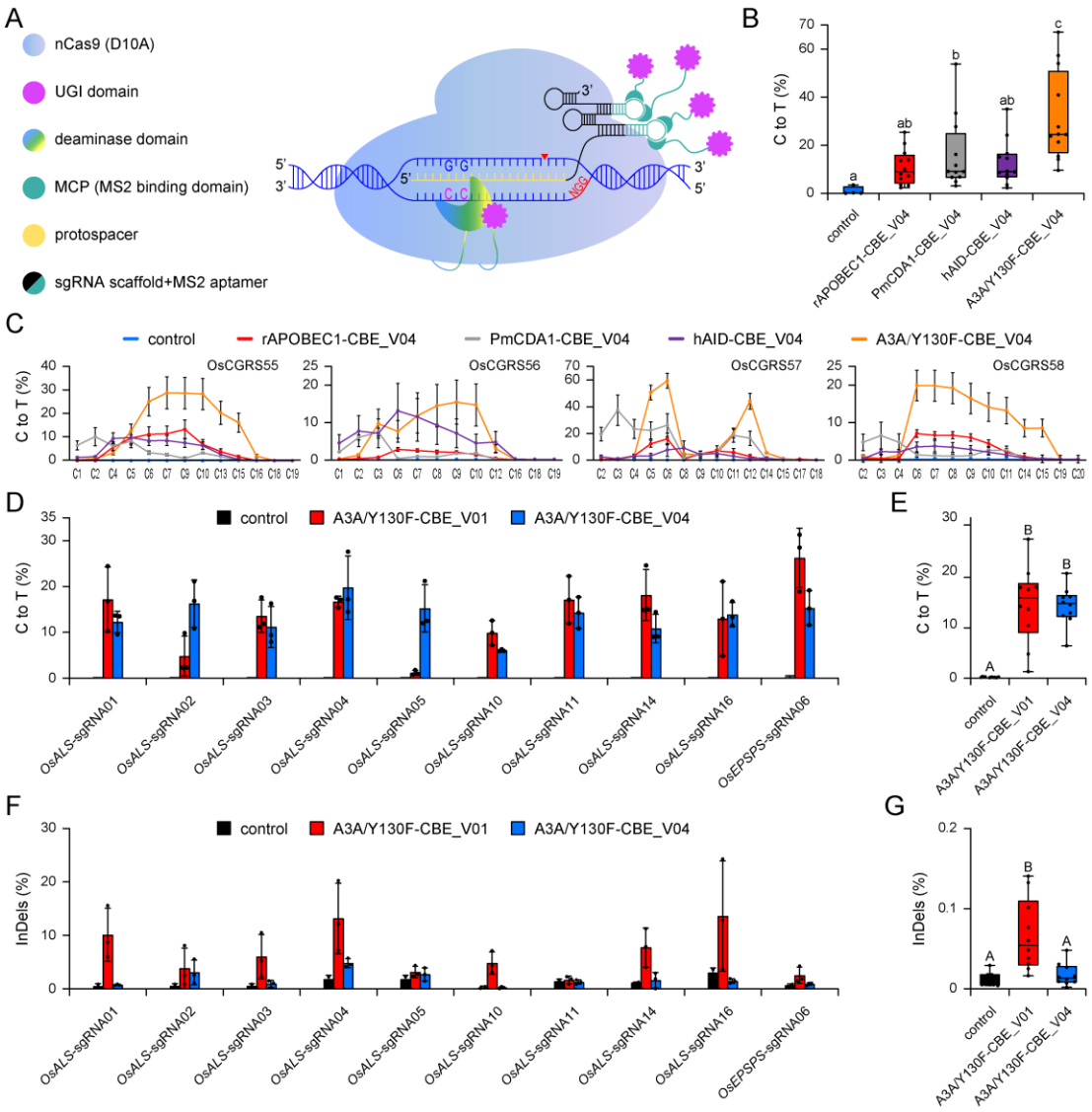


Figure 21: C to T base editing with CBE_V04 editors and A3A/Y130F-CBE_V01 in rice protoplasts. **A)** Schematic representation of the CBE_V04 architecture based on recruiting UGIs through MS2-MCP through modified sgRNA scaffold. **B)** C to T base editing at four target sites with seven CBE_V04s. Each dot represents a biological replicate and each target site was tested with at three biological replicates. Different lowercase letters indicate significant difference ($P<0.05$; one way ANOVA, Duncan test). The maxima, center and minima of each box refer to the upper quartile (Q3), median (Q2) and lower quartile (Q1). The maxima and minima of whiskers refer to the maximum and minimum values. **C)** Editing windows of four CBE_V04 editors at four target sites in rice protoplasts. Each target site was tested with three biological replicates. Data are presented as mean value ± standard error of the mean (SEM). **D, E)** Singular and grouped comparison of A3A/Y130F-CBE_V01 and A3A/Y130F-CBE_V04 at C-to-T editing frequencies at ten sites. **F, G)** Singular and grouped indel generation comparison by A3A/Y130F-CBE_V01 and A3A/Y130F-CBE_V04 C to T base editing at ten target sites. **D, F)** Each dot represents a biological replicate and each target site was tested with three biological replicates. Data are presented as mean value ± standard deviation. **E, G)** Each dot represents a target site. Different capital letters indicate significant differences ($p< 0.05$;

one-way ANOVA, Duncan test). The maxima, center and minima of each box refer to the upper quartile (Q3), median (Q2) and lower quartile (Q1). The maxima and minima of whiskers refer to the maximum and minimum values.

To evaluate whether CBE_V04 editors are superior to CBE_V01 editors, top three performing CBE_V01 editors (PmCDA1-CBE_V01, hAID-CBE_V01, and A3A/Y130F-CBE_V01) were compared with the corresponding CBE_V04 editors with the same cytidine deaminases in stable T0 rice lines. The six CBEs were tested at editing the same four target sites as in rice protoplasts (Figure 18A). For each C to T base editor, between 20 to 33 T0 lines were genotyped for each of the four target sites. Amongst the three CBE_V01 editors, A3A/Y130F-CBE_V01 showed the highest percentage (23.1%) of generated edited T0 rice lines on average between four target sites (Table 13). For all three CBE_V01 editors, indels appeared to be the major byproducts at all four target sites with average percentage of indels in T0 rice lines per CBE_V01 editors reaching over 50%, probably due to high GC content in these target sites. The average generation of pure C to T edited T0 lines by PmCDA1-CBE_V04 and hAID-CBE_V04 was nearly doubled in comparison to the CBE_V01 counterparts at 33.1% and 31.4% versus 18.7% and 16.8%, respectively (Table 13). In case of A3A/Y130F-based editors, the average generation of pure C to T edited T0 lines increased from 23.1 % with the CBE_V01 architecture to 34.1% with the CBE_V04 architecture. Remarkably, for all three cytidine deaminases a twofold reduction in generation of T0 lines with indels with the CBE_V04 architecture was observed compared to the CBE_V01 architecture (Table 13). In conclusion, the data from stable T0 rice lines highlighted CBE_V04 editors as outperforming CBE_V01 editors in C to T editing efficiency and purity, consistent with the observations in rice protoplasts.

Table 13: C to T base editing by V01 and V04 architectures at four target sites (OsCGRS55, OsCGRS56, OsCGRS57, and OsCGRS58) in T0 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; 83.3%
	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	-	23.1%	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%

4.2.2 Off-target assessment by WGS showcased PmCDA1 based CBEs as the most specific base editors

Thorough assessment was performed of genome-wide off-target effects of ten CBE editors, including six CBE_V01 based on rAPOBEC1, PmCDA1, hAID, A3A, eA3A, and A3A/Y130F deaminases, and four CBE_V04 editors based on rAPOBEC1, PmCDA1, hAID and A3A/Y130F deaminases. A total of 59 plants were selected for WGS including 31 edited lines by the ten CBEs, 18 control lines expressing base editors without sgRNAs, four WT plants and six control plants for tissue culture, *Agrobacterium* mediated transformation and Cas9 expression (Figure 22A). The

number of indels spanning from approximately 50 to 100 in all CBE-expressing plants with or without sgRNAs were comparable to those in control plants (Figure 22B), suggesting the indels were part of the somaclonal variation due to tissue culture procedure. Bbetween 100 and 130 single nucleotide variations (SNVs) were found in the three control plant groups, representing the baseline SNVs level caused by somaclonal variations do to tissue culture procedure (Figure 22C). PmCDA1 and rAPOBEC1 based V01 and V04 editors demonstrated slightly increased SNV count of up to approximately 200. Approximately 200 to 400 SNVs were detected in T0 plants in case of A3A/Y130F-based V01 and V04 editors. Higher levels of SNVs at approximately 300 to 1200 were detected in case of hAID based V01 and V04 editors (Figure 22C). The various elevated SNV levels for different CBEs appeared to be sgRNA independent, as they were equally spread amongst the rice genomes.

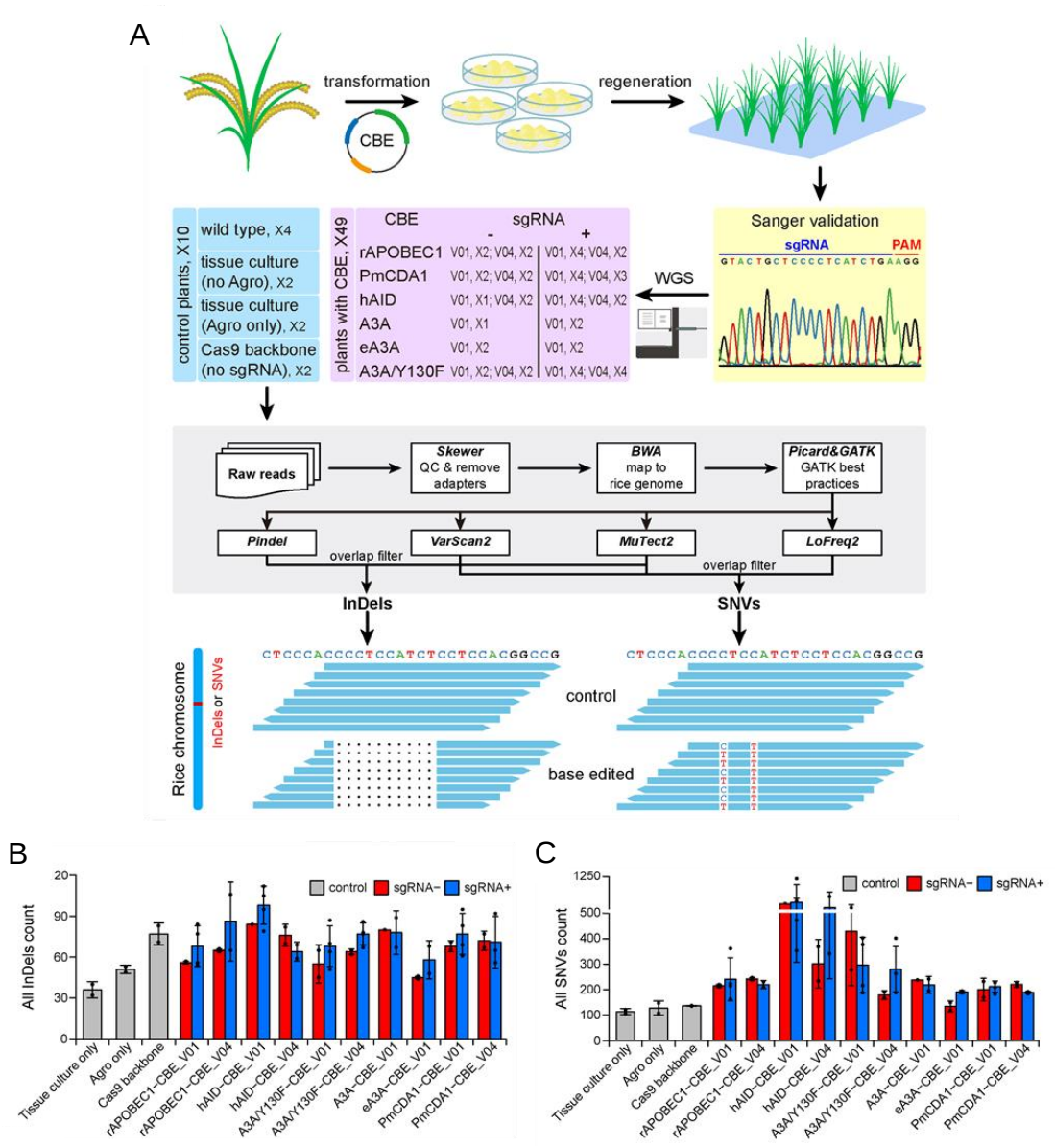


Figure 22: Off-targeting assessment of 10 CBEs detected by WGS of T0 rice plants. **A)** A flow chart listing T0 plants selected for WGS and WGS pipeline. **B)** Indels and **C)** single nucleotide variations identified by WGS in control plants for tissue culture, *Agrobacterium* mediated transformation, Cas9 expression as well as ten CBEs. **B, C)** Each dot represents independent T0 line. Data are presented as mean values ± standard deviation.

Off-target effects were further investigated by analyzing the detected SNVs for all six possible nucleotide substitutions. The frequency of C:G>T:A mutations was approximately 30% for the three tissue culture related control groups (Figure 23A-E).

The C:G>T:A mutation frequency was elevated to approximately 50% in rAPOBEC1-CBE_V01 T0 plants with or without sgRNAs (Figure 23A), indicating that a portion of C:G>T:A transitions derived from rAPOBEC1 mediated off-target effects as previously reported (102,103). The C:G>T:A mutation frequency rose to approximately 75% in case of hAID based V01 and V04 editors (Figure 23B), indicating higher levels of off-targeting. The C:G>T:A mutation frequencies for A3A/Y130F based V01 and V04 editors were approximately 60% and 50%, respectively (Figure 23C), while for A3A and eA3A based CBE-V01 editors were low at approximately 40% for both (Figure 23D). The minimal off-target effect was previously observed in case of eA3A-CBE in human cell lines (97), however, low off-target effect might be in correlation with its inherently low editing activity (Figure 18C) (174). Remarkably, PmCDA1 based V01 and V04 editors both displayed low C:G>T:A mutation frequencies comparable to levels of tissue culture related controls (Figure 23E), indicating PmCDA1 editors have the lowest tendency for C:G>T:A transition introductions compared to other high activity CBEs. Other non-C:G>T:A mutation frequencies were comparable to or slightly lower than the levels of tissue culture related controls for all ten tested CBEs (Figures 23A-E).

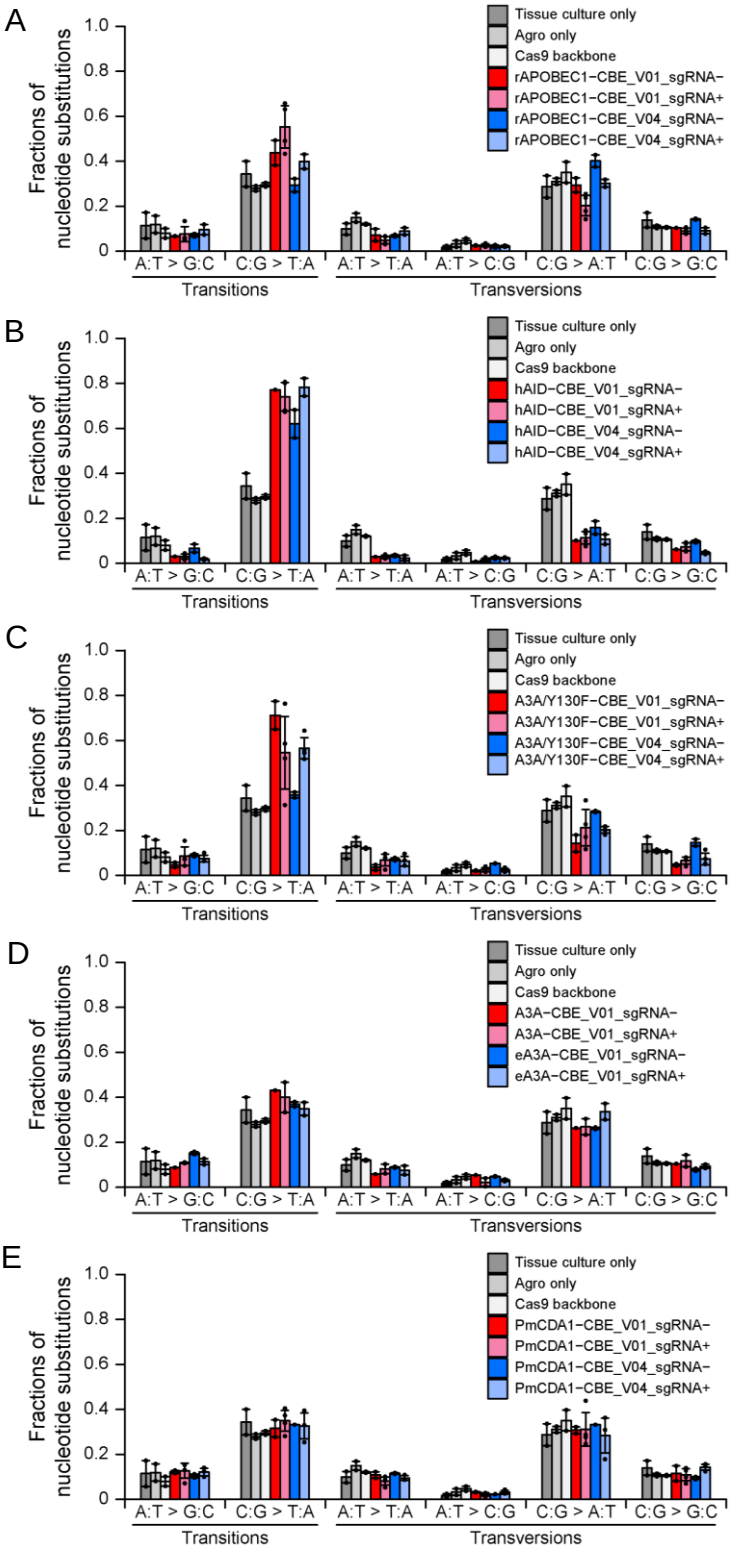


Figure 23: CBEs mediated off-target substitution introduction detected by WGS of T0 rice plants. Transitions and transversions mutations of **A)** rAPOBEC1 based CBEs, **B)** hAID based CBEs, **C)** A3A/Y130F based CBEs, **D)** A3A and eA3A based CBE_V01, **E)** PmCDA1 based

CBEs, detected by WGS. **A-E)** Controls include plants for tissue culture, *Agrobacterium* mediated transformation and Cas9 expression. Each dot represents independent T0 line. Data are presented as mean values ± standard deviation.

4.3 Developing novel C to T base editors with increased activity and refined editing windows

Inferred from results of testing CBEs with modified architecture to facilitate higher focal concentration of effectors at a target site or CBEs with engineered deaminases in the first part of this chapter, the base editing activity could be further improved by using novel cytidine deaminases sourced from diverse organisms. Using bioinformatics approach, 36 deaminases were chosen and were fused to N terminal end of Cas9n-UGI in a BE3 architecture (CBE_V01) (Figure 24A). 36 novel CBEs were screen for editing at OsCGRS55 target site (Figure 18A, 20B) in rice protoplasts along with previously tested PmCDA1, hAID, and hA3A/Y130F based CBE_V01 as positive controls. The preliminary results indicated that BasA3G deaminase (from minke whale) and DnA3C2 deaminase (from nine banded armadillo) based CBEs significantly outperformed the other tested CBEs with approximately 22% and 8% C to T editing efficiencies, respectively (Figure 24C). Based on the novel high performing deaminases' protein sequences, the search was expanded to additional 30 novel deaminases. Thus, additional 30 CBEs in BE3 architecture were screened for editing OsCGRS55 target site (Figure 18A, 24B) in rice protoplasts. Amongst the 30 novel CBEs in the second group, 18 CBEs demonstrated at least approximately equal C to T base editing efficiency compared to PmCDA1-CBE, the lowest of the positive controls (Figure 24D). The success rate of obtaining novel deaminases that would perform at

least as good as established controls increased from 2 out of 36 novel CBEs (5.6%) to 18 out of 30 novel CBEs (60.0%) (Figures 24C, D). Almost 10-fold increase in harnessing of high performing novel deaminases in the second group could be explained by searching for closely related deaminases to the two deaminases that performed well in the first group of deaminases originating from a phylogenetically more diverse group of organisms.

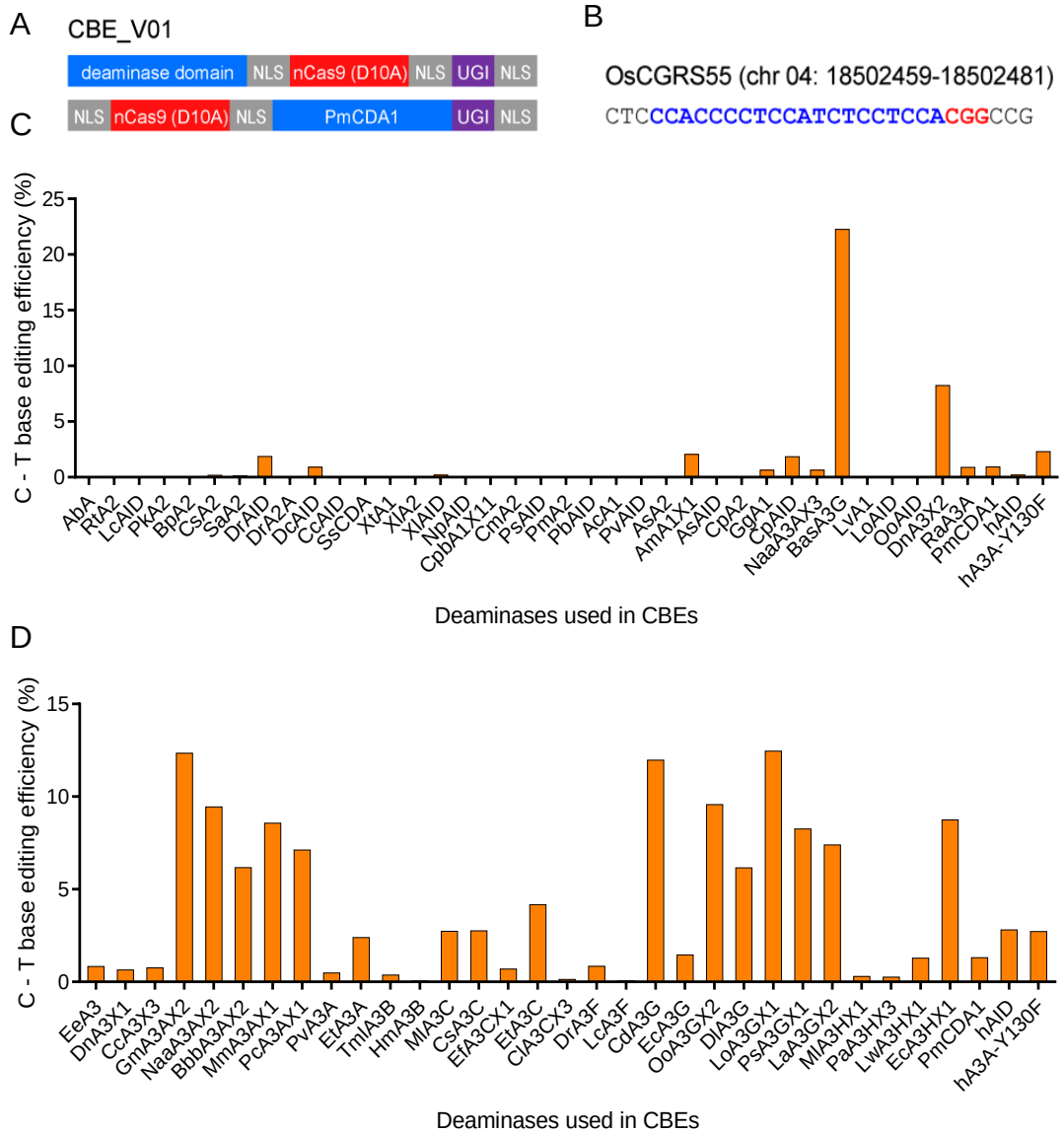


Figure 24: Preliminary screening of 66 novel cytidine deaminases for C to T base editing in rice protoplasts. **A)** Novel cytidine deaminases were fused to N terminal of Cas9n-UGI except PmCDA1 was fused to C terminal end of Cas9n. UGI=uracil glycosylase inhibitor and NLS=nuclear localization signal. **B)** Sequence of CG rich OsCGRS55 target site, at which we screened novel CBEs. **C)** Novel CBE testing at OsCGRS55 target site. Deaminases were sourced from a phylogenetically diverse group of organisms. **D)** Testing the second group of novel CBEs at OsCGRS55 target site. Deaminases in the second group CBEs had a closer protein similarity to top performing deaminases from the first group. **C, D)** For preliminary testing three biological replicates for each novel CBEs were pooled together prior to C to T editing validation with NGS due to cost savings. Presented are average C to T editing efficiencies without error estimations.

Based on the preliminary results of assessing novel cytidine deaminases in CBEs, 29 best-performing novel CBEs were chosen from both groups to be further assessed for base editing activity, base editing window as well as editing purity. Additionally, a fourth previously tested editor was tested as a positive control, hA3A-CBE. In general, the results demonstrated higher C to T base editing activities compared to editing activities during the preliminary testing (Figure 25A compared to Figure 25 C, D). Base editing activities of tested PmCDA1-CBE, hAID-CBE, and hA3A/Y130F-CBE were also noticeably more similar to previous testing (Figure 8C) compared to preliminary testing of novel CBEs (Figure 24 C, D). This could be explained by the omission of gradient centrifugation step during rice protoplasts isolation when performing preliminary testing of novel CBEs. The gradient centrifugation step allows for removal of broken protoplasts to increase the protoplast transformation rate. BasA3G-CBE, MmA3AX1-CBE, CdA3G-CBE, and LoA3GX1-CBE displayed the highest base editing at approximately 30% while AmA1X1, DnA3X2, GmA3AX2, NaaA3AX2, BbbA3AX2, EtA3C, OoA3GX2, DlA3G, PsA3GX1, and LaA3GC2 based CBEs showed editing activities between 10% and 25% (Figure 25A). The rest of base editors

showed base editing activities below 10% except for four previously tested CBEs (Figure 25A). Overall, 14 novel CBEs were identified that can achieve similar or higher C to T base editing activity compared to four previously tested CBEs (Figure 25A). Results of C to A (Figure 25B) and C to G (Figure 25C) editing efficiency at OsCGRS55 target site were low in case of novel CBEs as well as four previously tested CBEs, at levels comparable to NGS sequencing errors, indicating novel CBEs have high editing purity.

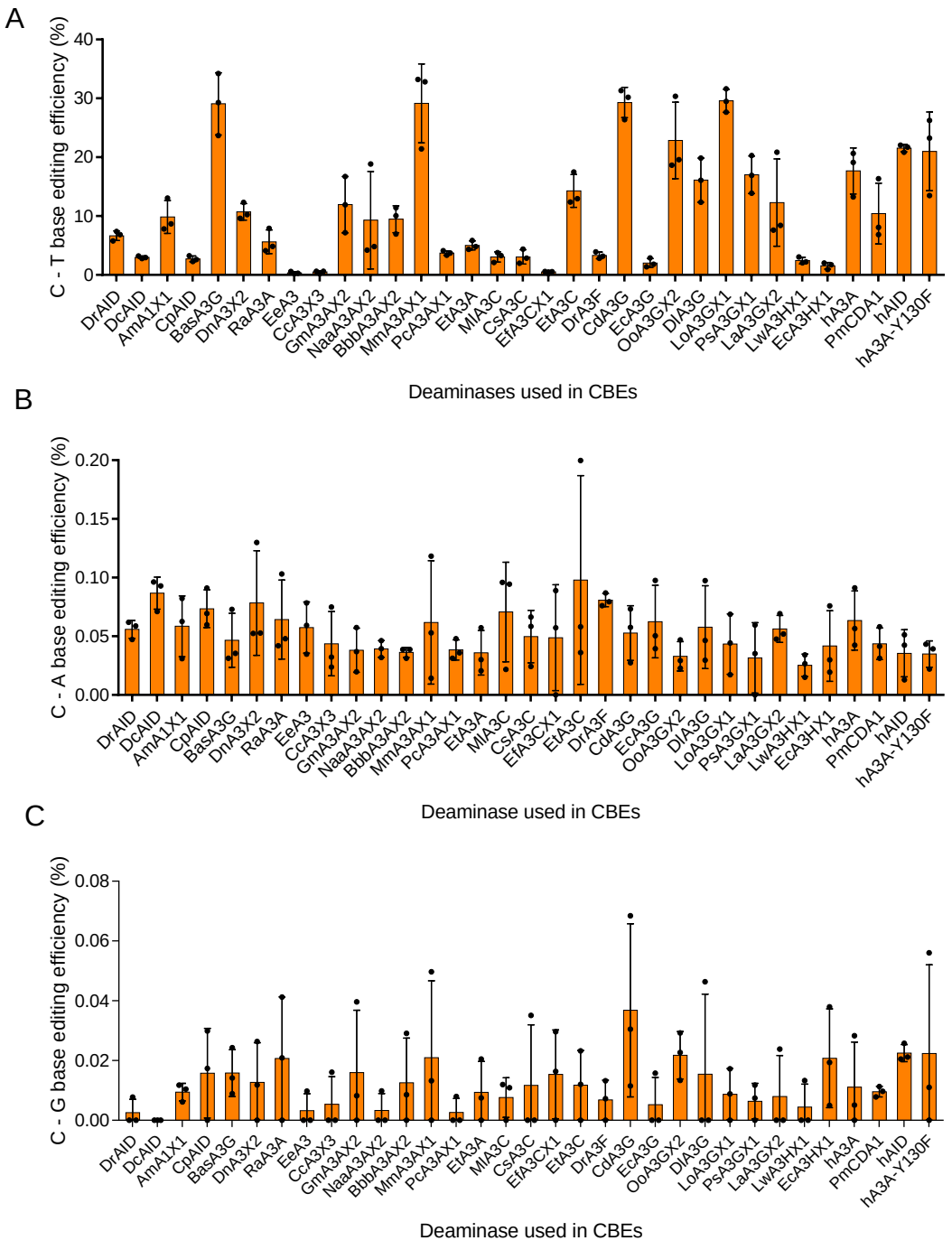
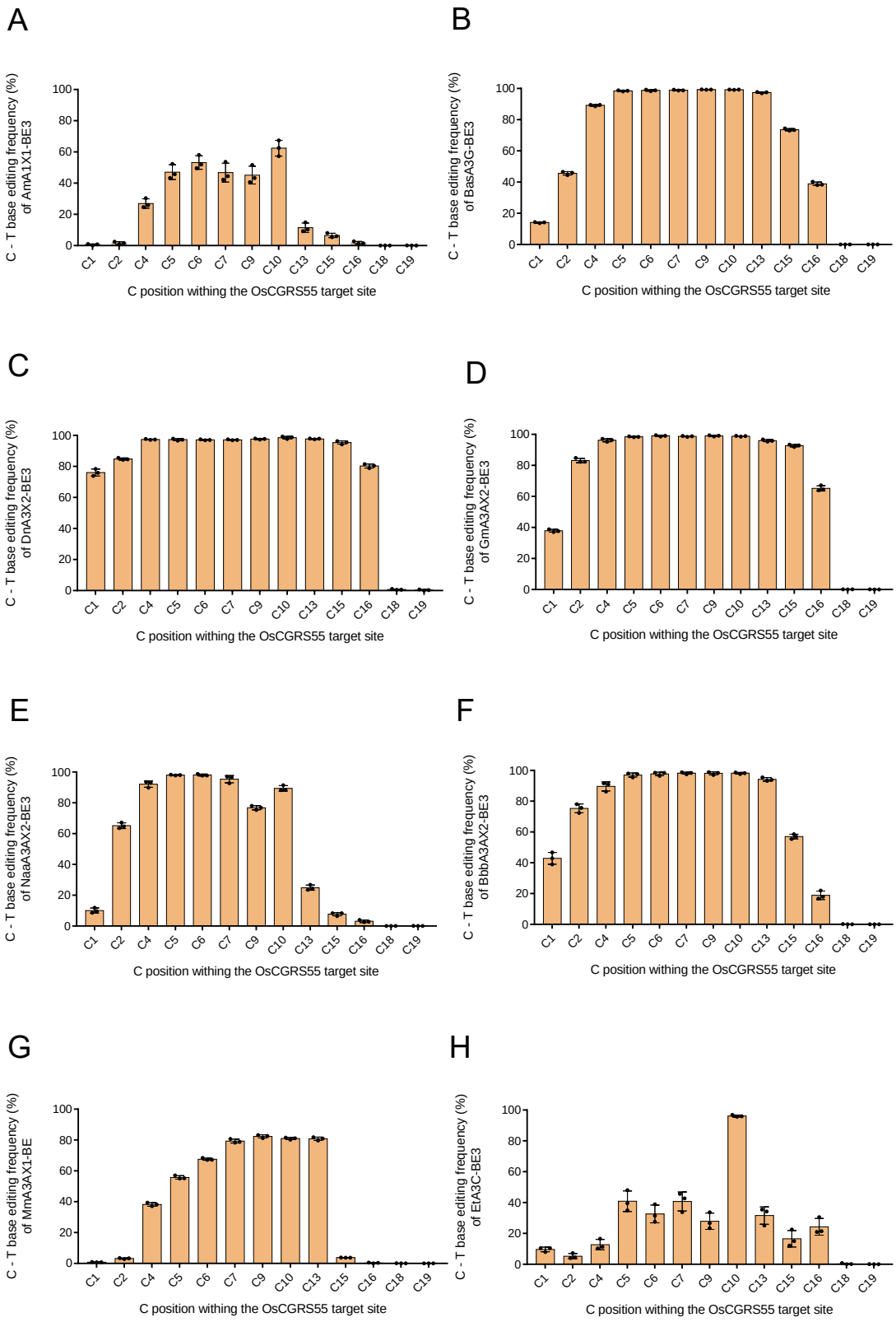
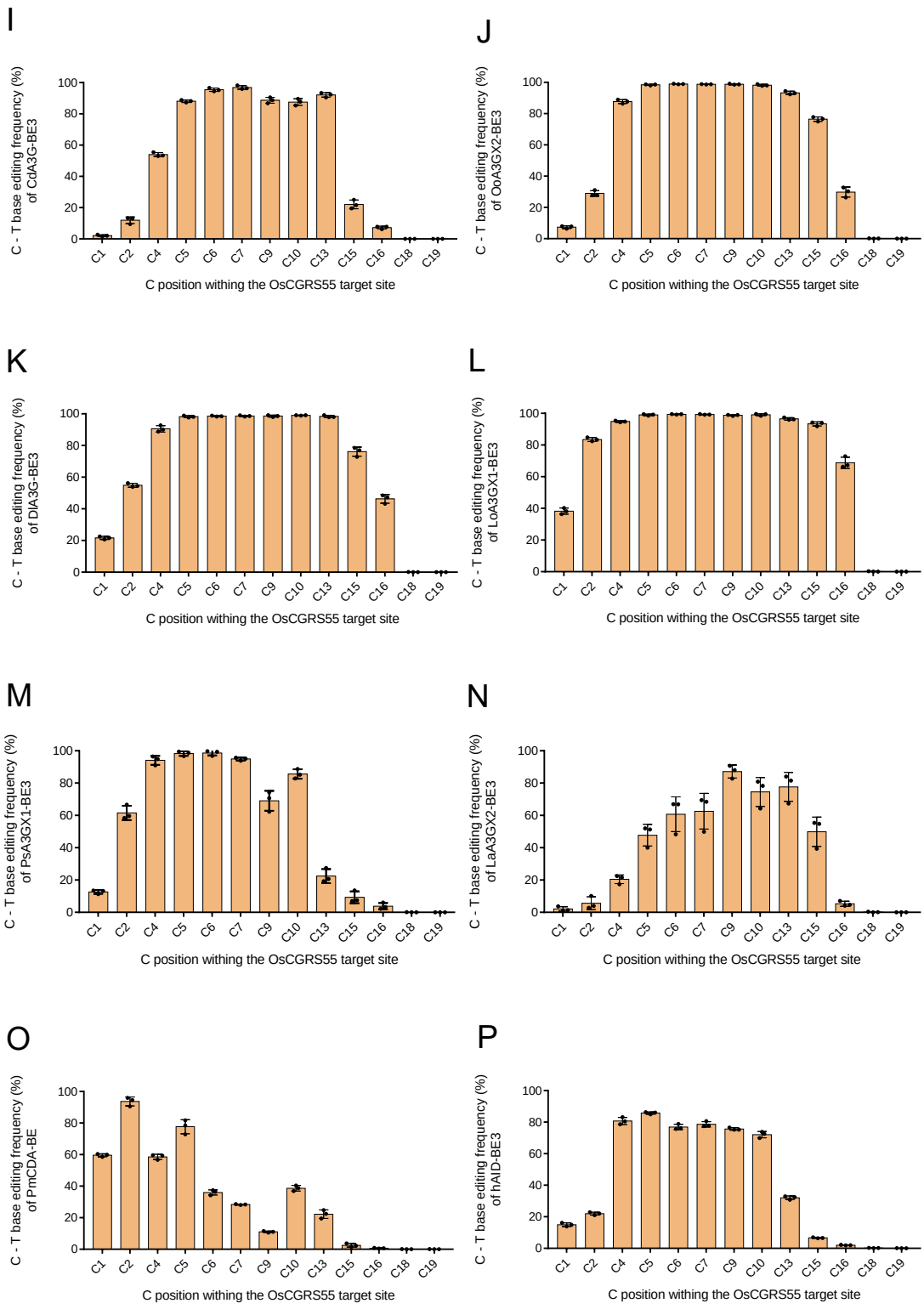


Figure 25: Testing of 29 best performing CBEs at OsCGRS55 target site in rice protoplasts. **A)** C to T, **B)** C to A, and **C)** C to G base editing activities. **A-C**) Each dot represents a biological replicate and each target site was tested with three biological replicates. Data are presented as mean value ± standard deviation.

Furthermore, analyzed were base editing windows of 14 novel CBEs that performed equally well or better compared to four CBEs previously tested at editing OsCGRS55 target site (Figure 25A). Most base editors had a rather wide base editing window with editing frequencies approximately following Gaussian distribution. Editing windows spanned on average from approximately C4 to C13 position of OSCGRS55 target site (Figure 26). PmCDA-CBE had editing window shifted towards 5' prime end of the protospacer (Figure 26O) like it was observed before (Figure 18C). Surprisingly, DnA3X2-CBE (Figure 26C), GmA3AX3-CBE (Figure 26D), and LoA3GX1-CBE (Figure 26L) showed wider Gaussian distribution of editing frequencies across the targetable cytosines in the protospacer compared to previously tested hA3A-CBE (Figure 18C, 26R). Wide editing window with high C to T base editing activity could be useful for directed evolution of a target gene due to its broad coverage of target bases. Furthermore, EtA3C-CBE (Figure 26H) displayed a very narrow editing window focused around C10 position within the OsCGRS55 target site. While testing at more target sites needs to be pursued, I envision a narrow editing window could be useful for introducing targeted C:G>T:A mutations without byproduct edits, facilitating SNV or stop codon introductions in reverse genetic experiments or cop breeding, especially when coupled with PAM-less SpRY Cas9n.


A
C - T base editing frequency (%) of AmA1X1-BE3
C position withing the OsCGRS55 target site
B
C - T base editing frequency (%) of BasA3G-BE3
C position withing the OsCGRS55 target site
C
C - T base editing frequency (%) of DnA3X2-BE3
C position withing the OsCGRS55 target site
D
C - T base editing frequency (%) of GmA3AX2-BE3
C position withing the OsCGRS55 target site
E
C - T base editing frequency (%) of NaaA3AX2-BE3
C position withing the OsCGRS55 target site
F
C - T base editing frequency (%) of BbbA3AX2-BE3
C position withing the OsCGRS55 target site
G
C - T base editing frequency (%) of MmA3AX1-BE
C position withing the OsCGRS55 target site
H
C - T base editing frequency (%) of EtA3C-BE3
C position withing the OsCGRS55 target site
0
20
40
60
80
100
C1
C2
C4
C5
C6
C7
C9
C10
C13
C15
C16
C18
C19


I
C - T base editing frequency (%) of CdA3G-BE3
J
C - T base editing frequency (%) of OoA3GX2-BE3
K
C - T base editing frequency (%) of DlA3G-BE3
L
C - T base editing frequency (%) of LoA3GX1-BE3
M
C - T base editing frequency (%) of PsA3GX1-BE3
N
C - T base editing frequency (%) of LaA3GX2-BE3
O
C - T base editing frequency (%) of PmCDA-BE
P
C - T base editing frequency (%) of hAID-BE3
0
20
40
60
80
100
C1
C2
C4
C5
C6
C7
C9
C10
C13
C15
C16
C18
C19
C position withing the OsCGRS55 target site

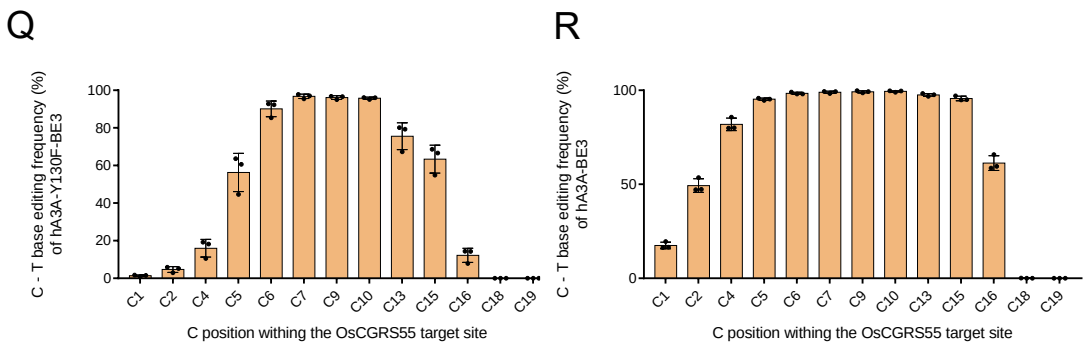


Figure 26: C to T base editing windows of CBEs at OsCGRS55 target site in rice protoplasts. Each dot represents a biological replicate and each target site was tested with three biological replicates. Data are presented as mean value ± standard deviation.

Chapter 5: Prime editing technology can be useful for introducing desired small-scale edits

5.1 Introduction

Prime editing is a CRISPR-based, precise genome editing technology that enables introduction of all possible combinations of nucleotide substitutions as well as small-scale deletions and insertions (40). Introduction of small-scale insertions into plant genomes could be especially enticing due to a low efficiency of homologous dependent repair (HDR) pathway in plants (101,177). In this chapter, it was explored whether prime editing technology could be used in rice protoplasts. The work presented in this chapter has been published (125). Members of Yong Zhang lab from the University of electronic science and technology of China performed the rice protoplasts assay and analyzed the data.

5.2 PE3 and PE3b systems introduced desired mutations, albeit with low efficiencies

The exciting prime editing technology was assessed in rice protoplasts. Since in mammalian cell lines the PE3 and PE3b strategies were best performing according to the seminal report (40), Plant prime editor 3 version 1 (PPE3-V01), optimized for plant expression was first evaluated (Figure 27A). In PE3 strategy, while the reverse transcriptase (RT) modifies the target DNA sequence, nicking single guide RNA (nsgRNA) would nick the non-edited strand within 100 bp from the editing site. Testing PPE3-V01 at five target sites with 13 nucleotide primer binding site (PBS) and RT templates resulted in successful prime editing at all tested target sites in rice protoplasts, although the editing efficiencies were extremely low, spanning merely from 0.05% to 0.15% (Figure 27B). The positive NGS reads were verified for precise incorporation of

desired edits and some reads revealed large deletions, probably due to paired nicking when using a Cas9-H840A nickase, pegRNA and an nsgRNA in the PE3 system (Figure 27C). To minimize these deletion byproducts, PE3b strategy was used conferring temporally separated nicking events of complementary strands in the vicinity of the target site as nsgRNA was designed to match the already edited strand, but not wild type sequence (40). Hence, five additional sites were targeted with PPE3b-V01 system with variable lengths of RT templates, ranging from 13nt to 23nt. Once again, successful prime editing was detected at all five tested target sites with low editing efficiencies in rice protoplasts, up to 0.4% (Figure 27D). Further verification of positive NGS reads revealed that most of the positive reads represented desired editing outcomes and deletion byproducts were greatly reduced with the PE3b strategy (Figure 27E) compared to with PE3 strategy (Figure 27C).

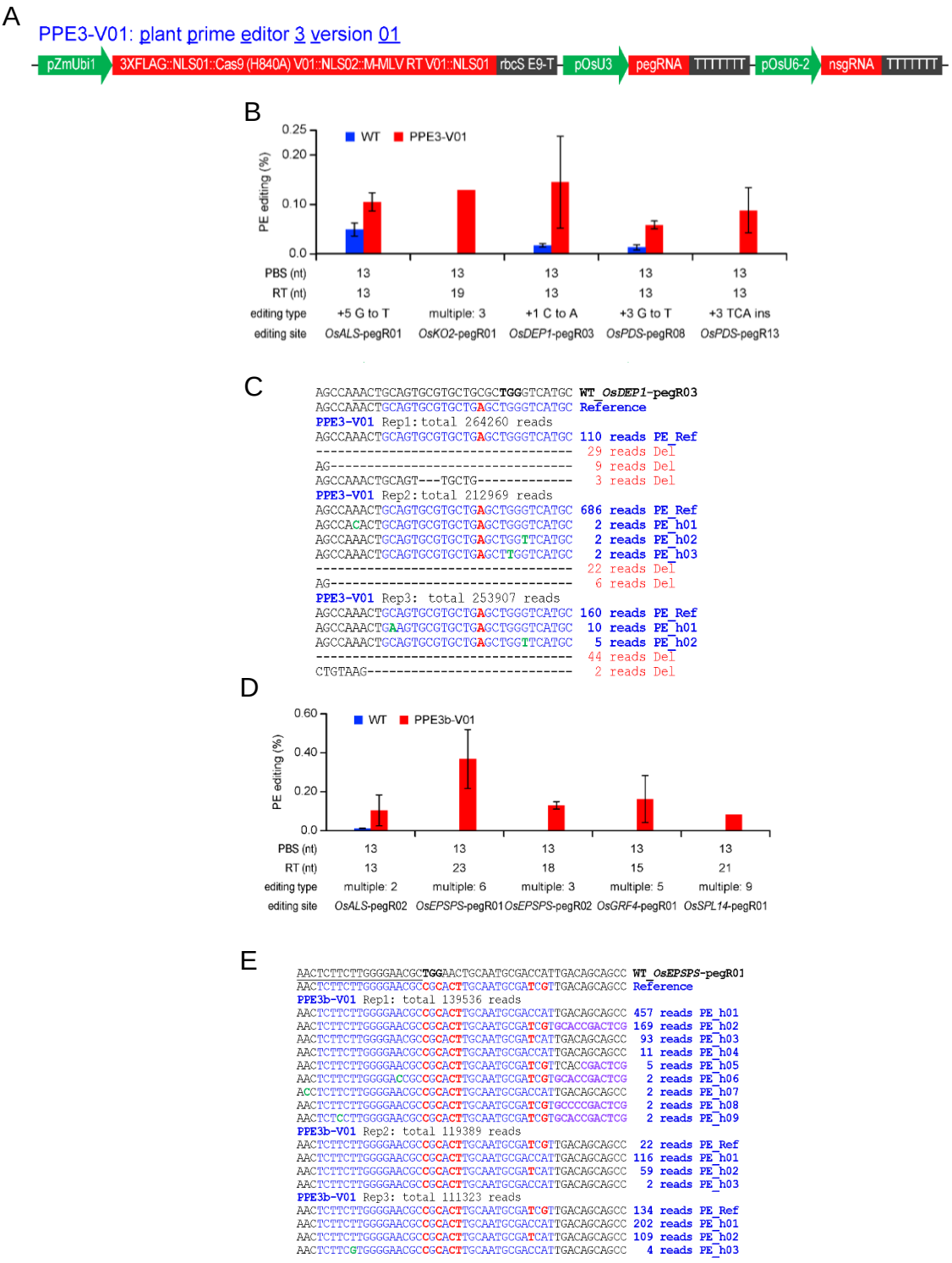


Figure 27: Assessing PE3-V01 and PE3b-V02 editors in rice protoplasts. **A)** Schematic representation of Plant Prime Editor 3-Version 1 (PPE3-V01). **B)** Prime editing frequencies by PPE3-V01 editor at 5 target sites. Primer binding site and reverse transcriptase template lengths, editing type and target sites are indicated under the abscissa. **C)** Prime editing events with PPE3-V01 as determined by NGS. **D)** Prime editing frequencies by PPE3b-V01 editor at

5 target sites. Primer binding site and reverse transcriptase template lengths, editing type and target sites are indicated under the abscissa. **E)** Prime editing events with PPE3b-V01 as determined by NGS. **B-E)** Each target site was tested with three biological replicates. **B, D)** Data are presented as mean value ± standard deviation. **C, E)** The sequences from top to bottom are the wild type sequence (protospacer is underlined, PAM is in bold), the desired prime editing outcome (Reference), confirmed precise prime editing events matching the desired prime editing outcome (PE_Ref), precise prime editing with additional single nucleotide polymorphisms (e.g. PE_h01; h stands for haplotype), and deletions resulted from NHEJ repair mechanism. The nucleotides edited by prime editing are in red.

5.3 PE2 strategy was comparable to PE3 and PE3b strategies in rice cells

PPE3 and PPE3b-V01 systems could have been responsible for the observed low prime editing activities at ten tested target sites in rice protoplasts (Figure 27B, D). To address this concern, a separate PPE3-V02 was prepared, which was different from PPE3-V01 in many aspects to avoid transcription and intercellular transport related inefficiencies including NLS configuration, promoters and terminators for prime editing as well as gRNA, strategy for multiplexing two gRNAs and vector backbone (Figure 28A). Additionally, a systematic approach was used for testing various PBS and RT template pairs of variable lengths to achieve higher prime editing efficiency at each target site. Thus, PPE3-V02 system was tested at three target sites with five different pairs of PBS and RT templates in rice protoplasts. The results demonstrated that prime editing frequencies varied at each target site, which was impacted by the different PBS-RT pairs (Figure 28B-D). Since the same three edits in *OsPDS* and *OsDEP1* genes were pursued earlier with PPE3-V01 (Figure 28B) the data indicate that PPE3-V02 demonstrated improved editing efficiency at these target sites with an optimal PBS-RT template combination (Figure 28B, C). Remarkably, PPE3-V02 resulted in the highest observed editing frequency of 1.55% at a different site in *OsDEP1* (Figure 28D).

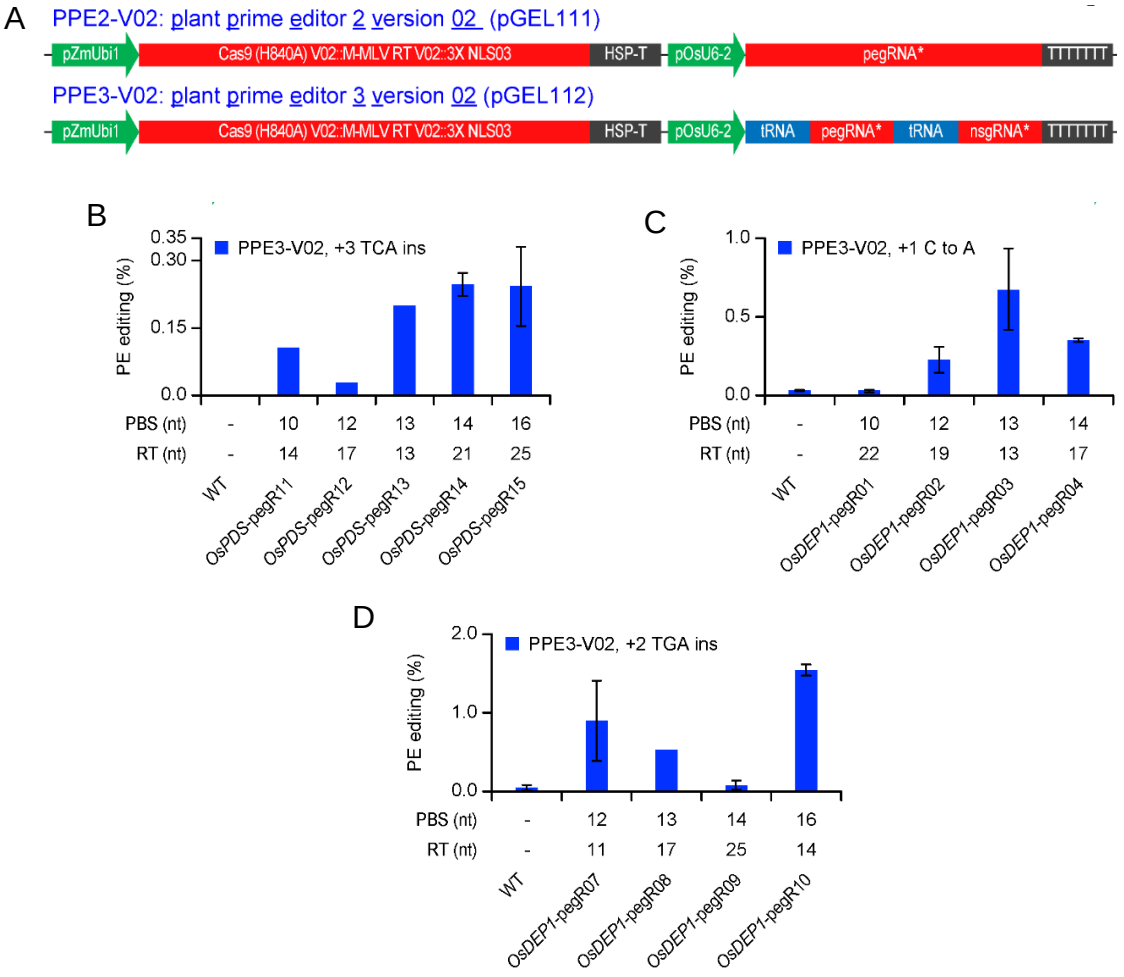


Figure 28: Assessing PPE2-V02 editor in rice protoplasts. **A)** Schematic representation of Plant Prime Editor 2-Version 2 (PPE2-V02) and Plant Prime Editor 3-Version 2 (PPE3-V02). **B-D)** Comparison of multiple PBS-RT template pairs of different lengths enabling TCA insertion at the OsPDS target site, C to A base change at the OsDEP1 site and TGA insertion at the OsDEP1 site, respectively, by PPE3-V02. Data are presented as mean value ± standard deviation of two biological repeats.

While PE3 and PE3b strategies were superior compared to PE2 strategy in mammalian cell lines (40), this might not be the case also in rice protoplasts due to possible differences in RT reactions and DNA repair pathways in plant cells. PE2 strategy does not employ a second sgRNA to nick the non-edited strand. Plant prime editor 2 version 2 (PPE2-V02) (Figure 28A) was tested and compared with PPE3-V02 at editing target

site in *OsPDS* gene with different PBS-RT template combinations. The results indicated PPE2-V02 had comparable or higher editing efficiencies than PPE3-V02 at this site in most cases (Figure 29A). Several other research groups have also demonstrated higher editing efficiencies with PE2 strategy compared to PE3 or PE3b strategies (126–128). However, testing PPE2-V02 at another site resulted in low editing frequencies despite of using multiple PBS-RT template combinations (Figure 29B). To conclude, the prime editing efficiencies that were achieved in rice protoplasts are an order of magnitude lower compared to editing efficiencies reported in seminal paper in mammalian cell lines (40). Therefore, significant improvements are needed to develop efficient plant PEs.

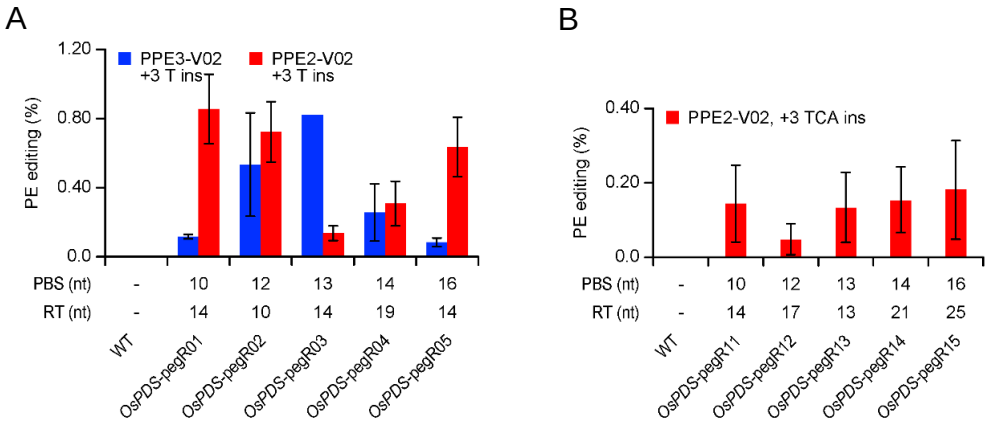


Figure 29: Assessing PPE2-V02 and PPE3-V02 editors in rice protoplasts. **A)** Comparison of multiple PBS-RT template pairs of different lengths enabling T insertion at the OsPDS target site by PPE3-V02 and PPE2-V02. **B)** PPE2-V02 prime editing at another site in the *OsPDS* gene with multiple PBS-RT template pairs of different lengths enabling TCA insertion. **A, B)** Data are presented as mean value ± standard deviation of two biological repeats.

Chapter 6: Conclusions and future perspectives

6.1 Introduction

The focus of the dissertation was on expanding the genome editing scope with engineered Cas9 variants as well as improving and developing novel precise technologies for genome editing in plants. The results showed that these CRISPR reagents can expand the targeting scope in plants. Additionally, precise genome editing technologies can be utilized for plant base editing and prime editing, albeit with varying efficiencies. In chapter 6, the conclusions are collated, and future perspectives outlined.

6.2 Conclusions

Genome editing with engineered xCas9, Cas9-NG, Cas9-NGv1, (i)SpyMacCas9, and SpRY Cas9 variants was tested in rice protoplast as well as T0 rice lines. Inferred from the results of testing genome editing at NGN PAM sites obtained in rice protoplasts and stable T0 lines (Figure 5, 6 and Table 1), xCas9 did not demonstrate altered PAM preference in rice compared to the Cas9-WT, which is different from the data previously published in mammalian cell lines (47). However, xCas9 showed enhanced targeting specificity in rice protoplasts, a feature that was also reported in mammalian cell lines. The data suggest that a conclusion made for a CRISPR-Cas9 system based on data in mammalian cell lines may not always apply to plant cells. It is hence necessary to carefully evaluate the reported Cas9 variants in plants for their genome editing potential in crops. Indeed, when assessing iSpyMacCas9's PAM preference in rice protoplasts (Figure 11B), iSpyMacCas9 showed high genome editing efficiency at NAAR (R=A, G) PAM sites, not NAA PAMs as claimed in the paper that explored iSpyMacCas9 in human cell lines (51). These discrepancies on PAM requirements

between mammalian cells and plant cells warrant future investigation, before xCas9 and iSpyMacCas9's wide adoption for genome editing in other plant species by the plant research community.

Testing both Cas9-NG engineered variants at NGN PAM sites in rice protoplasts (Figure 7) revealed that Cas9-NG variants demonstrated lower genome editing efficiencies compared to canonical Cas9-WT at canonical NGG PAM sites. Both Cas9-NG variants expanded the targeting scope to non-canonical PAM sites, especially sites associated with NGT PAMs. Comparing both Cas9-NG variants at generating edited T0 lines at four target sites (Table 2) showed Cas9-NG having superior performance over Cas9-NGv1 variant in generating monoallelic and biallelic T0 edited lines. The observation of the Cas9-NGv1 and Cas9-NG performance in rice seems to be consistent with the performance in human cell lines (48). Testing SpRY Cas9 engineered variant at NNN PAM sites in rice protoplasts (Figure 14) showed that SpRY Cas9 could edit sites associated with all NNN PAM groups with higher genome editing efficiencies at NRN PAM sites than NYN PAM sites. The data hence are largely consistent with the SpRY Cas9 data generated in human cell lines (49).

However, PAM-relaxed Cas9-NG and PAM-less SpRY Cas9 would face an interesting problem that was not mentioned in studies using mammalians cell lines. In plant genome editing, DNA-based delivery (e.g. plasmids for protoplast transformation and T-DNA for stable plant transformation) is the predominant method. This can lead to a vector self-editing. In fact, T-DNA self-editing was previously reported when applying

Cas9-NG in rice (178). Presumably, this self-editing problem will exacerbate with PAM-less SpRY Cas9. This was indeed what was discovered. Comparison of SpRY Cas9 and canonical Cas9-WT editing profiles in rice protoplasts (Figure 15) revealed SpRY Cas9 introduced a higher frequency of a few nucleotides larger deletions than Cas9-WT. The data suggest self-editing of the transfected plasmids could occur in rice protoplasts. Furthermore, SpRY Cas9's self-editing tendency was found in T0 rice lines with different gRNAs (Table 7). Interestingly, *de novo* generated sgRNAs contributed to a secondary off-target effect as demonstrated in T0 rice lines (Table 9). Fortunately, such off-target effects, albeit being detectable, are not widespread. In conclusion, SpRY Cas9 systems are useful CRISPR tools that can greatly expand genome editing scope in plants. Future work may focus on enhancing editing activity of SpRY Cas9 while minimizing its self-editing activity.

Base editing is constrained by the Cas9 PAM requirements as both CBEs and ABEs have their own editing windows, which in many cases are narrow. To broader the base editing scope, base editors with engineered Cas9-NG, (i)SpyMacCas9, and SpRY Cas9 variants were tested in rice. Cas9-NG derived C to T base editor with PmCDA1 deaminase demonstrated up to 20% base editing efficiency in rice protoplasts at four tested non-canonical PAM sites (Figure 9) and generated both monoallelic and biallelic edited T0 rice lines (Table 3) with variable base conversion profiles amongst the independent T0 lines (Figure 10A, B). z-iSpyMacCas9 derived CBEs with hAID, PmCDA1 and A3A/Y130F cytidine deaminases expanded the base editing scope to A-

rich regions as demonstrated at four target sites in rice protoplasts with up to 6% average editing efficiency (Figure 12). z-iSpyMacCas9 derived CBEs with PmCDA1 and A3A/Y130F cytidine deaminases generated up to 70% of base edited T0 rice lines at three tested YAAA PAM sites (Y=C/T) (Table 5). A to G base editing reached only up to half of a percent at four tested sites associated with A rich PAMs with z-iSpyMacCas9 ABE in rice protoplasts (Figure 13), however, 12.5% of tested T0 rice lines were base edited (Table 6). Testing SpRY Cas9n-PmCDA1 at 26 target sites in rice protoplasts revealed PAM-less base editing with higher editing efficiencies at NRN PAM sites compared to NTN PAM sites (Figures 16A-D). SpRY Cas9n-PmCDA1 base editor generated up to 34% C to T base edited T0 rice lines at three sites tested with favoring base editing over T-DNA self-editing (Table 10). Similarly, as with z-iSpyMacCas9 ABE, SpRY Cas9n ABE demonstrated low A to G editing efficiency at seven relaxed PAM sites in rice protoplasts (Figure 17A, B). SpRY Cas9n ABE generated up to 79% A to G base edited T0 rice lines at three relaxed PAM sites tested, albeit with frequent self-editing (Table 11). Collectively, multiple base editors with altered PAM requirements were developed and demonstrated in rice, and some of which have been successfully applied in other plant species in our lab and elsewhere.

It is critical for plant base editors to have a high editing activity, different editing windows, and broadened targeting scope. Amongst seven canonical Cas9n derived CBEs, A3A/Y130F-CBE_V01 and PmCDA1-CBE_V01 demonstrated the highest C to T base editing at four CG rich target sites with canonical PAMs (Figure 18A) in rice protoplasts (Figure 18C). A3A/Y130F-CBE_V01 favored a broad activity window

spanning from 4th to 15th nucleotide within the GC rich target sites while PmCDA1-CBE_V01 had editing window shifted towards the 5' end of GC rich target sites (Figure 18D). With CBE_V02 and CBE_V03, several different or the same cytidine deaminases were recruited to the GC rich target sites resulting in, unfortunately, diminished C to T base editing activities in rice protoplasts (Figures 19 and 20). Recruiting several uracil glycosylase inhibitors (UGI) to GC rich target sites in rice protoplasts did not decrease the C to T editing activity (Figure 21B compared to 18C) or modify base editing windows (Figure 21C compared to 18D). Comparing A3A/Y130F-CBE_V01 and A3A/Y130F-CBE_V04 at ten additional target sites in rice protoplast showed that A3A/Y130F-CBE_V04 was a more robust base editor due to having drastically smaller dynamic range of C to T editing efficiency (Figure 21E) and conferred purer base editing due to introducing lower amounts of indels (Figure 21F). The data from targeting four GC rich target sites in stable T0 rice lines showed CBE_V04 editors outperformed CBE_V01 editors in C to T editing efficiency and purity (Table 13). Off-target assessment by WGS of edited T0 rice lines showed PmCDA1 based CBE_V01 and CBE_V04 were the most specific base editors tested with high editing efficiency (Figures 18C and 21B) due to introducing low amount of indels and SNVs (Figures 22A, B). Higher levels of indels and SNVs were detected in case of hAID based CBE_V01 and CBE_V04 (Figures 22A, B). This work collectively provided highly efficient and versatile cytosine base editors to the plant research community.

In pursuit of additional highly efficient CBEs for plants, 66 CBEs with novel deaminases were tested at a CG rich target site in rice protoplasts (Figures 24C, D). Fourteen out of 66 novel CBEs demonstrated equal or higher C to T base editing activity compared to A3A/Y130F-CBE_V01 and PmCDA1-CBE_V01 (Figure 25A) with low C to A and C to G unintended editing (Figures 25B and C). DnA3X2-CBE (Figure 26C) demonstrated wider Gaussian distribution of editing frequencies across the targetable Cs in the protospacer compared to previously tested hA3A-CBE (Figures 18D, 26R) while EtA3C-CBE (Figure 26H) displayed a narrow editing window focused around C10 position within the target site in rice protoplasts. I anticipate that these novel cytidine deaminases will greatly enrich the precise base editing toolbox in plants upon further testing for C to T base editing in additional plant species.

Different plant prime editors were developed and tested in rice protoplasts. Prime editing 3 and 3b systems introduced desired edits to the target sites with canonical PAMs in rice protoplasts, however, with editing efficiencies reaching only up to 0.40 % on average (Figures 27B, D). PE3b system greatly reduced deletion byproducts at the target sites in rice protoplasts compared to PE3 system (Figure 27E compared to Figure 27C). Testing PE3 system at three target sites with PBS and RT template pairs of variable lengths revealed various prime editing efficiencies while introducing the same desired edits (Figures 28B-D). PE2 system demonstrated similar prime editing efficiencies compared to PE3 and PE3b systems in rice protoplasts (Figure 29A). On average, prime editing efficiencies achieved in rice protoplasts were an order of magnitude lower compared to editing efficiencies reported in seminal paper in

mammalian cell lines (40). While prime editors with improved activity were recently reported in plants (179), it is anticipated that further improvement is needed before prime editing becomes a reliable tool for plant researchers for introduction of precise changes in the plant genomes.

6.3 Future perspectives

CRISPR-Cas9 technology profoundly revolutionized the genome editing field after the 2012 seminal publication (10). Ever since, there has been a race to discover novel CRISPR systems and utilize them as well as CRISPR based technologies in various organisms. In the future, the CRISPR toolbox will become even more versatile in facilitating small-scale and large-scale genomic changes and epigenetic modifications. A major hindrance in implementing genome editing technologies is the delivery of molecular reagents that facilitate genome editing. While there are several viral vectors to be modified for plant tissue specific delivery, nanotechnology will also surely play a pivotal role for organism-wide as well as cell-specific delivery. A highly efficient delivery method could circumvent a tissue culture procedure that is time consuming and, most importantly, not yet established for many crops or industrial plants. Due to widespread bacterial, fungal and viral diseases in various crops (180), pathogen detection could be a major implementation of the versatile CRISPR toolbox. With the development of many useful genome editing tools, plant research and agriculture innovation will be further boosted. For instance, genome editing may pave a way for more sustainable agriculture with reduced water and fertilizer usage while obtaining high crop yields in light of increasing global population and looming climate change.

Appendices

Supplementary Table 1: List of cytidine deaminases adopted for base editing with NCBI protein database accessions

#	Assigned cytidine deaminase name	Origin	NCBI protein database accession
1	AbA	Acinetobacteria	WP_143213723.1
2	RtA2	Rhincodon typus	XP_020366596.1
3	LcAID	Latimeria chalumnae	XP_014350178.1
4	PkA2	Paramormyrops kingsleyae	XP_023669168.1
5	BpA2	Boleophthalmus pectinirostris	XP_020791572.1
6	CsA2	Cynoglossus semilaevis	XP_008319408.1
7	SaA2	Salvelinus alpinus	XP_023842724.1
8	DrAID	Danio rerio	NP_001008403.1
9	DrA2A	Danio rerio	NP_001013332.1
10	DcAID	Denticeps clupeoides	XP_028834826.1
11	CcAID	Cyprinus carpio	XP_018981523.1
12	SsCDA	Salmo salar	XP_014010073.1
13	XtA1	Xenopus tropicalis	XP_002941248.1
14	XlA2	Xenopus laevis	XP_018104881.1
15	XlAID	Xenopus laevis	NP_001089181.1
16	NpAID	Nanorana parkeri	XP_018426850.1
17	CpbA1X11	Chrysemys picta bellii	XP_023965938.1
18	CmA2	Chelonia mydas	XP_007063573.1
19	PsAID	Pelodiscus sinensis	XP_025038798.1

20	PmA2	Protobothrops mucrosquamatus	XP_015671004.1
21	PbAID	Python bivittatus	XP_025022614.1
22	AcA1	Anolis carolinensis	XP_008102031.1
23	PvAID	Pogona vitticeps	XP_020670354.1
24	AsA2	Alligator sinensis	XP_006025139.1
25	AmA1X1	Alligator mississippiensis	XP_019337862.1
26	AsAID	Alligator sinensis	XP_006032700.1
27	CpA2	Crocodylus porosus	XP_019402241.1
28	GgA1	Gavialis gangeticus	XP_019371538.1
29	CpAID	Crocodylus porosus	XP_019399695.1
30	NaaA3AX3	Neophocaena asiaeorientalis asiaeorientalis	XP_024617344.1
31	BasA3G	Balaenoptera acutorostrata scammoni	XP_007189136.1
32	LvA1	Lipotes vexillifer	XP_007469425.1
33	LoAID	Lagenorhynchus obliquidens	XP_026948891.1
34	OoAID	Orcinus orca	XP_012391905.1
35	DnA3X2	Dasypus novemcinctus	XP_004447910.1
36	RaA3A	Rousettus aegyptiacus	XP_016017136.1
37	EeA3	Elephantulus edwardii	XP_006890259.1
38	DnA3X1	Dasypus novemcinctus	XP_023445737.1
39	CcA3X3	Castor canadensis	XP_020029507.1
40	GmA3AX2	Globicephala melas	XP_030711791.1

41	NaaA3AX2	Neophocaena asiaeorientalis asiaeorientalis	XP_024617343.1
42	BbbA3AX2	Bison bison bison	XP_010843425.1
43	MmA3AX1	Monodon monoceros	XP_029060932.1
44	PcA3AX1	Physeter catodon	XP_028346448.1
45	PvA3A	Pteropus vampyrus	XP_011384362.1
46	EtA3A	Echinops telfairi	XP_030741607.1
47	TmlA3B	Trichechus manatus latirostris	XP_023584683.1
48	HmA3B	Hylobates moloch	XP_032004340.1
49	MlA3C	Mandrillus leucophaeus	XP_011834235.1
50	CsA3C	Chlorocebus sabaeus	NP_001332881.1
51	EfA3CX1	Eptesicus fuscus	XP_008159951.2
52	EtA3C	Echinops telfairi	XP_030741606.1
53	ClA3CX3	Columba livia	XP_021153311.1
54	DrA3F	Desmodus rotundus	XP_024435233.1
55	LcA3F	Lontra canadensis	XP_032707103.1
56	CdA3G	Camelus dromedarius	XP_031319156.1
57	EcA3G	Equus caballus	XP_023486964.1
58	OoA3GX2	Orcinus orca	XP_012392105.1
59	DlA3G	Delphinapterus leucas	XP_022452168.1
60	LoA3GX1	Lagenorhynchus obliquidens	XP_026963494.1
61	PsA3GX1	Phocoena sinus	XP_032501203.1
62	LaA3GX2	Loxodonta africana	XP_023415683.1
63	MlA3HX1	Mandrillus leucophaeus	XP_011834225.1

64	PaA3HX3	Papio anubis	XP_003905607.2
65	LwA3HX1	Leptonychotes weddellii	XP_030873488.1
66	EcA3HX1	Equus caballus	XP_005606535.1

Supplementary Table 2: List of final T-DNA vectors used for screening novel cytosine base editors used in Chapter 4

Vector	Purpose
pLR2918	For AbA-CBE editing at OsCGRS55 target site
pLR2919	For RtA2-CBE editing at OsCGRS55 target site
pLR2920	For LcAID-CBE editing at OsCGRS55 target site
pLR2921	For PkA2-CBE editing at OsCGRS55 target site
pLR2922	For BpA2-CBE editing at OsCGRS55 target site
pLR2923	For CsA2-CBE editing at OsCGRS55 target site
pLR2924	For SaA2-CBE editing at OsCGRS55 target site
pLR2925	For DrAID-CBE editing at OsCGRS55 target site
pLR2926	For DrA2A-CBE editing at OsCGRS55 target site
pLR2927	For DcAID-CBE editing at OsCGRS55 target site
pLR2928	For CcAID-CBE editing at OsCGRS55 target site
pLR2929	For SsCDA-CBE editing at OsCGRS55 target site
pLR2930	For XtA1-CBE editing at OsCGRS55 target site
pLR2931	For XlA2-CBE editing at OsCGRS55 target site
pLR2932	For XlAID-CBE editing at OsCGRS55 target site
pLR2933	For NpAID-CBE editing at OsCGRS55 target site
pLR2934	For CpbA1X11-CBE editing at OsCGRS55 target site
pLR2935	For CmA2-CBE editing at OsCGRS55 target site
pLR2936	For PsAID-CBE editing at OsCGRS55 target site
pLR2937	For PmA2-CBE editing at OsCGRS55 target site
pLR2938	For PbAID-CBE editing at OsCGRS55 target site
pLR2939	For AcA1-CBE editing at OsCGRS55 target site
pLR2940	For PvAID-CBE editing at OsCGRS55 target site
pLR2941	For AsA2-CBE editing at OsCGRS55 target site
pLR2942	For AmA1X1-CBE editing at OsCGRS55 target site
pLR2943	For AsAID-CBE editing at OsCGRS55 target site
pLR2944	For CpA2-CBE editing at OsCGRS55 target site
pLR2945	For GgA1-CBE editing at OsCGRS55 target site
pLR2946	For CpAID-CBE editing at OsCGRS55 target site
pLR2947	For NaaA3AX3-CBE editing at OsCGRS55 target site

pLR2948	For BasA3G-CBE editing at OsCGRS55 target site
pLR2949	For LvA1-CBE editing at OsCGRS55 target site
pLR2950	For LoAID-CBE editing at OsCGRS55 target site
pLR2951	For OoAID-CBE editing at OsCGRS55 target site
pLR2952	For DnA3X2-CBE editing at OsCGRS55 target site
pLR2953	For RaA3A-CBE editing at OsCGRS55 target site
pLR2954	For PmCDA1-CBE editing at OsCGRS55 target site
pLR2955	For hAID-CBE editing at OsCGRS55 target site
pLR2956	For hA3A-Y130F-CBE editing at OsCGRS55 target site
pLR3502	For EeA3-CBE editing at OsCGRS55 target site
pLR3503	For DnA3X1-CBE editing at OsCGRS55 target site
pLR3504	For CcA3X3-CBE editing at OsCGRS55 target site
pLR3505	For GmA3AX2-CBE editing at OsCGRS55 target site
pLR3506	For NaaA3AX2-CBE editing at OsCGRS55 target site
pLR3507	For BbbA3AX2-CBE editing at OsCGRS55 target site
pLR3508	For MmA3AX1-CBE editing at OsCGRS55 target site
pLR3509	For PcA3AX1-CBE editing at OsCGRS55 target site
pLR3510	For PvA3A-CBE editing at OsCGRS55 target site
pLR3511	For EtA3A-CBE editing at OsCGRS55 target site
pLR3512	For TmlA3B-CBE editing at OsCGRS55 target site
pLR3513	For HmA3B-CBE editing at OsCGRS55 target site
pLR3514	For MlA3C-CBE editing at OsCGRS55 target site
pLR3515	For CsA3C-CBE editing at OsCGRS55 target site
pLR3516	For EfA3CX1-CBE editing at OsCGRS55 target site
pLR3517	For EtA3C-CBE editing at OsCGRS55 target site
pLR3518	For ClA3CX3-CBE editing at OsCGRS55 target site
pLR3519	For DrA3F-CBE editing at OsCGRS55 target site
pLR3520	For LcA3F-CBE editing at OsCGRS55 target site
pLR3521	For CdA3G-CBE editing at OsCGRS55 target site
pLR3522	For EcA3G-CBE editing at OsCGRS55 target site
pLR3523	For OoA3GX2-CBE editing at OsCGRS55 target site
pLR3524	For DlA3G-CBE editing at OsCGRS55 target site
pLR3525	For LoA3GX1-CBE editing at OsCGRS55 target site
pLR3526	For PsA3GX1-CBE editing at OsCGRS55 target site
pLR3527	For LaA3GX2-CBE editing at OsCGRS55 target site
pLR3528	For MlA3HX1-CBE editing at OsCGRS55 target site
pLR3529	For PaA3HX3-CBE editing at OsCGRS55 target site
pLR3530	For LwA3HX1-CBE editing at OsCGRS55 target site
pLR3531	For EcA3HX1-CBE editing at OsCGRS55 target site

Glossary

(i)SpyMacCas9	Engineered SpyMacCas9 and iSpyMacCas9 variants suitable for editing NAAR PAM sites.
A3A	Cytidine deaminase that catalyzes hydrolytic deamination of cytidine to uracil. It is a part of cytosine base editor along with Cas9n and UGI.
A3A/Y130F	Cytidine deaminase with tyrosine (Y) to phenylalanine (F) mutation at 130 amino acid position that reduces the bystander mutations at the target site. A3A/Y130F catalyzes hydrolytic deamination of cytidine to uracil. It is a part of cytosine base editor along with Cas9n and UGI.
adenine base editors, ABE	Adenine base editors mediate the direct conversion of adenosine to inosine, thereby effecting an A to G (or T to C) substitution.
base editing, BE	A CRISPR based technology conferring the direct, irreversible conversion of one target DNA base into another in a programmable manner, without requiring dsDNA backbone cleavage or a donor template.
base editing activity window	The range of nucleotides that can be converted in an approximately 20 nucleotide long target site.
base editor architecture	The configuration and copy number of effector molecules that confer base editing.
callus (s), calli (pl)	A growing mass of unorganized plant parenchyma cells.
Cas9-NG	Engineered Cas9-NG variant suitable for editing NG PAM sites, especially NGT PAM sites. On average, Cas9-NG demonstrated higher genome editing activity compared to Cas-NGv1.
Cas9-NGv1	Engineered Cas9-NGv1 variant suitable for editing NG PAM sites, especially NGT PAM sites. On average Cas9-NGv1 demonstrated lower genome editing activity compared to Cas-NG.
Cas9-WT, SpCas9	Wild type or canonical Cas9 endonuclease isolated from Gram + bacterium *Streptococcus pyogenes*.
CRISPR-Act3.0	A third generation CRISPR activation system.
CRISPR-Combo	CRISPR-Cas9 Combo is suitable for simultaneous genome editing (targeted mutagenesis or base editing) and gene activation.
cytosine base editors, CBE	Cytosine base editors mediate the direct conversion of cytidine to uridine, thereby effecting a C to T (or G to A) substitution.
cytosine to guanine base editors, CGBE	Cytosine to guanine base editors mediate the direct conversion of cytidine to uridine and after the uridine is removed resulting

	in apyrimidinic site, G is incorporated. This results in C to G (or G to C) substitution.
eA3A	Engineered cytidine deaminase that reduces the bystander mutations at the target site. eA3A catalyzes hydrolytic deamination of cytidine to uracil. It is a part of cytosine base editor along with Cas9n and UGI.
eA3A/Y130F	Engineered cytidine deaminase with an additional tyrosine (Y) to phenylalanine (F) mutation at 130 amino acid position that reduces the bystander mutations at the target site. eA3A/Y130F catalyzes hydrolytic deamination of cytidine to uracil. It is a part of cytosine base editor along with Cas9n and UGI.
hAID	Cytidine deaminase that catalyzes hydrolytic deamination of cytidine to uracil. It is a part of base editor along with Cas9n and UGI.
MCP-MS2	The interaction between MS2 bacteriophage coat protein (MCP) and MS2 RNA aptamer can be used as a recruitment strategy to increase local concentration of desired effector molecules, fused to the MCP, at the target site. Through each MCP-MS2 interaction, 4 effector molecules can be recruited.
organellar base editors	Organellar base editors confer base editing in mitochondria and plastids.
protospacer adjacent motif, PAM	Protospacer adjacent motif is located downstream of a target site in case of Cas9 endonucleases.
PmCDA1	Cytidine deaminase that catalyzes hydrolytic deamination of cytidine to uracil. It is a part of cytosine base editor along with Cas9n and UGI.
prime editing, PE	A versatile and precise CRISPR-Cas9 based genome editing technology that directly writes small-scale desired edits to a DNA target site using a reverse transcriptase fused to a Cas9n.
protoplast	A plant cell without cell wall.
rAPOBEC1	Cytidine deaminase that catalyzes hydrolytic deamination of cytidine to uracil. It is a part of cytosine base editor along with Cas9n and UGI.
self-editing	Self-editing can happen when sgRNA is encoded as dsDNA. Due to relaxed PAM nature of engineered Cas9 endonucleases (like SpRY Cas9), the first tri nucleotides of tracrRNA GTT can serve as a PAM and such nuclease can edit its own sgRNA.
sgRNA-dependent off-target editing	Off-target editing that occurs at unintended target sites that are similar to intended target site.
sgRNA-independent off-target editing	Off-target editing that occurs at unintended target sites.

SpRY Cas9	Engineered SpRY Cas9 variant with near PAM-less nature. SpRY Cas9 showed higher mutation frequencies at NRN PAM sites (R = A, G) than NYN PAM sites (Y = C, T).
T0 rice lines	Regenerated edited rice line.
TadA-7.10, TadA8e	Evolved adenosine deaminases of 7th or 8th generation that catalyze hydrolytic deamination of adenosine to inosine. It is a part of adenine base editor along with Cas9n.
target site	Target site or protospacer is approximately 20 nucleotide sequence within double stranded DNA, where Cas9 introduces double stranded break, nick or Cas9 just binds to.
xCas9-3.7	Engineered Cas9 variant suitable for editing canonical PAM sites with superior targeting specificity compared to SpCas9.

Bibliography

1. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300,** 763 (2003).
2. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol* **15,** 1968–1973 (1995).
3. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12,** 224–228 (1996).
4. Puchta, H., Dujon, B. & Hohn, B. Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res* **21,** 5034–5040 (1993).
5. Epinat, J. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research* **31,** 2952–2962 (2003).
6. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93,** 1156–1160 (1996).
7. Boch, J. & Bonas, U. *Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function. *Annual Review of Phytopathology* **48,** 419–436 (2010).
8. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29,** 143–148 (2011).

9. Mojica, F. J. M., Juez, G. & Rodriguez-Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Molecular Microbiology* **9,** 613–621 (1993).
10. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (2012).
11. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
12. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
13. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315,** 1709–1712 (2007).
14. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. The basic building blocks and evolution of CRISPR-CAS systems. *Biochem Soc Trans* **41,** 1392–1400 (2013).
15. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–338 (2012).
16. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* **45,** 273–297 (2011).
17. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Curr Opin Microbiol* **14,** 321–327 (2011).
18. Deltcheva, E. *et al.* CRISPR RNA maturation by trans -encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

19. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–964 (2008).
20. Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139,** 945–956 (2009).
21. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol* **13,** 722–736 (2015).
22. Pausch, P. *et al.* CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369,** 333–337 (2020).
23. Xiao, R. *et al.* Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. *Molecular Cell* **81,** 4457-4466.e5 (2021).
24. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18,** 67–83 (2020).
25. Lees-Miller, S. P. & Meek, K. Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* **85,** 1161–1173 (2003).
26. Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences* **40,** 701–714 (2015).
27. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *PNAS* **95,** 5172–5177 (1998).
28. Jeggo, P. A. DNA breakage and repair. *Adv. Genet.* **38,** 185–218 (1998).

29. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **36,** 765–771 (2018).
30. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).
31. Chen, H., Choi, J. & Bailey, S. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J Biol Chem* **289,** 13284–13294 (2014).
32. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109,** E2579–E2586 (2012).
33. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31,** 833–838 (2013).
34. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154,** 1380–1389 (2013).
35. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533,** 420–424 (2016).
36. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551,** 464–471 (2017).
37. Chen, L. *et al.* Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat Commun* **12,** 1384 (2021).

38. Kurt, I. C. *et al.* CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* **39,** 41–46 (2021).
39. Zhao, D. *et al.* Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* **39,** 35–40 (2021).
40. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576,** 149–157 (2019).
41. Lowder, L. G. *et al.* A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiology* **169,** 971–985 (2015).
42. Zhang, Y., Iaffaldano, B. & Qi, Y. CRISPR ribonucleoprotein-mediated genetic engineering in plants. *Plant Communications* **2,** 100168 (2021).
43. Klimek-Chodacka, M., Oleszkiewicz, T., Lowder, L. G., Qi, Y. & Baranski, R. Efficient CRISPR/Cas9-based genome editing in carrot cells. *Plant Cell Rep* **37,** 575–586 (2018).
44. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186–191 (2015).
45. Müller, M. *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* **24,** 636–644 (2016).
46. Karvelis, T. *et al.* Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biology* **16,** 253 (2015).
47. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556,** 57–63 (2018).
48. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361,** 1259–1262 (2018).

49. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368,** 290–296 (2020).
50. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).
51. Chatterjee, P. *et al.* A Cas9 with PAM recognition for adenine dinucleotides. *Nat Commun* **11,** 2474 (2020).
52. Schellenberger, V. *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol* **27,** 1186–1190 (2009).
53. Palmenberg, A. C. *et al.* Proteolytic processing of the cardioviral P2 region: Primary 2A/2B cleavage in clone-derived precursors. *Virology* **190,** 754–762 (1992).
54. Lowder, L. G. *et al.* Robust transcriptional activation in plants using multiplexed CRISPR-Act2.0 and mTALE-Act systems. *Molecular Plant* **11,** 245–256 (2018).
55. Shao, S. *et al.* Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res* **44,** e86 (2016).
56. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).
57. Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* **12,** 595–600 (1993).

58. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159,** 635–646 (2014).

59. Pan, C. *et al.* CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. *Nat. Plants* **7,** 942–953 (2021).

60. Pan, C. *et al.* Boosting plant genome editing with a versatile CRISPR-Combo system. *Nat. Plants* **8,** 513–525 (2022).

61. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358,** 1019–1027 (2017).

62. Abudayyeh, O. O. *et al.* A cytosine deaminase for programmable single-base RNA editing. *Science* **365,** 382–386 (2019).

63. Mok, B. Y. *et al.* A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583,** 631–637 (2020).

64. Li, C. *et al.* Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat Biotechnol* **38,** 875–882 (2020).

65. Grünewald, J. *et al.* A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat Biotechnol* **38,** 861–864 (2020).

66. Zhang, X. *et al.* Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat Biotechnol* **38,** 856–860 (2020).

67. Sakata, R. C. *et al.* Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat Biotechnol* **38,** 865–869 (2020).

68. Xie, J. *et al.* ACBE, a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems. *BMC Biology* **18,** 131 (2020).

69. Carter, R. J. & Parsons, J. L. Base Excision Repair, a Pathway Regulated by Posttranslational Modifications. *Mol Cell Biol* **36,** 1426–1437 (2016).

70. Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. *Mol Plant* **10,** 526–529 (2017).

71. Zong, Y. *et al.* Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nature Biotechnology* **36,** 950–953 (2018).

72. Hua, K., Tao, X. & Zhu, J.-K. Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnology Journal* **17,** 499–504 (2019).

73. Zong, Y. *et al.* Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol* **35,** 438–440 (2017).

74. Lu, Y. & Zhu, J.-K. Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. *Mol Plant* **10,** 523–525 (2017).

75. Shimatani, Z. *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nature Biotechnology* **35,** 441–443 (2017).

76. Chen, Y. *et al.* CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. *Sci China Life Sci* **60,** 520–523 (2017).

77. Bastet, A. *et al.* Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. *Plant Biotechnol J* **17,** 1736–1750 (2019).

78. Li, Z., Xiong, X., Wang, F., Liang, J. & Li, J.-F. Gene disruption through base editing-induced messenger RNA missplicing in plants. *New Phytol* **222,** 1139–1148 (2019).

79. Xue, C., Zhang, H., Lin, Q., Fan, R. & Gao, C. Manipulating mRNA splicing by base editing in plants. *Sci China Life Sci* **61,** 1293–1300 (2018).
80. Zhang, R. *et al.* Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. *Nat. Plants* **5,** 480–485 (2019).
81. Veillet, F. *et al.* Transgene-Free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. *International Journal of Molecular Sciences* **20,** 402 (2019).
82. Veillet, F. *et al.* Expanding the CRISPR Toolbox in P. patens Using SpCas9-NG Variant and Application for Gene and Base Editing in Solanaceae Crops. *International Journal of Molecular Sciences* **21,** 1024 (2020).
83. Hunziker, J. *et al.* Multiple gene substitution by Target-AID base-editing technology in tomato. *Sci Rep* **10,** 20471 (2020).
84. Tian, S. *et al.* Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep* **37,** 1353–1356 (2018).
85. Qin, L. *et al.* High-efficient and precise base editing of C•G to T•A in the allotetraploid cotton (Gossypium hirsutum) genome using a modified CRISPR/Cas9 system. *Plant Biotechnol J* **18,** 45–56 (2020).
86. Target base editing in soybean using a modified CRISPR/Cas9 system - Cai - 2020 - Plant Biotechnology Journal - Wiley Online Library. at <https://onlinelibrary.wiley.com/doi/full/10.1111/pbi.13386>
87. Malabarba, J. *et al.* New Strategies to Overcome Present CRISPR/Cas9 Limitations in Apple and Pear: Efficient Dechimerization and Base Editing. *International Journal of Molecular Sciences* **22,** 319 (2021).

88. Xing, S. *et al.* Fine-tuning sugar content in strawberry. *Genome Biology* **21,** 230 (2020).

89. Wu, J. *et al.* Engineering herbicide-resistant oilseed rape by CRISPR/Cas9-mediated cytosine base-editing. *Plant Biotechnol J* **18,** 1857–1859 (2020).

90. Cheng, H. *et al.* Base editing with high efficiency in allotetraploid oilseed rape by A3A-PBE system. *Plant Biotechnology Journal* **19,** 87–97 (2021).

91. Guyon-Debast, A. *et al.* A blueprint for gene function analysis through Base Editing in the model plant Physcomitrium (Physcomitrella) patens. *New Phytol* **230,** 1258–1272 (2021).

92. Li, G., Sretenovic, S., Eisenstein, E., Coleman, G. & Qi, Y. Highly efficient C-to-T and A-to-G base editing in a Populus hybrid. *Plant Biotechnol J* **19,** 1086–1088 (2021).

93. Tang, X. *et al.* Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnology Journal* (2018). doi:10.1111/pbi.13068

94. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353,** (2016).

95. Ren, B. *et al.* Improved Base Editor for Efficiently Inducing Genetic Variations in Rice with CRISPR/Cas9-Guided Hyperactive hAID Mutant. *Mol Plant* **11,** 623–626 (2018).

96. Wang, X. *et al.* Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nature Biotechnology* **36,** 946–949 (2018).

97. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36,** 977–982 (2018).

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

99. Molla, K. A. & Yang, Y. CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. *Trends Biotechnol* **37,** 1121–1142 (2019).

100. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov* **19,** 839–859 (2020).

101. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics* **19,** 770–788 (2018).

102. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364,** 289–292 (2019).

103. Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364,** 292–295 (2019).

104. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569,** 433–437 (2019).

105. Molla, K. A., Shih, J. & Yang, Y. Single-nucleotide editing for zebra3 and wsl5 phenotypes in rice using CRISPR/Cas9-mediated adenine base editors. *aBIOTECH* **1,** 106–118 (2020).

106. Hua, K., Tao, X., Yuan, F., Wang, D. & Zhu, J.-K. Precise A·T to G·C Base Editing in the Rice Genome. *Mol Plant* **11,** 627–630 (2018).

107. Li, C. *et al.* Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biology* **19,** 59 (2018).

108. Yan, F. *et al.* Highly Efficient A·T to G·C Base Editing by Cas9n-Guided tRNA Adenosine Deaminase in Rice. *Mol Plant* **11,** 631–634 (2018).

109. Hao, L. *et al.* CRISPR/Cas9-Mediated Adenine Base Editing in Rice Genome. *Rice Science* **26,** 125–128 (2019).

110. Li, J. *et al.* Optimizing plant adenine base editor systems by modifying the transgene selection system. *Plant Biotechnol J* **18,** 1495–1497 (2020).

111. Kang, B.-C. *et al.* Precision genome engineering through adenine base editing in plants. *Nature Plants* **4,** 427–431 (2018).

112. Wang, Z. *et al.* ABE8e with Polycistronic tRNA-gRNA Expression Cassette Sig-Nificantly Improves Adenine Base Editing Efficiency in Nicotiana benthamiana. *Int J Mol Sci* **22,** 5663 (2021).

113. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38,** 883–891 (2020).

114. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol* **37,** 1041–1048 (2019).

115. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* **38,** 892–900 (2020).

116. Yan, D. *et al.* High-efficiency and multiplex adenine base editing in plants using new TadA variants. *Molecular Plant* **14,** 722–731 (2021).

117. Hua, K. *et al.* Simplified adenine base editors improve adenine base editing efficiency in rice. *Plant Biotechnol J* **18,** 770–778 (2020).

118. Wei, C. *et al.* Efficient generation of homozygous substitutions in rice in one generation utilizing an rABE8e base editor. *J Integr Plant Biol* **63,** 1595–1599 (2021).

119. Li, J. *et al.* Genome editing mediated by SpCas9 variants with broad non-canonical PAM compatibility in plants. *Mol Plant* **14,** 352–360 (2021).

120. Xu, Z. *et al.* SpRY greatly expands the genome editing scope in rice with highly flexible PAM recognition. *Genome Biology* **22,** 6 (2021).

121. Wu, Y. *et al.* Genome-wide analyses of PAM-relaxed Cas9 genome editors reveal substantial off-target effects by ABE8e in rice. *Plant Biotechnol J* **20,** 1670–1682 (2022).

122. Li, S., Liu, L., Sun, W., Zhou, X. & Zhou, H. A large-scale genome and transcriptome sequencing analysis reveals the mutation landscapes induced by high-activity adenine base editors in plants. *Genome Biology* **23,** 51 (2022).

123. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap Endonuclease 1: A Central Component of DNA Metabolism. *Annual Review of Biochemistry* **73,** 589–615 (2004).

124. Keijzers, G., Bohr, V. A. & Rasmussen, L. J. Human exonuclease 1 (EXO1) activity characterization and its function on flap structures. *Biosci Rep* **35,** (2015).

125. Tang, X. *et al.* Plant Prime Editors Enable Precise Gene Editing in Rice Cells. *Mol Plant* **13,** 667–670 (2020).

126. Lin, Q. *et al.* Prime genome editing in rice and wheat. *Nature Biotechnology* **38,** 582–585 (2020).

127. Xu, R. *et al.* Development of Plant Prime-Editing Systems for Precise Genome Editing. *Plant Communications* **1,** 100043 (2020).

128. Butt, H. *et al.* Engineering herbicide resistance via prime editing in rice. *Plant Biotechnology Journal* **18,** 2370–2372 (2020).

129. Li, H., Li, J., Chen, J., Yan, L. & Xia, L. Precise Modifications of Both Exogenous and Endogenous Genes in Rice by Prime Editing. *Mol Plant* **13,** 671–674 (2020).

130. Xu, W. *et al.* Versatile Nucleotides Substitution in Plant Using an Improved Prime Editing System. *Mol Plant* **13,** 675–678 (2020).

131. Wang, L. *et al.* Spelling Changes and Fluorescent Tagging With Prime Editing Vectors for Plants. *Frontiers in Genome Editing* **3,** (2021).

132. Xu, R., Liu, X., Li, J., Qin, R. & Wei, P. Identification of herbicide resistance OsACC1 mutations via in planta prime-editing-library screening in rice. *Nat. Plants* **7,** 888–892 (2021).

133. Jiang, Y.-Y. *et al.* Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biology* **21,** 257 (2020).

134. Perroud, P.-F. *et al.* Prime Editing in the model plant Physcomitrium patens and its potential in the tetraploid potato. *Plant Science* **316,** 111162 (2022).

135. Lu, Y. *et al.* Precise genome modification in tomato using an improved prime editing system. *Plant Biotechnol J* **19,** 415–417 (2021).

136. Veillet, F. *et al. Prime editing is achievable in the tetraploid potato, but needs improvement.* (Plant Biology, 2020). doi:10.1101/2020.06.18.159111

137. Jin, S. *et al.* Genome-wide specificity of prime editors in plants. *Nat Biotechnol* **39,** 1292–1299 (2021).

138. Lin, Q. *et al.* High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat Biotechnol* **39,** 923–927 (2021).

139. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol* **40,** 402–410 (2022).

140. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184,** 5635-5652.e29 (2021).

141. Zhong, Z. *et al.* Improving Plant Genome Editing with High-Fidelity xCas9 and Non-canonical PAM-Targeting Cas9-NG. *Molecular Plant* **12,** 1027–1036 (2019).

142. Sretenovic, S. *et al.* Expanding plant genome-editing scope by an engineered iSpyMacCas9 system that targets A-rich PAM sequences. *Plant Communications* **2,** 100101 (2021).

143. Ren, Q. *et al.* Improved plant cytosine base editors with high editing activity, purity, and specificity. *Plant Biotechnology Journal* **19,** 2052–2068 (2021).

144. Ren, Q. *et al.* PAM-less plant genome editing using a CRISPR–SpRY toolbox. *Nat. Plants* **7,** 25–33 (2021).

145. Tang, X. *et al.* A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat Plants* **3,** 17018 (2017).

146. Sretenovic, S., Pan, C., Tang, X., Zhang, Y. & Qi, Y. in *Rice Genome Engineering and Gene Editing: Methods and Protocols* (eds. Bandyopadhyay, A. & Thilmony, R.) 95–113 (Springer US, 2021). doi:10.1007/978-1-0716-1068-8_7

147. Sretenovic, S., Pan, C. & Qi, Y. in *CRISPR-Cas Methods: Volume 2* (eds. Islam, M. T. & Molla, K. A.) 83–101 (Springer US, 2021). doi:10.1007/978-1-0716-1657-4_7

148. Chen, H., Nelson, R. S. & Sherwood, J. L. Enhanced recovery of transformants of Agrobacterium tumefaciens after freeze-thaw transformation and drug selection. *Biotechniques* **16,** 664–668, 670 (1994).

149. Murray, M. G. & Thompson, W. F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8,** 4321–4325 (1980).

150. You, Q. *et al.* CRISPRMatch: An Automatic Calculation and Visualization Tool for High-throughput CRISPR Genome-editing Data Analysis. *Int J Biol Sci* **14,** 858–862 (2018).

151. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33,** 286–288 (2017).

152. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37,** 224–226 (2019).

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

154. Liu, W. *et al.* DSDecode: A Web-Based Tool for Decoding of Sequencing Chromatograms for Genotyping of Targeted Mutations. *Molecular Plant* **8,** 1431–1433 (2015).

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

156. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

157. Kawahara, Y. *et al.* Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6,** 4 (2013).

158. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

159. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20,** 1297–1303 (2010).

160. Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* **40,** 11189–11201 (2012).

161. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. 861054 Preprint at https://doi.org/10.1101/861054 (2019)

162. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22,** 568–576 (2012).

163. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25,** 2865–2871 (2009).

164. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

165. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).

166. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29,** 24–26 (2011).

167. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19,** 1639–1645 (2009).

168. Guo, M. *et al.* Structural insights into a high fidelity variant of SpCas9. *Cell Res* **29,** 183–192 (2019).

169. Endo, M. *et al.* Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. *Nature Plants* (2018). doi:10.1038/s41477-018-0321-8

170. Wang, J. *et al.* xCas9 expands the scope of genome editing with reduced efficiency in rice. *Plant Biotechnology Journal* **17,** 709–711 (2019).

171. Liu, Z. *et al.* Efficient base editing with expanded targeting scope using an engineered Spy-mac Cas9 variant. *Cell Discov* **5,** 58 (2019).

172. Zhou, J. *et al.* CRISPR-Cas9 Based Genome Editing Reveals New Insights into MicroRNA Function and Regulation in Rice. *Frontiers in Plant Science* **8,** (2017).

173. Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171,** 470-480.e8 (2017).

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

175. Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3,** (2017).

176. Wang, L. *et al.* Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Research* **27,** 1289–1292 (2017).

177. Molla, K. A., Sretenovic, S., Bansal, K. C. & Qi, Y. Precise plant genome editing using base editors and prime editors. *Nature Plants* **7,** 1166–1187 (2021).

178. Qin, R. *et al.* SpCas9-NG self-targets the sgRNA sequence in plant genome editing. *Nat. Plants* **6,** 197–201 (2020).

179. Sretenovic, S. & Qi, Y. Plant prime editing goes prime. *Nat. Plants* **8,** 20–22 (2022).

180. Nazarov, P. A., Baleev, D. N., Ivanova, M. I., Sokolova, L. M. & Karakozova, M. V. Infectious Plant Diseases: Etiology, Current Status, Problems and Prospects in Plant Protection. *Acta Naturae* **12,** 46–59 (2020).