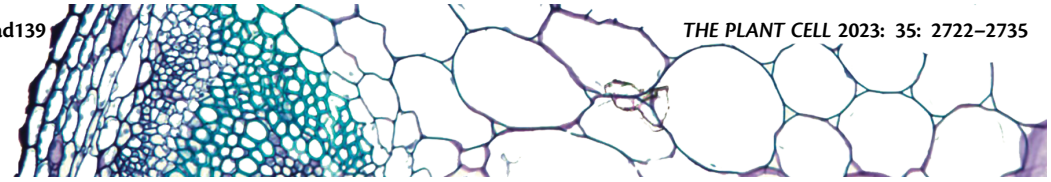
https://doi.org/10.1093/plcell/koad139 THE PLANT CELL 2023: 35: 2722–2735



Breakthrough Report

Efficient protein tagging and *cis*-regulatory element engineering via precise and directional oligonucleotide-based targeted insertion in plants

Jitesh Kumar [1,2,†] **Si Nian Char** [3,†] **Trevor Weiss** [1,2] **Hua Liu** [3] **Bo Liu** [3] **Bing Yang** [3,4,*] **and Feng Zhang** [1,2,*]

1 Department of Plant and Microbial Biology, University of Minnesota, USA
2 Center for Precision Plant Genomics, University of Minnesota, USA
3 Division of Plant Sciences and Technology, Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA
4 Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

* Author for correspondence: zhangumn@umn.edu (F.Z.), yangbi@missouri.edu (B.Y.)
†These authors contributed equally.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) is: Bing Yang (yangbi@missouri.edu) and Feng Zhang (zhangumn@umn.edu).

Abstract

Efficient and precise targeted insertion holds great promise but remains challenging in plant genome editing. An efficient non-homologous end-joining-mediated targeted insertion method was recently developed by combining clustered regularly interspaced short palindromic repeat (CRISPR)/*Streptococcus pyogenes* CRISPR-associated nuclease 9 (SpCas9) gene editing with phosphorothioate modified double-stranded oligodeoxynucleotides (dsODNs). Yet, this approach often leads to imprecise insertions with no control over the insertion direction. Here, we compared the influence of chemical protection of dsODNs on efficiency of targeted insertion. We observed that CRISPR/SpCas9 frequently induced staggered cleavages with 1-nucleotide 5′ overhangs; we also evaluated the effect of donor end structures on the direction and precision of targeted insertions. We demonstrate that chemically protected dsODNs with 1-nucleotide 5′ overhangs significantly improved the precision and direction control of target insertions in all tested CRISPR targeted sites. We applied this method to endogenous gene tagging in green foxtail (*Setaria viridis*) and engineering of *cis*-regulatory elements for disease resistance in rice (*Oryza sativa*). We directionally inserted 2 distinct transcription activator-like effector binding elements into the promoter region of a recessive rice bacterial blight resistance gene with up to 24.4% efficiency. The resulting rice lines harboring heritable insertions exhibited strong resistance to infection by the pathogen *Xanthomonas oryzae* pv. *oryzae* in an inducible and strain-specific manner.

Introduction

Recent advances in clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated nuclease 9 (Cas9)-mediated genome editing allow the introduction of intended genetic modifications at predefined target sites (Chen et al. 2019). The engineered CRISPR/Cas system involves searching, binding, and introducing double- or single-stranded DNA breaks within a target site directed by single-guide RNAs (sgRNAs). The resulting double-strand breaks (DSBs) induced by the Cas nuclease can be repaired via either error-prone DNA repair pathways, such as classical nonhomologous end-joining (c-NHEJ) and microhomology-mediated end-joining (MMEJ), or by a more precise homology-directed repair (HDR) (Nambiar et al. 2022). Depending on the intended

Received December 12, 2022. Accepted April 25, 2023. Advance access publication May 16, 2023

mutation type, the end-joining repair has been primarily used to introduce small insertions or deletions (InDels) at targeted sites with apparent stochastic characteristics. HDR is often the preferred pathway to generate precise targeted modifications such as targeted knock-in and site-directed nucleotide substitutions (Anzalone et al. 2020).

Targeted HDR-mediated modifications generally have a much lower efficiency compared to end joining-based modifications (Chen et al. 2019). In recent years, the NHEJ pathways have been tested for efficient targeted insertions, primarily in animal systems (Yamamoto and Gerbi 2018). These approaches however often suffer from imprecise outcomes with small InDels at the junction sequences and weak control over the direction of the insertion (Yamamoto and Gerbi 2018). To improve the directional control and precision of NHEJ-mediated targeted insertions, one strategy, also known as end capture, has been developed by generating compatible cohesive ends between target sites and donor DNA molecules (Orlando et al. 2010; Maresca et al. 2013). Indeed, improved targeted insertions were achieved by making compatible 3′ overhangs between target sites and donor DNA with a pair of nucleases, such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), in the fruit fly (*Drosophila melanogaster*), zebrafish (*Danio rerio*), mouse (*Mus musculus*), and human cell lines (Yamamoto and Gerbi 2018). In contrast to these recent efforts to develop the NHEJ-mediated targeted insertion approaches in mammalian systems, very few studies have explored their potentials in plants (Weinthal et al. 2013; Dong and Ronald 2021). NHEJ-based approaches may produce similar outcomes in plants, as the key components involved in the NHEJ pathway, such as Ku70/80, Ligase IV, and X-RAY REPAIR CROSS COMPLEMENTING 4 (XRCC4), are well conserved between vertebrates and plants. However, homologs for several key NHEJ proteins involved in processing and rejoining of the DNA ends have yet to be identified in flowering plants, such as DNA-dependent protein kinase (DNA-PK), XRCC4-like factor (XLF), and 3 members of X-family DNA polymerases (Manova and Gruszka 2015; Miller et al. 2021). Thus, it remains to be determined whether NHEJ-mediated approaches, such as the direct ligation-based end capture approach, might perform effectively in plants as in mammalian systems.

Recently, Lu et al. demonstrated NHEJ-mediated targeted insertions in rice (*Oryza sativa*) by using blunt-ended double-stranded oligodeoxynucleotides (dsODNs) that were chemically protected at both ends (Lu, Tian, et al. 2020). When combined with *Streptococcus pyogenes* Cas9 (SpCas9) (referred as Cas9 hereafter), efficient targeted insertion events were obtained but often with small InDels at the target sites. In addition, most insertions occurred in 2 opposite directions at nearly equal frequencies. Importantly, this study opened the avenue to improve NHEJ-based targeted insertions in plants. However, CRISPR/Cas9 was thought to be unsuitable for the NHEJ-based direct end capture approach because it was initially proposed to predominantly generate blunt-ended cleavage (Jiang and Doudna 2017). However, increasing evidence indicates that wild-type Cas9 frequently induces 1-nucleotide (1-nt) 5′ overhang at target sites (Zuo and Liu 2016; Lemos et al. 2018; Shi et al. 2019; Hussmann et al. 2021). This observation prompted us to evaluate whether a simple CRISPR/Cas9 system might be leveraged to achieve targeted insertions with improved direction control and higher precision through NHEJ.

In this study, we quantitatively compared the influence of chemical protection on the efficiency of dsODN-based insertion via the NHEJ pathway. We also evaluated the effect of the donor end structures (blunt or 5′ 1-nt overhang) on the direction and precision of targeted insertion events at different Cas9 target sites using green foxtail (*Setaria viridis*) protoplasts. Our results indicate that dsODN donors with 5′ 1-nt overhangs improve directional control and precision for NHEJ-based targeted insertions. This observation suggested that the staggered cleavage events induced by CRISPR/Cas9 can be harnessed to improve NHEJ-mediated targeted insertions in plant genomes. Furthermore, the use of dsODN in this approach brings at least 2 research applications within reach in plants. First, it could facilitate endogenous protein tagging. Fusing endogenous genes with sequences encoding luminescent or affinity tags, many of which are the size of dsODNs used in this study, would help characterize protein subcellular localization and function, as efficient approaches are underdeveloped in plant species (Lu, Ronald, et al. 2020). The second potential application is to engineer *cis*-regulatory elements to enable the generation of quantitative phenotypic variation by fine-tuning gene expression (Rodríguez-Leal et al. 2017). To date, most of these studies have deleted *cis*-regulatory regions (Li et al. 2020). While inserting new regulatory elements in the promoter regions offers great potential to generate novel traits through rewiring gene expression, such possibilities have not been tapped in plants largely due to the lack of efficient targeted insertion technique. In this study, we applied the NHEJ-based directionally targeted insertion approach to endogenous gene tagging in *S. viridis* and to engineering *cis*-regulatory element (CRE) to confer bacterial blight (BB) disease resistance in rice. Together, our study demonstrates the feasibility and efficacy of achieving directional and precise targeted insertions using Cas9 through the classic NHEJ pathway in plants. This approach should open many opportunities for basic and applied research in plant genome engineering.

Results

Quantitative comparisons of end protection for efficient dsODN-based targeted insertion in plant cells

While previous studies have reported improved targeted insertion efficiency using dsODN with both 5′ and 3′ phosphorothioate modifications in human cell lines and rice (Tsai et al. 2015; Li et al. 2016; Lu, Tian, et al. 2020), how these

modifications influence the efficacy of targeted insertions has not been thoroughly investigated in a quantitative manner. In this study, we quantified dsODN-based targeted insertion frequencies with distinct phosphorothioate modifications. As outlined in Fig. 1A, using an efficient protoplast transfection system in *S. viridis* (Weiss et al. 2020), we cotransfected mesophyll protoplasts with synthesized dsODN donors and a DSB-inducing *Cas9* construct (Supplemental Fig. S1). We evaluated the efficiencies of targeted insertion by the high-throughput sequencing after a 48-h incubation. We used 2 targets, located in the *MS26* and *MS45* genes described previously (Weiss et al. 2020), for insertion sites (Fig. 1B). We designed and constructed the 34-bp dsODN donor with the following features: (i) a 12-bp sequence sharing homology with the *MS26* site but not the *MS45* site at each side of the Cas9-cutting site and (ii) a 10-bp internal sequence (Fig. 1B). This dsODN donor design allowed us to test targeted insertions that resulted not only from the classical NHEJ pathway but also from the MMEJ pathway with the homologous sequences to the *MS26* site (Fig. 1B) (Sakuma et al. 2016; Wierson et al. 2020).

We synthesized the dsODNs with 5′ phosphorylation and 3 types of phosphorothioate linkage at either the 5′, 3′, or both 5′ and 3′ ends (Fig. 1C). When we tested the dsODNs without phosphorothioate modification or with only 5′ phosphorothioate linkage in protoplasts, we obtained very low levels of targeted insertion in the *MS26* and *MS45* targeted sites (0% to 1.2%; Fig. 1C). By contrast, when we delivered dsODNs with 3′ phosphorothioate modifications, the targeted insertion frequencies increased to 51.1% and 27.8% in *MS26* and *MS45* sites, respectively (Fig. 1C). Similarly, with dsODNs containing phosphorothioate modifications at the 5′ and 3′ ends, we obtained high frequencies of targeted insertions in *MS26* (45.3%) and *MS45* (34.3%) (Fig. 1C). Thus, the 3′ modification appears to be critical to achieving efficient dsODN-based targeted insertions. Further investigation would be required to understand the underlying mechanism. Nonetheless, there were no statistically significant differences for targeted insertion frequencies between dsODNs with 3′ alone or with 5′ and 3′ modifications at the *MS26* and *MS45* sites (*P*-values from a 2-tailed paired *t*-test: 0.49 at *MS26* and 0.54 at *MS45*). To minimize the possible complications that could be derived from potential 5′ DNA end resections, we chose 5′ and 3′ modifications for subsequent donor design.

Targeted insertions of chemically protected dsODN through the c-NHEJ pathway

We investigated the roles of the 2 end-joining pathways, c-NHEJ and MMEJ, in dsODN-based targeted insertion. The dsODN with the 12-bp homologous sequences at the *MS26* site allowed us to discriminate between targeted insertion events from the c-NHEJ and MMEJ-mediated products. If targeted insertions are generated through c-NHEJ, the size of insertions would be close to the full length, i.e. 34 bp, with a duplication of the 12-bp homologous sequence on each side. On the contrary, if a targeted insertion is mediated by the MMEJ pathway, the 12-bp homologous sequences would base pair with the targeted sequences, leading to the insertion of the 10-bp internal sequence without duplication of homologous sequences. We thus examined all insertion events from protoplasts transfected with dsODNs containing either the 3′ phosphorothioate alone or 5′ and 3′ phosphorothioate modifications at the *MS26* site (Fig. 1, D and E). In each experiment, we identified 2 classes of insertions: 1-bp insertions and dsODN-based targeted insertions (Fig. 1D). Most targeted insertion events had a size of 34 bp (Fig. 1D), with a few being 35-bp or smaller with InDels at the junction sites (Figs. 1, D and E, and S2). We noticed a 22-bp insertion event when using 3′ or 5′and 3′ modified dsODN, which was consistent with a DNA repair product via homology-dependent repair at the 5′ junction but via NHEJ at the 3′ junction (Supplemental Fig. S2). This structure can be explained by the synthesis-dependent strand annealing (SDSA) or 1-sided invasion (OSI) model, in which the combination of HDR and NHEJ occurs at different junctions (Puchta 2005). Alternatively, this event can be explained by the combination of MMEJ repair at 1 side and NHEJ repair at the other. In addition, we detected a low frequency of targeted 68-bp insertions using the 3′ phosphorothioate dsODN (Fig. 1D). Sequencing revealed 2 copies of dsODNs being inserted in a head-to-tail fashion at the target site. Taken together, our results suggest that most phosphorothioate modified dsODNs tested here appeared to be integrated into the targeted site via the c-NHEJ pathway.

CRISPR/Cas9 staggered cleavage enables directional dsODN-based targeted insertion (DOTI)

All dsODN described above were blunt ended. When we analyzed the orientations of targeted insertions from either 3′ or 5′ and 3′ phosphorothioate-modified dsODN groups, we detected the 2 possible orientations in nearly equal frequencies at the *MS26* and *MS45* sites (Fig. 2A). While *Streptococcus pyogenes* Cas9 was initially thought to generate blunt cleavage 3-bp upstream of the PAM sequence (−3 position), increasing evidence suggests that this Cas9 can also cleave the nontargeted strand at the −4 position, thus producing 5′ 1-nt overhangs in human cell lines (Fig. 2B) (Molla and Yang 2020; Schmid-Burgk et al. 2020). This staggered cleavage may lead to specific 1-bp insertions or 1-bp deletions at the −4 position via c-NHEJ (Fig. 2B). When we analyzed the sequencing reads from the transfected protoplasts without dsODN donor, we detected 1-bp InDels at the −4 position with frequencies of 56% (*MS26*) and 34% (*MS45*) (Fig. 2B). We also assessed the Cas9-induced mutation profiles from 8 additional targeted sites from the *S. viridis* genome. We obtained frequencies of 1-bp InDels at the −4 position ranging from 4.8% to 77.6%, with an average frequency at 37.8% (Supplemental Fig. S3A). These observations suggest that staggered cleavage at the Cas9-targeted sites also occurs prevalently in plants.

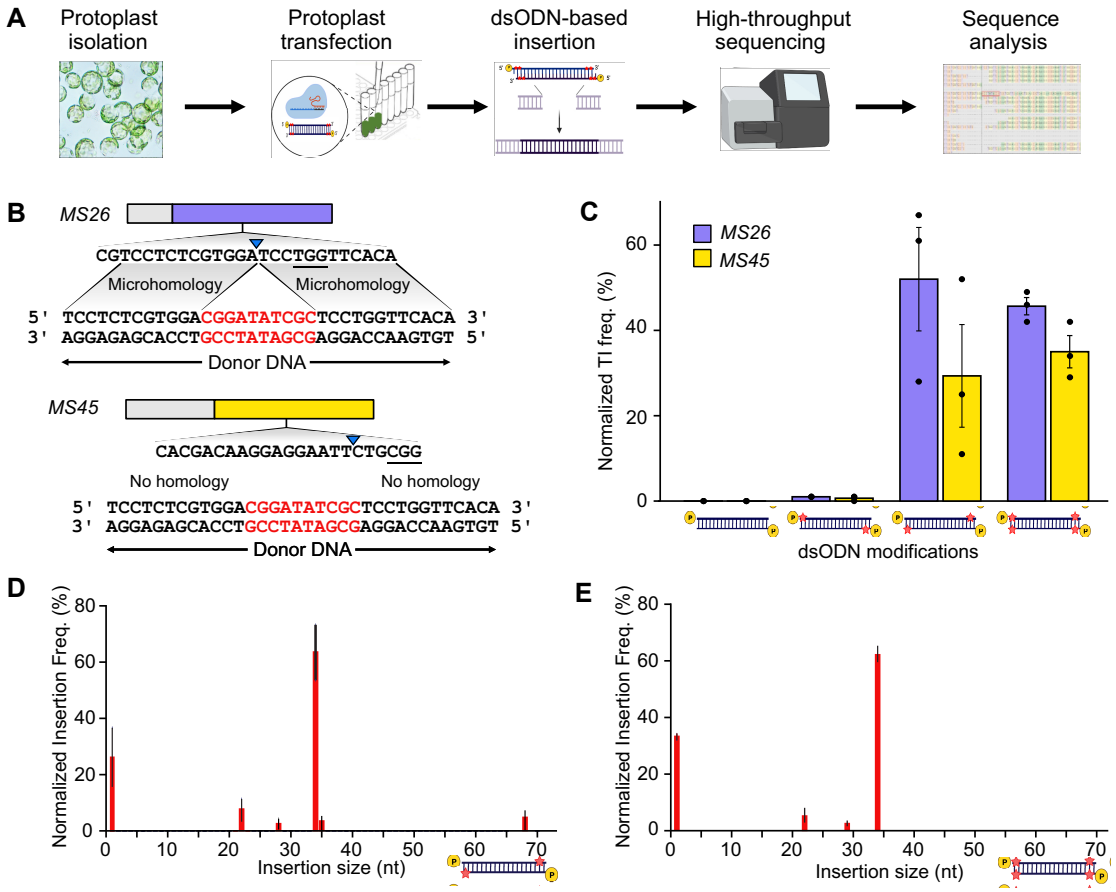


Figure 1. Influence of phosphorothioate modifications on dsODN-based targeted insertion efficiency. **A)** Diagram of protoplast-based assay for evaluating the outcome of targeted insertions (TIs). **B)** Design of a blunt-ended dsODN donor for the *MS26* and *MS45* sites. A 10-bp internal sequence flanked by 12-bp homologous to the *MS26* site is shown. The protospacer adjacent motif (PAM) is underlined; arrowheads are Cas9 cleavage sites. **C)** Normalized frequency of TIs at the *MS26* and *MS45* sites. Chemical modifications of individual dsODNs are shown under the *x*-axis with 5′ phosphorylation (circles) and phosphorothioate linkages at different ends (asterisk). The normalized frequencies were calculated by dividing the number of reads containing the intended insertion by the total number of reads with any mutation at the target site. **D**, **E)** Distribution of normalized frequencies for insertions of various sizes in the *MS26* site with dsODNs containing phosphorothioate modifications at the 3′ end alone **D)** or 5′ and 3′ ends **E)**. The normalized frequencies of individual insertion events were calculated by dividing the number of reads containing the intended insertion at each size by the total number of reads containing any insertion. Values are means ± standard error of mean (SEM) from 3 biological replicates.

However, 1-bp InDel events can be derived from blunt cleavage, followed by repair via c-NHEJ. To test the occurrence of Cas9-induced staggered cleavage, we hypothesized that the use of dsODN with a complementary 1-nt 5′ overhang would capture the staggered cleavage product and generate targeted insertion events in a directional manner (Fig. 2C). Accordingly, we designed 2 phosphorothioate-modified dsODN donors with 1-nt 5′ overhangs, 1 containing complementary A/T overhangs (A at the 5′ end, T at the 3′ end), and the other having noncomplementary C/G overhangs. Using the same protoplast transfection and sequencing procedures described above, we observed targeted insertion frequencies of 16.7% (*MS26*) and 27% (*MS45*) out of all mutations with the A/T-overhang dsODNs, while we detected lower targeted insertion frequencies with the C/G overhang donor at the *MS26* (4.0%) and *MS45* (5.7%) sites (Supplemental Fig. S3B). Nevertheless, when we examined the orientation of targeted insertions in these samples, most targeted insertions with the complementary A/T-overhang dsODNs were in the intended orientation at the *MS26* (78% in forward orientation) and *MS45* site (88% in reverse orientation) (Fig. 2D). By contrast, with the noncomplementary C/G-overhang donor, the majority of targeted insertions appeared to be in orientation opposite to those with the A/T-overhang donor, with 66% at *MS26* and 72% at *MS45* (Fig. 2D). This result can be explained by noncanonical base pairing, i.e. G:T and C:A, between the overhangs from the dsODN donor and the targeted sites. We hypothesize that this cleavage mode can be harnessed to enable directionally targeted insertions with 1-nt complementary 5′ overhang donors, which we call directional dsODN-based targeted insertion (DOTI).

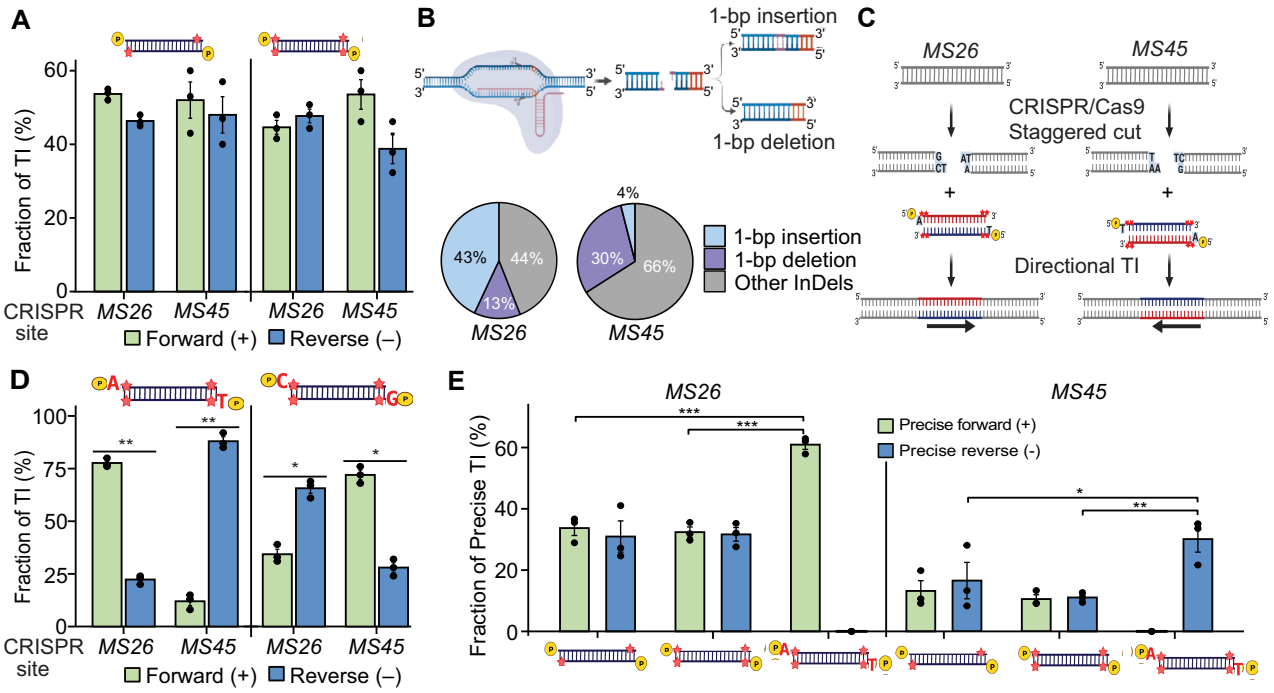


Figure 2. Precise and directional dsODN-based targeted insertion. **A)** Direction of targeted insertions (forward [+] and reverse [−]) with blunt-ended dsODNs. The dsODN modifications are indicated as in Fig. 1C. The fraction of targeted insertions in different orientation (forward [+]; reverse [−]) was calculated by dividing the number of reads with the intended insertion in each orientation by the total number of reads with the intended insertion in any orientation. **B)** Model for 1-bp InDels by Cas9-induced staggered cleavage. Diagram of 1-bp InDels showing the outcomes of staggered cleavage and c-NHEJ repair. The frequencies of 1-bp insertions and deletions were calculated by dividing the number of reads containing a1-bp InDel by the total number of reads containing a mutation. **C)** Diagram of directionally targeted insertions via Cas9-induced staggered cleavage. The predicted orientations are indicated by the arrows. **D)** Comparisons of targeted insertions with forward (+) and reverse (−) orientations. The dsODN donors with 1-bp A/T or C/G overhangs are shown on the top with phosphorylation and phosphorothioate linkages (asterisk). The fractions of targeted insertions in different orientations were calculated as above. Values are means ± SEM from 3 biological replicates. **E)** Frequencies of precise (seamless) targeted insertions in forward (+) and reverse (−) orientations. The dsODN donors with either blunt ends or 1-bp A/T overhangs are shown under the *x*-axis. The fraction of precise targeted insertions in different orientations (blue, forward [+]; red, reverse [−]) was calculated by dividing the number of reads containing a seamless insertion at the intended site in each orientation by the total number of reads with an insertion in any orientation. Values are means ± SEM from 3 biological replicates. *P*-values were calculated by 2-tailed paired Student's *t*-test between samples (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$).

DOTI improves the precision of targeted insertions in a directional manner

Previous studies reported that blunt-ended dsODNs often lead to high frequencies of imprecise targeted insertion events with small InDels at the junction sites (Lu, Tian, et al. 2020). We hypothesized that these small InDels result from the 1-nt 5′ overhangs generated by Cas9 staggered cleavage. We thus investigated whether the DOTI approach with its 5′ 1-nt complementary overhangs might affect the precision of dsODN-based targeted insertions. To this end, we analyzed 5′ and 3′ junctions of individual targeted insertions using end-protected dsODNs with either blunt or 5′ 1-nt A/T overhangs at the *MS26* and *MS45* sites (Fig. 2E). We examined the fraction of precise (seamless) targeted insertions in each orientation, designated as precise forward (PF) and precise reverse (PR), by dividing the number of sequencing reads with seamless junctions in each orientation by the total number of reads containing an insertion at the intended target site. We measured similar fractions of seamless insertions in the PF and PR orientations with blunt donors using 3′ or 5′ and 3′ modifications (33.7% [PF, 3′], 30.95% [PR, 3′], 32.4% [PF, 3′ and 5′], and 31.7% [PR, 3′ and 5′] at the *MS26* site; 13.2% [PF, 3′], 16.6% [PR, 3′], 10.55% [PF, 3′ and 5′], and 11.1% [PR, 3′ and 5′] at the *MS45* site; Fig. 2E). By contrast, the fraction of precise insertions with 1-nt 5′ A/T overhang dsODN increased by 90% (from 32.4% to 60.9%) in the PF direction at the *MS26* site and by 272% (from 11.1% to 30.1%) in the PR direction at the *MS45* site (Fig. 2E). Importantly, we barely detected precise insertions in the opposite orientation at both targeted sites (Fig. 2E). Taken together, we conclude that dsODNs with 1-nt 5′ complementary overhangs improve the efficiency of precise targeted insertion in a directional manner.

DOTI enables efficient endogenous protein tagging in *S. viridis*

We tested DOTI for endogenous protein tagging in *S. viridis* protoplasts. We designed single-guide RNAs (sgRNAs) to target either the 5′ end or 3′ end of 3 endogenous genes in *S.*

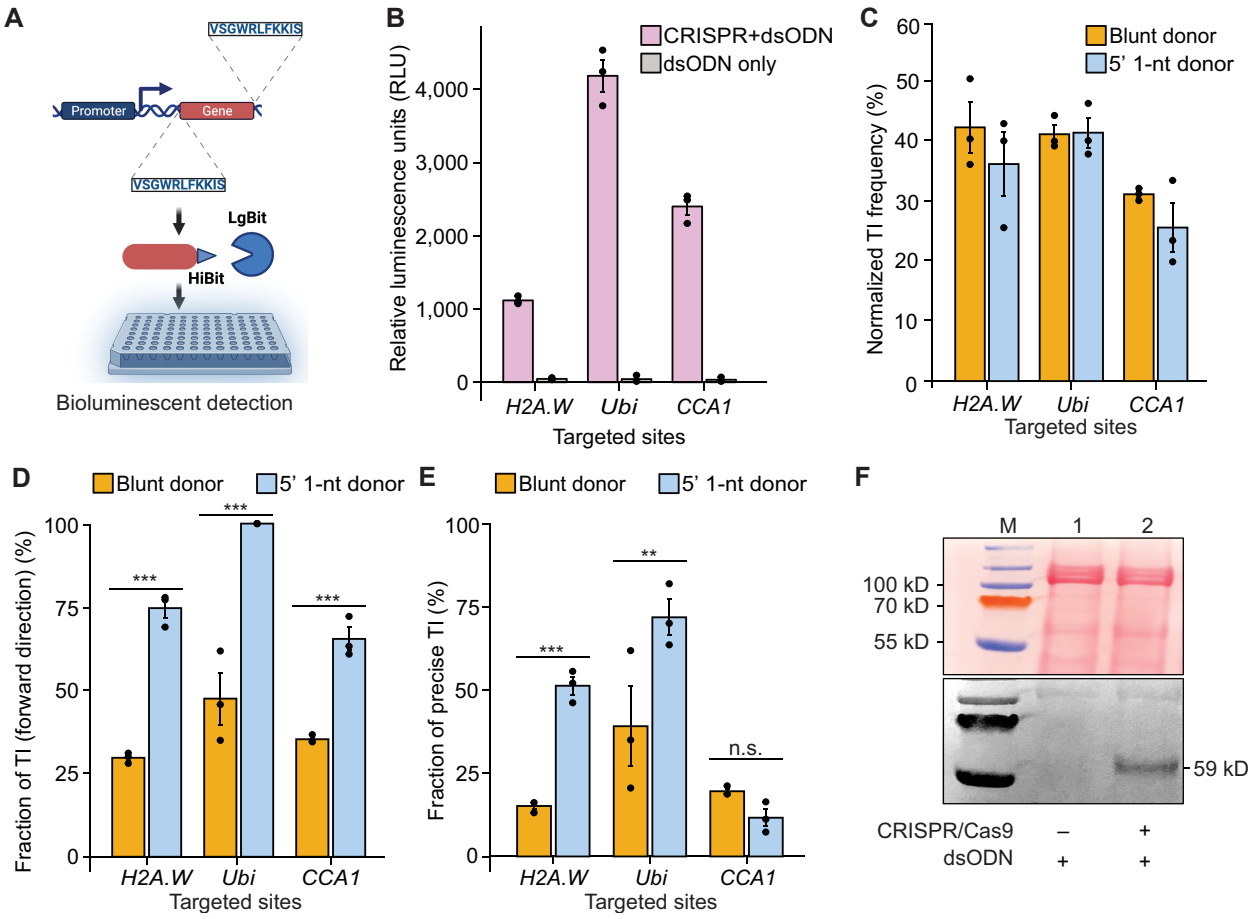


Figure 3. Endogenous protein tagging via DOTI. **A)** Diagram of endogenous gene tagging using the small subunit of NanoLuciferase (HiBiT). **B)** Luminescence levels of HiBiT-tagged endogenous proteins, encoded by *H2A.W*, *UBI*, and *CCA1*, in *Setaria* protoplasts. **C)** Normalized frequencies of targeted insertions between blunt and overhang donors at each of the 3 endogenous sites. The normalized frequencies were calculated by dividing the number of reads with the intended insertion by the total number of reads containing any mutation. **D)** Normalized frequencies of targeted insertions in the intended (forward) orientation with blunt and overhang donors. The fractions of targeted insertions with different orientation were calculated by dividing the number of reads containing the intended insertion in the forward orientation by the total number of reads with the intended insertion. **E)** Normalized frequencies of precise targeted insertions in the intended orientation with blunt and overhang donors at each of the 3 endogenous sites. The fraction of precise targeted insertions was calculated by dividing the number of reads containing seamless targeted insertion in the intended orientation by the total number of the reads containing the intended insertion. **F)** Immunoblot of the 3× FLAG-tagged CCA1, using an anti-FLAG antibody. **M)** Molecular weight marker; lane 1: total soluble proteins from protoplasts transfected with the dsODN only; lane 2: total soluble proteins from protoplasts cotransfected with the dsODN and the *Cas9* plasmid. Values are means ± SEM from 3 biological replicates. *P*-values were calculated using a 2-tailed paired Student's *t*-test between samples (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$, n.s., not significant).

viridis, histone H2A.W (*H2A.W*, Sevir.9G453900), *Ubiquitin* (*Ubi*, Sevir.5G079801), and a transcription factor, a homolog of *Arabidopsis* (*Arabidopsis thaliana*) *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*, Sevir.6G053000) (Supplemental Data Set 1). We tagged these genes with a sequence encoding the small subunit of NanoLuciferase (SmBiT or HiBiT) (Dixon et al. 2016). The HiBiT subunit consists of an 11-amino acid peptide that interacts with the large subunit (LgBiT) to form a functional luciferase protein (Fig. 3A). The tagged protein can be detected and quantified using a luminescence assay (Schwinn et al. 2018). Because of its small size and high affinity, the HiBiT peptide has been used to tag endogenous proteins in mammalian systems (Schwinn et al. 2018). To test the effectiveness of the HiBiT tagging system in plant cells, we codelivered the *Cas9*-expressing construct with a phosphorothioate-modified dsODN containing a 33-bp HiBiT sequence and 1-nt 5′ complementary overhangs for each target site (Supplemental Data Set 1). Compared to the donor-only control groups, we detected significant luciferase activity in the *H2A.W*, *Ubi*, and *CCA1* samples with both dsODN and *Cas9*, with values of 1,118, 4,186, and 2,401 relative luminescence units (RLUs), respectively (Fig. 3B). To directly compare the effects of blunt or 5′ 1-nt overhang donors on the orientation and precision of targeted insertions, we assessed the normalized frequencies of targeted insertions with each donor type by high-throughput

sequencing. We obtained comparable normalized frequencies of targeted insertions with blunt and overhang donors at each targeted site (Fig. 3C). An examination of the fractions of targeted insertions in the intended orientation (forward orientation) at each site revealed that the protoplasts transfected with overhang donors exhibit significantly higher rates of insertions in the intended orientation than those treated with the blunt-ended donors for all 3 sites (Fig. 3D). In addition, we compared the fractions of seamless targeted insertions between the blunt-ended and overhang dsODNs at all 3 sites. The dsODN with 1-nt 5′ complementary overhangs outperformed the blunt-ended donor with 3.3- and 1.8-fold higher precise targeted insertions at the *H2A.W* and *Ubi* sites, respectively (50.55% vs. 15.15% at *H2A.W*; 71.75% vs. 39.1% at *Ubi*; Fig. 3E). Taken together, DOTI enabled efficient targeted insertions at all 3 sites in the preferred orientation with improved precision.

To address whether 1-nt 5′ overhang donors might compromise the overall targeted insertion efficiencies, we compared the efficiencies of targeted insertions between the blunt and overhang donors at all 5 sites tested so far. We observed no significant difference at any of the tested sites except for the *H2A.W* site, for which the overall targeted insertion efficiencies using the overhang donor were significantly higher than those using the blunt donor (Supplemental Fig. S4). Thus, the 1-nt 5′ overhang donors do not appear to negatively affect targeted insertion efficiency.

Next, we used DOTI to tag CCA1 with a widely used epitope tag, 3× FLAG (corresponding to a 66-bp sequence). We produced the dsODN donor containing the sequence encoding 3× FLAG with a phosphorothioate linkage and 1-nt 5′ A/T overhangs (Supplemental Data Set 1) before codelivering it with the corresponding *Cas9* construct. After a 48-h incubation, we collected the transfected protoplasts for protein extraction and immunoblot analysis with an anti-FLAG antibody. Consequently, we detected a protein with the expected molecular weight protein (M.W. = ~59.03 kD) for CCA1 in protoplasts cotransfected with the *Cas9* construct and dsODN donor, in contrast to the donor-only control (Fig. 3F).

DOTI enables CRE engineering to confer BB disease resistance in rice

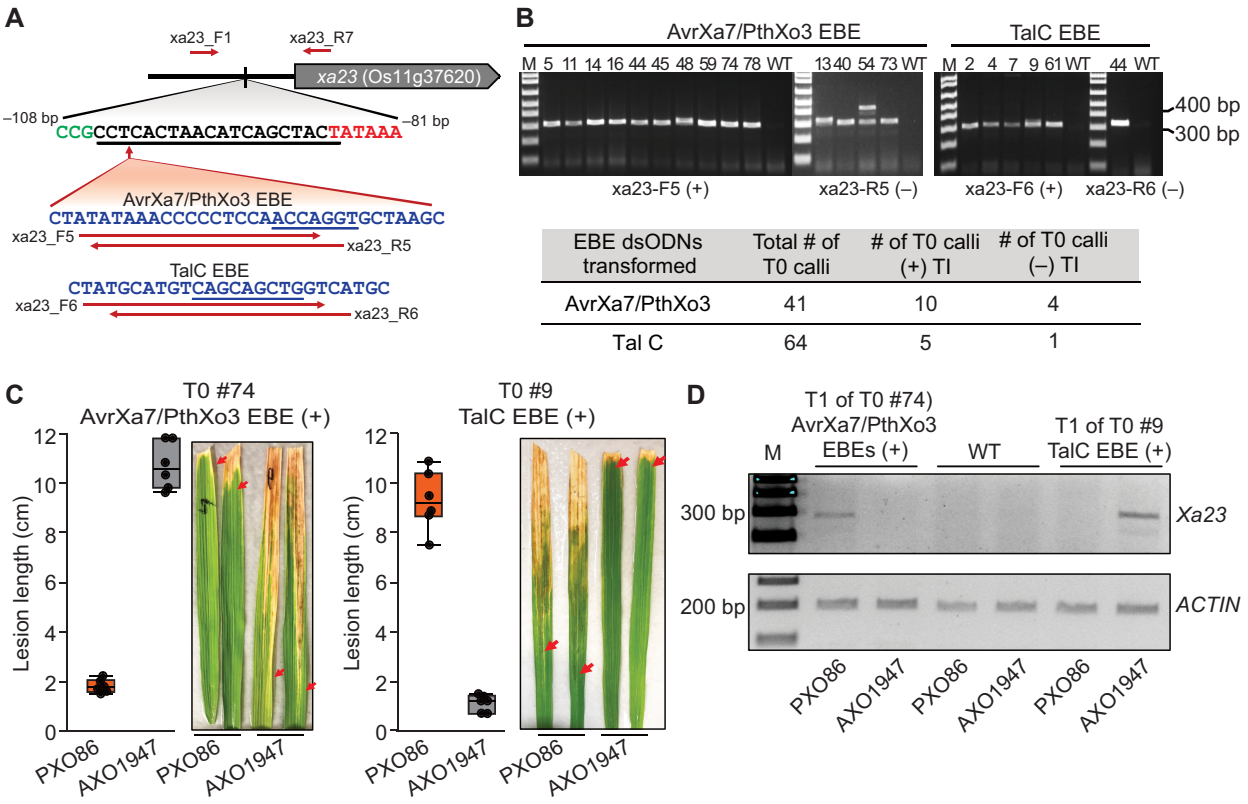
We applied DOTI to CRE engineering. Previous studies demonstrated that *xa23*, a recessive allele of the BB *R* gene *Xa23*, can be activated by inserting the TAL effector AvrXa23-binding element (EBE) of the *Xa23* promoter into the promoter region of the *xa23* allele using a low-efficiency HDR-based approach. This insertion via knock-in resulted in resistance to the pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) expressing *AvrXa23* (Wei et al. 2021). In this study, we designed 2 distinct dsODN donors based on 2 different TAL EBEs identified from the *SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTER 14* (*OsSWEET14*) gene: 1 (31 bp) recognized by the TAL effectors AvrXa7 and PthXo3 present in most Asian *Xoo* isolates and the other (25 bp) recognized by the TAL effector TalC present in all African *Xoo* isolates (Fig. 4A) (Antony et al. 2010; Chen et al. 2010; Oliva et al. 2019). We identified a Cas9 target site between 86- and 108-bp upstream of the transcription start site (upstream of a putative TATA box) of *xa23* (Fig. 4A). We synthesized dsODNs with 5′ phosphorylation together with 5′ and 3′ phosphorothioate modifications and 1-nt 5′ C/G overhangs complementary to the predicted overhangs induced by staggered cleavage at the targeted site (Fig. 4A and Supplemental Data Set 1).

We cobombarded individual dsODNs with the *Cas9*-expressing plasmid into rice calli derived from immature embryos of the Kitaake cultivar (a BB-susceptible rice cultivar). We selected hygromycin-resistant calli to identify stable transgenic events (T_0) and extracted genomic DNA to use as template for PCR amplification using EBE-specific and *xa23* gene-specific primers in both forward (+) and reverse (−) orientations (Fig. 4A and Supplemental Data Set 1). In the transformation group with the dsODN containing the AvrXa7/PthXo3 EBE, among 41 hygromycin-resistant calli screened, we determined that 10 calli (24.4%) have targeted insertions in the intended orientation (forward), with another 4 calli (9.7%) with targeted insertions in the reverse orientation (Fig. 4B). Notably, we identified 2 PCR products from T_0 callus #54 in the reverse orientation. The larger amplicon possessed 2 copies of the dsODN inserted at the targeted site. In the transformation group with the dsODN containing the TalC EBE, we detected targeted insertions in 5 out of 64 (7.8%) hygromycin-resistant calli with the intended orientation (forward), with 1 callus (1.6%) identified a targeted insertion in the reverse orientation (Fig. 4B). Taken together, targeted insertions from both transformation groups were inserted in the intended orientation at higher frequencies than those in the reverse orientation.

We regenerated T_0 plants from targeted insertion-containing calli and evaluated them for resistance to BB. We used 2 *Xoo* strains in this study, an Asian *Xoo* strain, PXO86, carrying the AvrXa7 TAL effector, and an African *Xoo* strain, AXO1947, harboring the TalC effector. After inoculation, we measured the length of lesions from the leaves of 2 T_0 plants per dsODN. When inoculated with PXO86, the leaves of plants carrying an insertion for the AvrXa7/PthXo3 EBE (T_0 plant #74) showed greater BB resistance (shorter lesions) but exhibited a susceptible phenotype against AXO1947 (longer lesions) (Fig. 4C; Oliva et al. 2019). Similarly, the leaves from the plant harboring an insertion for the TalC EBE (T_0 plant #9) developed short lesions when inoculated with AXO1947 but remained susceptible to PXO86 (Fig. 4C).

Inheritance of EBE insertions in T_1 rice plants

To investigate the inheritability of these targeted insertions, we collected seeds from 6 independent T_0 plants: 4 plants containing the AvrXa7/PthXo3 EBE, including plant #74, and 2 plants containing the TalC EBE, including plant #9. For each T_1 progeny, we performed PCR amplification using *Xa23* gene-specific primers followed by restriction enzyme



EBE dsODNs transformed	Total # of T0 calli	# of T0 calli (+) TI	# of T0 calli (–) TI
AvrXa7/PthXo3	41	10	4
Tal C	64	5	1

Figure 4. Targeted insertion of CREs for disease resistance engineering in rice. **A)** Diagram of the recessive *S* locus, *xa23* (LOC_Os11g37620). The gray bar indicates the *Xa23* coding region. The targeted insertion site is located 86- to 108-bp upstream of the transcription start site with the putative TATA box (-81 to -86 bp), the 20-bp CRISPR recognition sequences (underlined) and the PAM sequence (-106 to -108 bp). The AvrXa7/PthXo3 and TalC EBE sequences are indicated with the restriction sites *Sex*AI and *Alw*NI underlined. *Xa23* and insertion-specific primers used for PCR genotyping are indicated by the arrows above the *Xa23* gene model. **B)** Identification of T_0 calli with the intended insertion. T_0 calli with targeted insertions were identified using the EBE-specific primer (primer names indicated at the bottom) and the gene-specific primer xa23_R7. Targeted insertions in the forward or reverse orientation are indicated with (+) or (−). **C)** Assessment of disease resistance in EBE-inserted plants (right) and length of disease lesions, indicated by the red arrows (left). T_0 plants, #74 and #9, were used for the *Xoo* infection assay. **D)** Induction of *Xa23* expression upon *Xoo* infection. End-point RT-PCR products are shown for *Xa23* in response to PXO86 and AXO1947 infection, respectively. *OsACTIN* (LOC_Os03g50885) was used as reference. **M)** 1 kb plus DNA ladder; WT, wild-type control.

digestion. Insertion of the AvrXa7/PthXo3 EBE added a *Sex*AI site at the targeted site, while insertion of the TalC EBE added an *Alw*NI site. We identified homozygous and heterozygous insertion events at the target sites (Supplemental Fig. S5, A and B). We confirmed individual homozygous events by Sanger sequencing for the T_1 progeny of plants #74 (AvrXa7/PthXo3 EBE) and #9 (TalC EBE) (Supplemental Fig. S5, C and D).

We evaluated the induction of *Xa23* expression in homozygous T_1 progeny from T_0 plants #74 and #9. At 24-h post-infection with PXO86 and AXO1947 respectively, we collected the inoculated leaves for total RNA extraction and RT-PCR using *Xa23*-specific primers with the *ACTIN* gene as internal control (Fig. 4D and Supplemental Data Set 1). In T_1 plants containing the AvrXa7/PthXo3 EBE, we observed induction of *Xa23* expression following PXO86 infection, but not with AXO1947. Similarly, in the T_1 plants harboring the TalC EBE, *Xa23* expression was induced in response to AXO1947 but not PXO86 infection (Fig. 4D). In addition, we challenged a total of 69 T_1 plants with homozygous targeted insertions derived from the 6 T_0 edited plants with their respective *Xoo* strains to evaluate their resistance. All T_1 plants carrying the AvrXa7/PthXo3 EBE showed resistance to BB (short lesions) caused by PXO86 but were susceptible to AXO1947 (long lesions) (Supplemental Fig. S6, A and B); and all T_1 plants containing the TalC EBE were resistant to AXO1947 but susceptible to PXO86 (Supplemental Fig. S5, A and B). Taken together, our data demonstrate that DOTI can be used to effectively engineer CREs. Targeted insertion of TAL EBEs into the promoter region activated expression of the *R* gene *Xa23* in response to the *Xoo* strain expressing the cognate TAL effector.

Discussion

In this study, we sought to improve the precision and direction control of dsODN-based targeted insertion. To this end, we thoroughly evaluated the effects of phosphorothioate

modifications on the efficacy of dsODN-based targeted insertion in the model plant, *S. viridis*. The use of phosphorothioated dsODN was first reported in human cell lines to increase targeted insertion efficiency (Tsai et al. 2015; Malinin et al. 2021). The phosphorothioate modification in the phosphate backbone protects DNA molecules from degradation by cellular exonucleases. However, 2 major distinctions are noteworthy between our results and previous observations in nonplant systems. First, previous studies in human cell lines indicated that phosphorothioate modifications were required at both the 5′ and 3′ ends of dsODNs to achieve efficient insertions (Tsai et al. 2015; Malinin et al. 2021). Second, in animal and human cell line systems, the c-NHEJ and MMEJ pathways both mediate efficient targeted insertion (Sakuma et al. 2016; Wierson et al. 2020). By contrast, our study indicates that the phosphorothioate linkage at the 3′ rather than the 5′ end of dsODN donors is more critical to substantially improve targeted insertion efficiency. One hypothesis is that, in *S. viridis* protoplasts, 3′ to 5′ end resections, potentially mediated by a 3′ to 5′ exonuclease (such as the Meiotic Recombination 11 [Mre11] homolog) in the unprotected 3′ end of the dsODN donor may lead to extensive DNA resections and then dsODN degradation (Symington 2016). On the contrary, 5′ to 3′ end resections would be limited at the unprotected 5′ end. Thus, the 3′ rather than 5′ end must be protected in dsODNs to prevent them from degradation in plant cells. Further investigation will be required to address these questions.

Our results suggest that targeted integration of chemically protected dsODN donors appears to be primarily through the c-NHEJ pathway in plant cells. In our previous study on CRISPR/Cas9-induced mutagenesis in *S. viridis*, we demonstrated that c-NHEJ is the major pathway to generate small InDels at the *MS26* and *MS45* sites. We only observed MMEJ-mediated mutations at an average frequency of 3.5%, with microhomology sequences ranging from 2 to 12 bp (Weiss et al. 2020). Based on this observation, we designed dsODNs with 12 bp of microhomology sequences to evaluate MMEJ-mediated targeted insertions in this study. However, further investigation with various sizes of microhomology sequences will be needed to thoroughly study the choice in pathways for targeted insertion using chemically modified dsODNs. In addition, the current data cannot completely rule out the possibility that the 1-nt overhangs can be used as MMEJ substrates to facilitate targeted insertions. Such possibilities should be systematically tested using plants defective in the MMEJ or c-NHEJ pathway, such as mutants in *DNA Pol θ*, *Ku70*/*Ku80*, or *Ligase IV* (Miller et al. 2021).

SpCas9 was initially thought to only generate blunt cleavage (Jiang and Doudna 2017). Thus, all phosphorothioate-modified dsODNs used in previous studies have been blunt ended (Lu, Tian, et al. 2020; Malinin et al. 2021). However, in agreement with recent research in human cell lines, we determined that SpCas9 can frequently induce staggered cleavage and generate 1-nt 5′ overhangs at the −4 position upstream of the PAM sequence in plants (Molla and Yang 2020; Schmid-Burgk et al. 2020). An alternative model is that, instead of being produced from staggered cleavage, the overhang structure is generated by an end-processing protein through the addition of 1-nt at the Cas9-induced blunt end. Recently, the end-processing protein, DNA polymerase λ, was shown to be responsible for 1-bp insertions in yeast (*Saccharomyces cerevisiae*) and human cell lines (Lemos et al. 2018; Hussmann et al. 2021). Because this X-family DNA polymerase possesses nearly no activity to transfer nucleotides to DNA blunt ends (Stinson et al. 2020), the 1-bp insertions are most likely derived from Cas9-induced staggered cleavage. Moreover, sequence and epigenetic features were demonstrated to influence the 1-bp InDel frequencies in human cell lines and plants (Schep et al. 2021; Weiss et al. 2022). In this study, we observed highly variable frequencies of 1-bp InDels, ranging from 4.8% to 77.6%, at individual Cas9 target sites. Thus, Cas9 staggered cleavage does not appear to occur with the same frequencies at different target sites. The ratio between staggered and blunt cleavages likely depends on sequence and chromatin contexts. A large-scale analysis with additional Cas9 target sites would be helpful to uncover the principles determining these 2 distinct cleavage modes.

The prevalence of Cas9 staggered cleavage prompted us to design dsODN donors with 1-nt 5′ overhangs complementary to the cleaved targeted site. The use of sticky-ended dsODNs demonstrated directionally targeted insertions with improved precision compared to blunt-ended donors. In all sites tested in *S. viridis*, DOTI donors led to most targeted insertions being in the intended direction, ranging from 65.4% to 92.4%. This significantly improved precision was also observed in 4 out of 5 sites when DOTI donors were compared to blunt donors (Figs. 2E and 3E). The different frequencies of seamless directionally targeted insertions may result from varied staggered versus blunt cleavage rates at different sites. Understanding the factors determining these 2 cleavage modes would help further improve the frequencies of precise and direction-controllable NHEJ-mediated insertions. In addition, an alternative strategy would be to screen for Cas9 variants with increased staggered cleavage rates. In a recent study, the Cas9 variant Cas9_LZ3 was shown to mediate an increased frequency of staggered cleavage in human cell lines (Schmid-Burgk et al. 2020; Bermudez-Cabrera et al. 2021). It would be interesting to test whether this Cas9 variant can improve directionally targeted insertions using the DOTI donors in plants.

The development of the precise and directional dsODN-based targeted insertion approach, DOTI, opens opportunities for many applications. In this study, we successfully tagged 3 endogenous genes with 2 different tags, a 33-bp sequence (encoding the 11 amino acids of the small nanoLuciferase subunit [HiBiT]) and a 66-bp sequence (encoding the 22 amino acids of a 3× FLAG epitope). An efficient HiBiT tagging system would enable antibody-free protein detection and quantification that are not well developed in plant cells (Schwinn et al. 2018). In addition, efficient epitope tagging of endogenous transcription factors would

make it possible to directly study the gene regulatory network their control with their native expression levels as opposed to an overexpression approach (Tu et al. 2020). Moreover, we demonstrated that DOTI can be applied to CRE engineering to confer disease resistance in an inducible and pathogen-specific manner. Given a large number of CREs being identified in response to biotic and abiotic stresses (Schmitz et al. 2022), this approach offers opportunities to engineer new biotic and abiotic tolerance traits.

While *S. pyogenes* Cas9 was the only Cas system tested for directionally targeted insertions in this study, a similar approach can be readily extended to other overhang-producing Cas systems, such as Cas12a or paired Cas9 or Cas12a nucleases. One advantage of using 2 Cas nucleases is to reduce the formation of multimeric dsODNs due to their complementary ends. In addition, although only short DNA sequences (25 to 66 bp) were used in this study, this approach offers opportunities to integrate large DNA fragments in a precise and directional manner. Combined with recombinase technology, our approach would allow to first insert recombinase recognition sites (typically 30 to 200 bp; Coates et al. 2005) as dsODN donors precisely and with high efficiency, followed by integrating large DNA fragments at predefined target sites. One limitation of this study is that direct delivery of modified dsODNs requires either protoplast or biolistic-based transformation methods. The biolistic approach was previously shown to result in either simple or complex insertion events in transformed plants (Liu et al. 2019). Here, all targeted insertion events generated through biolistic transformation appeared to be simple insertions at the target site, with the exception of 1 example in which 2 copies of the dsODN sequence were inserted as a tandem repeat. However, it is very likely that dsODNs can be inserted elsewhere in a nontargeted fashion. The nontargeted insertions can be segregated out through crossing. Nevertheless, our study demonstrates the feasibility to achieve directionally targeted insertions with improved precision via the classic NHEJ pathway. We anticipate continuous improvement of this approach for higher efficiency and with potentially less destructive delivery methods, such as the nanoparticle-based transformation approach (Cunningham et al. 2018), in future studies.

Methods and materials

Plant materials and growth conditions

Green foxtail (*S. viridis*) accession ME034V was used for the isolation of mesophyll protoplasts. Seeds were sown on soil (BM2 seed germination mix from Berger Inc., QC, Canada), and seedlings were allowed to grow for 3 weeks in a 26 °C (day)/22 °C (night) temperature cycle under a 16-h light/8-h dark photoperiod (light intensity 100 μmol/m^2/s; fluorescent lamps from Conviron, CA), with 50% relative humidity in the plant growth chamber. The plants were watered and fertilized every other day.

The rice (*O. sativa*) *japonica* variety Kitaakewas was grown in the University of Missouri greenhouse in a 30 °C (day)/26 °C (night) temperature cycle under a 12-h light/12-h dark photoperiod, with 80% to 85% relative humidity. The soil for growing rice was collected from the field in Bradford Research Center, Columbia, MO, USA.

Plasmid constructs and dsODN preparation

The plasmids containing the CRISPR sgRNAs for *MS26* and *MS45*, the *mGFP* reporter, and the *Cas9* cassette were used as mentioned in the previous study (Weiss et al. 2020). The plasmids harboring the sgRNAs individually targeting 3 *Setaria* genes, *H2A.W* (pTW128), *CCA1* (pJK12), and *Ubiquitin* (pJK16), were assembled into the pMOD_B2303 construct using a Golden Gate assembly method (Čermák et al. 2017). Similarly, the plasmids expressing the other 5 sgRNAs listed in Supplemental Data Set 1 were assembled into the pMOD_B2303 construct. These plasmids were further assembled with pMOD_A1110 (containing SpCas9) and pMOD_C3003 (containing the *mGFP* coding sequence) into the destination vector pTRANS100, by following the Golden Gate assembly method (Čermák et al. 2017). The resulting constructs were used for protoplast transfection using a NucleoBond Xtra plasmid Midi kit (Macherey-Nagel GmbH & Co.KG, Duren, German) according to the manufacturer's protocol.

The sgRNA to target the *xa23* promoter was constructed according to the detailed protocol that was previously described (Char et al. 2019). Briefly, 2 oligonucleotides forming a double-stranded oligonucleotides with 19 base pairs to target *xa23* promoter as the spacer sequence of sgRNA and appropriate 4-nt 5′ overhangs were designed and cloned into pENTR-gRNA1 sequentially using *Btg*Z1and *Bsa*I restriction sites (Supplemental Table S1). The sgRNA cassette was mobilized into the JD633-ccdB SpCas9 binary vector (a gift generously provided by Jorge Dubcovsky's lab) (Debernardi et al. 2020) via the Gateway recombination reaction using LR clonase (Thermo Scientific, MA, USA). *Escherichia coli* strain EPI300 (LGC Biosearch Technologies) was used to transform the CRISPR construct and grown in Luria–Bertani medium containing 25 μg/mL of kanamycin (Thermo Scientific, MA, USA).

For dsODN preparation, complementary single-stranded ODNs were synthesized by Integrated DNA Technologies (Coralville, IA, USA) and mixed in equimolar concentration. The mixed oligos were incubated at 94 °C for 2 min and gradually cooled to room temperature for annealing. The oligonucleotide sequences used for dsODN preparation are listed in Supplemental Data Set 1.

Protoplast isolation, transfection and high-throughput sequencing assay

S. viridis leaves from ∼3-wk-old seedlings were used for protoplast isolation and transfection by following a protocol as described previously with slight modifications (Weiss et al. 2020). For protoplast transfection, 20 pmol of dsODN and 15 μg of *Cas9* plasmid were incubated with 200,000 cells in 20%

(*w*/*v*) polyethylene glycol (PEG) solution (pH 5.7). Transfected cells were washed twice and incubated in W5 buffer (2 mM MES pH 5.7, 150 mM NaCl, 125 mM $CaCl_2$, and 5 mM KCl) in the dark at room temperature for 48 h. Transfection efficiency of the transfected protoplast was monitored through the fluorescence derived from the *mGFP* reporter gene. All transfection experiments were performed in 3 independent replicates for each treatment.

After 48 h of incubation, the protoplasts were collected by centrifugation at 100 × *g* for 5 min at room temperature for genomic DNA extraction and high-throughput sequencing as described previously (Weiss et al. 2020). In brief, PCR amplification was performed using primers flanking each target site. A GoTaq Green Master Mix (Promega Corp., Madison, WI, USA) was used following the manufacturer's instruction, with an annealing temperature of 55 to 58 °C and an extension time of 1 min. PCR amplicons were sequenced using the Illumina paired-end amplicon sequencing service at Genewiz Inc. (South Plainfield, NJ, USA). The raw sequencing reads were processed and analyzed using CRISPRESSO2 with the parameters: (−max_paired_end_reads_overlap 200 −min_average_read_quality 30 −amplicon_min_alignment_score 60 −ignore_substitutions −plot_window_size 60 −min_frequency_alleles_around_cut_to_plot 0.05 −max_rows_alleles_around_cut_to_plot 100) (Clement et al. 2019). The sequencing reads with different mutation types were called and categorized in the CRISPRESSO output files with a cutoff value of 0.2% (the minimum percentage of the total read numbers required for a mutation-containing sequence to be classified as a true editing event). In each sample, to calculate the normalized target insertion frequencies, the number of reads containing the intended insertion was divided by the total number of reads with any mutation at the target site.

Bioluminescence detection and immunoblot analysis

The protoplast samples transfected with constructs carrying the HiBiT tag were analyzed using a Nano-Glo HiBiT Lytic Detection System (Promega Corp., Madison, WI). As per the manufacturer's instructions, protoplasts were resuspended in HiBiT lytic reagent and homogenized for 10 min on an orbital shaker, followed by another 10-min incubation without shaking. The luminescence activity was recorded with an integration time of 10 s with a GloMax Explorer System (Promega Corp., Madison, WI, USA).

Immunoblots were used to detect protein harboring a 3× FLAG epitope tag. Protoplasts were resuspended in protoplast lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.1% [*v*/*v*] Tween 20, 1 mM phenylmethylsulfonyl fluoride [PMSF], 10% [*v*/*v*] glycerol, and 1% [*v*/*v*] ReadyShield Protease Inhibitor Cocktail [Sigma-Aldrich, St. Louis, MO, USA]) for 20 min on a shaker at 4 °C. Total protein content in the supernatant was estimated using a Pierce BCA protein assay kit (Thermo Scientific, Rockford, IL, USA). The protein samples were denatured at 65 °C in 1× LDS sample buffer (Invitrogen, Waltham, MA, USA) with 2.5% (*v*/*v*) β-mercaptoethanol for 5 min. Equal amount of protein was electrophoresed on 10% NuPAGE Bis-Tris gels in 1× MES (2-[N-morpholino] ethane sulfonic acid) SDS buffer (Invitrogen, Waltham, MA, USA) and transferred to a nitrocellulose membrane (BioRad, Hercules, CA, USA). The membrane was stained with Ponceau S for 1 min and washed immediately with TBS. The membrane was then blocked with 5% (*w*/*v*) BSA and incubated with a primary antibody (Monoclonal ANTI-FLAG M2 antibody, 1:1,000; from Sigma-Aldrich, St. Louis, MO, USA) overnight at 4 °C. Membranes were washed and incubated with a secondary antibody (goat antimouse AP, 1:15,000; Sigma Aldrich, St. Louis, MO, USA). The membrane was washed with TBS with 0.1% (*v*/*v*) Tween 20 (TBS-T) and developed using freshly dissolved NBT/BCIP (NBT/BCIP Ready-to-Use Tablets, Roche, 11697471001) solution until a band appeared.

Rice transformation and genotyping for targeted insertion events in transgenic calli and plants

The rice cultivar Kitaake was used for biolistic-mediated transformation based on the protocol described previously (Frame et al. 2000; Banakar and Wang 2020). Briefly, Kitaake callus cells were induced on callus induction medium (Murashige and Skoog [MS] salts) (Cat. #, M519, PhytoTech Labs, Lenexa, Kansas, USA) and 2,4-D (2 mg/L, Cat. #, D309, PhytoTech Labs) for 10 d before particle bombardment with a biolistic gun. Four hours before particle bombardment, the calli were cultured on osmotic MSD (MS + 2,4-D) medium (MS + vitamin, 4.44 g/L, sucrose, 30 g/L, 2,4-D, 2 mg/L, Gelzan, 4.6 g/L, and pH, 5.8, all from PhytoTech Labs). Callus cells derived from embryo scutellum were cobombarded with a mixture of Cas9/sgRNA and dsODNs at a ratio of 0.1 pmol:10 pmol coated onto gold particles (Cat. #, 165232, Bio-Rad) (Li et al. 2016). DNA-coated gold particles were bombarded onto rice callus cells using a PDS-1000 He system (BioRad, Hercules, CA, USA) at the University of Missouri plant transformation facility. One shot with 650-psi rupture disk was performed for each plate of callus cells. The bombarded callus cells were kept in the dark for 20 h before transfer onto the MSD medium containing 50 mg/L of hygromycin (Cat. #, H7772-1G, Sigma, MO, USA).

To screen for targeted insertion events, genomic DNA was extracted from rice calli or leaves from T_0 and T_1 generation plants using the CTAB method (Murray and Thompson 1980). Insertion-specific and *Xa23* gene-specific primers (xa23_F1, xa23_F5, xa23_R5, xa23_F6, xa23_R6, and xa23_R7; Supplemental Data Set 1) were used to detect the candidate AvrXa7/PthXo3 and TalC EBEs knock-in events. PCR products were sequenced by Sanger sequencing at the University of Missouri DNA core facility to confirm the presence of insertion events. Restriction digestion was also used to confirm the insertion events by digesting the PCR amplicons derived from the flanking primers with *Sex*AI for AvrXa7/PthXo3 EBE insertion or *Alw*NI for TalC EBE insertion.

Bacterial resistance assays

Xoo strain PXO86 and AXO1947 were used for BB disease assay using the leaf tip clipping method (Yang and Bogdanove

2013). *Xoo* cells were grown on TSA (1% [*w*/*v*] tryptone, 1% [*w*/*v*] sucrose, 0.1% [*w*/*v*] glutamic acid, 1.5% [*w*/*v*] agar, and pH 6.8) plates for 2 to 3 d at 28 °C. Bacterial cells were scraped from the plates and resuspended in sterile water to an OD_{600} of 0.5. Rice leaves of 2-mo-old T_0 plants were inoculated. Lesion length (cm) was measured 14 d after inoculation.

RNA isolation and gene expression analysis

A syringe without a needle was used to infiltrate bacterial inocula into newly emerged leaves of T_1 plants. The infiltrated leaf segments were collected 24 h after *Xoo* inoculation for total RNA extraction using TRIZOL reagent (Cat. #, 15596018, Thermo Scientific, MA, USA) according to the manufacturer's instructions. Total RNA samples (1 μg each) were treated with DNase I (Cat. #, 7326828, Bio-Rad, CA, USA) before reverse transcription in a 20-μL reaction using the iScript cDNA synthesis kit (Cat. #, 1708891, Bio-Rad, CA, USA). The resulting first-strand cDNA samples were diluted 10 times with RNAse-free water and used for end-point PCR to determine the induction of *Xa23* in the EBE knock-in plants.

Accession numbers

All sequencing data in this study were deposited to GenBank under accession number PRJNA874417. Gene IDs: *SvMS26*, Sevir.9G530800; *SvMS45*, Sevir.9G459500; *SvH2A.W*, Sevir.9G453900; *SvUbi*, Sevir.5G079801; *SvCCA1*, Sevir.6G053000; *Xa23*, KP123634; *xa23*, LOC_Os11g37620; *OsSWEET14*, LOC_Os11g31190; and *Osactin*, LOC_Os03g50885.

Acknowledgments

We thank Shannen Griffiths and Ella Anderson for their technical assistance and Dr. Nathan Springer for helpful discussions.

Author contributions

F.Z. and B.Y. conceived and planned the study. J.K., S.N.C., T.W., H.L., and B.L. planned and performed the experiments. F.Z., B.Y., J.K., and S.N.C. analyzed the data. J.K., F.Z., and B.Y. wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. Map of the CRISPR/Cas9 plasmids used for *S. viridis* protoplast transformation.

Supplemental Figure S2. Sequence alignment from the CRISPRESSO output showing distinct insertion events.

Supplemental Figure S3. The 1-bp InDels frequencies across 10 individual sties in *S. viridis* and comparisons of targeted insertion frequencies between A/T and G/C overhang-containing donors.

Supplemental Figure S4. Comparison of overall targeted insertion frequencies between blunt and overhang donors at 5 targeted sites in *S. viridis*.

Supplemental Figure S5. Inheritance of EBE targeted insertion events in the T_1 generation.

Supplemental Figure S6. Disease phenotypes of T_1 plants derived from 6 EBE targeted insertion lines.

Supplemental Data Set 1. Summary of CRISPR target site and oligonucleotide sequence information

Supplemental Data Set 2. Summary of statistical analyses.

Funding

J.K. and F.Z. were partially supported by the The National Institute of Food and Agriculture (2021-67013-34565) and the National Science Foundation (IOS-2040218) awards. T.W. was supported by a Bernard and Jean Phinney Graduate Fellowship in Plant Biology and a Doctoral Dissertation Fellowship from the University of Minnesota. S.N.C., H.L., B.L., and B.Y. were partially supported by a subaward to the University of Missouri from the Heinrich Heine University of Dusseldorf funded by the Bill and Melinda Gates Foundation (OPP1155704) and the National Science Foundation award (IOS-1936492).

Conflict of interest statement. None declared.

References

Antony G, Zhou J, Huang S, Li T, Liu B, White F, Yang B. Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-*11N3*. Plant Cell 2010:**22**(11): 3864–3876. https://doi.org/10.1105/tpc.110.078964

Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol. 2020:**38**(7):824–844. https://doi.org/10.1038/s41587-020-0561-9

Banakar R, Wang K. Biolistic transformation of Japonica rice varieties. Methods Mol Biol. 2020:**2124**:163–176. https://doi.org/10.1007/978-1-0716-0356-7_8

Bermudez-Cabrera HC, Culbertson S, Barkal S, Holmes B, Shen MW, Zhang S, Gifford DK, Sherwood RI. Small molecule inhibition of ATM kinase increases CRISPR/Cas9 1-bp insertion frequency. Nat Commun. 2021:**12**(1):5111. https://doi.org/10.1038/s41467-021-25415-8

Čermák T, Curtin SJ, Gil-Humanes J, Čegan R, Kono TJY, Konečná E, Belanto JJ, Starker CG, Mathre JW, Greenstein RL, et al. A multipurpose toolkit to enable advanced genome engineering in plants. Plant Cell 2017:**29**(6):1196–1217. https://doi.org/10.1105/tpc.16.00922

Char SN, Li R, Yang B. CRISPR/Cas9 for mutagenesis in rice. Methods Mol Biol. 2019:**1864**:279–293. https://doi.org/10.1007/978-1-4939-8778-8_19

Chen L-Q, Hou B-H, Lalonde S, Takanaga H, Hartung ML, Qu X-Q, Guo W-J, Kim J-G, Underwood W, Chaudhuri B, et al. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature 2010:**468**(7323):527–532. https://doi.org/10.1038/nature09606

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**(1):667–697. https://doi.org/10.1146/annurev-arplant-050718-100049

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol. 2019:**37**(3):224–226. https://doi.org/10.1038/s41587-019-0032-3

Coates CJ, Kaminski JM, Summers JB, Segal DJ, Miller AD, Kolb AF. Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. Trends Biotechnol. 2005:**23**(8):407–419. https://doi.org/10.1016/j.tibtech.2005.06.009

Cunningham FJ, Goh NS, Demirer GS, Matos JL, Landry MP. Nanoparticle-mediated delivery towards advancing plant genetic engineering. Trends Biotechnol. 2018:**36**(9):882–897. https://doi.org/10.1016/j.tibtech.2018.03.009

Debernardi JM, Tricoli DM, Ercoli MF, Hayta S, Ronald P, Palatnik JF, Dubcovsky J. A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. Nat Biotechnol. 2020:**38**(11):1274–1279. https://doi.org/10.1038/s41587-020-0703-0

Dixon AS, Schwinn MK, Hall MP, Zimmerman K, Otto P, Lubben TH, Butler BL, Binkowski BF, Machleidt T, Kirkland TA, et al. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. ACS Chem Biol. 2016:**11**(2):400–408. https://doi.org/10.1021/acschembio.5b00753

Dong OX, Ronald PC. Targeted DNA insertion in plants. Proc Natl Acad Sci U S A. 2021:**118**(22):1–9. https://doi.org/10.1073/pnas.2004834117

Frame BR, Zhang H, Cocciolone SM, Sidorenko LV, Dietrich CR, Pegg SE, Zhen S, Schnable PS, Wang K. Production of transgenic maize from bombarded type II callus: effect of gold particle size and callus morphology on transformation efficiency. In Vitro Cell Dev Biol Plant. 2000:**36**(1):21–29. https://doi.org/10.1007/s11627-000-0007-5

Hussmann JA, Ling J, Ravisankar P, Yan J, Cirincione A, Xu A, Simpson D, Yang D, Bothmer A, Cotta-Ramusino C, et al. Mapping the genetic landscape of DNA double-strand break repair. Cell 2021:**184**(22):5653–5669.e25. https://doi.org/10.1016/j.cell.2021.10.002

Jiang F, Doudna JA. CRISPR–Cas9 structures and mechanisms. Annu Rev Biophys. 2017:**46**(1):505–529. https://doi.org/10.1146/annurev-biophys-062215-010822

Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo J, Anand RP, Waterman DP, Haber JE. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. Proc Natl Acad Sci U S A. 2018:**115**(9):E2040-E2047. https://doi.org/10.1073/pnas.1716855115

Li J, Meng X, Zong Y, Chen K, Zhang H, Liu J, Li J, Gao C. Gene replacements and insertions in rice by intron targeting using CRISPR/Cas9. Nat Plants. 2016:**2**(10):16139. https://doi.org/10.1038/nplants.2016.139

Li Q, Sapkota M, van der Knaap E. Perspectives of CRISPR/Cas-mediated *cis*-engineering in horticulture: unlocking the neglected potential for crop improvement. Hortic Res. 2020:**7**(1):36. https://doi.org/10.1038/s41438-020-0258-8

Liu J, Nannas NJ, Fu F-F, Shi J, Aspinwall B, Parrott WA, Dawe RK. Genome-scale sequence disruption following biolistic transformation in rice and maize. Plant Cell 2019:**31**(2):368–383. https://doi.org/10.1105/tpc.18.00613

Lu Y, Ronald PC, Han B, Li J, Zhu J-K. Rice protein tagging project: a call for international collaborations on genome-wide in-locus tagging of rice proteins. Mol Plant. 2020:**13**(12):1663–1665. https://doi.org/10.1016/j.molp.2020.11.006

Lu Y, Tian Y, Shen R, Yao Q, Wang M, Chen M, Dong J, Zhang T, Li F, Lei M, et al. Targeted, efficient sequence insertion and replacement in rice. Nat Biotechnol. 2020:**38**(12):1402–1407. https://doi.org/10.1038/s41587-020-0581-5

Malinin NL, Lee G, Lazzarotto CR, Li Y, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Iafrate AJ, Le LP, et al. Defining genome-wide CRISPR–Cas genome-editing nuclease activity with GUIDE-seq. Nat Protoc. 2021:**16**(12):5592–5615. https://doi.org/10.1038/s41596-021-00626-x

Manova V, Gruszka D. DNA damage and repair in plants—from models to crops. Front Plant Sci. 2015:**6**(885):1–26. https://doi.org/10.3389/fpls.2015.00885

Maresca M, Lin VG, Guo N, Yang Y. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. Genome Res. 2013:**23**(3):539–546. https://doi.org/10.1101/gr.145441.112

Miller V, Beying N, Schmidt C, Puchta H. Double strand break (DSB) repair pathways in plants and their application in genome engineering. Genome editing for precision crop breeding. Cambridge, UK: Burleigh Dodds Science Publishing; 2021. p. 27–61.

Molla KA, Yang Y. Predicting CRISPR/Cas9-induced mutations for precise genome editing. Trends Biotechnol. 2020:**38**(2):136–141. https://doi.org/10.1016/j.tibtech.2019.08.002

Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980:**8**(19):4321–4326. https://doi.org/10.1093/nar/8.19.4321

Nambiar TS, Baudrier L, Billon P, Ciccia A. CRISPR-based genome editing through the lens of DNA repair. Mol Cell. 2022:**82**(2):348–388. https://doi.org/10.1016/j.molcel.2021.12.026

Oliva R, Ji C, Atienza-Grande G, Huguet-Tapia JC, Perez-Quintero A, Li T, Eom J-S, Li C, Nguyen H, Liu B, et al. Broad-spectrum resistance to bacterial blight in rice using genome editing. Nat Biotechnol. 2019:**37**(11):1344–1350. https://doi.org/10.1038/s41587-019-0267-z

Orlando SJ, Santiago Y, DeKelver RC, Freyvert Y, Boydston EA, Moehle EA, Choi VM, Gopalan SM, Lou JF, Li J, et al. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Res. 2010:**38**(15):e152. https://doi.org/10.1093/nar/gkq512

Puchta H. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot. 2005:**56**(409):1–14. https://doi.org/10.1093/jxb/eri025.

Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. Engineering quantitative trait variation for crop improvement by genome editing. Cell 2017:**171**(2):470–480.e8. https://doi.org/10.1016/j.cell.2017.08.030

Sakuma T, Nakade S, Sakane Y, Suzuki K-IT, Yamamoto T. MMEJ-assisted gene knock-in using TALENs and CRISPR/Cas9 with the PITCh systems. Nat Protoc. 2016:**11**(1):118–133. https://doi.org/10.1038/nprot.2015.140

Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, van Schaik T, Manzo SG, Peric-Hupkes D, van den Berg J, et al. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. Mol Cell. 2021:**81**(10):2216–2230.e10. https://doi.org/10.1016/j.molcel.2021.03.032

Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, Zhang F. Highly parallel profiling of Cas9 variant specificity. Mol Cell. 2020:**78**(4):794–800.e8. https://doi.org/10.1016/j.molcel.2020.02.023

Schmitz RJ, Grotewold E, Stam M. *Cis*-regulatory sequences in plants: their importance, discovery, and future challenges. Plant Cell 2022:**34**(2):718–741. https://doi.org/10.1093/plcell/koab281

Schwinn MK, Machleidt T, Zimmerman K, Eggers CT, Dixon AS, Hurst R, Hall MP, Encell LP, Binkowski BF, Wood KV. CRISPR-mediated tagging of endogenous proteins with a luminescent peptide. ACS Chem Biol. 2018:**13**(2):467–474. https://doi.org/10.1021/acschembio.7b00549

Shi X, Shou J, Mehryar MM, Li J, Wang L, Zhang M, Huang H, Sun X, Wu Q. Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di- and tri-nucleotide CRISPR insertions without template donor. Cell Discov. 2019:**5**(1):53. https://doi.org/10.1038/s41421-019-0120-z

Stinson BM, Moreno AT, Walter JC, Loparo JJ. A mechanism to minimize errors during non-homologous end joining. Mol Cell. 2020:**77**(5):1080–1091.e8. https://doi.org/10.1016/j.molcel.2019.11.018

Symington LS. Mechanism and regulation of DNA end resection in eukaryotes. Crit Rev Biochem Mol Biol. 2016:**51**(3):195–212. https://doi.org/10.3109/10409238.2016.1172552

Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, et al. GUIDE-seq enables

genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol. 2015:**33**(2):187–197. https://doi.org/10.1038/nbt.3117

Tu X, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu P-Y, Shen W, Wei Y, Dai X, Li P, Buckler ES, et al. Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. Nat Commun. 2020:**11**(1):50895101. https://doi.org/10.1038/s41467-020-18832-8

Wei Z, Abdelrahman M, Gao Y, Ji Z, Mishra R, Sun H, Sui Y, Wu C, Wang C, Zhao K. Engineering broad-spectrum resistance to bacterial blight by CRISPR/Cas9-mediated precise homology directed repair in rice. Mol Plant. 2021:**14**(8):1215–1218. https://doi.org/10.1016/j.molp.2021.05.012

Weinthal DM, Taylor RA, Tzfira T. Nonhomologous end joining-mediated gene replacement in plant cells. Plant Physiol. 2013:**162**(1):390–400. https://doi.org/10.1104/pp.112.212910

Weiss T, Crisp PA, Rai KM, Song M, Springer NM, Zhang F. Epigenetic features drastically impact CRISPR–Cas9 efficacy in plants. Plant Physiol. 2022:**190**(2):1153–1164. https://doi.org/10.1093/plphys/kiac285

Weiss T, Wang C, Kang X, Zhao H, Elena Gamo M, Starker CG, Crisp PA, Zhou P, Springer NM, Voytas DF, et al. Optimization of multiplexed CRISPR/Cas9 system for highly efficient genome editing in *Setaria viridis*. Plant J. 2020:**104**(3):828–838. https://doi.org/10.1111/tpj.14949

Wierson WA, Welker JM, Almeida MP, Mann CM, Webster DA, Torrie ME, Weiss TJ, Kambakam S, Vollbrecht MK, Lan M, et al. Efficient targeted integration directed by short homology in zebrafish and mammalian cells. eLife. 2020:**9**(e53968):1–25. https://doi.org/10.7554/eLife.53968

Yamamoto Y, Gerbi SA. Making ends meet: targeted integration of DNA fragments by genome editing. Chromosoma 2018:**127**(4):405–420. https://doi.org/10.1007/s00412-018-0677-6

Yang B, Bogdanove A. Inoculation and virulence assay for bacterial blight and bacterial leaf streak of rice. Methods Mol Biol. 2013:**956**:249–255. https://doi.org/10.1007/978-1-62703-194-3_18

Zuo Z, Liu J. Cas9-catalyzed DNA cleavage generates staggered ends: evidence from molecular dynamics simulations. Sci Rep. 2016:**5**(1):37584. https://doi.org/10.1038/srep37584