Plant Molecular Biology
https://doi.org/10.1007/s11103-018-00819-1

CRISPR/Cas9-mediated targeted T-DNA integration in rice

Keunsub Lee[1,2] · Alan L. Eggenberger[1,2] · Raviraj Banakar[1,2] · Morgan E. McCaw[1,2] · Huilan Zhu[2,3] · Marcy Main[2,3] · Minjeong Kang[1,2,4] · Stanton B. Gelvin[5] · Kan Wang[1,2]

Received: 18 November 2018 / Accepted: 27 December 2018

Abstract

Key message **Combining with a CRISPR/Cas9 system, *Agrobacterium*-mediated transformation can lead to precise targeted T-DNA integration in the rice genome.**

Abstract *Agrobacterium*-mediated T-DNA integration into the plant genomes is random, which often causes variable transgene expression and insertional mutagenesis. Because T-DNA preferentially integrates into double-strand DNA breaks, we adapted a CRISPR/Cas9 system to demonstrate that targeted T-DNA integration can be achieved in the rice genome. Using a standard *Agrobacterium* binary vector, we constructed a T-DNA that contains a CRISPR/Cas9 system using SpCas9 and a gRNA targeting the exon of the rice AP2 domain-containing protein gene *Os01g04020*. The T-DNA also carried a red fluorescent protein and a hygromycin resistance (*hptII*) gene. One version of the vector had *hptII* expression driven by an *OsAct2* promoter. In an effort to detect targeted T-DNA insertion events, we built another T-DNA with a promoterless *hptII* gene adjacent to the T-DNA right border such that integration of T-DNA into the targeted exon sequence in-frame with the *hptII* gene would allow *hptII* expression. Our results showed that these constructs could produce targeted T-DNA insertions with frequencies ranging between 4 and 5.3% of transgenic callus events, in addition to generating a high frequency (50−80%) of targeted indel mutations. Sequencing analyses showed that four out of five sequenced T-DNA/gDNA junctions carry a single copy of full-length T-DNA at the target site. Our results indicate that *Agrobacterium*-mediated transformation combined with a CRISPR/Cas9 system can efficiently generate targeted T-DNA insertions.

Keywords *Agrobacterium*-mediated transformation · T-DNA integration · *Oryza sativa* · CRISPR/Cas9

Introduction

Plant genome engineering technologies using sequence-specific nucleases (SSNs) have created great opportunities for both basic research and crop improvement (Voytas and Gao 2014; Collonnier et al. 2017). Recently developed Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated system 9 (CRISPR-Cas9) systems have been successfully used for targeted genome engineering of many plant species (Carroll 2014; Kim and Kim 2014; Sander and Joung 2014; Yin et al. 2017). These genome engineering reagents are usually delivered to plant cells via two major platforms, *Agrobacterium*-mediated transformation using transferred DNA (T-DNA), and biolistic (particle bombardment) delivery systems.

Agrobacterium-mediated delivery utilizes the bacterium's natural ability to transfer and integrate a fragment of its own DNA (the T-DNA) into the plant genomes (Gelvin 2003). As a plant pathogen, *Agrobacterium* species cause

Keunsub Lee and Alan L. Eggenberger contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11103-018-00819-1) contains supplementary material, which is available to authorized users.

✉ Kan Wang
kanwang@iastate.edu

[1] Crop Bioengineering Center, Iowa State University, Ames, IA 50011, USA

[2] Department of Agronomy, Iowa State University, Ames, IA 50011, USA

[3] Plant Transformation Facility, Iowa State University, Ames, IA 50011, USA

[4] Interdepartmental Plant Biology Major, Iowa State University, Ames, IA 50011, USA

[5] Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Published online: 15 January 2019

crown gall and hairy root disease by inserting T-DNAs, which encode biosynthetic enzymes for plant hormones (i.e., oncogenes) and opines, into host plant genomes (Gelvin 2003, 2017; Tzfira and Citovsky 2006; Nester 2015). Only the border sequences are required for T-DNA transfer, thus oncogenes can be replaced with virtually any DNA sequence (Gelvin 2003).

Compared to the biolistic delivery system, *Agrobacterium*-mediated delivery has many advantages. First, *Agrobacterium* has a large cargo capacity and can efficiently deliver an intact T-DNA molecule over 100 kbp in size (Miranda et al. 1992; Hamilton et al. 1996). The large cargo capacity is ideal to introduce the multiple genes specifying a complex trait for crop improvement or a novel biosynthetic pathway for synthetic biology purposes (Field and Osbourn 2008; Purnick and Weiss 2009; Baltes and Voytas 2015). Recently, Collier et al. (2018) developed a highly efficient gene stacking system which allows assembly of multiple genes into a single T-DNA within *Agrobacterium* cells. Second, *Agrobacterium*-mediated transformation is more likely to deliver single or low copy number T-DNA integration events, thus minimizing undesired insertional mutagenesis or genome rearrangements. Third, highly virulent *Agrobacterium* strains are readily available and the protocols do not require expensive equipment or supplies for transformation. Therefore, *Agrobacterium*-mediated transformation is the most widely used tool for many laboratories, especially those with limited resources.

Agrobacterium-mediated transformation, however, still has constraints that limit utilization of this tool. T-DNA randomly integrates into plant genomes (Kim et al. 2007), and this can lead to poor transgene expression (Hobbs et al. 1990; Peach and Velten 1991; Bochardt et al. 1992; Kilby et al. 1992; Mlynarova et al. 1994; Iglesias et al. 1997) or undesired insertional mutagenesis (Alonso et al. 2003; Sallaud et al. 2004). Therefore, it is highly desirable to develop an efficient approach for targeted T-DNA integration into a specific genomic location to ensure a high level of transgene expression without insertional gene disruption. Double-stranded DNA break (DSB) induction using meganucleases (Salomon and Puchta 1998; Chilton and Que 2003; Tzfira et al. 2003) or X-ray irradiation (Köhler et al. 1989) can significantly enhance T-DNA integration, suggesting that T-DNAs preferentially integrate into DSBs. More recently, zinc finger nucleases (ZFN) (Wright et al. 2005; Cai et al. 2009; Shukla et al. 2009; Townsend et al. 2009; Ainley et al. 2013; De Pater et al. 2013), transcription activator-like effector nucleases (TALEN) (Zhang et al. 2013; Forsyth et al. 2016), and CRISPR/Cas9 systems (Li et al. 2013; Shan et al. 2013; Svitashev et al. 2015; Endo et al. 2016; Dahan-Meir et al 2018) have been adopted for targeted DNA insertion in various plant species.

The CRISPR/Cas9 system introduces highly specific DSBs at target sites, which are specified by the guide RNAs (Jinek et al. 2012). Gene knock-ins or knock-outs can be generated during DSB repair by two dominant pathways, homology-directed repair (HDR) and non-homologous end joining (NHEJ) (Cong et al. 2013; Bibikova et al. 2002). HDR is precise and can lead to gene replacements and knock-ins in the presence of appropriate donor templates (Cong et al, 2013), but it is cell cycle dependent and has a low efficiency in plants (Collonnier et al. 2017). Unlike HDR, NHEJ is not cell cycle dependent and can join DNA ends with or without short insertions or deletions (indels) (Bibikova et al. 2002). The small indels occasionally introduced during NHEJ break repair can generate a premature stop codon or cause a frameshift in the coding sequence, resulting in gene knock-outs. NHEJ can also generate gene knock-ins by stimulating recombination between the DNA ends induced by the CRISPR/Cas9 nuclease and those of a donor template (Voytas and Gao 2014; Li et al. 2017). For example, Li et al. (2017) observed that double-strand DNA (dsDNA) donors without homology arms were integrated into the CRISPR/Cas9 target site in rice. Li et al. (2016) demonstrated that an exon can be replaced by targeting the flanking introns with a pair of gRNAs and providing a donor template containing the same gRNA target sites, further suggesting that NHEJ can facilitate targeted gene knock-ins in plants. Recently, Zhang et al. (2018) observed that T-DNAs could integrate into Cas9 cleavage sites at a high frequency in *Arabidopsis*. Although the precise mechanism of T-DNA integration into plant genomes is not fully understood, integration often resembles that of NHEJ or micro-homology-mediated end joining (MMEJ) DSB repair processes (Gelvin 2017). Thus, the highly specific CRISPR/Cas9 system is very promising for facilitating targeted T-DNA integration into plant genomes. In this study, we demonstrate that CRISPR/Cas9-mediated DSB induction is a simple and efficient approach to achieve targeted T-DNA integration into the rice genome.

Materials and methods

Target gene selection and promoter activity assay

We chose six candidate genes for targeted T-DNA integration. Criteria for selecting candidate genes for targeting were (1) high expression in callus, allowing sufficient expression of the target gene exon-*hptII* gene fusion for transformed callus to survive selection, and (2) existence of a T-DNA disruption line, providing evidence that the gene is not essential for plant survival in the event that homozygous or biallelic mutations occur. Gene expression information was from the Michigan State University Rice Genome Annotation Project (Kawahara et al. 2013; http://rice.plantbiology.msu.edu/)

and T-DNA insertion information was from the Rice Mutant Database (Zhang et al. 2006).

Total RNA was isolated from wild-type Japonica rice (cv. Nipponbare) callus samples using a Qiagen RNeasy Plant Mini Kit according to the manufacturer's instructions. Contaminating genomic DNA was removed by incubation with RNase-free DNase (Qiagen, Hilden, Germany). First-strand cDNA was synthesized from one microgram of total RNA using a SuperScript IV First-Strand cDNA Synthesis System kit (ThermoFisher Scientific, MA, USA). qPCR reactions were carried out using a Qiagen RT^2 SYBR Green Mastermix (Qiagen, Hilden, Germany) in a Mx3005P qPCR System (Agilent Technologies, CA, USA). Each 25 µl reaction contained 1X qPCR master mix, 0.5 µl first-strand cDNA reaction, and 0.3 µM forward and reverse primers. The cycling conditions were: initial activation for 10 min at 94 °C, followed by 40 cycles of 15 s at 94 °C, 30 s at 55 °C, and 30 s at 72 °C. The relative transcript abundance of each candidate gene product was estimated by the $2^{-\Delta\Delta CT}$ method (Livak and Schmitten 2001) and normalized to the two reference genes *OsAct1* and *OsUbi* (Narsai et al. 2010). Oligonucleotides used for RT-qPCR are listed in Table S1.

Vector construction

The vector pPT5, used to create R285, R288, and R289 (Fig. 1), was designed with a promoterless *hptII* gene adjacent to the T-DNA right border so that integration into an exon with the *hptII* gene in-frame would be required for expression. A sequence encoding the flexible

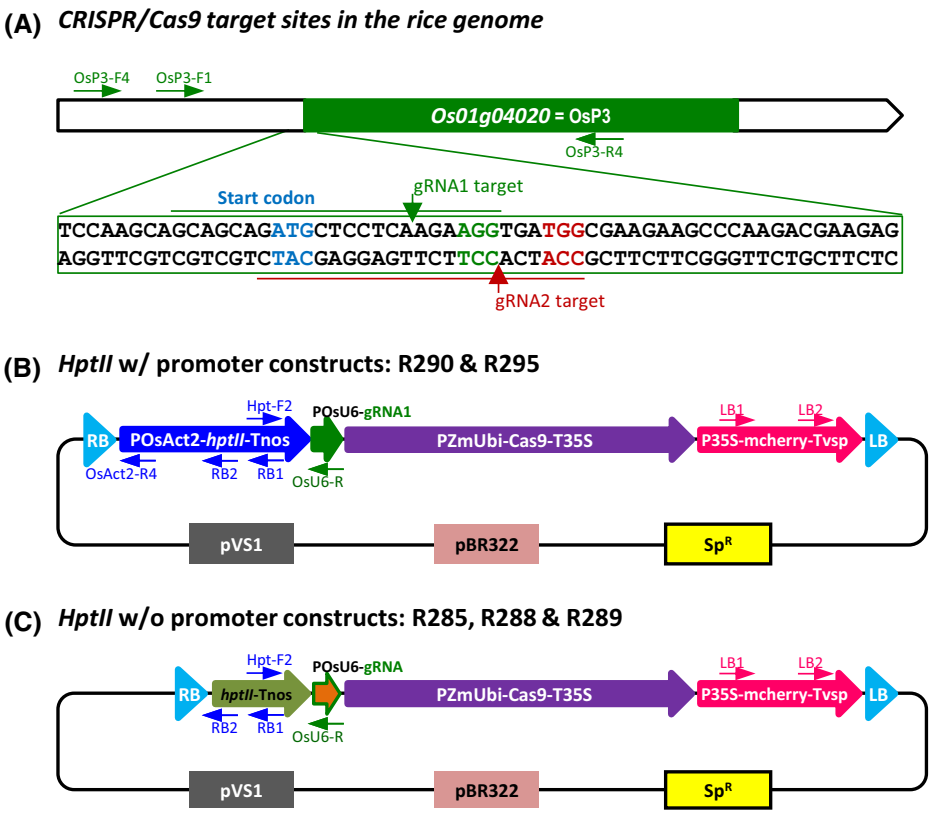


Fig. 1 Schematic illustration of CRISPR/Cas9-mediated targeted T-DNA integration into the rice genome. **a** CRISPR/Cas9 target sites in the rice genome (*Os01g04020* = *OsP3*). The green box represents the *OsP3* exon. Two gRNA target sites are indicated by the green and red lines, respectively. The expected DSB positions, three base pairs upstream of each corresponding 'NGG' PAM, are indicated by arrows. **b** Two pPT5-Act2 constructs for *Agrobacterium*-mediated rice transformation. The hygromycin resistance gene (*hptII*) is driven by the rice *Actin2* promoter. These constructs are identical except for the gRNA sequences. R290, no gRNA (control); R295, POsU6 driving gRNA1 corresponding to the 5′-end of *Os01g04020*. **c** Three pPT5 constructs for *Agrobacterium*-mediated rice transformation. These constructs carry an *hptII* cassette without a promoter, and are identical except for the gRNA sequences. R285, no gRNA (control); R288, POsU6 driving gRNA1 corresponding to the 5′-end of *Os01g04020*; R289, POsU6 driving gRNA2 corresponding to the 5′-end of *Os01g04020*. RB, right border; LB, left border; *hptII*, hygromycin resistance gene; Tnos, *Agrobacterium* nopaline synthase terminator; POsU6, *Oryza sativa* U6 small RNA promoter; PZmUbi, *Zea mays* Ubiquitin promoter; T35S, cauliflower mosaic virus 35S terminator; P35S, cauliflower mosaic virus 35S RNA gene promoter; mCherry, a red fluorescent protein; Tvsp, soybean vegetative storage protein terminator; pVS1, replication origin from *Pseudomonas aeruginosa*; pBR322, replication origin from pMB1; Sp^R, spectinomycin resistance gene. Primers for PCR: OsP3-F4, OsP3-F1, OsP3-R4, Hpt-F2, LB1, LB2, OsAct2-R4, RB2, RB1, and OsU6-R

peptide linker Gly-Gly-Ser-Gly-Gly was inserted immediately upstream of the start codon of the *hptII* gene to reduce the possibility that the protein fusion would interfere with the function of the HPTII protein. The vector was assembled from five components using Gibson assembly (Gibson et al. 2009). Three components: the plasmid backbone, including origins of replication and left and right T-DNA borders; the *hptII*-Tnos (nopaline synthase polyA addition signal) cassette; and the rice-codon-optimized Cas9 cassette (PZmUbi-Cas9-T35S) were from pDW3586 (unpublished-D.A. Wright and M.H. Spalding, Iowa State University, Ames, IA, USA). The P35S (Cauliflower Mosaic Virus 35S promoter)-*mCherry* cassette was from pAN583 (Nelson et al. 2007), and the Tvsp terminator was from pTF101.1 (Paz et al. 2004). R288 and R289 contain gRNA expression cassettes targeting the AP2 domain-containing protein gene *Os01g04020*, whereas R285, a negative targeting control, does not encode a gRNA.

The vector pPT5-Act2 was created using Gibson Assembly to insert the *Act2* promoter from pCOR113 (McElroy et al. 1991) upstream of the *hptII* gene in pPT5 (Fig. 1) so that expression of hygromycin resistance does not require integration of the T-DNA into an actively transcribed coding region. R295 encodes a gRNA1 identical to that of R288, targeting *Os01g04020*. R290, lacking a gRNA, served as a negative control for R295. The sequence information of each component is included in Appendix 1.

Agrobacterium-mediated rice transformation

Agrobacterium-mediated rice transformation was performed as described by Main et al. (2015). Briefly, rice calli initiated from mature dry seeds of cultivar Nipponbare were used as the target material for transformation and regeneration. *A. tumefaciens* EHA101 (Hood et al. 1986) harboring R285, R288, R289, R290, or R295 was grown on YEP medium amended with kanamycin (50 µg/ml) and spectinomycin (100 µg/ml) at 28 °C for 1–2 days. A small loop of cells was removed from the plates and suspended in infection medium containing 100 µM acetosyringone and the cell density was adjusted to OD_{600} of 0.03–0.05. Rice calli were immersed with *Agrobacterium* cultures, air-dried, and incubated for 2 days at 25 °C in the dark. After co-cultivation, infected calli were transferred to fresh selection medium containing 50 mg/l hygromycin B. Three weeks later, hygromycin resistant proliferating calli were subcultured to fresh selection plates for another round of selection. After another 3-week incubation in the dark, proliferating calli were cultured on regeneration medium under light to produce shoots and roots.

Screening for targeted T-DNA integration events

Genomic DNA was isolated from transgenic calli as described (Edwards et al. 1991). Approximately 50 mg of callus was placed in a microfuge tube, frozen with liquid nitrogen, and ground into a powder using a plastic pestle and drill. Five hundred microliters of extraction buffer (200 mM Tris buffer, pH 7.5; 250 mM NaCl; 25 mM Na_2-EDTA; 0.5% sodium dodecyl sulfate [SDS]) was added and the material was vortexed. Following a phenol/chloroform extraction, a chloroform extraction, and precipitation with 0.7 volumes of isopropanol, DNA pellets were suspended in 50 µl Milli-Q H_2O and quantified using a NanoDrop spectrophotometer.

PCR screening of genomic DNAs was performed using high-fidelity PrimeStar GXL DNA Polymerase (Takara Bio, CA, USA) following the manufacturer's protocol. Primers for T-DNA right border (RB) sequences were used to identify targeted T-DNA insertions because integration of the right border is much more precise than that of the left border (LB) (Kleinboelting et al. 2015). Three primer pairs were used to screen for targeted insertion of the promoterless *hptII* gene of R288 and R289 (Fig. 1) into the *Os01g04020* gene: one for detection of targeted insertion, another for detection of T-DNA, and the third for detection of target gene without T-DNA insertion. OsP3-F1 and OsP3-R4 flanked the target site and were expected to yield a 0.5 kbp amplicon in the absence of a T-DNA insertion. Hpt-F2 and OsU6-R primers were used to detect the presence of T-DNA and were designed to amplify a 0.5 kbp fragment which includes part of the *hptII* gene. Targeted insertion of T-DNA was detected with the primer OsP3-F1, upstream of the target sites, and RB1 in the *hptII* gene, which together were expected to yield a 0.5 kbp amplicon if a targeted insertion had occurred.

A similar strategy was used to detect targeted integration of the pPT5-Act2 T-DNA, which had an *Act2* promoter driving expression of the *hptII* gene, allowing T-DNA to integrate anywhere in the genome in either orientation. Target site flanking primers OsP3-F4 and OsP3-R4 generated a 0.5 kbp amplicon in the absence of a biallelic or homozygous T-DNA insertion in the target site, and Hpt-F2 and OsU6-R primers were used to detect the presence of T-DNA as described above. OsP3-F4 and OsAct2-R4 primers were used to detect T-DNA integration in the "forward" orientation, shown in Fig. 1, yielding an amplicon of 0.9 kbp; whereas the primer combination OsP3-R4 and OsAct2-R4 could detect T-DNA integration in the "reverse" orientation, yielding an amplicon of 1.0 kbp. The oligonucleotide sequences are listed in Table S1.

Sequencing was performed by the Iowa State University DNA Facility. Some of the PCR products were sequenced directly employing primers used for the initial PCR reaction. Other PCR products were first cloned into the vector pJET1.2 using a CloneJET PCR Cloning Kit (ThermoFisher

Scientific, MA, USA) and sequenced using primers annealing to insert flanking sites in the vector.

Targeted mutagenesis analysis

CRISPR/Cas9-induced mutations were detected by sequencing 0.5 kbp fragments amplified by PCR using the OsP3-F4 and OsP3-R4 primers (Table S1). PCR conditions were as described above for screening targeted T-DNA integration events. PCR products were treated with ExoSAP-IT reagent (ThermoFisher Scientific, MA, USA) and sequenced using the OsP3-R4 primer. The trace files were analyzed using the tools TIDE (Brinkman et al. 2014) and ICE (Hsiau et al. 2018).

T-DNA/rice genomic (g)DNA junction sequencing

T-DNA/rice gDNA junctions were sequenced using the protocol of Thole et al. (2009). Briefly, 200 ng of gDNA was digested with *Bfa*I overnight at 37 °C and the enzyme was inactivated by incubating for 20 min at 80 °C. *Bfa*I adapters were ligated overnight at room temperature and the T4 DNA ligase was inactivated by incubation at 65 °C for 10 min. First-round PCR was performed using the adapter-specific primer AP1 and T-DNA specific primers RB1 or LB1 (Table S1). Each 20 µl PCR mixture contained 1X PrimeStar buffer, 125 µM dNTPs, 0.2 µM primers, and 1 unit of PrimeStar GXL DNA polymerase. The thermocycling conditions were as follows: initial denaturation for 2 min at 98 °C, followed by 30 cycles of 10 s at 98 °C, 15 s at 55 °C, 30 s at 72 °C, and a final extension for 5 min at 72 °C. Second-round PCR was carried out using the primers AP2 and LB2 or RB2 (Table S1) as described above except that 2 µl of first-round PCR product was used as a template in a 50 µl reaction. PCR amplifications were verified by 1% agarose gel electrophoresis and the PCR products were cloned into the vector pJET1.2 (ThermoFisher Scientific, MA, USA). At least eight clones were sequenced for each gDNA sample, and the sequencing results were analyzed by a BLAST search against the rice genome sequence from the GenBank database (*Oryza sativa* subsp. japonica).

T-DNA copy number estimation

To estimate T-DNA copy numbers in the transgenic rice samples, we selected a single copy reference gene *OsUBC* (*Os02g42314*; Jain et al. 2006). A control vector, pKL1026, was constructed by inserting both the T-DNA *hptII* gene fragment (395 bp) and an *OsUBC* gene fragment (518 bp) into pJET1.2. PCR amplifications were carried out using Phusion High-Fidelity DNA polymerase (ThermoFisher Scientific, MA, USA), and Hpt-F1/R1 and OsUBC-F1/R1 primers (Table S1). The *hptII* fragment was directly cloned into pJET1.2 and the *OsUBC* fragment was inserted using two restriction enzymes *Not*I and *Xho*I. The resulting vector pKL1026 was sequenced to verify the insertion of both *hptII* and *OsUBC* fragments.

Quantitative PCR (qPCR) was carried out as described above using a Qiagen RT^2 SYBR Green master mix in an Mx3005P qPCR system (Agilent Technologies, CA, USA). Serially diluted pKL1026 DNA was used to generate standard curves for both *hptII* and *OsUBC*. Because there are two alleles for the single copy reference gene (Jain et al. 2006), a 1:2 ratio of *hptII* to *OsUBC* represents a single copy T-DNA.

Results

Experimental design and constructs for *Agrobacterium*-mediated rice transformation

We adopted a CRISPR/Cas9 system to generate DSBs to facilitate targeted T-DNA integration into the rice genome. We initially selected six candidate target genes based on their known expression in rice plant tissues and their lack of reported known essential role for growth and development based on insertional mutagenesis (Kawahara et al. 2013; Zhang et al. 2006) (Fig. S1). To ensure strong expression of the *hptII* selection marker gene following T-DNA integration, we assayed the relative activity of the promoters for these six genes in rice calli using reverse transcription quantitative PCR (RT-qPCR) of their transcripts. We chose *OsP3* (*Os01g04020*, AP2 domain containing protein) as our target for T-DNA integration because the transcripts of *OsP3* were most abundant (Fig. S1) and all the T1 T-DNA insertional mutant lines exhibited normal phenotypes (mutant ID: 04Z11EB52 in Rice Mutant Database, rmd.ncpgr.cn; Zhang et al. 2006). Overall, transcript abundance data estimated by RT-qPCR were comparable to the RNA-seq data obtained from the rice genome database (http://rice.plantbiology.msu.edu/), and there was a high level of correlation between them ($R^2 = 0.86$; Fig. S1b).

Using the open reading frame of *OsP3* as the target site to generate DNA breaks, we tested if CRISPR/Cas9-mediated DSB induction could lead to targeted T-DNA integration at this site. We generated two pPT5-Act2 constructs, R295 and R290, containing or lacking gRNA1, respectively, to target the *OsP3* exon (Fig. 1). In both of these constructs, the hygromycin-resistance gene *hptII* was expressed from a rice *Actin2* promoter (*Act2*). R290 lacking the gRNA1 served as a negative control to monitor random T-DNA insertions into the genome.

In addition, we used another approach that requires T-DNA to integrate downstream of a promoter and in-frame with an open reading frame (ORF) (Lindsey et al. 1993) to enrich for targeted T-DNA insertion events at this locus.

We made three constructs using a promoterless *hptII* gene: R288 and R289 are two constructs with gRNA1 or gRNA2, respectively; and R285 is a negative control construct lacking a gRNA (Fig. 1c). Because the *hptII* gene in these constructs has no promoter, we expect that any hygromycin resistant calli recovered from transformation should have its T-DNA integrated in-frame with an ORF under the *OsP3* gene promoter or an off-target promoter. Random T-DNA integration into an expressed gene, and in-frame with that ORF, should be minimal.

We infected a similar number of rice calli with *A. tumefaciens* EHA101 harboring each construct and obtained 26−82 hygromycin-resistant transgenic events per construct (Table 1). Overall, the three promoterless *hptII* constructs (R285, R288, and R289) yielded fewer transgenic events (23−26 events; Table 1) than did the pPT5-Act2 constructs (R290 and R295; 60−76 events). This was expected because only a portion of T-DNAs carrying the promoterless *hptII* gene could lead to successful expression of the selectable marker gene. The expression of the *hptII* gene in these three constructs requires an in-frame T-DNA integration into an expressed gene with a correct orientation. In this case, only 1/6 of targeted events at a particular site would be recovered under the hygromycin selection.

Characterization of targeted T-DNA integration events

PCR analyses identified a total of four targeted T-DNA integration events from 76 transgenic callus events generated by the pPT5-Act2-gRNA1 construct with a frequency of 5.3% (R295; Table 1). As expected, no targeted insertion events were observed in the control construct R290 without the gRNA, indicating that the DSBs generated by the Cas9 nuclease indeed facilitated T-DNA integration into the targeted site (Table 1). Thus, a simple CRISPR/Cas9 system with highly specific gRNAs can effectively generate targeted T-DNA integration. The pPT5 series constructs R288 and R289 resulted in one targeted T-DNA integration event each from 25 to 26 T0 transgenic events (3.8−4.0%), respectively (Table 1). The promoterless constructs require an in-frame translational fusion to express the *hptII* gene. A vector designed to express the *hptII* gene as a transcriptional fusion would be expected to generate three times this number of targeted events.

Because gRNA efficiencies affect DSB induction and subsequently targeted T-DNA integration, we checked if the callus events without targeted T-DNA integrations carry mutations at the target sites, an indication of Cas9 cleavage activity. As summarized in Table 2, 52.4−85.5% of the callus events carried indel mutations at the target site. The gRNA1 of R295 and R288 produced higher mutation frequencies than did the gRNA2 of R289: 71.4−85.5% vs. 52.4% in overall mutation frequency and 64.3−70.9% vs. 19.0% in homozygous (Homo) and biallelic (Bi) mutation frequency. The high gRNA1 efficiency, 71.4−85.5%, suggests that DSB induction by the Cas9 nuclease was not a limiting factor for generating targeted T-DNA integration events. Consistent with previous studies (Jiang et al. 2013; Zhang et al. 2014; Endo et al. 2015), 1 bp insertion and short deletions (< 10 bp) were the most commonly observed mutations (Table S2).

Random T-DNA integration is still predominant as 95−96% of the transgenic callus samples did not carry

Table 1 Summary of Agrobacterium-mediated rice transformation and targeted T-DNA insertion

Construct ID	Construct name	# Callus events examined	# Callus events with targeted insertion	% Targeted insertion
R290	pPT5-Act2	60	0	0.0
R295	pPT5-Act2-gRNA1	76	4	5.3
R285	pPT5	23	0	0.0
R288	pPT5-gRNA1	25	1	3.8
R289	pPT5-gRNA2	26	1	4.0

Table 2 Targeted mutagenesis frequencies in OsP3-targeted callus events

Construct	Mutant callus events					WT	Total analyzed	Mutation frequency (%)	
	Homozygous	Bi-allelic	Mono-allelic	Chimera	Mutant total			Homozygous + Bi-allelic	Overall
R295 (gRNA1)	11	28	4	4	47	8	55	70.9%	85.5%
R288 (gRNA1)	5	4	0	1	10	4	14	64.3%	71.4%
R289 (gRNA2)	3	1	3	4	11	10	21	19.0%	52.4%

targeted insertions, and an average of five T-DNA copies were integrated per genome as determined by qPCR (Table 3). Six events with a targeted insertion had 1−16 copies of T-DNA with an average of 5.1 copies, whereas nine events without a targeted insertion had 2−16 copies of T-DNA with an average of 5.5 copies (Table 3). The relatively high T-DNA copy numbers were due to two events with about 16 copies of T-DNAs (R288-12 and R287-30; Table 3); without these two events, the average T-DNA copy numbers were 2.8 and 4.2 for events with and without a targeted insertion, respectively, which were similar to previous reports (Hiei et al. 1994; Breitler et al. 2004; Yang et al. 2005). The slightly higher T-DNA copy number without targeted T-DNA integrations was not unexpected, given that they were generated from the promoterless *hptII* constructs. Out of four targeted T-DNA integration events of construct R295, one event, R295-61, carried a single copy T-DNA insertion at the Cas9 cut site without additional T-DNAs in the genome (Table 3). Whereas the overall number is small, the fact that one in four events is a single copy event at the desired site is encouraging for targeted T-DNA integration using this approach.

T-DNA/genomic DNA junction sequencing confirms targeted integration

To confirm the targeted T-DNA integration events, we sequenced T-DNA/genomic DNA (gDNA) junctions for all six targeted insertion events (Fig. 2; Table S3). We were unable to sequence the T-DNA-LB/gDNA junction of one event, R295-11, because the PCR approach that we used could not amplify that region. This may have resulted from a large deletion or insertion (Zhang et al. 2018). While all four targeted integration events from R295 had a reverse T-DNA orientation (Fig. 2a), each of the two promoterless *hptII* constructs, R288 and R289, yielded one targeted integration event with a forward T-DNA orientation (R288-12 and R289-9; Fig. 2b). Overall, the 5′ end of the T-DNAs (RB) were relatively well conserved (0–37 bp deletion) compared to the 3′ ends (LB; 1–926 bp deletion); these types of deletions are consistent with previous reports (Gelvin 2017). Three of the six T-DNAs had intact RBs (R288-12, R295-1, and R295-61), two had a one bp deletion (R295-11, R295-54) and one had a 37 bp truncation (R289-9). In contrast, none of the five sequenced LBs had intact sequences as there were at least 43 bp truncated at each sequenced junction (Fig. 2; Table S3). Interestingly, the five T-DNAs integrated at the target site had larger deletions than did the randomly integrated T-DNAs at the LBs: 42–926 bp vs. 1–10 bp (Table S3). Despite the small sample sizes, the difference was statistically significant at $P = 0.05$ level (One-tailed Student's two sample *t*-test; $t = 1.92$, $P = 0.048$).

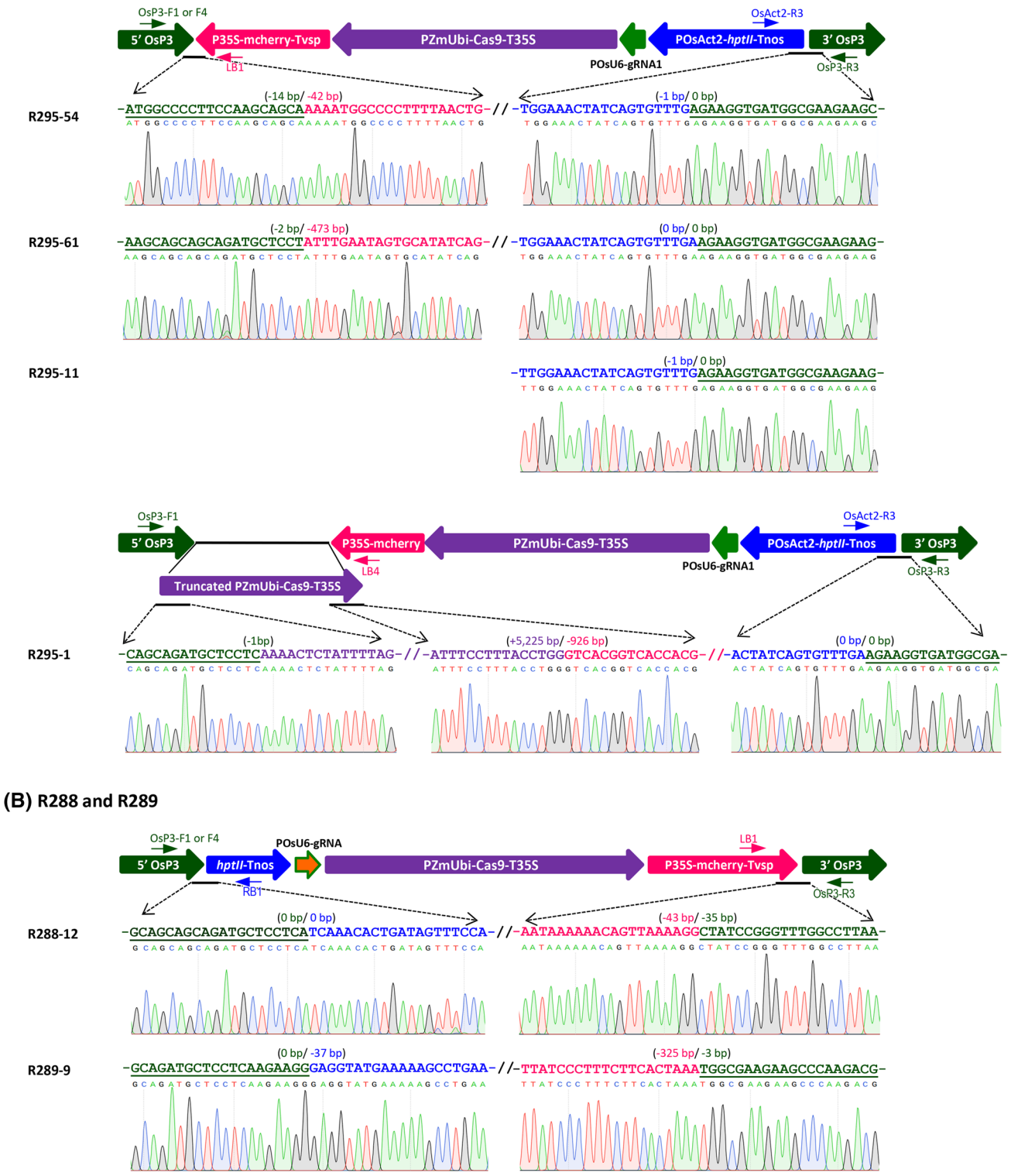
Rice genomic DNA sequences at the T-DNA insertion site were well conserved, as all six events had no DNA loss at the RB-gDNA junctions and 1–35 bp deletions at the LB-gDNA junctions (Fig. 2; Table S3). The length of gDNA deletions at the T-DNA insertion site was within the previously reported ranges (Kleinboelting et al. 2015). R295-1 had an additional partial T-DNA integrated between the 5′ *OsP3* and the truncated LB of the first T-DNA (Fig. 2a). Overall, four of the six targeted integration events had a single-copy near-full-length T-DNA at the Cas9 cut site.

Sequence results also revealed that the 5′ end of the R288-12 T-DNA did not have any DNA sequence loss, and that the hygromycin resistance gene, *hptII*, was not in frame with *OsP3* coding sequence, indicating that this callus event must carry other T-DNAs which allow the expression of functional hygromycin resistance gene. As summarized in Table 3, R288-12 indeed has ~16 T-DNA copies and one of these T-DNAs might have been inserted in-frame into an ORF downstream of a promoter. R289-9 has a 37 bp truncation from the 5′ end and was in frame with the *OsP3* coding sequence. This event had an additional T-DNA inserted in the same chromosome (Chromosome 1) where no known or predicted genes exist (Table S3), suggesting that the *hptII* gene inserted at the target site was responsible for hygromycin resistance of the calli. Notably, the first T-DNA inserted at the target site was on the short arm of Chromosome 1, whereas the second T-DNA was inserted into the long arm, separating the two T-DNAs in R289-9 by ~36 Mbp (locus positions 1,747,127 vs. 37,881,175; Table S3). Given that

Table 3 Estimated T-DNA copy number in callus events

Targeted insertion	Event ID	T-DNA copy #
With targeted insertion	R288-12	16.3
	R289-9	1.7
	R295-1	2.7
	R295-11	3.6
	R295-54	5.3
	R295-61	0.9
	Avg ± SD	**5.1 ± 5.7**
Without targeted insertion	R285-19	5.4
	[a]R286-4	4.5
	[a]R286-11	4.6
	[a]R286-15	4.9
	[a]R287-1	7.1
	[a]R287-9	2.7
	[a]R287-30	15.9
	R288-3	1.6
	R288-4	2.6
	Avg ± SD	**5.5 ± 4.3**

[a] These events were generated from two additional pPT5-based constructs carrying a gRNA targeting OsP3



Chromosome 1 is ~43 Mbp (Sasaki et al. 2002), the two T-DNAs of R289-9 can be segregated in the next generation.

We were curious whether any non-targeted T-DNA integrations, especially from the two constructs carrying the promoterless *hptII* gene, were located in sites with DNA sequences similar to the target site. Sequencing analyses revealed that the genomic locations with non-targeted T-DNA insertions did not have sequences similar to the

◂**Fig. 2** Verified T-DNA/gDNA junctions of targeted insertion events. The junctions of T-DNA and rice genomic DNA were determined by PCR and Sanger sequencing. Primers used to screen targeted insertion events are indicated. Numbers in the parentheses indicate the size of indel mutations at the junction. **a** pPT5-Act2 construct R295 and **b** two pPT5 constructs, R288 and R289. 5′ *OsP3*, the 5′-end of *Os01g04020*; P35S-mCherry-Tvsp, mCherry gene regulated by CaMV 35S promoter and soybean vegetative storage protein terminator; PZmUbi-Cas9-T35S, SpCas9 gene regulated by maize Ubiquitin promoter and CaMV 35S terminator; POsU6-gRNA1, rice U6 promoter driving gRNA1; POsAct2-*hptII*-Tnos, hygromycin resistant gene regulated by rice *Actin2* gene promoter and *Agrobacterium* nopaline synthase gene terminator; 3′ *OsP3*, the 3′-end of *Os01g04020; hptII*-Tnos, promoterless hygromycin resistant gene regulated by *Agrobacterium* nopaline synthase gene terminator; POsU6-gRNAs, rice U6 promoter driving gRNA1 (for R288) or gRNA2 (for R289). Primers for PCR: OsP3-F1 or F4, OsAct2-R3, LB1, OsP3-R3, LB4, RB1

target sites, i.e., potential off-targets of the Cas9-gRNAs, further suggesting that they were random insertions (Fig. S2).

Discussion

We have demonstrated that targeted T-DNA integration can be achieved using a CRISPR/Cas9 system and an *Agrobacterium*-mediated T-DNA delivery method in rice. The overall targeted integration frequency was 4.7% (6 out of 127 total events; Table 1), which is higher than the previously reported targeted T-DNA insertion frequency of 1–3% at the DSBs introduced by meganucleases, such as I-*Sce*I (Salomon and Puchta 1998; Tzfira et al. 2003) and I-*Ceu*I (Chilton and Que 2003) in tobacco plants. With one construct, R295, one of the four targeted events had the desirable single copy T-DNA insertion.

Our results are also comparable to recently reported CRISPR/Cas9-mediated gene knock-in studies in rice. For example, Li et al. (2017) achieved a 2% gene knock-in frequency using *Agrobacterium*-mediated transformation with a 442 bp repair template harboring three substitutions. In that same study, a 1.6 kbp donor DNA lacking homology arms was inserted into the target site with a frequency of 2.2% (Li et al. 2017). Likewise, Wang et al. (2017) observed a 3.9% (6 out of 153 T0 lines) targeted gene knock-in frequency using two constructs delivered by *Agrobacterium*-mediated transformation. Zhang et al. (2018) observed that 6.7−83.3% of *Arabidopsis* T1 lines carried T-DNAs inserted at the Cas9 cut site. Therefore, a CRISPR/Cas9 system combined with an *Agrobacterium*-mediated T-DNA delivery method is a simple and efficient approach for targeted T-DNA integration.

Sequence analyses of the T-DNA/gDNA junctions suggest that the mechanism of targeted T-DNA integration into the CRISPR/Cas9-induced DSBs resembles that of natural T-DNA integrations. Most of the targeted integration events (5/6) had none or only 1 bp of homology in the RB/gDNA junctions suggesting that they were repaired by canonical NHEJ, whereas that of R289-9, which had a 37 bp deletion (Fig. 2b), might have been repaired by micro-homology-mediated end-joining (MMEJ) (Bétermier et al. 2014). The length of deletions in both gDNA and T-DNA borders are within the range of previous reports (Salomon and Puchta 1998; Chilton and Que 2003; Tzfira et al. 2003; Kleinboelting et al. 2015). However, the five T-DNAs integrated at the target site had larger deletions than did the four randomly inserted T-DNAs, which had 1–10 bp deletions (Table S3). The larger deletions might have resulted from delayed joining of LB-gDNAs due to persistent binding of Cas9 to DSBs (Jinek et al 2014; Nishimasu et al. 2014; Richardson et al. 2016; Clarke et al. 2018). It has been shown that Cas9 binding to DSB is persistent and can last ~ 5.5 h in vitro (Richardson et al. 2016). Clarke et al. (2018) also showed that persistent binding of Cas9 protein interferes with end joining by repair enzymes.

Persistent Cas9-DSB binding might have affected T-DNA orientation as well. Intriguingly, all four targeted insertion events using the R295 construct had a reverse orientation, whereas R288 and R289 T-DNAs had a forward orientation (Fig. 2). Unlike the promoterless *hptII* constructs (R288 and R289; Fig. 1c), which require forward T-DNA orientation for selectable marker gene expression, the pPT5-Act2 construct (R295; Fig. 1b) does not have such a constraint as *hptII* expression was directed by the rice *Act2* promoter. Although not statistically significant due to the small sample size, these data further suggest that only a portion of the inserted R288 and R289 T-DNAs would have resulted in in-frame fusion of the *hptII* coding sequence with the *OsP3* ORF. Clarke et al. (2018) showed that Cas9 can be efficiently dislodged by RNA polymerase only when a gRNA targets the template strand, indicating that Cas9 has a stronger association to the protospacer side than to the PAM at the DSB site. It is thus likely that the PAM side of the DSB end (i.e., 3′ of *OsP3*) was released first and joined with the 5′ end of the T-DNA (RB) by NHEJ. Designing gRNAs to target the non-template strand of the *OsP3* gene might facilitate directional T-DNA integration. Taken together, our results suggest that CRISPR/Cas9-induced DSBs can achieve targeted T-DNA integration with a reasonable efficiency.

Our promoterless *hptII* constructs, R288 and R289, did not yield the predicted higher targeted T-DNA integration frequency compared to the pPT5-Act2-gRNA1 construct, R295 (Table 1). This result might be attributable to two main factors. First, these constructs were designed using a translational fusion strategy, which requires in-frame insertion of the *hptII* coding sequence with the target gene ORF. With the translational fusion design only one in six of the targeted integration events would be predicted to have the

correct orientation and reading frame for translation of the desired *OsP3*/*hptII* fusion protein. Any insertion of T-DNA with expression of *hptII* from the *OsAct2* promoter, on the other hand, would confer resistance to hygromycin selection. Second, as explained above, persistent Cas9 binding to the target site might affect the orientation of T-DNAs integrated at the target site. Based on the observation that all four targeted integration events generated by R295 contained reverse orientated T-DNAs, persistent Cas9 binding to the protospacer sequence at the target site might further hinder forward T-DNA integration, which is required for both translational and transcriptional fusion constructs. Therefore, a vector designed to express an *hptII* gene as a transcriptional fusion and CRISPR/Cas9-gRNA targeting the non-template strand of a target gene would be expected to yield an enhanced targeted T-DNA integration frequency.

One of the advantages of *Agrobacterium*-mediated transformation is the large T-DNA cargo capacity. In this study, we were able to deliver ~ 11 kbp full-length T-DNAs into a target site in the rice genome. Although we did not test the T-DNA cargo capacity, Hamilton et al. (1996) showed that a 150 kbp T-DNA could be delivered to plants via *Agrobacterium*-mediated transformation. Miranda et al. (1992) showed that even larger DNA molecules could be introduced into plants. Recently, Collier et al. (2018) developed an efficient in vivo gene stacking system called "GA*A*NTRY" (Gene Assembly in *Agrobacterium* by Nucleic acid Transfer using Recombinase technologY) and assembled 10 genes into a 28.5 kbp T-DNA. Combining CRISPR/Cas9-mediated DSB induction with an in vivo gene stacking system for T-DNA delivery has great potential for crop improvement. This approach brings two distinct advantages. First, targeted T-DNA integration into a verified genomic location would ensure optimal expression of the transgenes (Matzke and Matzke 1998) without unwanted insertional mutagenesis. Second, complex traits which require multiple genes, such as complex biosynthesis pathways (Purnick and Weiss 2009; Boyle et al. 2017), or multiple traits could be efficiently assembled into a single T-DNA (Collier et al. 2018) and transferred to plants. Such a platform would abrogate the necessity for sequential transformations (Ainley et al. 2013), thus dramatically reducing the time and resources to obtain the final products.

Although we obtained targeted integration, random T-DNA insertion events were still predominant; ~ 95% of the callus events did not carry targeted T-DNA integration events (121 out of 127 total events; Table 1). Because Cas9 and gRNA cassettes were encoded by the T-DNA, it is possible that some T-DNAs might have integrated into the rice genome before the DSB was induced by Cas9 nuclease at the target site. However, we did obtain a single copy T-DNA integration event at the target site without additional T-DNAs integrating into the genome (R295-61; Table 3), indicating that a desired targeted T-DNA integration could be achieved within one generation without subsequent breeding procedures to segregate other T-DNAs inserted elsewhere in the genome.

In summary, this study indicates that a CRISPR/Cas9 system can generate targeted T-DNA integration events in rice; this system can be easily applied to other plants. Combined with existing in vivo gene stacking systems such as "GA*A*NTRY" (Collier et al. 2018), CRISPR/Cas9 and *Agrobacterium*-mediated T-DNA delivery systems are poised to bring valuable applications for crop improvement via targeted genome engineering.

Acknowledgements We thank David Wright for providing the binary vector pDW3586. This project was partially supported by the Agriculture and Food Research Initiative Competitive Grant no. 2016–06247 from the USDA National Institute of Food and Agriculture (NIFA) to K.W., National Science Foundation Plant Genome Research Program Grant no IOS 1725122 to S.B.G. and K.W., by the USDA NIFA Hatch project # IOW04341, by State of Iowa funds, and by Crop Bioengineering Center of Iowa State University.

Author contributions KW, ALE and KL designed the experiments; HZ and MM performed the *Agrobacterium*-mediated rice transformation; KL, ALE, RB, MEM, and MK analyzed the transgenic plants; KL, ALE, HZ, SBG and KW analyzed the data, and wrote and edited the manuscript. All authors contributed to discussion and revision of the manuscript

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interests.

References

Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R et al (2013) Trait stacking via targeted genome editing. Plant Biotechnol J 11:1126–1134

Alonso JM, Stepanova AN, Jeisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653–657

Baltes NJ, Voytas DF (2015) Enabling plant synthetic biology through genome engineering. Trends Biotechnol 33:120–131

Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. Genetics 161:1169–1175

Bochardt A, Hodal L, Palmgren G, Mattsson O, Okkels FT (1992) DNA methylation is involved in maintenance of an unusual expression pattern of an introduced gene. Plant Physiol 99:409–414

Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenetic to omnigenic. Cell 15:1177–1186

Breitler JC, Meynard D, Van Boxtel J, Royer M, Bonnot F, Cambillau L, Guiderdoni E (2004) A novel two T-DNA binary vector allows efficient generation of marker-free transgenic plants in three elite cultivars of rice (*Oryza sativa* L.). Transgenic Res 13:271–287

Brinkman EK, Chen T, Amendola M, van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res 42:e168

Cai CQ, Doyon Y, Ainley WM, Miller JC, Dekelver RC, Moehle EA et al (2009) Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol Biol 69:699–709

Carroll D (2014) Genome engineering with targetable nucleases. Annu Rev Biochem 83:409–439

Chilton M-D, Que Q (2003) Targeted integration of T-DNA into the tobacco genome at double stranded breaks: New insights on the mechanism of T-DNA integration. Plant Physiol 133:956–965

Clarke R, Heler R, MacDougall MS, Yeo NC, Chavez A, Regan M, Hanakahi L, Church GM, Marraffini LA, Merrill BJ (2018) Enhanced bacterial immunity and mammalian genome editing vir RNA-polymerase-mediated dislodging of Cas9 from double-strand DNA breaks. Mol Cell 71:42–55.e8

Collier R, Thomson JG, Thilmony R (2018) A versatile and robust *Agrobacterium*-based gene stacking system generates high quality transgenic *Arabidopsis* plants. Plant J. https://doi.org/10.1111/tpj.13992

Collonnier C, Guyon-Debast A, Maclot F, Mara K, Charlot F, Nogué F (2017) Towards mastering CRISPR/induced gene knock-in in plants: survey of key features and focus on the model *Physcomitrella patens*. Methods 121–122:103–117

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823

Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A, Levy AA (2018) Efficient *in planta* gene targeting in tomato using gemini viral replicons and the CRISPR/Cas9. Plant J 95:5–16

De Pater S, Pinas JE, Hooykaas PJ, van der Zaal BJ (2013) ZFN-mediated gene targeting of the *Arabidopsis* protoporphyrinogen oxidase gene through *Agrobacterium*-mediated floral dip transformation. Plant Biotechnol J 11:510–515

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res 19:1349

Endo M, Mikami M, Toki S (2015) Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. Plant Cell Physiol 56:41–47

Endo M, Mikami M, Toki S (2016) Biallelic gene targeting in rice. Plant Physiol 170:667–677

Field B, Osbourn AE (2008) Metabolic diversification-independent assembly of operon like gene clusters in different plants. Science 320:543–547

Forsyth A, Weeks T, Richael C, Duan H (2016) Transcription activator-like effector nucleases (TALEN)-mediated targeted DNA insertion in potato plants. Front Plant Sci 7:1572. https://doi.org/10.3389/fpls.2016.01572

Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67:16–37

Gelvin SB (2017) Integration of *Agrobacterium* T-DNA into the Plant Genome. Annu Rev Genet 51:195–217

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA III, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345

Hamilton CM, Frary A, Lewis C, Tanksley SD (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci U S A 93:9975–9979

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6:271–282

Hobbs SLA, Kpodar P, DeLong CMO (1990) The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. Plant Mol Biol 15:851–864

Hood EE, Helmer GL, Fraley RT, Chilton MD (1986) The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T-DNA. J Bacteriol 168:1291–1301

Hsiau T, Maures T, Waite K, Yang J, Kelso R, Holden K, Stoner R (2018) Inference of CRISPR edits from Sanger trace data. bioRxiv. https://doi.org/10.1101/251082

Iglesias VA, Moscone EA, Papp I, Neuhuber F, Michalowski S, Phelan T, Spiker S, Matzke M, Matzke AJM (1997) Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. Plant Cell 9:1251–1264

Jain M, Nijhawan A, Tyagi AK, Khurana JP (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. Biochem Biophys Res Commun 345:646–651

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41:e188

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S et al (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 343:1247997

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice (N Y) 6:4. https://doi.org/10.1186/1939-8433-6-4

Kilby NJ, Leyser HMO, Furner IJ (1992) Promoter methylation and progressive transgene inactivation in *Arabidopsis*. Plant Mol Biol 20:103–112

Kim H, Kim JS (2014) A guide to genome engineering with programmable nucleases. Nat Rev Genet 15:321–334

Kim SI, Veena, Gelvin SB (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. Plant J 51:779–791

Kleinboelting N, Huep G, Appelhagen I, Viehoever P, Li Y, Weisshaar B (2015) The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. Mol Plant 8:1651–1654

Köhler F, Cardon G, Pöhlman M, Gill R, Schieder O (1989) Enhancement of transformation rates in higher plants by low-dose irradiation: are DNA repair systems involved in the incorporation of exogenous DNA into the plant genome? Plant Mol Biol 12:189–199

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. Nat Biotechnol 31:688–691

Li J, Meng X, Zong Y, Chen K, Zhang H, Liu J, Li J, Gao C (2016) Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. Nat Plants 2:16139

Li J, Sun Y, Du J, Zhao Y, Xia L (2017) Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. Mol Plant 6:526–529

Lindsey K, Wei W, Clarke MC, McArdle HF, Rooke LM, Topping JF (1993) Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. Transgenic Res 2:33–47

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408

Main M, Frame B, Wang K (2015) Rice, Japonica (*Oryza sativa* L.). In: Wang K (ed) *Agrobacterium* protocols, 3rd edn. Springer, New York, pp 169–180

Matzke AJM, Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol 1:142–148

McElroy D, Blowers AD, Jenes B, Wu R (1991) Construction of expression vectors based on the rice actin 1 (*Act1*) 5′ region for use in monocot transformation. Mol Gen Genet 231:150–160. https://doi.org/10.1007/BF00293832

Miranda A, Janssen G, Hodges L, Peralta EG, Reem W (1992) *Agrobacterium tumefaciens* transfers extremely long T-DNAs by a unidirectional mechanism. J Bacteriol 174:2288–2297

Mlynarova L, Loonen A, Heldens J, Jansen RC, Keizer P, Stiekema WJ, Nap JP (1994) Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. Plant Cell 6:417–426

Narsai R, Ivanova A, Ng S, Whelan J (2010) Defining reference genes in *Oryza sativa* using organ, development, biotic and abiotic transcriptome datasets. BMC Plant Biol 10:56. https://doi.org/10.1186/1471-2229-10-56

Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. Plant J 51:1126–1136

Nester EW (2015) *Agrobacterium*: Nature's genetic engineer. Front Plant Sci 5:730

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156:935–949

Paz MM, Shou H, Guo Z, Zhang Z, Banergee AK, Wang K (2004) Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. Euphytica 136:67–179

Peach C, Velten J (1991) Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. Plant Mol Biol 17:49–60

Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422

Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol 34:339–344

Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, Piegu B, Droc G, Regad F, Bourgeois E, Meynard D et al (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. Plant J 39:450–464

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J 17:6086–6095

Sander JD, Joung JK (2014) CRISPR/Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:347–355

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y et al (2002) The genome sequence and structure of rice chromosome 1. Nature 420:312–316

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, Gao C (2013) Targeted genome modification of crop plants using a CRISPR/Cas system. Nat Biotechnol 31:686–688

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE et al (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature 459:437–441

Svitashev S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM (2015) Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. Plant Physiol 169:931–945

Thole V, Alves SC, Worland B, Bevan MW, Vain P (2009) A protocol for efficiently retrieving and characterizing flanking sequence tags in *Brachypodium distachyon* T-DNA insertional mutants. Nature Protocol 4:650–661

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK et al (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459:442–445

Tzfira T, Citovsky V (2006) *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. Curr Opin Biotechnol 17:147–154

Tzfira T, Frankman LR, Vaidya M, Citovsky V (2003) Site-specific integration of *Agrobacterium tumefaciens* T-DNA via double-stranded intermediates. Plant Physiol 133:1011–1023

Voytas DF, Gao C (2014) Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biol 12:e1001877

Wang M, Lu Y, Botella JR, Mao Y, Hua K, Zhu JK (2017) Gene targeting by homology-directed repair in rice using a geminivirus-based CRPSPR/Cas9 system. Mol Plant 10:1007–1010

Wright DA, Townsend JA, Winfrey RJ Jr, Irwin PA, Rajagopal J, Lonosky PM et al (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44:693–705

Yang L, Ding J, Zhang C, Jia J, Weng H, Liu W, Zhang D (2005) Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR. Plant Cell Rep 23:759–763

Yin K, Gao C, Qiu JL (2017) Progress and prospects in plant genome editing. Nat Plants 3:17107

Zhang J, Li C, Wu C, Xiong L, Chen G, Zhang Q (2006) RMD: a rice mutant database for functional analysis of the rice genome. Nucleic Acids Res 34:D745–D748

Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG et al (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol 161:20–27

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N et al (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J 12:797–807

Zhang Q, Xing HL, Wang ZP, Zhang HY, Yang F, Wang XC, Chen QJ (2018) Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in *Arabidopsis* and its prevention. Plant Mol Biol 96:445–456