

METHODOLOGY

Open Access



Engineering a bacterial toxin deaminase from the DYW-family into a novel cytosine base editor for plants and mammalian cells

Dingbo Zhang^{1,2}, Fiona Parth¹, Laura Matos da Silva¹, Teng-Cheong Ha^{3,4}, Axel Schambach^{3,4,5} and Jens Boch^{1*}

*Correspondence:
jens.boch@genetik.uni-hannover.de

¹ Institute of Plant Genetics, Leibniz Universität Hannover, Herrenhäuser Str. 2, Hannover 30419, Germany

² Research Institute of Biology and Agriculture, University of Science and Technology, Beijing 100083, China

³ Institute of Experimental Hematology, Hannover Medical School, Hannover, Germany

⁴ REBIRTH - Research Center for Translational Regenerative Medicine, Hannover Medical School, Hannover, Germany

⁵ Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

Abstract

Base editors are precise editing tools that employ deaminases to modify target DNA bases. The DYW-family of cytosine deaminases is structurally and phylogenetically distinct and might be harnessed for genome editing tools. We report a novel CRISPR/Cas9-cytosine base editor using SsdA, a DYW-like deaminase and bacterial toxin. A G103S mutation in SsdA enhances C-to-T editing efficiency while reducing its toxicity. Truncations result in an extraordinarily small enzyme. The SsdA-base editor efficiently converts C-to-T in rice and barley protoplasts and induces mutations in rice plants and mammalian cells. The engineered SsdA is a highly efficient genome editing tool.

Keywords: Crop improvement, Gene therapy, Base editing, CRISPR

Background

The discovery of site-specific and programmable DNA binding domains has revolutionized the targeted manipulation of genomes in many different species. First, site-specific nucleases like CRISPR/Cas9, TALEN (transcription activator-like effector nucleases), and ZFN (zinc-finger nucleases) were applied to generate deletions or random mutations at target sites mainly to inactivate specific genes [1–3]. Based on these initial achievements, researchers worldwide endeavored to develop more precise tools that allow for specific nucleotide exchanges to correct human diseases or to generate allelic variants for important crop traits. Targeted base editing is an important development in this direction that resulted in the first CRISPR-based pharmaceuticals approved for clinical treatments in humans to correct single-nucleotide polymorphisms (SNPs) [4–7].

Base editors are typically fusions of Cas9-nickase (nCas9, D10A) as the targeting domain for a specific DNA sequence and an enzymatic domain to deaminate target DNA bases. Based on the hydro-deamination chemistry executed by deaminase enzymes, two main types of base editors: cytosine base editors (CBEs) and adenine



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

base editors (ABEs) have been developed [8, 9]. For CBEs, binding of the CRISPR/Cas complex to a target locus enables the deamination of cytosine to uracil in the displaced single-stranded DNA (the R-loop) [8]. To prevent immediate uracil removal from DNA, an uracil glycosylase inhibitor (UGI) domain is typically fused to the CBE [8, 10]. During replication and repair, the uracil is base-pairing with adenine in the second strand and subsequently replaced by thymine resulting in C-to-T transitions. To trigger transversions (C-to-G) instead of transitions, the base excision repair of uracil can be stimulated by fusing an uracil glycosylase instead of UGI to the nCas9 resulting in abasic sites which can cause random base exchanges [11–13].

In mammalian cells, first APOBEC1 from rat has been used, and more recently, the TadA8e variant of the *E. coli* adenine deaminase TadA has been engineered to alter the substrate preference from deoxyadenosine to deoxycytidine for cytosine base editing [14–16]. In contrast, in plants, APOBEC1 shows particularly low activity [17] and has been replaced by human APOBEC3A (hA3A) [18]. In addition, two other cytosine deaminases, i.e. hAID from human and PmCDA1 from sea lamprey, have been used in plants with variable efficiencies [19–21].

Interestingly, suitable deaminase domains can not only be found in higher eukaryotes but also in bacteria. Recently, a bacterial toxin secreted by *Burkholderia cenocepacia* has been described as a double-strand DNA (dsDNA)-specific cytosine deaminase (DddA). This novel CBE domain allowed the development of TALE-based CBEs [22, 23]. In contrast to the CRISPR/Cas complex, TALEs do not unwind the DNA upon binding to target sites and only produce suitable substrates for dsDNA-specific deaminases as DddA. In addition, TALE-base editors can be imported into cellular organelles using an N-terminal targeting signal, resulting in efficient chloroplast and mitochondrial base editing in plant and mammalian cells, respectively [24–26]. To circumvent DddA toxicity when used in a genome editing tool, DddA has been split into two halves that only complement each other to a functional enzyme upon binding of two matching TALE-base editors with one half of DddA each at a target locus [22, 23, 27]. Non-toxic DddA variants have also been developed to be used with only one TALE [27]. One caveat of DddA is its target requirement for TC motifs [22, 28]. This constraint has been partially relieved through protein evolution into variants with HC (H = A, C, or T) specificity [23]. Applying bioinformatic sequence homology searches, different members of the DddA-family have been identified as CBEs, including ssDNA and dsDNA-specific cytosine deaminases with a broader target specificity [29–31]. Such studies indicate the ongoing need to develop novel base editing domains.

Recently, another bacterial cytosine deaminase, SsdA (single-strand DNA deaminase toxin A), from plant-pathogenic *Pseudomonas syringae* has been described, which is also a type VI-secreted protein that is highly toxic upon expression in bacterial cells (Additional file 1: Fig. S1) [32]. In *P. syringae*, the inhibitor protein, SsdA_i binds to SsdA to prevent self-toxicity. SsdA is predominantly active on ssDNA, but has residual dsDNA deaminating activity in vitro and the precise conformations of the deaminase domain SsdA_{tox} in complex with the inhibitor SsdA_i has been solved by crystallography [32]. Interestingly, SsdA_{tox} showed a target specificity for Cs with relaxed preference for

neighboring pyrimidines, making this a very promising candidate to develop it into a novel genome editing tool.

Here, we have applied SsdA_{tox} as ssDNA-specific deaminase fusion to nCas9-UGI to constitute a novel CBE tool for genome editing. When used for transient expression in *Nicotiana benthamiana*, we experienced toxicity in *A. tumefaciens* even though the used 35S promoter should only have residual activity in bacteria. Interestingly, we were able to isolate the spontaneous amino exchange mutant SsdA^{G103S}, which shows significantly reduced toxicity, but enhanced target deamination. Our study demonstrates that this member of a novel deaminase family can be engineered into a highly efficient CBE tool to generate allelic variants in rice plants. Such enhanced genome editing tools are pivotal for crop improvement to address the challenge of feeding an expanding global population in the looming global warming crisis.

Results

Isolation of less toxic SsdA_{tox} variants

SsdA is a bacterial toxin from *P. syringae* with cytosine deaminating activity [32]. Unlike the cytosine deaminases that have been used in CBEs, SsdA is classified into the DYW-family (Fig. 1a) which is a promising class of enzymes to develop genome editing tools. To explore the potential use of SsdA in a CBE, we cloned the catalytic domain of SsdA (SsdA_{tox}) from *Pseudomonas syringae* pv. *aptata* and fused it to nCas9-UGI using our genome editing MoClo kit (Fig. 1b; Additional file 1: Fig. S2). We assembled the SsdA_{tox}-CBE under control of the 2 × 35S promoter into a level-M binary vector to test its base editing-activity in a GUS reporter system in plants [33]. We noticed that there was a significant (approx. 100x) decrease in colony-forming units (CFU) following transformation of the SsdA_{tox}-CBE construct into *A. tumefaciens* in comparison to a hA3A-CBE construct (Fig. 1c). This difference in transformation efficiency was not visible in *E. coli* (Fig. 1d). We speculate that the 2 × 35S promoter has a low background activity in *A. tumefaciens*, but not *E. coli*, producing the SsdA_{tox}-CBE protein which is then toxic to the cells.

To understand why some transformants survive this possible toxic effect, we randomly selected four surviving *A. tumefaciens* single colonies and amplified the SsdA_{tox}-encoding region. Sanger sequencing showed C-to-T transitions within the SsdA_{tox} domain which led to amino acid changes of D77N, S87F, G103S, and Q113stop, respectively (Fig. 1e). To validate these variants, we introduced D77N and G103S into wild-type SsdA_{tox} by site-directed mutagenesis, and assembled SsdA^{D77N}-CBE and SsdA^{G103S}-CBE, respectively, in level-M binary vectors. We found that both of them yielded comparable CFU to the hA3A-CBE following transformation in *E. coli* and *A. tumefaciens* (Fig. 1c, d), indicating that the toxicity of SsdA_{tox} is abolished in these variants. Previous studies have shown that SsdA_{tox} introduced high levels of random C-to-T editing in the bacterial genome when heterologously expressed in *E. coli* [32]. We could show that the toxicity of SsdA_{tox} is reduced by specific amino acid changes that possibly occur via spontaneous endogenous C-to-T editing by SsdA_{tox} itself.

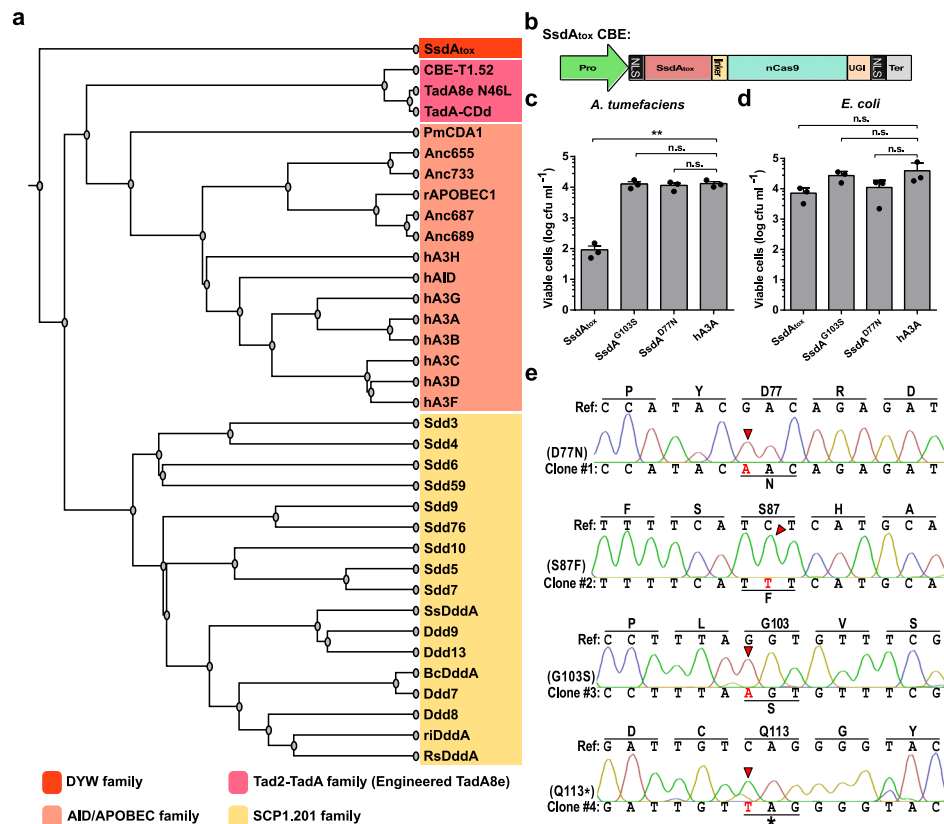


Fig. 1 SsdA_{tox}-CBEs exhibit mutagenic activity. **a** Phylogenetic tree analysis of ssDNA and dsDNA cytidine deaminases used in cytosine base editors (CBEs). **b** Architecture of the SsdA_{tox} CBE system. Pro: promoter, NLS: nuclear localization sequence, nCas9: SpCas9 D10A nickase, UGI: uracil glycosylase inhibitor, Ter: terminator. **c** Viability in colony-forming units of *A. tumefaciens* GV3101 strains after transformation of binary vectors containing SsdA_{tox}-CBE or hA3A-CBE. **d** Viability in colony-forming units of *E. coli* Top10 strains after transformation of binary vectors containing SsdA_{tox}-CBE or hA3A-CBE. Values and error bars indicate the mean \pm SEM, $n = 3$ independent experiments. ** $P < 0.01$; n.s. (not significant) using Student's two-tailed unpaired *t*-test. **e** Sanger sequencing of SsdA_{tox} coding region. Mismatches are highlighted in red and are indicated by red triangles. Encoded amino acids are indicated in single letter code above the chromatograms for the reference sequence and below the chromatograms for the mutant variant

Introduction of an intron into SsdA_{tox} reduces toxicity in *A. tumefaciens*

To reduce the toxicity of SsdA in bacteria, we designed two different variants (SsdA_v1 and SsdA_v2) with an intron inserted into the coding sequence of SsdA_{tox} (Additional file 1: Fig. S2). The intron terminates translation in the absence of RNA splicing in bacteria, but allows correct translation after splicing in eukaryotes. To quantify base editing, we used a GUS-assay in *N. benthamiana*. The assay is based on an inactive GUS^{G537} allele with a missense mutation of glutamic acid (GAA) to glycine (GGA) in one of the catalytic residues. A C-to-T (G-to-A on the opposite DNA strand) conversion can revert the glycine residue to glutamic acid and restore GUS enzymatic activity [33] (Fig. 2a). We mixed the *A. tumefaciens* strains carrying SsdA_v1-CBE or SsdA_v2-CBE with a strain carrying the GUS^{G537} reporter and infiltrated *N. benthamiana* leaves. Quantification of GUS activity showed that both base editors result in base editing with SsdA_v1 being more efficient than SsdA_v2 (Additional File 1: Fig. S3). This demonstrates that SsdA can be used as the catalytic domain in a novel CBE tool.

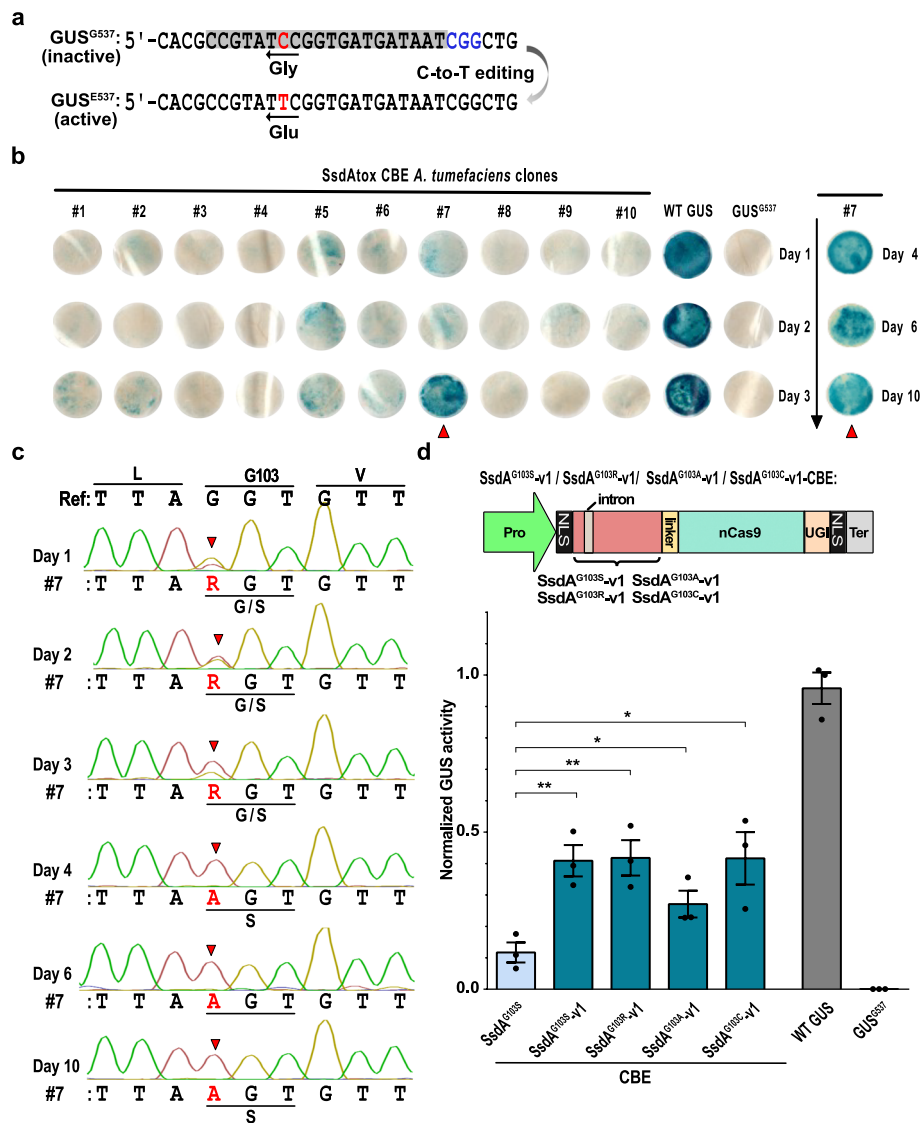


Fig. 2 SsdA_{tox}-CBE variants enable C-to-T editing of a GUS^{G537} reporter in *N. benthamiana*. **a** Schematic of the GUS^{G537} cytosine base editing reporter. The C-to-T (highlighted in red) editing in GUS^{G537} can alter the glycine codon (GGA) to a glutamic acid codon (GAA) and restore GUS activity. The protospacer is shown with gray background, and the PAM is in blue. **b** Ten different *A. tumefaciens* transformants were co-inoculated together with the GUS^{G537} reporter into *N. benthamiana* leaves. Leaf disks were harvested 2 dpi then stained in GUS staining solution and de-stained in ethanol. Blue color indicates restored GUS activity. Dark blue leaf disk from clone #7 are marked by triangles. WT GUS: wild-type GUS, positive control. GUS^{G537}: negative control. **c** Sanger sequencing results from *A. tumefaciens* clone #7 between days 1 and 10. Mismatches to the wild-type sequence are highlighted in red and indicated by triangles. **d** Top: Architectures of CBEs containing SsdA^{G103S-v1}, SsdA^{G103R-v1}, SsdA^{G103A-v1}, or SsdA^{G103C-v1}. An intron was introduced into the coding sequences of these SsdA variants to prohibit translation in bacteria. Bottom: C-to-T editing efficiencies of SsdA-CBE variants of the GUS^{G537} reporter. GUS activities were measured and normalized to 2 × 35S::GUS (WT GUS, positive control). Values and error bars indicate the mean ± SEM, $n = 3$. * $P < 0.05$; ** $P < 0.01$ using Student's two-tailed unpaired t -test

An engineered SsdA-CBE induces efficient C-to-T editing

To investigate whether the SsdA_{tox} variants with reduced toxicity can also function as catalytic domains in cytosine deaminases, we randomly selected 10 *A. tumefaciens* single colonies after transformation with the original SsdA_{tox}-CBE binary vector. We mixed the

10 *A. tumefaciens* transformants containing SsdA_{tox}-CBE individually with a strain containing the GUS^{G537} reporter, and infiltrated the mixtures into *N. benthamiana* leaves. After 48 h, leaf discs were harvested and stained for GUS activity. All 10 clones exhibited very low overall GUS activity, but some of the leaf areas showed a darker blue color suggesting that the infiltrated bacteria were a mixture of mutant variants with and without base editing activity, respectively. The 10 *A. tumefaciens* strains were further sub-cultured, and infiltrations were repeated for three consecutive days with the same *A. tumefaciens* clones. Leaf staining showed that the GUS activity of clone #7 increased from day 1 to day 3 (Fig. 2b). Sanger sequencing of clone #7 revealed a stepwise accumulation of the SsdA^{G103} variant over the wild type version (Fig. 2c).

We further sub-cultured clone #7 over 10 days and used *A. tumefaciens* suspensions from days 4, 6, and 10 for leaf infiltrations. Leaf staining showed stable GUS activity, and sequencing results showed a pure amino acid mutation of G103S in the SsdA_{tox} domain (Fig. 2b, c). These results suggest that the amino acid G103 of SsdA_{tox} plays a crucial role for the activity of the enzyme. On the one hand, the G103S mutation results in a low-toxic protein variant, and on the other hand, this amino acid change results in a highly active SsdA^{G103S}-CBE.

To further investigate the relevance of this specific amino acid change of SsdA-CBE, we introduced four different types of amino acid substitutions at the G103 position (G103S, G103R, G103A, and G103C) in SsdA_{v1} containing an intron in the coding sequence. These variants are named SsdA^{G103S}-v1, SsdA^{G103R}-v1, SsdA^{G103A}-v1, and SsdA^{G103C}-v1, respectively (Fig. 2d). We then tested the SsdA-CBE variants in *N. benthamiana* using the GUS^{G537} reporter. The GUS activity was normalized to a constitutive 2 × 35S::GUS construct. SsdA^{G103S}-v1-CBE had approximately a fourfold increase in GUS activity compared to SsdA^{G103S} without intron.

Moreover, SsdA^{G103S}-v1-CBE, SsdA^{G103R}-v1-CBE, and SsdA^{G103C}-v1-CBE exhibited ~ 40% normalized GUS activity, while SsdA^{G103A}-v1-CBE displayed ~ 26% normalized GUS activity (Fig. 2d). The highly active SsdA^{G103S}-v1-CBE was chosen for further analyses. Taken together, SsdA_{tox} can be engineered into an efficient cytosine base editing tool.

SsdA^{G103S} is a ssDNA-specific deaminase

In previous in vitro studies, purified SsdA_{tox} protein efficiently deaminated cytosines in all four sequence contexts (AC, CC, TC, and GC) using ssDNA as substrate. However, it still exhibited some residual catalytic activity toward dsDNA [32]. To investigate if SsdA can deaminate cytosine in dsDNA in vivo, we fused SsdA_{tox}-v1 or SsdA^{G103S}-v1 (both containing an intron within the coding sequence) to a single TALE (transcription activator-like effector) array protein (Fig. 3a). Unlike the SpCas9 nickase, the TALE array guides the deaminase to a designated dsDNA sequence without unwinding and nicking the dsDNA. Four different TALE arrays targeting the GUS^{G537} reporter were used separately. This allowed the target cytosine to be positioned at C6 or C12 downstream of the TALE-binding DNA strand (TALE1 or TALE2), or at C5 or C8 (TALE3 or TALE4) downstream of the opposite strand (Fig. 3b). When testing these fusions using the GUS^{G537} reporter in *N. benthamiana*, all the TALE-SsdA_{tox}-v1 and TALE-SsdA^{G103S}-v1 fusions

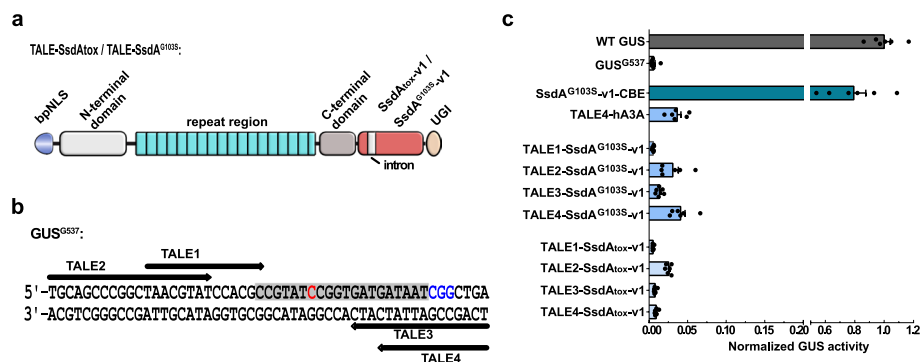


Fig. 3 SsdA^{G103S} cannot use dsDNA as substrate. **a** Architectures of TALE-SsdA_{tox} or TALE-SsdA^{G103S} fusions. bpNLS: bipartite nuclear localization sequence; UGI: uracil glycosylase inhibitor. **b** Binding sites of four TALE arrays to the GUS^{G537} reporter. TALE binding sites are indicated by arrows in N- to C-terminal orientation. A Cas9 protospacer is in gray background, and the PAM is in blue. The target C is in red. **c** C-to-T editing efficiencies of the TALE-SsdA_{tox} and TALE-SsdA^{G103S} fusions using the GUS^{G537} reporter in *N. benthamiana*. SsdA^{G103S}-v1-CBE (with nCas9-UGI): positive control. GUS activities were measured and normalized to 2 × 35S::GUS (WT GUS). GUS^{G537}: reporter alone (negative control). Values and error bars indicate the mean ± SEM, n = 6

showed background GUS activity compared to a TALE-hA3A fusion. hA3A is a ssDNA-specific deaminase and was used as a negative control. In contrast, the SsdA^{G103S}-v1 fusion to nCas9-UGI (positive control) shows approximately 80% GUS activity (Fig. 3c). These results strongly suggest that the original SsdA_{tox} as well as SsdA^{G103S} are ssDNA-specific cytosine deaminases.

Rational truncation of SsdA^{G103S}-v1 to generate a very small deaminase

Crystal structure analysis revealed that SsdA contains the fundamental features of deaminase enzymes, including histidine and cysteine residues near the active site to coordinate a zinc ion, as well as three α -helices and five β -strands that make up the core fold of the enzymes in the deaminase superfamily [32].

We wondered whether we could rationally truncate the SsdA protein to shorten its size. We generated five differently truncated SsdA^{G103S} variants (named SsdA^{G103S}-T1 to SsdA^{G103S}-T5) with truncation at the N-terminal or/and C-terminal end of SsdA^{G103S} (Fig. 4a and Additional File 1: Fig. S2).

We tested these variants as CBE fusions and quantified C-to-T editing efficiency in the GUS^{G537} reporter in *N. benthamiana*. Compared to the full-length SsdA^{G103S} protein (SsdA^{G103S}-v1), SsdA^{G103S}-T1, SsdA^{G103S}-T2, and SsdA^{G103S}-T3 with truncation at the N-terminal end show comparable GUS activities, while SsdA^{G103S}-T4 and SsdA^{G103S}-T5 with truncation at the C- and C/N-terminal end, respectively, result only in background GUS activity (Fig. 4b). These results indicate that the β -strand located at the C-terminal end of the SsdA^{G103S} protein is crucial for the enzyme activity, and that the two α -helical and one β -strand at the N-terminal region can be truncated without reducing the enzymatic activity. Taken together, we could successfully reduce the protein size of SsdA^{G103S} to a length of only 114 amino acids.

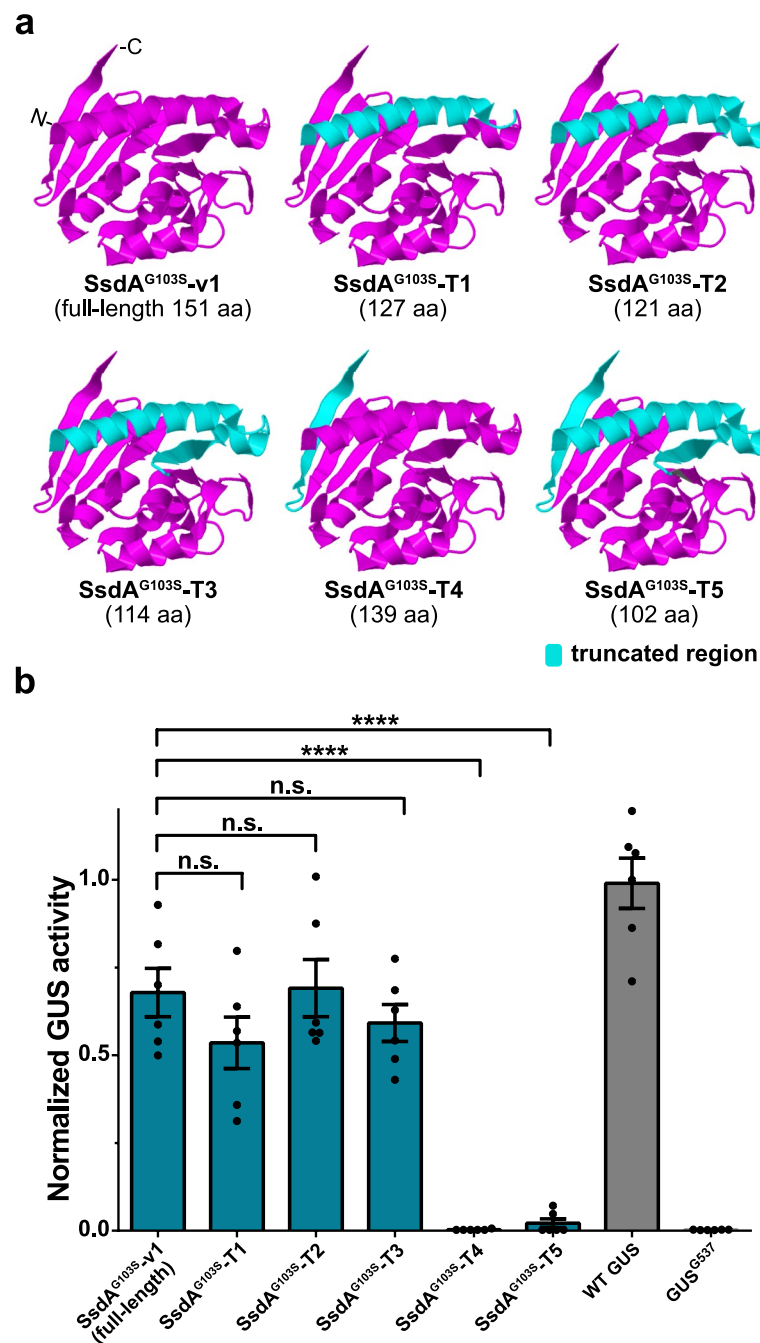


Fig. 4 Editing activities of truncated SsdA^{G103S} variants. **a** Engineering truncations of the SsdA^{G103S} protein. Cryo-EM structure of SsdA_{tox} (PDB: 7JTU). The truncated sequences are in cyan. The final length of truncated SsdA^{G103S} proteins is listed. **b** C-to-T editing efficiencies of the truncated SsdA^{G103S} variants using the GUS^{G537} reporter in *N. benthamiana*. GUS activities were measured and normalized to 2 × 35S::GUS (WT GUS). GUS^{G537} reporter alone (negative control). Values and error bars indicate the mean ± SEM, *n* = 6. **** *P* < 0.0001; n.s. (not significant) using Student's two-tailed unpaired *t*-test

SsdA^{G103S}-v1-CBE exhibits a narrow editing window in rice and barley protoplasts

To investigate the editing window of SsdA^{G103S}-v1-CBE, we targeted five genomic loci in rice (*OsALS*, *OsPDS*, *OsFBX109*, *Os01g40290*, and *Os11g26790*) and two genomic loci in barley (*HvSTP13* and *HvFLS2A*) using protoplasts. These sites contain possible target

cytosines at different positions of the protospacer. Amplicon deep sequencing results revealed that all five rice target sites were successfully edited by SsdA^{G103S}-v1-CBE with varying efficiencies depending on the individual target site. In comparison to a hA3A-CBE, it had a somewhat lower activity, but also a narrower activity window at the PAM-distal protospacer (Fig. 5a–e). Specifically, the SsdA^{G103S}-v1-CBE showed comparable C-to-T editing efficiency at C5 and C6 at the *OsFBX109* target site, as well as C7 at the *Os11g26790* target site when compared to the hA3A-CBE. In contrast, the SsdA_{tox}-v1-CBE (wild-type SsdA_{tox} with intron in the coding sequence) showed very low to no activity at these five rice target sites, demonstrating that the G103S amino acid change particularly enhances activity. Besides from rice, SsdA^{G103S}-v1-CBE exhibited efficient editing at the two barley target sites (Fig. 5f, g). Overall, our protoplast results show that the SsdA^{G103S}-v1-CBE can efficiently convert C-to-T in plant genomes, with optimal target Cs ranging from C5 to C8 across the protospacer sequences.

Efficient base editing with SsdA^{G103S}-v1 in rice plants

To investigate the use of the SsdA^{G103S}-v1-CBE for base editing in plants, we targeted five rice genomic loci via *Agrobacterium*-mediated transformation of rice calli. These were *OsPDS*, *Os11g26790*, and *OsFBX109*, as well as two different sites within the *OsALS* gene (*OsALS*-T1, *OsALS*-T2). Genotyping of regenerated T0 plants showed that SsdA^{G103S}-v1-CBE could effectively induce C-to-T conversions at all of these target sites with an editing efficiency ranging from 30 to 58.8% (Fig. 6a and Additional file 2: Table S1). In comparison to the SsdA^{G103S}-v1-CBE, we used the hA3A-CBE targeting the *OsALS*-T1 site. The hA3A-CBE achieved a precise C-to-T editing rate of 30.8% (8/26 plants) at the *OsALS*-T1 site, whereas the SsdA^{G103S}-v1-CBE exhibited a higher editing efficiency of 58.8% (10/17 plants). Consistent with the protoplast results, SsdA^{G103S}-v1-CBE showed a narrow editing window in these edited plants (Fig. 6b). Editing byproducts like indels and C-to-G changes were generated by both CBEs at these target sites (Fig. 6a). To investigate the heritability of the mutations that were generated

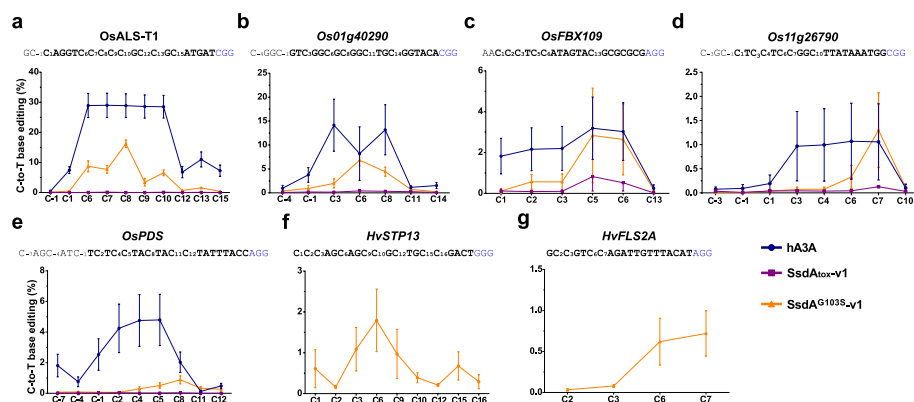
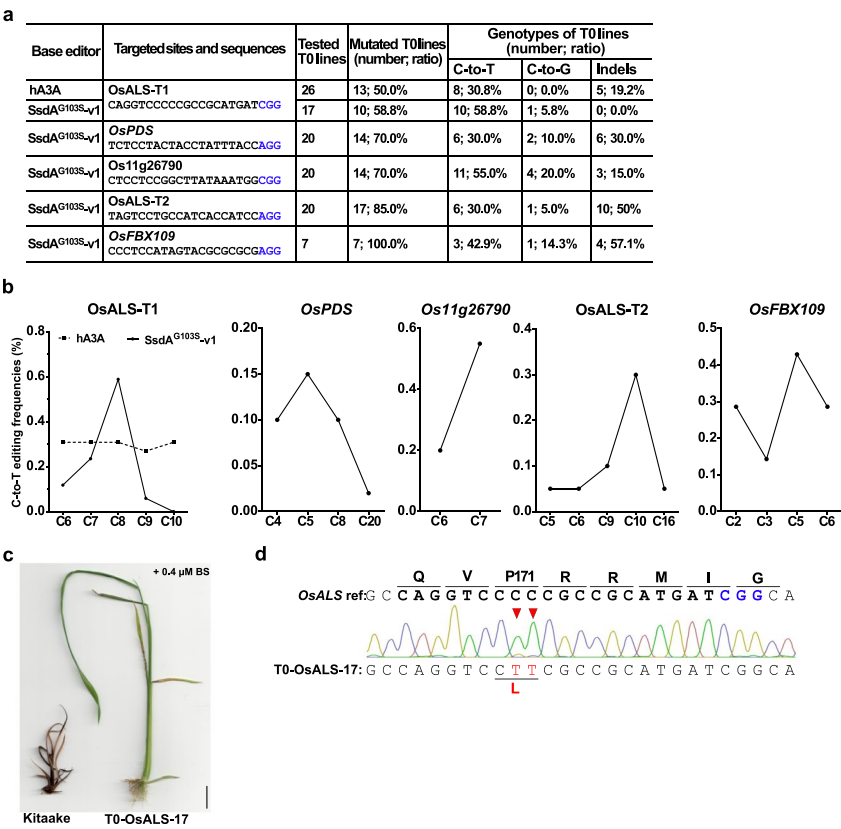


Fig. 5 SsdA^{G103S}-v1-CBE editing of genomic loci in rice and barley protoplasts. **a–e** Comparison of C-to-T editing frequencies of SsdA^{G103S}-v1-CBE, SsdA_{tox}-v1-CBE, and hA3A-CBE at four genomic rice loci (**a–e**). **f, g** C-to-T editing frequencies of SsdA^{G103S}-v1-CBE at two barley loci. The protospacer sequence is shown in bold with putative target C labelled with numbers indicating their position in the protospacer. The PAM is in blue. Error bars indicate the mean \pm SEM, $n = 3$



by SsdA^{G103S}-v1-CBE, a total of 32 individual T1 seedlings from two different T0 lines were subjected to genotyping. Sanger sequencing results showed stable inheritance of C-to-T editing in these T1 lines (Additional file 2: Table S2). Furthermore, 11 transgene-free edited rice plants were identified among these T1 lines.

The OsALS-T1 site is positioned within the rice acetolactate synthase gene (*OsALS*), specifically the region encoding the proline 171 (P171). *OsALS* confers resistance to the herbicide bispyribac-sodium (BS) when P171 is substituted with phenylalanine (F) or leucine (L) after editing of the coding sequence by CBEs [34]. To examine the BS-resistance of rice edited by SsdA^{G103S}-v1-CBE, the T0 line T0-OsALS-17 was incubated on MS medium containing 0.4 μM BS (Fig. 6c). This line harbors a P171L substitution at the OsALS-T1 locus which is caused by two cytosines having been converted to thymine (Fig. 6d). Compared to wild-type Kitaake rice plants, T0-OsALS-17 exhibited a strong tolerance to BS (Fig. 6c). Taken together, we conclude that SsdA^{G103S}-v1-CBEs can efficiently introduce C-to-T editing in rice and generate herbicide-tolerant rice lines.

To assess sgRNA-dependent off-target editing in the SsdA^{G103S}-v1-CBE-edited rice plants, we analyzed potential off-target sites that are predicted by CRISPR-P [35]. For the OsALS-T1 site, four predicted off-target sites with 1–4 mismatches

Table 1 Analyzing potential Cas-dependent off-target editing of hA3A-CBE and SsdA^{G103S}-v1-CBE in T0 rice plants

	Sequence ^a	Editor used	Mutation type	Frequency (edited plant / tested plant)
On-target	<i>OsALS</i> (OsALS-T1) CAGGTCCCCGCGCATGAT CGG	-	-	-
Off-target site 1	<i>Os04g106500</i> CAGGTCCCCGCGCATGAT CGG	hA3A	Indels	100% (2/2)
		SsdA ^{G103S} -v1	Indels & C-to-T	100% (3/3)
Off-target site 2	<i>Os04g106900</i> CAGGCCCCGCGCGCATGAT CGG	hA3A	C-to-T	100% (2/2)
		SsdA ^{G103S} -v1	C-to-T	66.7% (2/3)
Off-target site 3	<i>Os05g247500</i> CCGGTTCGCCGCTCATGAT TGG	hA3A	WT	0% (2/2)
		SsdA ^{G103S} -v1	WT	0% (2/2)
Off-target site 4	<i>Os08g020200</i> CAGATCCGCCGCGCACGCT CGG	hA3A	WT	0% (2/2)
		SsdA ^{G103S} -v1	WT	0% (2/2)

^a : Mismatched sequences in the protospacer are in italics, and the PAMs are in bold

to the on-target sequence were examined (Table 1). Sequencing results from three SsdA^{G103S}-v1-CBE-edited T0 plants detected off-target editing in three lines at the off-target site 1 (1 mismatch), and two out of three lines had off-target editing at the off-target site 2 (2 mismatches). No off-target editing was identified at the off-target sites 3 and 4 in two plants, respectively. At the same OsALS-T1 site, we also analyzed two hA3A-CBE-edited T0 plants and found 100% (2/2) off-target editing at both off-target site 1 and site 2, but no editing at the other two predicted off-target sites. We further tested the top predicted off-target sites at the *OsPDS*, *Os11g26790*, and OsALS-T2 target sites. The sequencing results showed that no off-target editing occurred (Additional file 2: Table S3). These results indicate that both hA3A- and SsdA^{G103S}-v1-CBEs can induce off-target editing, especially at near identical sgRNA target sites with only 1–2 mismatches.

SsdA^{G103S}-v1 base editing activity in mammalian cells

To determine if SsdA^{G103S}-v1-CBE is active in mammalian cells, we cloned a human codon-optimized SsdA^{G103S}-v1-CBE (HucoSsdA^{G103S}-v1-CBE) into an all-in-one lentiviral vector (LV; Fig. 7a). sgRNAs were selected to target two genes for knockout by generating premature stop codons in the transcripts: IL7RA on exon 2 and CD5 on exons 5 and 6. These sgRNAs were cloned into the HucoSsdA^{G103S}-v1-CBE and BE4max all-in-one LVs for comparison. The LVs were packaged using 293 T cells, and the viral supernatant was used to transduce K562 cells. Genomic DNA was then extracted from both the transfected 293 T cells and the LV-transduced K562 cells.

In both the LV plasmid transfected 293 T cells and LV-transduced K562 cells, we observed 3 to 48% total mutation in the targeted regions, including C-to-T conversions that resulted in de novo stop codons in all targets edited with either CBE (Fig. 7b). We also observed a low level of indels within the target windows generated by both CBEs.

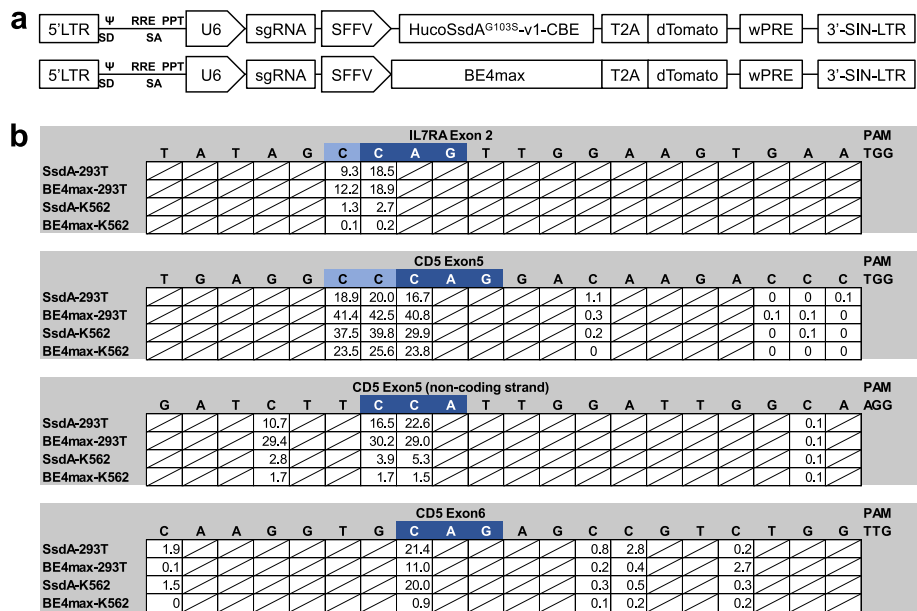


Fig. 7 Lentiviral vector design and SsdA^{G103S}-v1-CBE activity in mammalian cells. **a** The human codon-optimized SsdA^{G103S}-v1-CBE (HucoSsdA^{G103S}-v1-CBE) or BE4max-CBE were cloned into a self-inactivating lentiviral vector (SIN-LV) and expressed under a spleen focus-forming virus (SFFV) promoter. A ribosomal skipping sequence (T2A) was fused to the C-terminus of both CBEs to co-express a dTomato fluorescent protein. The sgRNA is expressed under a human U6 promoter. **b** A summary of both CBEs used in human 293 T and K562 cells, including the targeting sequence, along with the frequency of C-to-T conversions. Sequences in dark blue represent bases where C-to-T conversions lead to de novo stop codon formation. LTR, long terminal repeat; Ψ, packaging element; RRE, Rev response elements; PPT, polypurine tract; SD, splice donor; SA, splice acceptor; wPRE, Woodchuck hepatitis virus posttranscriptional regulatory element

Our findings confirm that the newly developed SsdA^{G103S}-v1-CBE is active in both mammalian and plant cells, highlighting its potential applications in crop improvement and clinical settings.

A structure-based phylogeny of SsdA_{tox} homologs

The common methodology for annotating and analyzing proteins relies on a one-dimensional (1D) amino acid similarity search, which is unable to fully reveal the functional characteristics of proteins [36, 37]. Comparing protein structures using three-dimensional (3D) superposition is more sensitive in identifying distantly related proteins with common structural or enzymatic functions^{31,38}. Here, we used the protein structure search tool Foldseek [38] for mining proteins that might be functionally related to SsdA_{tox} (Fig. 8a). The structure of SsdA_{tox} was aligned to the AlphaFold Protein Structure Database (AFDB50) [39] and the top 50 highest-scored candidates (SsdA_{tox}-Homolog, SH1 to SH50) as well as six manually selected related proteins from plants (SH51 to SH56) were used for phylogenetic clustering (Fig. 8b). We further scanned these candidates for conserved motifs that are a hallmark of the BaDTF2 (SsdA_{tox}) deaminase sub-family and differentiate it from the mRNA-targeting members of the DYW-family [32]. A HAE motif is believed to contain a catalytic glutamate and its histidine likely coordinates together with two cysteines of a CxDC motif a Zn²⁺ ion. In addition, a Ser-Gly-Trp

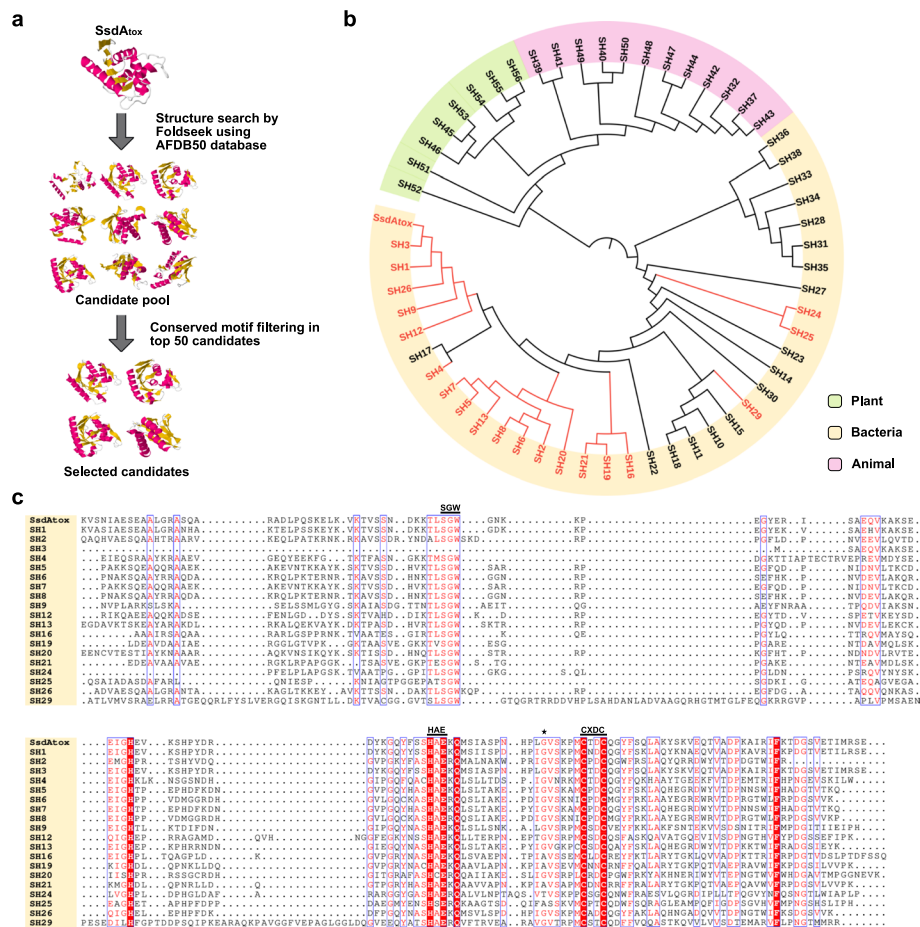


Fig. 8 Identifying structurally related proteins of SsdA_{tox} using Foldseek. **a** Workflow of the SsdA_{tox} structure search based on Foldseek and filtering for conserved motifs. **b** Phylogenetic clustering of the top 50 highest-scored candidates and six manually selected candidates from plants. Eighteen candidates labelled in red contain the conserved motifs from SsdA_{tox}. **c** Sequence alignment of the 18 candidates to SsdA_{tox} with conserved motifs shown above the alignment. Amino acid G103 from SsdA_{tox} is indicated by an asterisk

(SGW) motif lies in a loop region in close proximity to the active site of SsdA_{tox}. All three motifs were identified in 18 of the 50 candidates (Fig. 8c). We speculate that these 18 candidates, which exhibit structural similarity to SsdA_{tox}, may possess cytosine deaminase catalytic activities and could possibly also be engineered into CBEs.

Discussion

Members of the cytosine deaminase superfamily in prokaryotes and eukaryotes catalyze base deamination in a wide variety of contexts [31, 40]. Deaminases that have been incorporated into genome editing tools either use DNA as a substrate (ssDNA [8] or dsDNA [22]) or have been artificially evolved to accept DNA instead of RNA [9]. In mammalian systems, rat APOBEC1 and derivatives are commonly used in CBEs [8], but human APOBEC3A and APOBEC3B are significantly more active [41, 42], which results in a dramatic difference for plants [17, 43].

Here, we have developed a novel CBE, which is highly active in plants. For this, we developed an engineered variant of the bacterial SsdA_{tox} deaminase from *P. syringae*.

SsdA_{tox} is one of the interspecies bacterial toxin system-dependent deaminases and is classified into the DYW family whose evolutionary separation from all other currently used enzymes in CBEs predates the origin of eukaryotes [32, 40]. Plant and fungal enzymes from this family are often fused to PPR or ankyrin repeats and contribute to editing of organellar mRNA transcripts [44, 45]. Accordingly, synthetic PPR-DYW proteins have been used for targeted base editing of RNA [46]. In contrast, for SsdA_{tox}, DNA and not RNA is the only physiological substrate with a very high preference for ssDNA over dsDNA [32], which is desired for its use in a CBE. Bacterial type VI secreted toxins are often fused to additional domains like PAAR or Rhs to facilitate their assembly at the tip of a type VI secretion system (Additional file 1: Fig. S1) [47–49]. For our approach, we truncated SsdA_{tox} to the predicted catalytic domain (151 aa).

Expression of free SsdA_{tox} is toxic and resulted in numerous C•G-to-T•A transitions in *E. coli* [32], but we reasoned that the gene should be tolerated when placed under control of a plant-specific promoter and fused to nCas9, which guides it to specific genomic locations and produces a local ssDNA substrate. This was the case for our clonings in *E. coli*, but we experienced a significant toxicity in *A. tumefaciens*, which possibly resulted from a low level of background expression by the plant 2 × 35S-promoter in *A. tumefaciens*. On the other hand, this negative selective pressure allowed us to isolate surviving *A. tumefaciens* colonies that contained suppressor mutations in SsdA_{tox} resulting in amino acid changes or stop codons. Conspicuously, the isolated mutations are predominantly C-to-T transitions, rendering it possible that the mutagenic activity of SsdA_{tox} itself caused them. From these transformants, we readily identified clones with high in planta editing activity. The SsdA^{G103S} variant and others with amino acid exchanges of G103 showed no toxicity while enabling efficient C-to-T conversion. This is counterintuitive because mutations with low or abolished deaminase activity should have a selective advantage if this enzyme is causing the toxicity. The critical residue (G103) lies in the 3rd β-strand in relative proximity of a postulated catalytic glutamate residue. Structural data suggest that residue G103 of SsdA_{tox} interacts with residues S34 and S35, while in SsdA^{G103S}, the larger side chain of S103 interacts with four residues: S34, S35, V104, and A130 (Additional file 1: Fig. S4). The structure of SsdA_{tox} exhibits an anti-parallel arrangement of beta-strands 4 and 5, whereas in the APOBEC family, these β-strands are parallel [32]. We speculate that the interaction between residue S103 and A130 in β-strand 4 might affect the activity pocket of the enzyme, thereby restricting the activity of SsdA. This could result in lower toxicity toward untargeted regions and a more specific affinity to the target R-loop region generated by nCas9.

In a parallel study, a non-peer-reviewed preprint described SsdA as a deaminase in CBEs for genome editing in mammalian cells [50]. In that report, three missense mutations (P282S, Y335R, K392E) within SsdA enhance cytosine base editing efficiency, but they are located at different positions to G103 identified by us (which would correspond to G362 in [50]). Future studies will reveal how such mutations influence the activity of DYW proteins.

We further exploited the use of introns to reduce toxicity by blocking translation of the toxic SsdA_{tox} in bacteria, which was successful. Interestingly, we found that an intron-containing SsdA^{G103S} variant (SsdA^{G103S}-v1) shows also significantly higher C-to-T editing efficiency than SsdA^{G103S} lacking the intron (SsdA^{G103S}) in planta. Beneficial effects

of introns to expression efficiencies of transgenes are known and are possibly due to increased export from the nucleus to the cytoplasm in plant cells [42]. Accordingly, we propose SsdA^{G103S}-v1-CBE as the preferred tool for plants.

CBEs exhibit different editing windows. In plants, APOBEC1 (nCas9-PBE) shows an editing window from C3 to C9 within the protospacer [17], whereas hA3A (A3A-PBE) has a broader editing window spanning from C1 to C17 [18]. In this study, we demonstrate that the new SsdA^{G103S}-v1-CBE prefers a narrower editing window of C5 to C8, which is beneficial if a specific target site shall be edited, but multiple cytosines are present within the protospacer. For example, hA3A edited five Cs (C6 to C10) simultaneously at the OsALS-T1 site, while the SsdA^{G103S}-v1-CBE preferred to edit C7 and C8 (Fig. 6b and Additional file 2: Table S1). A change from C7 and/or C8 to T results in Pro171 (CCC) being altered to Phe (TTC), Leu (CTC), or Ser (TCC), which confers resistance to the BS-herbicide in rice [34].

CBEs can induce undesired off-target base substitutions [51, 52]. In this study, we found Cas-dependent off-target editing induced by hA3A-CBE or SsdA^{G103S}-v1-CBE in T0 rice plants. To address the Cas-dependent off-targets, high-fidelity Cas9 variants can be incorporated into the CBE architectures to replace the wild-type nickase Cas9 [53–58]. Cas-independent off-target effects of SsdA^{G103S}-v1-CBE were not analyzed in this study. Considering the comparable Cas-dependent off-target editing rates of hA3A-CBE and SsdA^{G103S}-v1-CBE, and our efficiency of achieving fully edited plants with no apparent toxicity using either editing tool, we estimate that both CBEs possible cause similar off-target rates overall.

In addition to plant cells, we have also shown that our SsdA^{G103S}-v1-CBE is functional in mammalian cells, as tested on two human cell lines: the 293T epithelial cell line and K562, a myeloid leukemia cell line, with comparable efficiency to BE4max.

Cytosine deamination by deaminase enzymes can be sequence context-specific. For example, APOBEC1 and APOBEC3 show preferential deamination of cytosines in a 5'-TC sequence context [8, 59]. According to structural studies, hA3A binds ssDNA substrates in a U-shaped conformation, with the target cytosine inserted deep into the zinc-coordinating active site pocket, while the – 1 thymine base flipped out and fits into a groove between flexible loops, making direct hydrogen bonds with the protein [42]. Previous in vitro studies indicated that SsdA_{tox} enables the deamination of cytosines in all 5'-NC sequence contexts with a slight preference for neighboring pyrimidines [32]. Our editing results also show that SsdA^{G103S}-v1 has a flexible substrate sequence context in plants. This suggests that SsdA^{G103S}-v1 can edit all cytosines within the editing windows without a sequence context restriction.

The SsdA_{tox} (151 aa) we used in this study is already smaller than currently used cytosine deaminases, such as APOBEC1 (227 aa) [8], hA3A (198 aa) [41], AID (182 aa) [8], PmCDA1 (207 aa) [60], TadA8e variants (166 aa) [14–16, 61], and mini-Ssd7 (158 aa) [31]. Moreover, we could further truncate SsdA^{G103S} to a length of 114 aa without sacrificing activity, making SsdA^{G103S}-derived base editors a promising candidate for delivery into cells using size-limited methods (e.g., adeno-associated virus). Using a state-of-the-art structure search algorithm, we discovered a group of related proteins that could potentially function as deaminases and are candidates for additional future cytosine base editor systems.

Conclusions

In summary, the new SsdA^{G103S}-derived base editors expand the genome editing toolbox and can precisely and efficiently target C-to-T conversion in plants and mammalian cells. They exhibit comparable activity to highly optimized deaminases in established CBEs while providing a narrower editing window which allows more precise base targeting. The truncated SsdA^{G103S}-T3 is with 114 aa length by far the smallest deaminase in a genome editing tool to date. Taken together, SsdA^{G103S}-derived CBEs provide valuable alternatives for crop improvement and human gene therapy.

Methods

Plasmid construction

The SsdA_{tox} was amplified from *Pseudomonas syringae* pv. *aptata* GSPB1067. SsdA variants were generated by point mutation PCR. All plasmids used in this work were assembled based on Modular Cloning-compatible (MoClo) vectors (Additional file 2: Table S4 and Additional file 1: Supplementary sequences) [62–64]. The 2 × 35S promoter was used to express CBEs in the leaf infiltration and protoplast assays. The ZmUbi promoter was used to express SsdA^{G103S}-v1-CBE in rice transformation. All the components were subcloned in individual modules that can be assembled using Golden-Gate Cloning [65]. Oligos used in this study are listed in Additional file 2: Table S4. In addition, a human codon-optimized SsdA^{G103S}-v1-CBE (HuCoSsdA^{G103S}-v1-CBE) was cloned into an all-in-one self-inactivating (SIN) lentiviral vector (LV) for gene transfer into mammalian cells. The SIN-LV contains a U6 promoter to express the sgRNA and uses a spleen focus-forming virus (SFFV) promoter to express the HuCoSsdA^{G103S}-v1-CBE and dTomato fluorescent protein, linked by a T2A ribosomal skipping sequence. sgRNA targeting human IL7RA and CD5 were cloned into the all-in-one SIN-LV with HuCoSsdA^{G103S}-v1-CBE or BE4Max [66].

Lentiviral vector production and transduction

Lentiviral vectors were generated as previously described [67, 68]. Briefly, subconfluent 293 T cells were cultured in DMEM supplemented with 10% fetal calf serum, 1% sodium pyruvate, and penicillin/streptomycin and transfected using the calcium phosphate method. The all-in-one HuCoSsdA^{G103S}-v1-CBE lentiviral vector plasmids were co-transfected with the helper plasmids HIV-Rev, HIV-gag/pol, and VSV-g glycoprotein. Supernatants were collected at 36 and 48 h post-transfection and filtered through 0.22-μm filters. The supernatant was then added to K562 cells for transduction in the presence of 1 mg/ml Synperonic® F 108 (Sigma-Aldrich) [69], and incubated for 48 h before genomic DNA extraction.

Nicotiana benthamiana infiltration and GUS reporter assay

N. benthamiana plants were grown in a greenhouse with 16 h of light, a relative humidity of 40–60%, and temperatures of 23 °C/19 °C during the day/night, respectively. Four- to six-week-old plants were used for *A. tumefaciens* inoculation experiments. GUS reporter assays were performed as previously described [70]. Briefly, *A. tumefaciens* GV3101 strains containing a CBE construct and the GUS^{G537} reporter construct, respectively, were mixed 1:1 with an OD₆₀₀ of 0.8 and inoculated into *N. benthamiana* leaves.

After 2 days, two leaf discs (diameter 0.8 cm) were harvested from the inoculation spots. For leaf staining, leaf disks were stained in X-Gluc solution and de-stained in ethanol. For qualitative GUS assays, leaf tissues were homogenized and incubated with 4-methyl-umbelliferyl- β -D-glucuronide. GUS activities were measured using a TECAN reader (360 nm excitation and 465 nm emission). Proteins were quantified by NanoDrop™ One (Thermo Fisher Scientific).

Protoplast isolation and transformation

Two-week-old leaves from rice cultivar Kitaake or barley cultivar Golden Promise were used to prepare protoplasts. Rice and barley protoplast isolation and transformation were performed as previously described [71]. Twenty micrograms plasmid DNA per construct were introduced into protoplasts by PEG-mediated transfection. The transfected protoplasts were incubated at room temperature. After 48 h, the protoplasts were collected and the genomic DNA extracted.

Rice stable transformation

Rice cultivar Kitaake was used for genetic transformation *in this study*, as previously described [72]. Briefly, *A. tumefaciens* strains EHA105, containing SsdA^{G103S}-v1-CBE and sgRNA, as well as a *hygromycin* resistance gene as a selection marker, were used to transform calli. Then the transformed calli were transferred to selection plates containing 50 mg/l hygromycin. Regenerated calli were moved to rooting medium, then subjected to genotyping (Additional file 2: Table S1). For herbicide-resistance assay, 3-week-old regenerated seedlings were cultured on MS medium containing 0.4 μ M bispyribac-sodium, and grown at a temperature of 28 °C under a photoperiod of 16 h light and 8 h dark per day.

DNA extraction and identification of mutants

We used the innuPREP Plant DNA Kit (Analytik Jena) to extract genomic DNA from regenerated plants and protoplasts DNA. The targeted sequences were amplified with site-specific primers (Additional file 2: Table S5). The PCR products were purified with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and sequenced by Sanger sequencing (plants) or amplicon deep sequencing (protoplasts). Off-target sites were predicted by the online tool CRISPR-P [32]. Based on the off-target score, the top three or four predicted off-target sites of the used sgRNAs were selected as potential off-target sites (Table 1 and Additional file 2: Table S3). Site-specific primers (Additional file 2: Table S5) were used to amplify the potential off-target sites and T-DNA. The PCR products were purified and Sanger sequenced.

Amplicon deep sequencing and data analysis

PCR amplicons were purified with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) or QIAquick Gel Extraction Kit (Qiagen), then quantified using a Qubit™ 1X dsDNA High Sensitivity Kit or NanoDrop (Thermo Fisher Scientific). Equal amounts of PCR products were pooled and sequenced (GENEWIZ, AMPLICON-EZ). Amplicon deep sequencing was performed three times for each target location using genomic

DNA isolated from three different protoplast transformation experiments. For mammalian cells, deep sequencing for each targeted location was performed once from two different cell types. The target sites in the sequenced reads were analyzed for mutations using CRISPResso2 [73] (crispresso2.pinellolab.org).

Identifying structurally related proteins of SsdA_{tox}

The structure of SsdA (PDB: 7JTU) was used for a structure search using the Foldseek webserver (<https://search.foldseek.com>) [38]. Top 50 candidates from the AFDB50 database [39] were selected for conserved motif filtering. Details of the identified candidate proteins are listed in Additional file 2: Table S6. Sequences were aligned using ClustalW and displayed by ESPript 3.0 [74]. The SsdA homologs were used to construct a phylogenetic tree using Geneious Prime (version 2019) Tree Builder (default parameters), and visualized in ITOL (<https://itol.embl.de/>).

Statistical analysis

All values are shown as mean \pm SEM (standard error of the mean). Statistical differences between the values were tested using two-tailed unpaired Student's *t* tests by GraphPad (Prism; www.graphpad.com). No other scripts and software were used than those mentioned in the “Methods” section.

Peer review information

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer review history is available in the online version of this article.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-025-03478-w>.

Additional file 1: Supplementary figures S1–S4. Fig. S1. SsdA domain structure, function, and phylogenetic family. Fig. S2. Schematic diagrams of cytosine base editors. Fig. S3. C-to-T editing efficiencies of the intronized SsdA-CBEs and hA3A-CBE in *N. benthamiana*. Fig. S4. Cryo-EM structure of SsdAtox and AlphaFold2-predicted structure of SsdAG103S. Supplementary Sequences. Sequences of cytosine base editor architectures.

Additional file 2: Supplementary tables S1–S6. Table S1. Genotyping results of regenerated T0 plants transformed with SsdAG103S-v1-CBE or hA3A-CBE. Table S2. Inheritance of mutations in rice T1 lines. Table S3. Analyzing potential Cas-dependent off-target editing of SsdAG103S-v1-CBE in T0 plants. Table S4. Plasmids used in this study. Table S5. Oligos used in this study. Table S6. Top 50 candidates from the AlphaFold Protein Structure Database (AFDB50) with structural similarity to SsdAtox.

Acknowledgements

The authors thank Jana Streubel for material and discussions and Beate Meyer for assistance.

Authors' contributions

D.Z., T.-C.H., A.S., and J.B. designed the experiments. D.Z., F.P., L.M.S., and T.-C.H. performed the experiments. D.Z., F.P., L.M.S., and T.-C.H. analyzed the data. D.Z., T.-C.H., A.S., and J.B. wrote the manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Deutsche Forschungsgemeinschaft, BO 1496/9-1

Availability of data and materials

The amplicon sequencing data have been deposited in the NCBI BioProject database: PRJNA1087292 [75] and PRJNA1205495 [76].

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 24 May 2024 Accepted: 21 January 2025

Published online: 03 February 2025

References

- Gao C. Genome engineering for crop improvement and future agriculture. *Cell*. 2021;184:1621–35.
- Chen K, Wang Y, Zhang R, Zhang H, Gao C. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu Rev Plant Biol*. 2019;70:667–97.
- Zhang Y, Pribil M, Palmgren M, Gao C. A CRISPR way for accelerating improvement of food crops. *Nat Food*. 2020;1:200–5.
- Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol*. 2020;38:824–44.
- Porto EM, Komor AC, Slaymaker IM, Yeo GW. Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov*. 2020;19:839–59.
- Porto EM, Komor AC. In the business of base editors: evolution from bench to bedside. *PLoS Biol*. 2023;21: e3002071.
- Li J, Zhang C, He Y, Li S, Yan L, Li Y, et al. Plant base editing and prime editing: the current status and future perspectives. *J Integr Plant Biol*. 2022;65:444–67.
- Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533:420–4.
- Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*. 2017;551:464–71.
- Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, 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*. 2017;3: eaao4774.
- Koblan LW, Arbab M, Shen MW, Hussmann JA, Anzalone AV, Doman JL, et al. Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat Biotechnol*. 2021;39:1414–25.
- Zhao D, Li J, Li S, Xin X, Hu M, Price MA, et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol*. 2021;39:35–40.
- Chen L, Park JE, Paa P, Rajakumar PD, Prekop H-T, Chew YT, et al. Programmable C: G to G: C genome editing with CRISPR–Cas9-directed base excision repair proteins. *Nat Commun*. 2021;12:1384.
- Zhang E, Neugebauer ME, Krasnow NA, Liu DR. Phage-assisted evolution of highly active cytosine base editors with enhanced selectivity and minimal sequence context preference. *Nat Commun*. 2024;15:1697.
- Lam DK, Feliciano PR, Arif A, Bohnuud T, Fernandez TP, Gehrke JM, et al. Improved cytosine base editors generated from TadA variants. *Nat Biotechnol*. 2023;41:686–97.
- Chen L, Zhu B, Ru G, Meng H, Yan Y, Hong M, et al. Re-engineering the adenine deaminase TadA-8e for efficient and specific CRISPR-based cytosine base editing. *Nat Biotechnol*. 2023;41:663–72.
- Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol*. 2017;35:438–40.
- Zong Y, Song Q, Li C, Jin S, Zhang D, Wang Y, et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat Biotechnol*. 2018;36:950–3.
- Ren B, Yan F, Kuang Y, Li N, Zhang D, Zhou X, et al. Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. *Mol Plant*. 2018;11:623–6.
- Tang X, Ren Q, Yang L, Bao Y, Zhong Z, He Y, et al. Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnol J*. 2018;17:1431–45.
- Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, et al. Targeted base editing in rice and tomato using a CRISPR–Cas9 cytidine deaminase fusion. *Nat Biotechnol*. 2017;35:441–3.
- Mok BY, De Moraes MH, Zeng J, Bosch DE, Kotrys AV, Raguram A, et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*. 2020;583:631–7.
- Mok BY, Kotrys AV, Raguram A, Huang TP, Mootha VK, Liu DR. CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat Biotechnol*. 2022;40:1378–87.
- Barrera-Paez JD, Moraes CT. Mitochondrial genome engineering coming-of-age. *Trends Genet*. 2022;38:869–80.
- Kang B-C, Bae S-J, Lee S, Lee JS, Kim A, Lee H, et al. Chloroplast and mitochondrial DNA editing in plants. *Nat Plants*. 2021;7:899–905.
- Maliga P. Engineering the plastid and mitochondrial genomes of flowering plants. *Nat Plants*. 2022;8:996–1006.
- Mok YG, Lee JM, Chung E, Lee J, Lim K, Cho S-I, et al. Base editing in human cells with monomeric DddA-TALE fusion deaminases. *Nat Commun*. 2022;13:4038.
- Yin L, Shi K, Aihara H. Structural basis of sequence-specific cytosine deamination by double-stranded DNA deaminase toxin DddA. *Nat Struct Mol Biol*. 2023;30:1153–9.
- Mi L, Shi M, Li Y-X, Xie G, Rao X, Wu D, et al. DddA homolog search and engineering expand sequence compatibility of mitochondrial base editing. *Nat Commun*. 2023;14:874.
- Guo J, Yu W, Li M, Chen H, Liu J, Xue X, et al. A DddA ortholog-based and transactivator-assisted nuclear and mitochondrial cytosine base editors with expanded target compatibility. *Mol Cell*. 2023;83:1710–1724.e7.

31. Huang J, Lin Q, Fei H, He Z, Xu H, Li Y, et al. Discovery of deaminase functions by structure-based protein clustering. *Cell*. 2023;186:3182–3195.e14.
32. De Moraes MH, Hsu F, Huang D, Bosch DE, Zeng J, Radey MC, et al. An interbacterial DNA deaminase toxin directly mutagenizes surviving target populations. *eLife*. 2021;10:e62967.
33. Zhang D, Boch J. Development of TALE-adenine base editors in plants. *Plant Biotechnol J*. 2024;22:1067–77.
34. Kuang Y, Li S, Ren B, Yan F, Spetz C, Li X, et al. Base-editing-mediated artificial evolution of OSALS1 in planta to develop novel herbicide-tolerant rice germplasms. *Mol Plant*. 2020;13:565–72.
35. Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen L-L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant*. 2014;7:1494–6.
36. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017;35:1026–8.
37. Mahlich Y, Steinegger M, Rost B, Bromberg Y. HFSP: high speed homology-driven function annotation of proteins. *Bioinformatics*. 2018;34:i304–12.
38. Van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*. 2023;42:243–6.
39. Barrio-Hernandez I, Yeo J, Jänes J, Mirdita M, Gilchrist CLM, Wein T, et al. Clustering predicted structures at the scale of the known protein universe. *Nature*. 2023;622:637–45.
40. Iyer LM, Zhang D, Rogozin IB, Aravind L. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res*. 2011;39:9473–97.
41. Wang X, Li J, Wang Y, Yang B, Wei J, Wu J, et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol*. 2018;36:946–9.
42. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol*. 2017;24:131–9.
43. Jin S, Fei H, Zhu Z, Luo Y, Liu J, Gao S, et al. Rationally designed APOBEC3B cytosine base editors with improved specificity. *Mol Cell*. 2020;79:728–740.e6.
44. Hayes ML, Santibanez PI. A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing in vitro. *J Biol Chem*. 2020;295:3497–505.
45. Oldenkott B, Yang Y, Lesch E, Knoop V, Schallenberg-Rüdinger M. Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. *Commun Biol*. 2019;2:85.
46. Ichinose M, Kawabata M, Akaiwa Y, Shimajiri Y, Nakamura I, Tamai T, et al. U-to-C RNA editing by synthetic PPR-DYW proteins in bacteria and human culture cells. *Commun Biol*. 2022;5:968.
47. Hernandez RE, Gallegos-Monterrosa R, Coulthurst SJ. Type VI secretion system effector proteins: effective weapons for bacterial competitiveness. *Cell Microbiol*. 2020;22: e13241.
48. Ma L-S, Hachani A, Lin J-S, Filloux A, Lai E-M. *Agrobacterium tumefaciens* deploys a superfamily of Type VI secretion DNase effectors as weapons for interbacterial competition in planta. *Cell Host Microbe*. 2014;16:94–104.
49. Mariano G, Trunk K, Williams DJ, Monlezun L, Strahl H, Pitt SJ, et al. A family of type VI secretion system effector proteins that form ion-selective pores. *Nat Commun*. 2019;10:5484.
50. Kweon J, Park S, Jeon MY, Lim K, Jang G, Jang AH, et al. Efficient DNA base editing via an optimized DYW-like deaminase. *bioRxiv* (Cold Spring Harbor Laboratory). 2024. <https://doi.org/10.1101/2024.05.15.594452>.
51. Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*. 2019;364:292–5.
52. Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*. 2019;364:289–92.
53. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2015;351:84–8.
54. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016;529:490–5.
55. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*. 2017;550:407–10.
56. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol*. 2018;36:265–71.
57. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y-H, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun*. 2018;9:3048.
58. Vakulskas CA, Dever DP, Rettig GR, Turk R, Jacobi AM, Collingwood MA, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med*. 2018;24:1216–24.
59. Logue EC, Bloch N, Dhuey E, Zhang R, Cao P, Herate C, et al. A DNA sequence recognition loop on APOBEC3A controls substrate specificity. *PLoS One*. 2014;9: e97062.
60. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. 2016;353:aaf8729.
61. Neugebauer ME, Hsu A, Arbab M, Krasnow NA, McElroy AN, Pandey S, et al. Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat Biotechnol*. 2023;41:673–85.
62. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One*. 2011;6: e16765.
63. Geißler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens S-E, et al. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*. 2011;6: e19509.
64. Grütznert R, Marillonnet S. Generation of MoClo standard parts using golden gate cloning. *Methods Mol Biol*. 2020;2205:107–23.
65. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*. 2008;3: e3647.

66. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol.* 2018;36:843–6.
67. Ha T-C, Stahlhut M, Rothe M, Paul G, Dziadek V, Morgan M, et al. Multiple genes surrounding BCL-XL, a common retroviral insertion site, can influence hematopoiesis individually or in concert. *Hum Gene Ther.* 2021;32:458–72.
68. Schambach A, Galla M, Modlich U, Will E, Chandra S, Reeves L, et al. Lentiviral vectors pseudotyped with murine ecotropic envelope: Increased biosafety and convenience in preclinical research. *Exp Hematol.* 2006;34:588–92.
69. Ha T-C, Morgan MA, Thrasher AJ, Schambach A. Alpharetroviral vector-mediated gene therapy for IL7RA deficient SCID. *Hum Gene Ther.* 2024;35:669–79.
70. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-Type III effectors. *Science.* 2009;326:1509–12.
71. Shan Q, Wang Y, Li J, Gao C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc.* 2014;9:2395–410.
72. Sallaud C, Meynard D, Van Bortel J, Gay C, Bès M, Brizard JP, et al. Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theor Appl Genet.* 2003;106:1396–408.
73. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol.* 2019;37:224–6.
74. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 2014;42:W320–4.
75. Zhang D, Parth F, Silva LM, Ha TC, Schambach A, Boch J. Engineering a bacterial toxin deaminase from the DYW-family into a cytosine base editor. Datasets. Sequence read archive data. 2024. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1087292>.
76. Zhang D, Parth F, Silva LM, Ha TC, Schambach A, Boch J. Engineering a bacterial toxin deaminase from the DYW-family into a novel cytosine base editor for plants and mammalian cells. Datasets. Sequence read archive data. 2025. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1205495>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.