nature biotechnology

LETTERS

https://doi.org/10.1038/s41587-020-0581-5

Check for updates

Targeted, efficient sequence insertion and replacement in rice

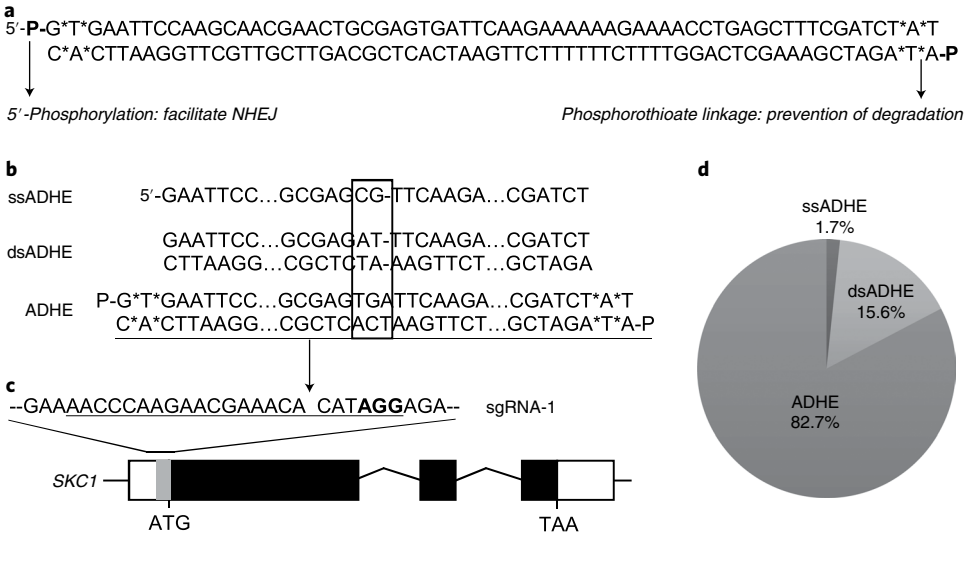
Yuming Lu[1,4], Yifu Tian[1,4], Rundong Shen[1], Qi Yao[1], Mugui Wang[1], Mei Chen[1], Jinsong Dong[1], Tongen Zhang[1], Feng Li[2], Mingguang Lei[1] and Jian-Kang Zhu[1,3]✉

CRISPR–Cas9 methods have been applied to generate random insertions and deletions, large deletions, targeted insertions or replacements of short sequences, and precise base changes in plants[1–7]. However, versatile methods for targeted insertion or replacement of long sequences and genes, which are needed for functional genomics studies and trait improvement in crops, are few and largely depend on the use of selection markers[8–11]. Building on methods developed in mammalian cells[12], we used chemically modified donor DNA and CRISPR–Cas9 to insert sequences of up to 2,049 base pairs (bp), including enhancers and promoters, into the rice genome at an efficiency of 25%. We also report a method for gene replacement that relies on homology-directed repair, chemically modified donor DNA and the presence of tandem repeats at target sites, achieving replacement with up to 130-bp sequences at 6.1% efficiency.

In mammalian cells, the use of a blunt, 5′-phosphorylated, double-stranded oligodeoxynucleotide (dsODN), bearing two phosphorothioate linkages at the 5′- and 3′-ends of both DNA strands, led to robust targeted integration of the oligodeoxynucleotide[12]. The phosphorothioate-linkage modification was designed to stabilize the oligos in cells and the 5′-phosphorylation could facilitate nonhomologous end joining (NHEJ) which acts as a major pathway to repair double-stranded breaks (DSBs), especially in cultured cells. In cultured plant cells for regeneration of plantlets, such as rice callus cells, NHEJ is also the predominant DSB repair pathway[10,13]. Therefore, it is possible that this type of modified dsODNs can improve the efficiency of targeted insertion in plant cells. To test this hypothesis, a 60-bp translational enhancer (ADHE) from the 5′ untranslated region (UTR) of rice *ADH1* (alcohol dehydrogenase 1)[14] was used as the donor DNA for insertion into the major salt tolerance locus *SKC1* in rice (Supplementary Table 1)[15]. As shown in Fig. 1a, the in vitro synthesized ADHE donor DNA was flanked by two additional nucleotides with phosphorothioate-linkage and 5′-phosphorylation modifications (ADHE; see Supplementary Fig. 1b). To compare with traditional donor DNA, both unmodified single- and double-stranded oligodeoxynucleotides (ssADHE and dsADHE) were also synthesized, bearing three-nucleotide polymorphisms for detection (Fig. 1b and Supplementary Fig. 1b). A single guide RNA (sgRNA) targeting the 5′ UTR was designed (sgRNA-1) and constructed into the CRISPR–Cas9 vector pCBSG032 (Fig. 1c and Supplementary Fig. 1a). The three donor DNA oligos were mixed in equimolar proportions and introduced into the rice calli of Zhonghua11 (ZH11), together with the CRISPR–Cas9 plasmid DNA (sgRNA-1) using particle bombardment. Hundreds of hygromycin-resistant calli were obtained after two rounds of selection on hygromycin. Approximately 200 of these hygromycin-resistant calli were mixed together for genomic DNA extraction. DNA fragments were directly amplified using primers (SKC1-F and SKC1-R) flanking the target site of sgRNA-1 (see Supplementary Fig. 2). Targeted insertion of ADHE is expected to produce a larger amplicon (188 bp) than without the insertion (124 bp), which was then sequenced using next-generation sequencing (NGS). As shown in Fig. 1d,e, targeted insertion of ADHE was detected in the NGS sequences. Among the NGS reads containing donor sequences, most of them (82.7%) were produced by the modified donor DNA (ADHE), indicating a substantial positive effect of the modifications on targeted insertion in plant cells. Consistent with the known characteristics of NHEJ, bidirectional insertions were detected. Most of the sequences harbored indels at the 5′- and/or 3′-junctions of the insertion, which may not affect the functions of the UTR and ADHE. We found that 10.9% of the ADHE insertions were seamless. These results suggested that it is feasible to use modified donor DNA for efficient targeted insertion in plant cells.

To assess the targeted insertion efficiency in stable transgenic plants, the *DRO1* (Deeper Rooting 1) gene was chosen as a target. *DRO1* is a major quantitative trait locus (QTL) controlling root growth angle in rice[16]. An sgRNA target site within the 5′ UTR of *DRO1* was selected (sgRNA-2) and corresponding CRISPR–Cas9 plasmid was constructed (Fig. 2a). Calli of ZH11 were transformed with ssADHE, dsADHE or the chemically modified ADHE, together with the constructed CRISPR–Cas9 plasmid. Dozens of plantlets were regenerated from hygromycin-resistant calli after 10 weeks of selection. As targeted insertion of ADHE could be easily identified using PCR, three pairs of primers were designed for genotyping. Primers (DRO1-F + DRO1-R) flanking the sgRNA-2 target site was used for detecting targeted insertion events and the other two pairs of primers (DRO1-F2 + ADHE-R and DRO1-F2 + ADHE-F) were used to determine the direction of ADHE insertion. Successful insertion of ADHE would result in a larger amplicon, and would also produce a PCR amplicon using ADHE-specific primers, either ADHE-F or ADHE-R (see Supplementary Fig. 3). Mutant plants possessing both types of PCR amplicons were counted as targeted insertion plants. According to the genotyping results, only one targeted insertion mutant (DRO-ADHE no. 40) was identified from 24 transgenic lines using the unmodified dsADHE, and none was identified from 23 transgenic lines when the ssADHE was used. In contrast, 10 of 22 of the T0 transgenic lines were identified as targeted insertion plants when the modified ADHE donor was used (Fig. 2b,c and Supplementary Fig. 4). Two lines

[1]Shanghai Center for Plant Stress Biology and Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China. [2]Shandong Shunfeng Biotechnology Co. Ltd, Jinan, China. [3]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN, USA. [4]These authors contributed equally: Yuming Lu, Yifu Tian. ✉e-mail: jkzhu@psc.ac.cn



e

Summary of targeted insertion frequency in rice callus identified using NGS

Donor	Number of NGS reads			
	Total No. (%)	Forward No. (%)	Reverse No. (%)	Seamless No. (%)
ssADHE	3,368 (1.7)	758 (0.4)	2,610 (1.3)	0
dsADHE	30,372 (15.6)	14,712 (7.5)	15,660 (8.1)	5,398 (2.8)
ADHE	161,454 (82.7)	102,374 (52.4)	59,080 (30.3)	21,280 (10.9)

Fig. 1 | Optimization of donor DNA for targeted insertion in rice. a, Sequence of chemically modified dsODN for ADHE. **b**, Nucleotide polymorphism (boxed) among ssADHE, dsADHE and ADHE. **c**, Schematic diagram of *SKC1* with its sgRNA target (underlined), PAM sequences (emboldened) and inserted donor (gray box) in 5′ UTR (white box). **d**,**e**, Comparison of relative targeted insertion frequencies (**d**) using ssADHE, dsADHE and ADHE in rice calli identified by NGS (**e**). 'Forward' and 'reverse' stand for the directions of ADHE at the target site, and 'seamless' means no indels at the junctions between ADHE and its flanking genomic sequence.

(DRO-ADHE nos. 74 and 76) showed a larger amplicon than the predicted 226 bp, which was probably caused by multiple ADHE insertions at the target site. To further confirm the results, the 226-bp PCR fragments from lines DRO-ADHE nos. 40, 64, 79 and 81 were recovered, and sequencing of these fragments using individual bacterial colonies confirmed the presence of ADHE at the target site in these plants (Fig. 2d and Supplementary Table 2). Three of these sequenced lines contained indels at the junctions between ADHE and the flanking genomic sequence, and one line (no. 79) contained a seamless insertion of ADHE.

The high efficiency of targeted insertion of ADHE in stable transgenic rice plants encouraged us to test the method at another three loci in rice plants. One was *SKC1*, described above, and the other two were *SLR1* and *SOS1*. *SLR1* encodes a DELLA protein regulating rice plant height, and *SOS1* encodes a plasma membrane Na^+/H^+ antiporter critical for plant salt tolerance[17,18]. Target sites within the 5′ UTR of these loci were designed (Fig. 2e,h,k). Calli of ZH11 were transformed with ADHE and the corresponding CRISPR–Cas9 plasmid. Hundreds of plantlets were regenerated from hygromycin-resistant calli. Similar to the genotyping method described above for *DRO1*, three pairs of primers were designed for each locus to identify targeted insertion mutants (Fig. 2f,i,l and Supplementary Figs. 2 and 3). As listed in Table 1, the targeted insertion frequencies at the *SKC1*, *SLR1* and *SOS1* loci were 31.7%, 26.5% and 35.7%, respectively. For each target site, the presence of ADHE in three representative mutants were further confirmed with Sanger sequencing using individual bacterial colonies (Fig. 2g,j,m and Supplementary Table 2).

To assess the insertion of other donors, another two chemically modified dsODNs were prepared: AMVE, a 57-bp translational enhancer from Alfalfa mosaic virus[19], and P1BS, a 26-bp cis-element for the binding of transcription factor PHR1 that regulates plant response to low phosphate stress (see Supplementary Fig. 5)[20]. As listed in Table 1, we targeted AMVE to insert into the 5′ UTRs of four loci, and simultaneously targeted P1BS to insert into the promoters of another four loci (see Supplementary Table 1). Genotyping results of 184 T0 seedlings using donor-specific primers (see Supplementary Fig. 6) showed that targeted insertion frequencies of AMVE ranged from 23.5% to 47.3%. For P1BS, 53.2% of the T0 rice plants (100 out of 188) had P1BS insertion in at least 1 of the target genes, among which 45 lines had P1BS insertions at multiple target genes. The targeted insertion frequencies for each of the four loci ranged from 10.6% to 31.4% (Table 1). Taken together, these results show that targeted insertion of a short DNA fragment could be achieved efficiently using chemically modified dsODNs in rice plants.

To test longer donor sequences for targeted insertion, the strong constitutive promoters CmYLCV9.11 (526 bp) and UBI (2,049 bp)

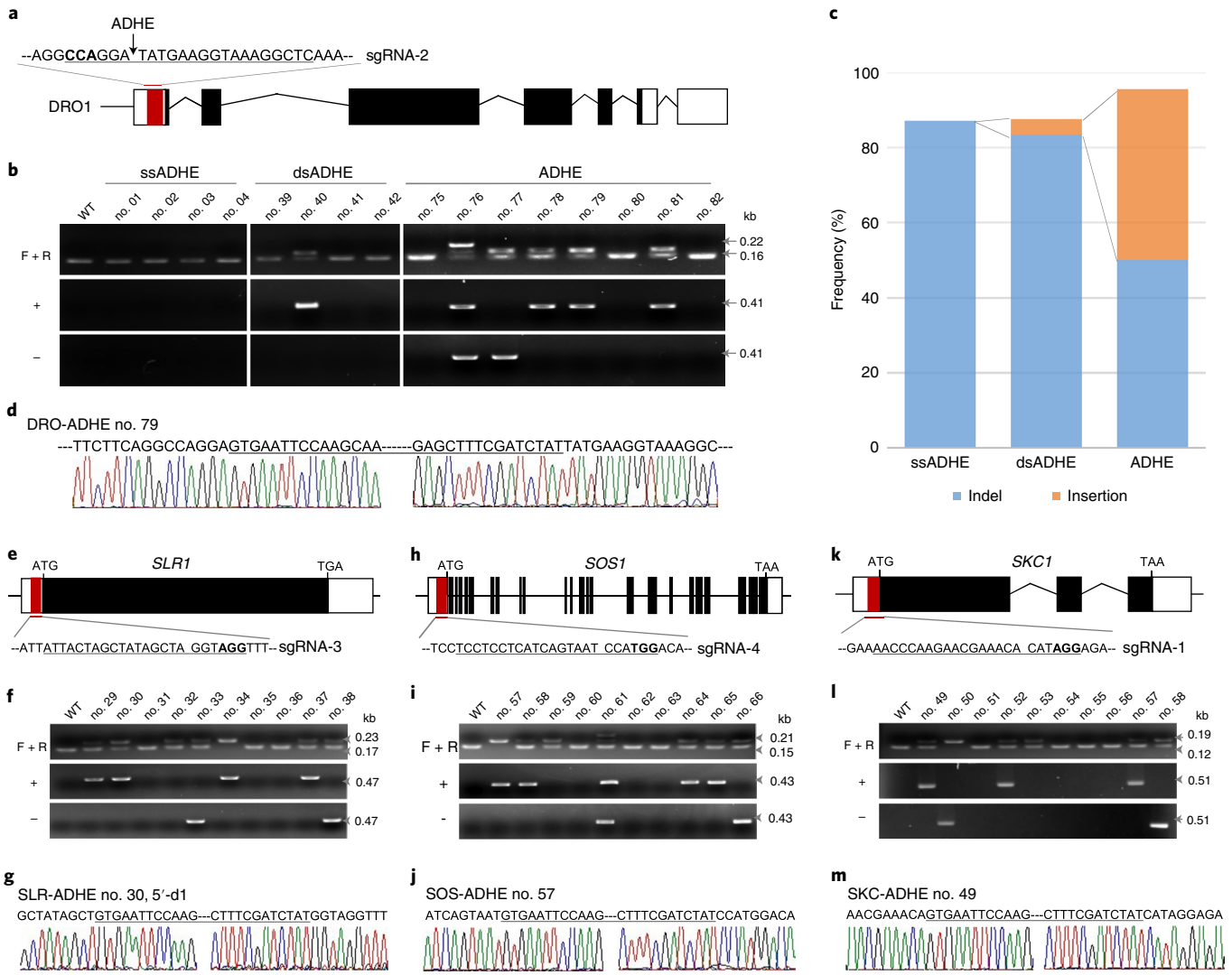


Fig. 2 | Targeted insertion of ADHE at four endogenous rice loci. a,**e**,**h**,**k**, Schematic diagram of *DRO1* (**a**), *SLR1* (**e**), *SOS1* (**h**) and *SKC1* (**k**) with their targets (underlined), PAM sequences (emboldened) and inserted donors (red box) in the 5′ UTR (white box). **b**,**f**,**i**,**l**, Genotyping of selected T0 transgenic rice plants for *DRO1* (**b**), *SLR1* (**f**), *SOS1* (**i**) and *SKC1* (**l**). Genomic DNA was amplified using primers flanking target sites (F + R, upper). The directions of ADHE at target sites were identified using ADHE specific primers. Forward (+, middle) and reverse (–, lower) insertion mutants were identified using primers F2 + ADHE-R and F2 + ADHE-F, respectively (Supplementary Fig. 3). The targeted insertion mutants are indicated in bold. One biological experiment was performed. **c**, Comparison of targeted insertion frequencies using ssADHE, dsADHE and chemically modified ADHE on *DRO1*. **d**,**g**,**j**,**m**, Sanger sequencing chromatograms for transgenic lines DRO-ADHE no. 79 (**d**), SLR-ADHE no. 30 (**g**), SOS-ADHE no. 57 (**j**) and SKC-ADHE no. 49 (**m**) containing forwardly inserted ADHE (underlined). '5′-d1' in **g** stands for one nucleotide deletion at the 5′-junction. One biological experiment was performed.

were selected[21]. Due to the length limitation of oligo synthesis, the long donors were prepared using PCR with corresponding chemically modified primers (see Supplementary Fig. 5). Each of them was designed to simultaneously insert into the 5′ UTR of SLR1 and SKC1, intending to overexpress these two genes. As listed in Table 1, the combined insertion frequency at the target loci was 25.5% for CmYLCV9.11 and 10.5% for UBI. The insertion of the two constitutive promoters substantially increased the transcript levels of SLR1 and SKC1 in some of the lines. The overexpression lines of SLR1 exhibited a semi-dwarf phenotype (see Supplementary Fig. 7). According to the sequencing results, PCR-amplified donors seemed more prone to produce deletions at both junctions of the inserted sequences (see Supplementary Table 2). This is probably due to the fact that only the 5′-ends of the PCR-amplified donors contained the phosphorothioate linkages, consistent with the notion that this chemical modification protects DNA from degradation.

To investigate whether these genome modifications could be transmitted to the next generation, we self-pollinated the T0 plants that carried the intended insertions, and genotyped individual T1 progeny using PCR. The T1 plants all contained multiple copies (2–10 copies per plant) of the donor inserts (see Supplementary Table 3), which suggested frequent off-target insertions besides the intended target site insertions. Some of the T1 plants contained no or a single copy of the Cas9 sequence, whereas many contained multiple copies (see Supplementary Table 3). In the genetic transmission assessment, only the intended target site insertions detected in T0 parent plants (see Supplementary Table 2) were analyzed. For the T0 homozygous mutants, transmission rates were 100%. Most T0 plants appeared to be in chimeric states, because, although their targeted-insertion modifications could be transmitted to the T1 progenies, the transmission did not happen in a mendelian fashion. On average, ~48% of T0 progenies contained the examined

Table 1 | Summary of genome editing results in T0 transgenic rice

Experiment	Purpose	Target			Donor		No. of T0 plants	Targeted insertion		TR-HDR		
		Locus	Region	sgRNA	Name	Length (bp)		Total	Forward	Seamless		Precise editing
								No. (%)		5′	3′	No. (%)
DRO1-ssADHE	Insertion of translational enhancers	DRO1	5′ UTR	sgRNA-2	ssADHE	59 nt	23	0	0	NA	NA	NA
DRO1-dsADHE		DRO1		sgRNA-2	dsADHE	59	24	1	1			
DRO1-ADHE		DRO1		sgRNA-2	ADHE	64	22	10 (45.5)	6			
SLR1-ADHE		SLR1		sgRNA-3			83	22 (26.5)	10			
SOS1-ADHE		SOS1		sgRNA-4			112	40 (35.7)	25			
SKC1-ADHE		SKC1		sgRNA-1			119	32 (31.7)	13			
WRKY71-AMVE		WRKY71.1		sgRNA-11	AMVE	57	16	4 (25)	2			
DEP1-AMVE		DEP1		sgRNA-12			79	34 (43)	20			
BZR1-AMVE		BZR1		sgRNA-13			55	26 (47.3)	9			
bZIP5-AMVE		bZIP5		sgRNA-14			34	8 (23.5)	3			
P1BS-KI	Insertion of *cis*-elements	OsPAP10a	Promoter	sgRNA-15	P1BS	26	188	49 (26.1)	NA			
		OsPAP10c		sgRNA-16				34 (18.1)	NA			
		OsPAP21b		sgRNA-17				20 (10.6)	NA			
		AVP1		sgRNA-18				59 (31.4)	NA			
CMP-KI	Insertion of promoters	SLR1	5′ UTR	sgRNA-3	CmYLCV9.11	526	55	5 (9.1)	3			
		SKC1		sgRNA-1				9 (16.4)	4			
UBI-KI		SLR1		sgRNA-3	UBI	2,049	76	3 (3.9)	2			
		SKC1		sgRNA-1				5 (6.6)	2			
SLR1-BS	Base substitution	SLR1	CDS	sgRNA-6	DN06	96	109	NA	13	4	4	4 (3.7)
TT1-BS		TT1		sgRNA-7	DN07	99	88	NA	13	5	3	3 (3.4)
NRT-BS		NRT1.1b		sgRNA-8	DN08	98	112	NA	20	11	7	6 (5.4)
UBQ6-Flag	Flag-tag fusion	UBQ6		sgRNA-9	DN09	130	140	NA	27	9	3	9 (6.4)
TT1-Flag		TT1		sgRNA-10	DN10	130	105	NA	21	5	12	12 (11.4)

NA, not available

modifications. Sequencing results of several T1 progenies of WRKY-AMVE no. 16 and BZR1-AMVE no. 31 (see Supplementary Table 2) confirmed the transmission of the seamless targeted insertion modifications (see Supplementary Table 4). For the progenies of DRO1 and SKC1 mutants, we further analyzed the transmission of T1 and T2 plants to their offspring. As expected, the homozygous ones were 100% transmitted, whereas the heterozygous ones segregated in a mendelian fashion (se Supplementary Table 4). These results show that the targeted insertions observed in T0 plants can be stably transmitted to subsequent generations.

Sequence replacement may be achieved through the homology-directed repair (HDR) mechanism, which is generally thought to occur at a very low frequency in plants[1,7]. However, we noticed that the HDR frequency can be very high (>50%) between two tandemly arranged repeat elements (such as the YFFP and GUUS reporters that contain a tandem repeat of the middle part (F and U, respectively) of the YFP and GUS reporters) in plant cells when a DSB is generated[22], probably because the repair template is present right at the DSB site. Having achieved robust targeted insertion as described above, we devised a tandem repeat-HDR strategy (TR-HDR) to try to achieve sequence replacement. As shown in Fig. 3a, an sgRNA site for targeted cleavage is designed, and a dsODN with homology to the target sequence and desired base substitutions is synthesized. This oligo is then inserted into the target site, presumably through NHEJ, forming a tandem-repeat structure with the flanking genomic sequence. The oligo is designed to also form a target site for the same sgRNA upon insertion. A DSB at the newly formed sgRNA target site is expected to trigger HDR between the tandem repeats, resulting in replacement of the target sequence with the inserted homologous sequence.

To test this strategy, the SLR1 gene was selected for editing. As shown in Fig. 3b, a 96-bp, modified dsODN, homologous to the genomic sequence downstream of the cleavage site (sgRNA-6), was designed. Nucleotide polymorphisms (five-base substitutions and a three-base deletion) were introduced into the modified dsODN. These polymorphisms would produce a one-amino acid deletion in SLR1 and would also make a clear distinction from random indels caused by CRISPR–Cas9. In addition, a 6-nt sequence was designed at the 3′-end of the oligo to form the cleavage site for sgRNA-6, together with the genomic sequence downstream of the insertion (see Supplementary Fig. 8). Therefore, precise insertion of this dsODN would generate an 87-bp tandem repeat with its downstream genomic sequence, and a DSB at the newly formed sgRNA site would trigger HDR to produce the designed sequence replacement. We synthesized the modified dsODN and transformed rice calli together with the corresponding CRISPR–Cas9 plasmid (sgRNA-6) using particle bombardment. A total of 109 T0 transgenic rice seedlings were regenerated from hygromycin-resistant calli in two experiments. Thirteen mutant plants containing a forward dsODN insert were identified using donor-specific primers (F1 + R2, 153 bp). As HDR between the tandem repeat is expected to result in base substitutions and seamless insertion of the dsODN,

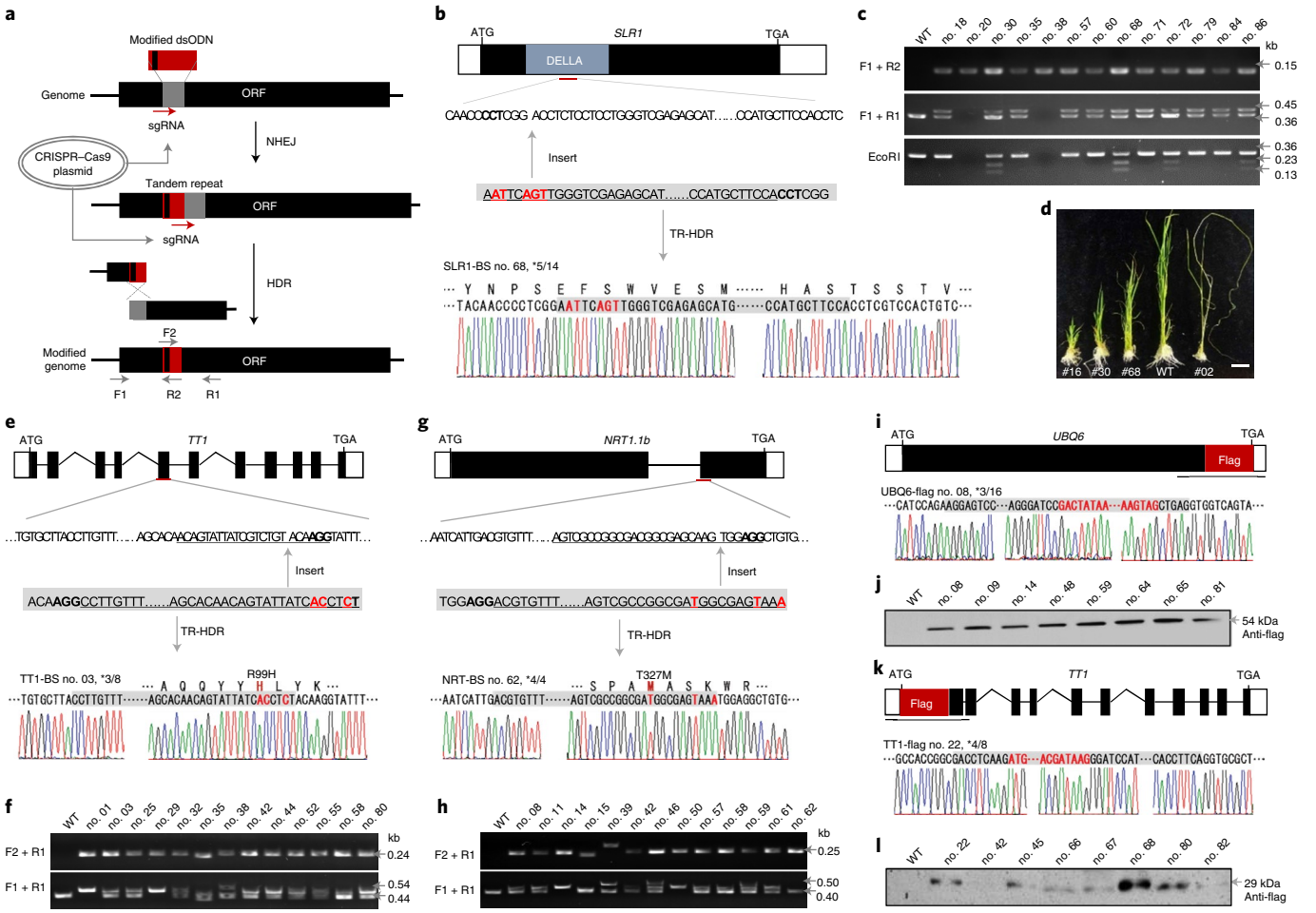


Fig. 3 | Precise genome editing in rice using the TR-HDR method. a, Schematic overview of the TR-HDR method. **b**,**e**,**g**,**i**,**k**, Design of base substitutions for SLR1 (**b**), TT1 (**e**) and NRT1.1b (**g**) and in-locus, flag-tag fusion for UBQ6 (**i**) and TT1 (**k**). Target sequences of sgRNA are underlined and the donor sequences were shadowed in gray. The homology sequences in the donor are underlined with red nucleotide polymorphisms or flag sequences. Their consequent sequences were shown below with sequencing chromatograms from representative T0 plants. The asterisk indicates the numbers of independent clones sequenced. One biological experiment was performed. **c**,**f**,**h**, Genotyping of T0 plants using PCR for SLR1 (**c**), TT1 (**f**) and NRT1.1b (**h**). T0 plants were identified using donor-specific primers (upper) and flanking primers F1 + R1. Mutants containing expected base substitutions are marked in bold. The PCR fragment of SLR1 (0.36 kb) was also further digested by EcoRI (**c**, lower). One biological experiment was performed. **d**, Phenotypes of T0 transgenic seedlings of SLR1. Scale bar, 2 cm. WT, wild type. **j**,**l**, Detection of the in-locus tagged UBQ6 (**j**) and TT1 (**l**) proteins using western blotting. One biological experiment was performed.

which would generate an EcoRI endonuclease recognition site, the target region was amplified using primers (F1 + R1) flanking the cleavage site. The expected HDR products (361 bp) were then identified using restriction fragment length polymorphism in which successful digestion by EcoRI cut the 361-bp PCR amplicon into 2 fragments (135 bp and 226 bp). The RFLP results showed that 4 of the 13 mutant plants contained the seamless junctions (Fig. 3c and Supplementary Fig. 9). In addition to the 361-bp expected HDR product, a 457-bp fragment was also amplified from these lines, possibly due to imprecise insertion that prevented HDR. The presence of an imprecise insertion would indicate that the seamless insertion event was probably chimeric. To validate the four sequence replacement mutants (nos. 30, 68, 72 and 86), their 361-bp amplicons were cloned and sequenced. The results revealed that all four mutants contained the designed base substitutions (Fig. 3c and Supplementary Fig. 9c). Thus, we successfully introduced designed base substitutions into the genome of rice plants. It was reported that amino acid variations near the TVHYNP motif of SLR1 can cause dominant dwarf mutant phenotypes[18]. Indeed, these base-substituted lines exhibited an obvious dwarf phenotype (nos. 30 and 68). As a comparison, a typical slender phenotype was observed in a line with a frameshift indel (no. 02) and an extreme dwarf phenotype was seen in a homozygous in-frame mutated line (no. 16; Fig. 3d and Supplementary Fig. 9c). These results indicated that the tandem-repeat strategy can be used to achieve HDR for base substitutions.

To further test the feasibility of TR-HDR, another two loci, TT1 and NRT1.1b, were chosen for base substitutions[23,24]. Two modified dsODNs homologous to the genomic sequence upstream of the cleavage sites (sgRNA-7 and sgRNA-8) were respectively designed for these two genes (Fig. 3e,g, and Supplementary Fig. 10). Nucleotide polymorphisms that cause amino acid substitutions were also introduced. Genotyping results of T0 transgenic seedlings, using donor-specific primers, found 33 out of 200 tested lines had the dsODN insertion in the correct orientation (Fig. 3f,h). Sequencing results using individual colonies for these mutants showed that three of the TT1 mutants and six of the NRT1.1b mutants contained the expected base substitutions without any indels (Table 1). These results further demonstrated that base substitutions could be achieved in rice using this TR-HDR strategy.

In-locus tagging of endogenous genes is very useful for functional genomics research, but has been extremely difficult in plants. We attempted to use our TR-HDR method to achieve in-locus fusion of the flag–tag to the N- or C-terminus of two proteins, TT1 and UBQ6 (see Supplementary Table 1). First, a dsODN was designed to include the flag sequence, a ~40-bp homologous sequence and a 6-nt tail to form the cleavage site to trigger HDR (see Supplementary Fig. 10). The dsODN donors were synthesized and used to transform rice calli together with the corresponding CRISPR–Cas9 plasmids. In a total of 245 regenerated T0 plants, we identified 12 TT1 and 9 UBQ6 mutants where the flag–tag was precisely fused to the CDS of UBQ6 or TT1 (Fig. 3i,k and Table 1). Differing from base substitutions, indels within the UTR at the other junction may not disrupt the target gene. The flag-tagging results were confirmed using western blotting (Fig. 3j,l), demonstrating the feasibility of this method for in-locus tagging.

Most of the TR-HDR lines were chimeric/heterozygous in the T0 generation. Only one mutant for NRT1.1b was homozygous (NRT-BS no. 62; see Supplementary Table 2). These T0 plants were self-pollinated and their T1 progenies were genotyped individually. For the homozygous mutant, the transmission rate was 100%. For the other plants, 8 of the 11 tested lines transmitted the mutations to T1 progenies, with the frequencies of transmitted mutations in T1 plants ranging from 2.4% to 62.5% (see Supplementary Table 4). Although the heritability varies among the T0 mutants, probably because of the different chimeric states, most of the genome modifications generated through TR-HDR could be stably transmitted to the offspring.

Collectively, making use of chemically modified donor DNA, we achieved targeted insertions at 14 loci with an average frequency of 25% in stable transgenic plants. The inserted sequences included translational enhancers, a cis-element and gene promoters, ranging in size from 26 to 2,049 bp. This efficient method would make targeted sequence insertion a routine practice in plant research and breeding. Building on the efficient target insertion method, we also devised the TR-HDR approach to achieve precise sequence replacement and insertion. By transforming rice calli with designed donors and CRISPR–Cas9 constructs, precise base substitutions and protein tag fusions were succeeded at five loci with an average frequency of 6.1% for precise editing in our experiments. As our strategy does not depend on the use of specifically designed selection markers at the target loci, any genomic sequences accessible to CRISPR–Cas9 or other engineered endonucleases can be precisely modified.

Although the method can be used to insert long sequences, it is particularly useful for insertion of short regulatory elements to simultaneously manipulate the expression levels of multiple, agronomically important genes. High donor concentration used for biolistic delivery could enhance the on-target insertion efficiency, but may also increase off-target insertion rates (see Supplementary Table 3). The high copy numbers of off-target insertions and the Cas9 transgene is a potential problem for use of this method in breeding. However, this problem could be alleviated both by generating a large number of T0 mutants and selecting the ones with fewer copies of insertions and by backcrossing. In addition, chimerism seems common in T0 plants, which indicates that the edits did not happen very early after biolistic delivery of the donor DNA and Cas9 construct to the rice callus cells. Nevertheless, homozygous plants with stable transmission can be isolated in the subsequent generations. Overall, the approaches presented in the present study provide new avenues for precision genome editing in crop plants, enabling not only efficient insertion of DNA fragments into target genomic loci, but also precise sequence replacements and fusion of protein tags. The simplicity and robustness of these methods will help advance precise genome editing for plant research and breeding.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0581-5.

Received: 27 June 2018; Accepted: 28 May 2020;
Published online: 06 July 2020

References

1. Zhang, H., Zhang, J., Lang, Z., Botella, J. R. & Zhu, J.-K. Genome editing—principles and applications for functional genomics research and crop improvement. *Crit. Rev. Plant Sci.* **36**, 291–309 (2017).
2. Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H. & Yang, B. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* **42**, 10903–10914 (2014).
3. Wang, Y. et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951 (2014).
4. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
5. Lu, Y. & Zhu, J. K. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 System. *Mol. Plant* **10**, 523–525 (2017).
6. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
7. Mao, Y., Botella, J. R., Liu, Y. & Zhu, J.-K. Gene editing in plants: progress and challenges. *Natl Sci. Rev.* **6**, 421–437 (2019).
8. Sun, Y. et al. Engineering herbicide-resistant rice plants through crispr/cas9-mediated homologous recombination of acetolactate synthase. *Mol. Plant* **9**, 628–631 (2016).
9. Sauer, N. J. et al. Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol.* **170**, 1917–1928 (2016).
10. Li, J. et al. Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nat. Plants* **2**, 16139 (2016).
11. Wang, M. et al. Gene targeting by homology-directed repair in rice using a geminivirus-based CRISPR/Cas9 system. *Mol. Plant* **10**, 1007–1010 (2017).
12. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
13. Puchta, H. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* **56**, 1–14 (2005).
14. Sugio, T., Satoh, J., Matsuura, H., Shinmyo, A. & Kato, K. The 5′-untranslated region of the Oryza sativa alcohol dehydrogenase gene functions as a translational enhancer in monocotyledonous plant cells. *J. Biosci. Bioeng.* **105**, 300–302 (2008).
15. Ren, Z. H. et al. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat. Genet.* **37**, 1141–1146 (2005).
16. Uga, Y. et al. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nat. Genet.* **45**, 1097–1102 (2013).
17. Shi, H., Lee, B. H., Wu, S. J. & Zhu, J. K. Overexpression of a plasma membrane Na^+/H^+ antiporter gene improves salt tolerance in *Arabidopsis thaliana*. *Nat. Biotechnol.* **21**, 81–85 (2003).
18. Ikeda, A. et al. slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. *Plant Cell* **13**, 999–1010 (2001).
19. Jobling, S. A. & Gehrke, L. Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. *Nature* **325**, 622–625 (1987).
20. Rouached, H., Secco, D., Arpat, B. & Poirier, Y. The transcription factor PHR1 plays a key role in the regulation of sulfate shoot-to-root flux upon phosphate starvation in *Arabidopsis*. *BMC Plant Biol.* **11**, 19 (2011).
21. Sahoo, D. K., Sarkar, S., Raha, S., Maiti, I. B. & Dey, N. Comparative analysis of synthetic DNA promoters for high-level gene expression in plants. *Planta* **240**, 855–875 (2014).
22. Feng, Z. et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **111**, 4632–4637 (2014).
23. Li, X. M. et al. Natural alleles of a proteasome $alpha_2$ subunit gene contribute to thermotolerance and adaptation of African rice. *Nat. Genet.* **47**, 827–833 (2015).
24. Hu, B. et al. Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* **47**, 834–838 (2015).

Methods

Construction of CRISPR–Cas9 plasmids. CRISPR–Cas9 plasmids were constructed as previously described[25]. Briefly, the 23-bp targeting sequences (including the point accepted mutation (PAM)) were selected within the target regions and their targeting specificity was analyzed using CRISPR-P[26]. Targeting sequences were synthesized and annealed to form the oligo adaptors. Vector pCBSG032 (see Supplementary Fig. 1a) was digested with BsaI and purified using DNA purification kit (Tiangen). A ligation reaction (10 µl) containing 10 ng of the digested pCBSG032 vector and 0.05 pmol oligo(adaptor) was carried out and directly transformed to Escherichia coli-competent cells to produce CRISPR–Cas9 plasmids.

Preparation of donor DNA. 5′-Phosphorylation and phosphorothioate-modified oligos[12] were synthesized (Sangon Biotech) and diluted to 50 pmol µl^{-1}. Paired oligos were then annealed in annealing buffer (50 mM NaCl and 1 mM ethylenediaminetetraacetic acid) at 5 pmol µl^{-1} to produce the dsODN. The modified oligos for donor preparation are listed in Supplementary Table 6. Due to the length limitation of oligo synthesis, the long donors were prepared using PCR with corresponding chemically modified primers (see Supplementary Fig. 5 and Supplementary Table 6). PCR fragments were produced with KOD DNA polymerase (Takara) and then purified using a DNA purification kit (Tiangen).

Biolistic transformation. CRISPR–Cas9 plasmids were extracted using a midipreps kit (Tiangen) and diluted to 1 µg µl^{-1} (about 0.1 pmol µl^{-1}). Plasmid and donor DNA were mixed (0.1 pmol plasmid + 10 pmol donor for ADHE insertion and SLR1 substitution, 0.05 pmol plasmid + 2 pmol donor for other experiments) and added to 50-µl gold microparticles solution (60 mg ml^{-1}) for bombardment. Embryogenic calli of rice ZH11 that were 1-month old were used for transformation. Biolistic transformation was performed using a PDS1000/He particle bombardment system (Bio-Rad) with a target distance of 6.0 cm from the stopping plate at a helium pressure of 7.5 MPa. Plantlets were regenerated from hygromycin-resistant calli using the routine rice transformation method described previously[27].

Genotyping. Genomic DNA was extracted using the CTAB method. For NGS analysis, the hygromycin-resistant calli were picked out after 4 weeks of selection on 50 µg l^{-1} of hygromycin and were further selected by subculturing for 2 weeks on 50 µg l^{-1} of hygromycin. Approximately 200 hygromycin-resistant calli were collected together for genomic DNA extraction. The amplified PCR products containing the target site were then subjected to NGS. A custom PHP script was used to analyze the NGS data to calculate the types of ADHE insertion in the UTR of SKC1 (https://github.com/zhulab-ge/knockin).

To genotype the T0 transgenic lines and their progenies, genomic DNA was extracted from leaves. To assess the mutagenesis frequency, the targeted sequences of sgRNAs were amplified for Sanger sequencing and then analyzed using Dsdecode[28]. To identify the targeted insertion events, three pairs of primers were used for each sample. Primers (F + R) flanking the target site were used for detecting targeted insertion events and the other two pairs of primers that contain donor-specific primer (F2 + Donor-R and F2 + Donor-F) were used to determine the direction of donor insertion. Successful insertion would result in a larger amplicon, and it would also produce a PCR amplicon using these donor-specific primers. Mutant plants possessing both types of PCR amplicons were counted as targeted insertion plants. These PCR products containing the donor of expected length were further selectively cloned into the TA cloning vector pEasy (TransGen Biotech) and sequenced, individually. To genotype the TR-HDR mutants, transgenic plant containing forward donor insert were first identified using donor-specific primers. Primers (F1 + R1) flanking the target site were used to amplify the target sequence. PCR fragments were then cloned into the TA cloning vector pEasy (TransGen Biotech), and 6–16 positive colonies for each sample were sequenced, individually. PCR primer sets are listed in Supplementary Table 5.

RNA extraction and RT–qPCR. Total RNA was extracted from leaves of transgenic seedlings using TRIzol reagent (Life Technologies) and treated with RNase-free DNase I (Invitrogen). RNA was reverse transcribed using oligo(dT) primer and M-MLV reverse transcriptase (TransGen Biotech). Quantitative PCR (qPCR) experiments were performed. Each qPCR assay with reverse transcription (RT–qPCR) was replicated at least three times with three independent RNA preparations. The rice ACT1 gene was used as an internal control. The primers used are listed in Supplementary Table 5.

Protein extraction and protein gel blot analysis. Protein was extracted from leaves of T0 transgenic seedlings with an extraction buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2% sodium dodecylsulfate and 1 mM phenylmethylsulfonyl fluoride. Protein gel blot analysis was performed with an anti-flag antibody (Sigma, 1:1,000 dilution). The secondary antibody was a goat anti-rat antibody conjugated to horseradish peroxidase (Sangon Biotech), and reaction signals were visualized using an enhanced chemiluminescence solution (Millipore).

Statistical analysis. The relevant statistical test, sample size and replicate type for each figure and table are found in the figure or table and/or the corresponding figure legends.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The authors declare that all data supporting the findings of the present study are available in the article and its supplementary figures and tables, or from the corresponding author upon request. For sequence data, rice LOC_Os IDs listed in Supplementary Table 1 are available on the Rice Genome Annotation Project site (http://rice.plantbiology.msu.edu/). The deep sequencing data were deposited with the National Center for Biotechnology Information BioProject database under the accession code PRJNA608130. Source data are provided with this paper.

Code availability

Custom script for analyzing NGS data is available at https://github.com/zhulab-ge/knockin. Source data are provided with this paper.

References

25. Lu, Y. et al. Genome-wide targeted mutagenesis in rice using CRISPR/Cas9 system. *Mol. Plant* **10**, 1242–1245 (2017).
26. Liu, H. et al. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant* **10**, 530–532 (2017).
27. Nishimura, A., Aichi, I. & Matsuoka, M. A protocol for *Agrobacterium*-mediated transformation in rice. *Nat. Protoc.* **1**, 2796–2802 (2006).
28. Liu, W. et al. DSDecode: a web-based tool for decoding of sequencing chromatograms for genotyping of targeted mutations. *Mol. Plant* **8**, 1431–1433 (2015).

Acknowledgements

This work was financially supported by the Chinese Academy of Sciences, including the CAS Strategic Priority Research Program grant no. XDB27040101 to J.-K.Z., and by the Major Project of China on New Varieties of GMO Cultivation (grant no. 2019ZX08010-003 to F.L.).

Author contributions

Y.L. and J.-K.Z. designed the experiments. M.C., Y.T., R.S., J.D., F.L. and T.Z. performed the rice transformations. Y.L., Y.T., R.S., Q.Y., M.C., M.W., J.D., T.Z. and M.L. performed all the other experiments. Y.L. and J.-K.Z. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0581-5.

Correspondence and requests for materials should be addressed to J.-K.Z.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Jian-Kang Zhu

Last updated by author(s): Apr 9, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina HiSeq 2500 platform was used to collect the amplicon deep sequencing data.
Data analysis	The software Dsdecode Version 1.0 software was used to analyze Sanger sequencing data. Custom script for analyzing NGS data is available at https://github.com/zhulab-ge/knockin.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data supporting the findings of this study are available in the article and its supplementary figures and tables or are available from the corresponding author on request. Dataset for the high-throughput sequencing experiments has been deposited in the National Center for Biotechnology Information (NCBI) BioProject database with accession number PRJNA608130.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In our study, we focused on whether the modified donor DNA could achieve targeted insertion or replacement. An average of 80 transgenic plants per experiment were analyzed by PCR and/or DNA sequencing. No need to pre-determine the sample size in this case. The work is more qualitative and does not rely on statistical significance.
Data exclusions	No data was excluded from the study.
Replication	Dozens of rice transformations were conducted and a total of 1440 transgenic plants were obtained and analyzed. In our study, all attempts at sequence insertion or replacement were successful, including 14 sequence insertions and 6 sequence replacements. These results were independently verified by different students or laboratories.
Randomization	The transgenic events investigated were selected randomly. The transgenic plants were grouped based on different experiments.
Blinding	Not applicable. As samples were processed identically through standard and in some cases automated procedures (DNA sequencing, transfection, DNA isolation) that should not bias outcomes.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-FLAG (Sigma, SAB4200071), HRP-conjugated Goat anti-rat IgG (Sangon Biotech, D110261)
Validation	The antibodies used have been verified by the manufacturer and the validation can be found via the following linker: Anti-FLAG: https://www.sigmaaldrich.com/catalog/product/sigma/sab4200071?lang=zh&region=CN HRP-conjugated Goat anti-rat IgG: https://www.sangon.com/productDetail?productInfo.code=D110261