nature biotechnology

Article

https://doi.org/10.1038/s41587-023-01769-w

Precise integration of large DNA sequences in plant genomes using PrimeRoot editors

Received: 9 January 2023

Accepted: 28 March 2023

Published online: 24 April 2023

Check for updates

Chao Sun[1,2,5], Yuan Lei[1,2,5], Boshu Li[1,2,5], Qiang Gao[3], Yunjia Li[1,2], Wen Cao[4], Chao Yang[4], Hongchao Li[1], Zhiwei Wang[3], Yan Li[3], Yanpeng Wang[1,2], Jun Liu[4], Kevin Tianmeng Zhao[3] ✉ & Caixia Gao[1,2] ✉

A technique for chromosomal insertion of large DNA segments is much needed in plant breeding and synthetic biology to facilitate the introduction of desired agronomic traits and signaling and metabolic pathways. Here we describe PrimeRoot, a genome editing approach to generate targeted precise large DNA insertions in plants. Third-generation PrimeRoot editors employ optimized prime editing guide RNA designs, an enhanced plant prime editor and superior recombinases to enable precise large DNA insertions of up to 11.1 kilobases into plant genomes. We demonstrate the use of PrimeRoot to accurately introduce gene regulatory elements in rice. In this study, we also integrated a gene cassette comprising *PigmR*, which confers rice blast resistance driven by an Act1 promoter, into a predicted genomic safe harbor site of Kitaake rice and obtain edited plants harboring the expected insertion with an efficiency of 6.3%. We found that these rice plants have increased blast resistance. These results establish PrimeRoot as a promising approach to precisely insert large segments of DNA in plants.

Introducing new genetic elements into crops has been instrumental in increasing agricultural output to support the growing global population. Many agronomic traits, such as herbicide resistance, pest and disease resistance and high nutritional value, are a result of introducing new genes into crop genomes[1–4]. Targeted gene insertions are also needed to create precise metabolic controls in plant synthetic biology systems. The most common form of plant transgenesis used today relies upon *Agrobacterium tumefaciens* and its inherent ability to insert T-DNA randomly into plant genomes[5]. However, *Agrobacterium*-mediated transformation events are prone to transcriptional and post-transcriptional gene silencing[6,7] caused by random insertion effects[8,9]. Therefore, insertion events have to be extensively screened over several generations to identify the most stable derivatives[9]. Recent advances in genome editing now offer new avenues for genome engineering.

CRISPR genome editing relies on a single guide RNA (sgRNA) to program Cas nuclease binding and DNA cleavage at specific genomic sites in living cells[10–12]. Recently developed precision genome editing technologies, such as base editing and prime editing, permit precise targeted genome modification without DNA double-strand break (DSB) intermediates[13,14]. These technologies have greatly accelerated genetic research and molecular breeding in agricultural crop research[15]. Targeted DNA insertions can also be generated in plants by CRISPR-based methods. Donor DNA can be inserted at Cas9 DNA DSB sites by non-homologous end joining (NHEJ) or by homology-directed repair (HDR)[16]. However, the NHEJ pathway is extremely imprecise, even if using a chemically modified donor DNA, as the ends of the inserted donor fragment comprise random DNA base insertions and deletions (indels)[17]. In HDR, although some insertion events at a DSB can be precise due to homology between the donor DNA and the host genome, HDR methods are extremely inefficient in higher plants and generally rely on a selectable marker to enrich for rare insertion events[18]. Moreover, as both NHEJ and HDR rely on DSBs, methods that

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China. [2]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China. [3]Qi Biodesign, Life Science Park, Beijing, China. [4]State Key Laboratory of Agrobiotechnology and MOA Key Laboratory for Monitoring and Green Management of Crop Pests, China Agricultural University, Beijing, China. [5]These authors contributed equally: Chao Sun, Yuan Lei, Boshu Li. ✉e-mail: kzhao@qi-biodesign.com; cxgao@genetics.ac.cn

bypass DSBs are needed as DSBs cause many undesired and unpredictable events, such as large deletions, chromosomal translocations and chromothripsis[19–22].

Prime editing (PE) is a precision genome editing technology capable of generating base changes and short DNA indels without the formation of DSBs. Although prime editors have been used widely across species, such as in human cells, mice, rice, wheat, maize and more, it is limited by its inability to insert large DNA fragments[14,23–25].

An alternative approach to generating large insertions relies on the use of site-specific recombinases (SSRs). SSRs first recognize specific sequences known as recombinase sites (RSs) and use these to undergo synapsis formation. During the recombination process, the recombinase creates a recombinant structure and performs strand exchange, which ultimately enables SSRs to perform DNA inversions, translocations, deletions and insertions between pairs of RSs[26,27]. SSRs can be mostly classified into two families: tyrosine SSRs, such as Cre from phage P1 and FLP from *Saccharomyces cerevisiae*, or serine SSRs, such as PhiC31 from phage PhiC31 and Bxb1 from *Mycobacterio phage*[28]. Although the recombination events catalyzed by tyrosine SSRs are generally reversible, whereas those of serine SSRs are usually unidirectional, mutant recombinase sites for Cre reduce reversibility, which makes the application of tyrosine SSRs more flexible and widely applicable[29]. Because recombination mediated by SSRs does not involve endogenous DNA repair, editing efficiencies are generally good, even in eukaryotic cells, so these tools are used extensively to engineer gene insertions and rearrangements[30,31]. However, SSR complexes recognize specific recombinase sites, which are absent from most desired insertion sites[32], so the need for an RS severely limits the utility of SSRs in many higher species[33–35]. TwinPE+Bxb1 and PASTE are new approaches that rely on PE and recombination to insert large DNA donor segments into mammalian genomes, but these approaches require further optimization[36,37].

CRISPR-associated transposases are new genome editing tools that have been shown to create targeted insertions in prokaryotes, but this process is extremely inefficient in eukaryotes[38,39]. Thus, despite the rapid expansion of the genome editing toolbox[40–42], it is important to be able to generate precise (defined as predictable and as specified) large DNA insertions in plants. Here we introduce PrimeRoot (Prime editing-mediated Recombination Of Opportune Targets) as a new genome editing technology capable of precisely inserting large DNA donors in plants without DSB intermediates (Fig. 1a). Furthermore, PrimeRoot uses many optimized components specifically to enhance overall editing efficiencies in plant genomes, in contrast to other approaches mostly optimized in mammalian cells.

We describe below the use of our efficient PrimeRoot system to insert an Actin1 promoter into the 5′ untranslated region (UTR) of *OsHPPD*. This illustrates an attractive genome editing approach to regulate gene expression levels. We further predicted and identified 30 genomic safe harbor (GSH) regions in the Kitaake rice variety genome suitable for receiving targeted gene insertions, and we used PrimeRoot to precisely integrate a 4.9-kilobase (kb) donor cassette comprising *pigmR*, which confers rice blast disease resistance, driven by an Actin1 promoter, into one GSH site. Notably, the resulting edited plants were disease resistant. These results highlight PrimeRoot as an effective molecular breeding technology capable of generating precise targeted DNA insertions in rice, further expanding the application of genome editing in plants.

Results

Evaluating SSR activity in rice

We speculated that SSRs in combination with a programmable and efficient plant prime editor might be able to generate precise targeted large DNA insertions in plants. We first constructed a fluorescence system to report the recombination efficiencies of commonly used SSRs[43] in rice protoplasts. We split green fluorescent protein (GFP) into N-terminal (GFP-N) and C-terminal (GFP-C) domains, with each half encoded on separate plasmids (Fig. 1b). P1 comprises a maize ubiquitin promoter driving GFP-N, followed by an intron and a specific recombinase site, whereas P2 contains a recombinase site followed by an intron and GFP-C terminated by the cauliflower mosaic virus terminator. Specific recombinase sites located on P1 and P2, and the corresponding recombinases, were co-transfected into rice protoplasts. After recombinase expression and subsequent recombination, GFP-N and GFP-C were joined by an intronic linker, leading to protoplast GFP expression. GFP fluorescence was then quantified by flow cytometry as a measure of recombinase activity in the protoplasts.

We constructed independent fluorescence reporter systems for six different tyrosine recombinases and two serine recombinases (all recombinases were codon optimized for expression in rice); among these were two recombinase sites previously shown to be effective with the Cre recombinase[29] (Extended Data Fig. 1a). Microscopic visualization of GFP fluorescence showed that the Cre and FLP recombinase systems yielded the strongest fluorescence (Extended Data Fig. 1a), and this was further confirmed by flow cytometry (Fig. 1c). This indicates that the Cre and FLP recombinase systems are the most effective of the recombinases evaluated here for use in plants.

Development of a dual-enhanced plant prime editor system for efficient introduction of recombinase sites in rice

To enable efficient targeted insertion of RSs in plant cells, we used the dual-prime editing guide RNA (pegRNA) approach that we had shown to enhance editing in plants and that has proven effective in creating targeted DNA insertions in mammalian cells[37,44]. We used two adjacent pegRNAs, each containing an RT template with homology only to the other pegRNA's template (Extended Data Fig. 2a). We first compared the insertion efficiencies achieved by plant prime editor (PPE) (PPE2), enhanced plant prime editor (ePPE)[45] and ePPE-wtCas9 (Extended Data Fig. 2b) using dual-pegRNAs to insert a Lox66 (34 base pairs (bp)) or FRT1 (48 bp) sequence with a roughly 30-bp overlap of the RT template between the two pegRNAs at five endogenous sites in rice protoplasts. We found that ePPE had the highest editing efficiency (Extended Data Fig. 2c).

To enhance insertion efficiency, we incorporated tevoPreQ1 toeholds, which enhance prime editing efficiency[46], to generate dual epegRNAs, and we evaluated the ability of combinations of PPE+pegRNA, PPE+epegRNA, ePPE+pegRNA and ePPE+epegRNA to precisely insert Lox66 and FRT1 RS sequences at eight endogenous sites. The combination of ePPE+epegRNA (hereafter referred to as 'dual-ePPE') displayed the highest insertion efficiencies (up to 50%) (Fig. 1d and Extended Data Fig. 2d). To expand the targeting scope of dual-ePPE, we engineered SpG and SpRY Cas9 variants into ePPE and evaluated their editing efficiencies at sites comprising NGN-containing PAMs (Extended Data Fig. 2e). We confirmed the efficacy of dual-ePPE by using this editing approach at five sites in rice plants, and we found that up to 46% of the regenerated plants obtained harbored the precise RS insertion (Extended Data Fig. 2f). These results indicate that the dual-ePPE system can perform efficient targeted insertion of RS sequences in plants.

Dual-ePPE insertion efficiencies drop rapidly with increasing fragment size

We wondered whether our optimized dual-ePPE could generate larger DNA fragment insertions at endogenous genomic sites in rice. We used digital droplet polymerase chain reaction (ddPCR) to identify DNA fragment insertions to minimize any PCR amplification bias generated through sequencing. We evaluated inserting LNT (Linker-NLS-T2A, 150 bp), t35S promoter (truncated 35S promoter, 222 bp), MCP (adapter of MS2, 300 bp), U3 promoter (401 bp), DD (degradation domain, 501 bp), tGFP (truncated GFP, 600 bp) and GFP (720 bp) across multiple genomic sites. In addition to using the canonical U3 promoter to drive epegRNA expression, we also explored using the type II promoter

https://doi.org/10.1038/s41587-023-01769-w

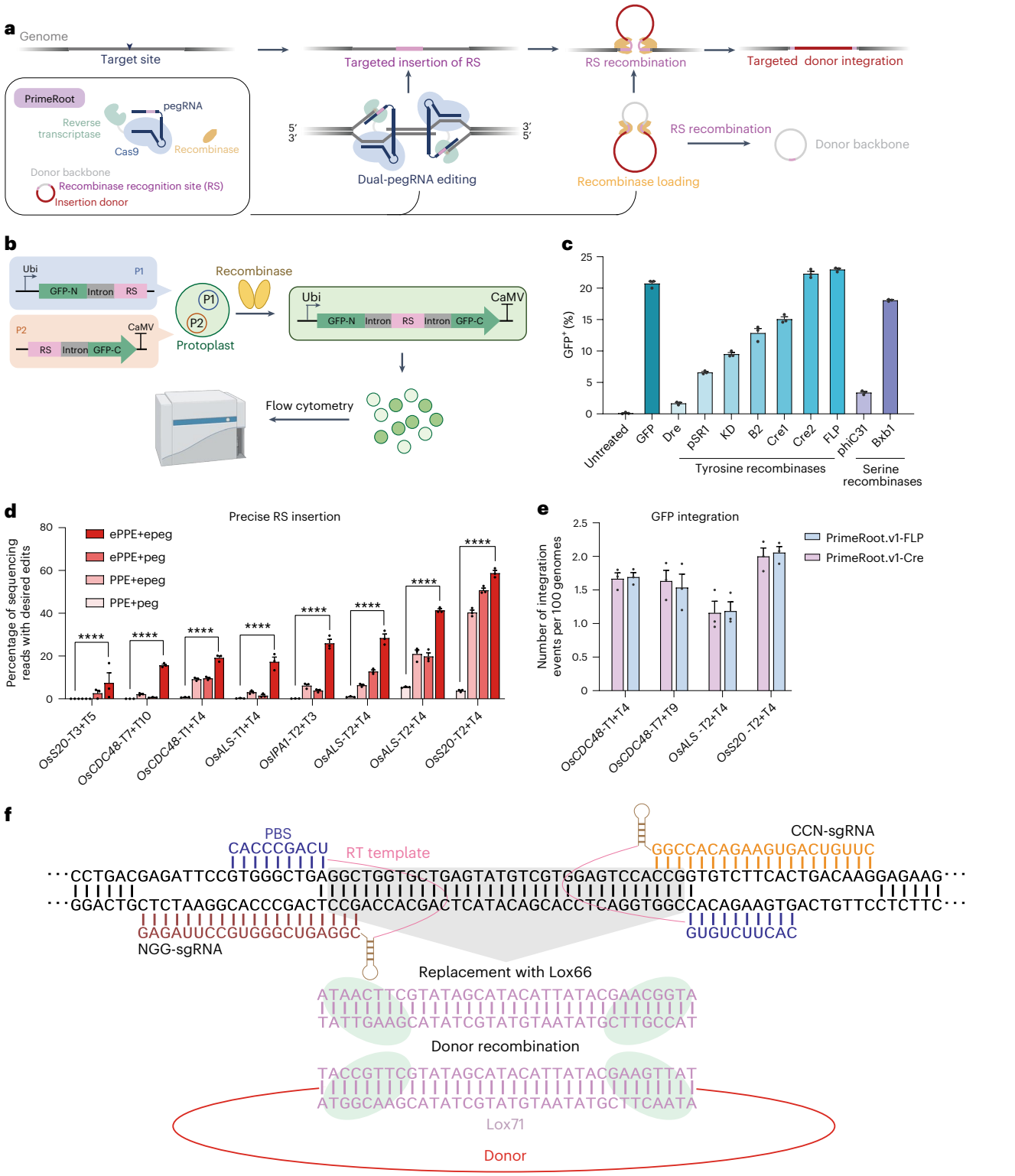


Fig. 1 | PrimeRoot combines plant-optimized recombinases and enhanced plant PE to create targeted DNA insertions. a, Schematic overview of how PrimeRoot creates precisely targeted large DNA insertions in plants. **b**, Schematic diagram of the fluorescence reporter for evaluating the integration activity of SSRs in plant protoplasts. **c**, Percentage of GFP$^+$ plant protoplasts, reflecting recombinase activity, as measured by flow cytometry. Seven tyrosine recombinases and two serine recombinases were evaluated, and GFP was used as a positive control. Cre1 and Cre2 use different recombinase sites with the same Cre recombinase, as noted in the Supplementary Methods. Values and error bars represent means and standard errors of means for three independent biological replicates. **d**, Percentages of precise insertions of recombinase sites generated by PPE+peg, PPE+epeg, ePPE+peg and ePPE+epeg at seven endogenous sites as measured by high-throughput sequencing. Detailed editing efficiencies at each site are shown in Extended Data Fig. 2d. Values represent editing efficiencies across the seven sites, and error bars represent means and standard errors of means for three independent biological replicates. P values were obtained using the two-tailed Student's *t*-test: ****$P < 0.0001$. **e**, Percentages of GFP insertions across four endogenous sites induced by PrimeRoot.v1-Cre and PrimeRoot.v1-FLP measured by ddPCR. Values and error bars represent means and standard errors of means for three independent biological replicates. **f**, Scheme of PrimeRoot integration at *OsS20*, showing dual-ePPE-mediated RS insertion followed by donor recombination.

pGS to drive epegRNA expression and evaluate if promoter structure would improve longer fragment insertion efficiencies[23] (Extended Data Fig. 3a). We observed that, although there were no obvious differences between using a U3 promoter or a pGS promoter for small fragment insertions (Extended Data Fig. 3b), the pGS-driven epegRNAs were up to two-fold more effective than U3-driven epegRNAs when generating larger insertions up to 300 bp in length (Extended Data Fig. 3c,d). However, pGS-driven insertion efficiencies dropped markedly when inserting even larger fragments (Extended Data Fig. 3e). Although we were able to detect 7.9% editing when inserting a 401-bp fragment, this dropped to 2.6% with a 501-bp fragment (three-fold decrease) and dropped further to 0.65% when inserting a 720-bp fragment (12-fold decrease). These results suggest that, although large insertion PE frequencies can be improved using type II promoters to drive pegRNA expression, editing efficiency drops sharply with increasing fragment sizes.

PrimeRoot enables targeted insertion of large DNA sequences without DSBs in plants

We combined our optimized dual-ePPE with the highly active Cre-Lox66/Lox71 or FLP-FRT1 recombinase systems to generate PrimeRoot.v1-Cre and PrimeRoot.v1-FLP, respectively. PrimeRoot.v1 employs dual-ePPE to efficiently insert an RS sequence at a target site, whereas the recombinase in parallel recognizes and excises two identical RS sites on the donor vector to generate an intermediate donor containing only the desired inserted DNA fragment with one corresponding RS (Fig. 1a). These two components come together when the pre-processed donor is recombined into the newly incorporated RS, ultimately resulting in a precise large DNA insert without incorporating any donor backbone components (Fig. 1a,f).

We first measured GFP (720 bp) integration frequencies mediated by PrimeRoot.v1-Cre and PrimeRoot.v1-FLP across four endogenous sites by ddPCR. We demonstrated that, although we could obtain up to 2% precise targeted insertions in rice protoplasts, this efficiency was quite low (Fig. 1e). To further optimize the system, we created two constructs in which ePPE was fused to the recombinase: in PrimeRoot.v2N, the recombinase is fused to the N-terminus of the ePPE system via an SV40 NLS and a 32-amino-acid flexible linker; in PrimeRoot.v2C, it is linked by the same linker to the C-terminus of the ePPE system (Fig. 2a). We also developed a new GFP all-in-one reporter (AR) to report insertions of RS sequences by dual-ePPE followed by recombination (Fig. 2b). Visual inspection of fluorescent cells and flow cytometry showed that both PrimeRoot.v2N and PrimeRoot.v2C were more efficient than PrimeRoot.v1 (Fig. 2c,d). When we compared GFP insertion efficiencies across four endogenous target sites, PrimeRoot.v2C proved to be superior to PrimeRoot.v2N, generating up to 6% targeted precise insertions at these targets (Fig. 2e). We next evaluated the efficiencies of integrating larger DNA donors using PrimeRoot.v2C-Cre. We generated constructs containing any of one to combined vectors of three genes—*pigmR*, *OsMYB30* and *OsHPPD*—driven by an Act1 or ubiquitin promoter, yielding donors of 1.4 kb, 4.9 kb, 7.7 kb and 11.1 kb (Fig. 2f). We tested the insertion efficiencies of the four donors at four endogenous sites by ddPCR, and we found only a minor decline with gradually increasing donor lengths (Fig. 2g).

To further expand the utility of PrimeRoot in plant applications, we next evaluated the editing efficiency of PrimeRoot in maize. We first tested the efficiency of dual-ePPE at six endogenous genomic sites in maize protoplasts and identified precise RS insertions edits up to 40% mediated by a construct using pGS to drive the expression of the pegRNA (Fig. 2h). We then used PrimeRoot.v2C-Cre and obtained up to 4% GFP integration at these endogenous sites (Fig. 2i), which is similar in editing efficiency as with rice. These results demonstrate PrimeRoot as a promising tool for plant synthetic biology and gene stacking. Notably, all insertions generated by PrimeRoot, as expected, incorporated only the desired donor DNA and no donor backbone sequences (Fig. 1a,f).

Engineering FRT recombinase sites improves recombination efficiencies

While we were using PrimeRoot-FLP, we noted the presence of short repeat sequences in the FRT1 RS (F1) and wondered whether these might reduce insertion frequencies (Extended Data Fig. 2d). We, therefore, generated three mutants of FRT1 (F1m1, F1m2 and F1m3) and two mutants of a truncated FRT1 (tFRT1) sequence (tF1m2 and tF1m3), each carrying different point mutations within the RS sequence based on previously identified key residues, to see if we could enhance FLP recombination[47,48] (Extended Data Fig. 3f). When we evaluated these RS variants using the AR, we indeed identified variant combinations that were more efficient than wild-type FRT1 (Extended Data Fig. 3g) and confirmed this by using ddPCR to examine the integration of GFP at a site in *OsALS* (Extended Data Fig. 3h). Whereas the wild-type combination comprising F1+F1 resulted in 1.4% integration, F1m2+F1m3 and tF1m1+F1m3 resulted in 3.0% and 4.0% integration efficiencies, reflecting a 2.1-fold and 2.8-fold improvement, respectively.

PrimeRoot is more predictable and precise than CRISPR–Cas9-mediated NHEJ

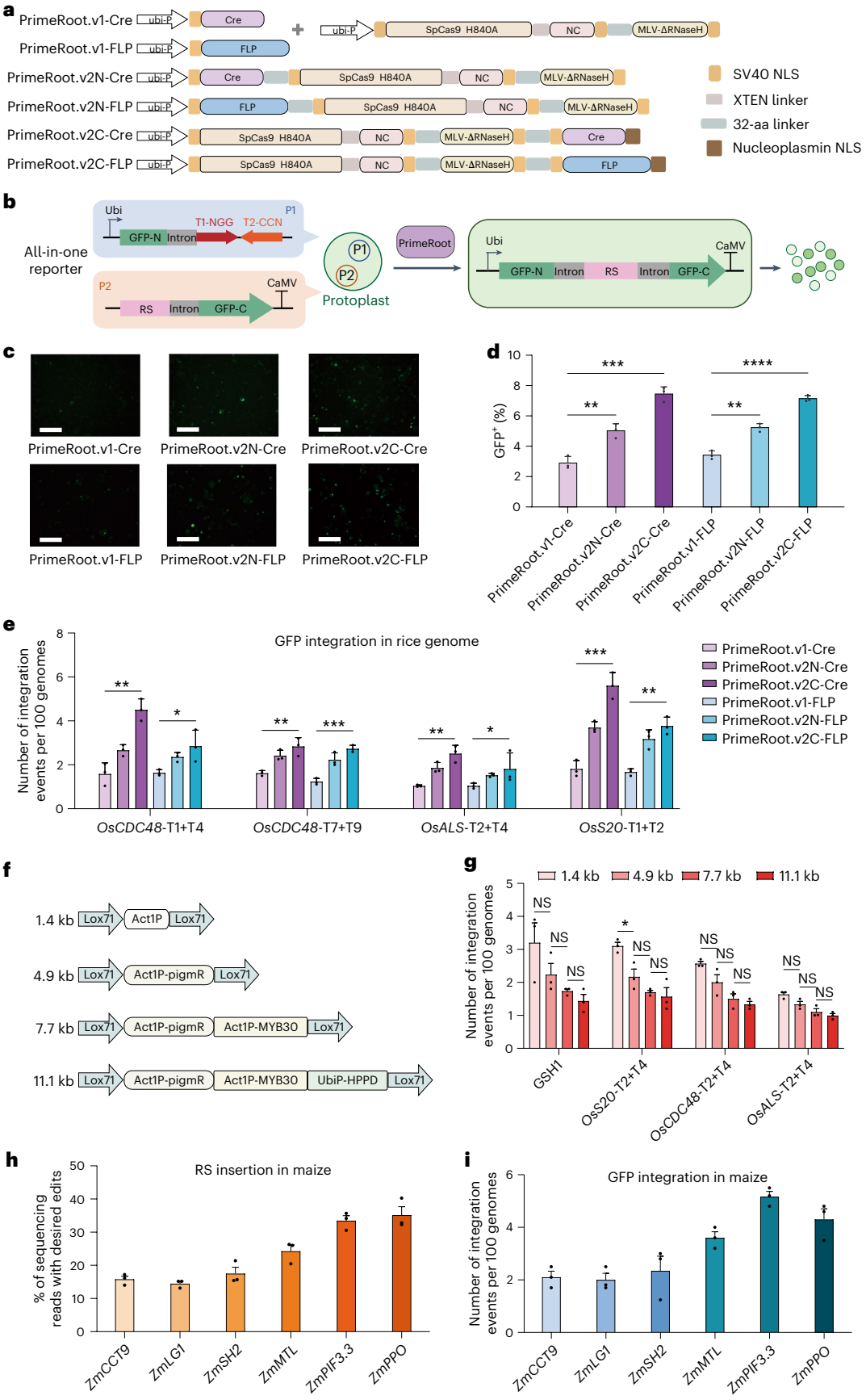
We next compared PrimeRoot with CRISPR-mediated NHEJ, which can also create targeted large DNA insertions in plants[49,50]. We used both systems to perform targeted insertions of GFP (720 bp), an Act1 promoter (Act1P, 1.4 kb), an Act1P-*pigmR* gene cassette (4.9 kb) and an Act1P-*pigmR*-Act1P-*OsMYB30* gene cassette (7.7 kb) at three endogenous sites (Extended Data Fig. 4a). Although the insertion efficiencies for GFP and Act1P were similar, PrimeRoot was, on average, 2–4-fold more effective than NHEJ for longer donor inserts, as measured across the three sites (Extended Data Fig. 4b). Notably, we obtained unambiguous Sanger sequencing traces for the Act1P insertion events mediated by PrimeRoot but mixed peaks when using NHEJ (Fig. 3a). This highlights PrimeRoot's superior editing precision compared to traditional CRISPR-mediated NHEJ insertion.

To further analyze individual genotypes, we subcloned edited insertion events from protoplasts into bacterial constructs and sequenced the junctions between the endogenous genome and the

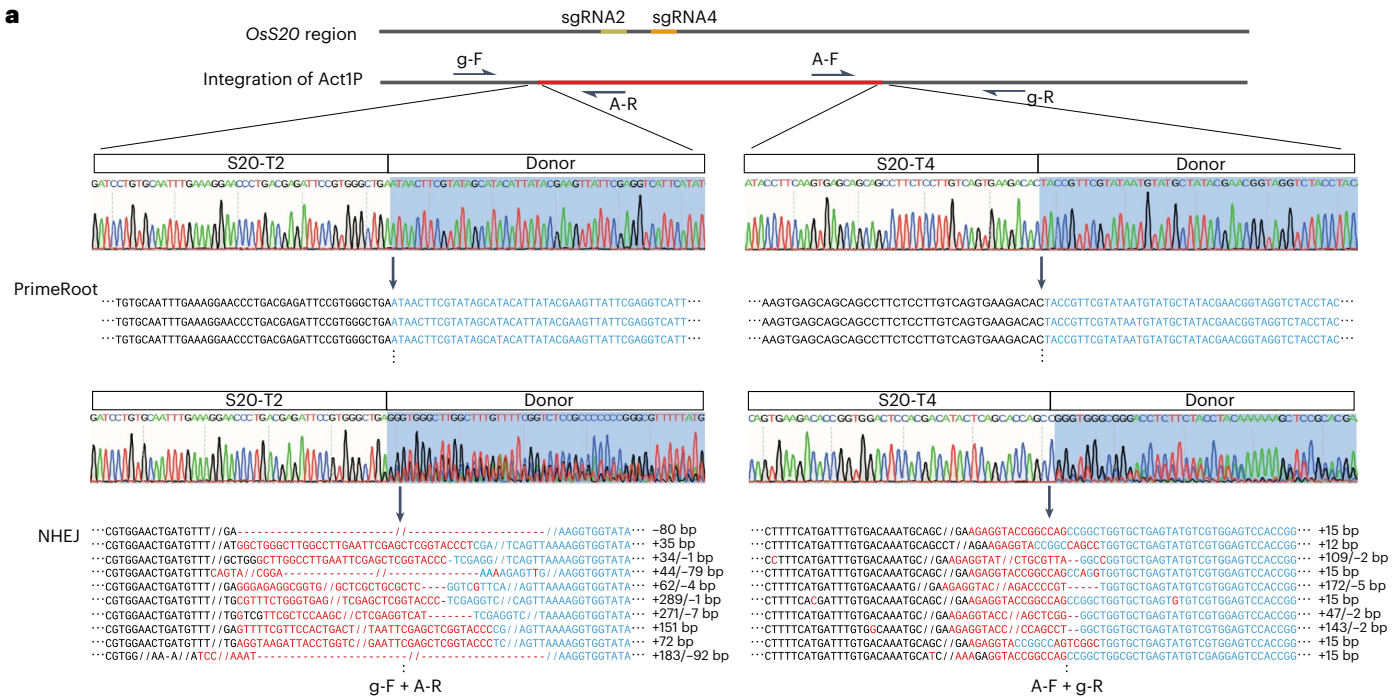
Fig. 2 | Development of improved PrimeRoot systems. a, Schematic overview of PrimeRoot.v1, PrimeRoot.v2N and PrimeRoot.v2C constructs. **b**, Schematic diagram of the all-in-one fluorescence reporter (AR) for evaluating PrimeRoot activity in plant protoplasts. **c**, Microscopic fluorescence images of protoplasts transformed with the AR and six different PrimeRoot editor constructs. Scale bars, 800 µm. **d**, Percent GFP$^+$ plant protoplasts as evaluated using the AR and measured by flow cytometry. Values and error bars represent the means and standard errors of the mean for three independent biological replicates. *P* values were obtained using the two-tailed Student's *t*-test: $^{**}P < 0.01$, $^{***}P < 0.001$. **e**, Comparison of the GFP insertion efficiencies of PrimeRoot.v1-Cre/FLP, PrimeRootv.2N-Cre/FLP and PrimeRoot.v2C-Cre/FLP at four endogenous sites as measured by ddPCR. Values and error bars represent the means and standard errors of the mean for three independent biological replicates. *P* values were obtained using the two-tailed Student's *t*-test: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. **f**, Schematic overview of four donor constructs. **g**, Percentages of donor insertions across four endogenous sites induced by PrimeRoot.v2C-Cre measured by ddPCR. Values and error bars represent means and standard errors of means for three independent biological replicates. *P* values were obtained using the two-tailed Student's *t*-test: $^{NS}P > 0.05$, $^{*}P < 0.05$. **h**, Percentages of precise insertions of recombinase sites generated by dual-ePPE at six maize endogenous genomic sites as measured by high-throughput sequencing. Values and error bars represent means and standard errors of means for three independent biological replicates. **i**, Percentages of GFP insertions across six maize endogenous sites induced by PrimeRoot.v2C-Cre measured by ddPCR. Values and error bars represent means and standard errors of means for three independent biological replicates. aa, amino acid; NS, not significant.

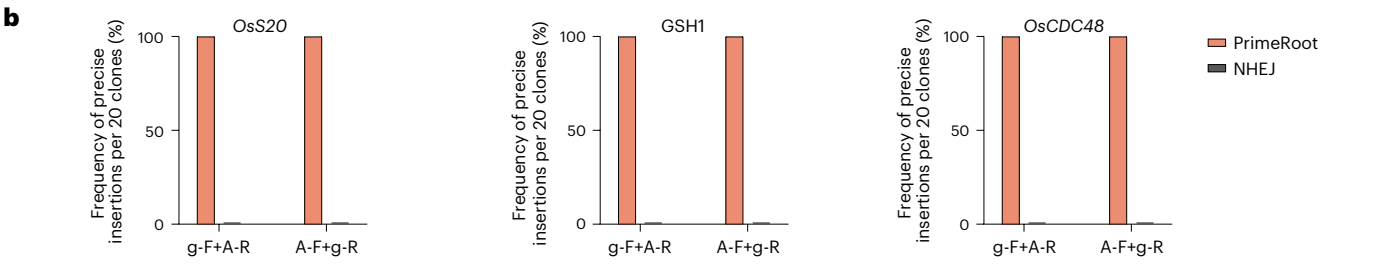
https://doi.org/10.1038/s41587-023-01769-w

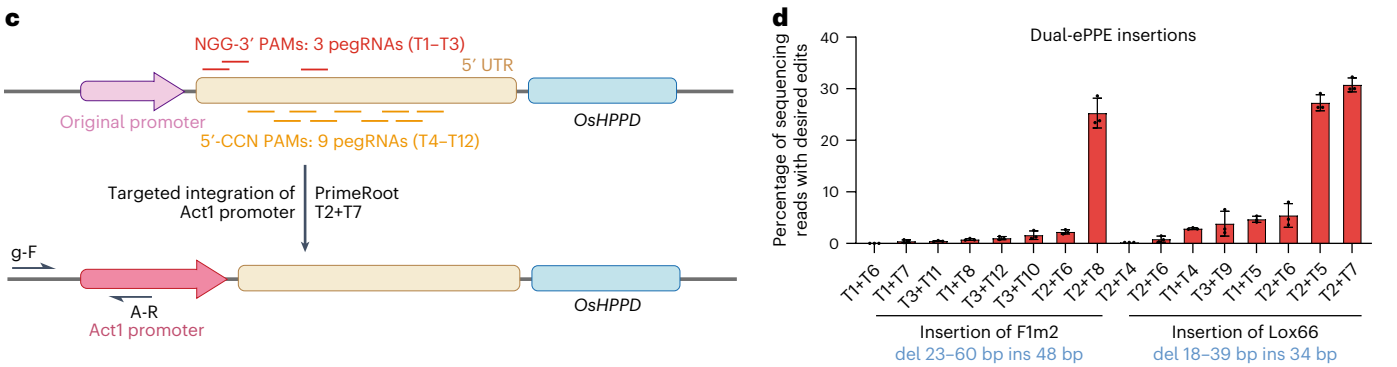
inserted segment of individual clones. When we selected 20 clones at random from the PrimeRoot-treated and the NHEJ-treated Act1P insertion samples, we found that all 20 generated by PrimeRoot contained the precisely inserted sequences as expected, whereas all 20 NHEJ inserts contained random DNA base indels at their junctions (Fig. 3a,b).



https://doi.org/10.1038/s41587-023-01769-w







e

Target site	Donor inserted	Number of regenerated plants	Precise insertion events (%)
OsHPPD-T2+T7	Act1P (1.4 kb)	507	12 (2.4%)

Fig. 3 | Comparison of targeted DNA insertions mediated by PrimeRoot and NHEJ. a, Sanger sequencing traces of the junctions between *OsS20* and the Act1P donor in individual edited mutants of protoplast created by PrimeRoot or CRISPR–Cas9-mediated NHEJ. sgRNA2 is the Cas9 target site used for NHEJ insertion, and sgRNA2 and sgRNA4 are the paired targeting sites used in PrimeRoot. g-F, A-R, A-F and g-R are the forward and reverse primers used for PCR and Sanger sequencing. Bases in blue are donor sequences, and red bases are sporadic indels. **b**, Insertion precision frequencies of PrimeRoot and NHEJ as measured by the number of individual clones in 20 randomly selected clones that contain the correct junctions between the donor and the endogenous genome. **c**, Schematic diagram of the PrimeRoot design for inserting a promoter into the 5′ UTR of a rice gene. **d**, Recombinase site insertion frequencies by dual-ePPE with different combinations of pegRNA designs in the 5′ UTR of *OsHPPD* as measured by high-throughput sequencing. Values and error bars represent means and standard errors of the mean of three independent biological replicates. **e**, Summary of the statistics for PrimeRoot-mediated Act1P-targeted insertion into the 5′ UTR of *OsHPPD*. del, deletion; ins, insertion.

We next used PrimeRoot and CRISPR-mediated NHEJ to insert Act1P and Act1P-*pigmR* sequences into genomic sites in rice calli (Extended Data Fig. 4c). After delivery and callus induction, we analyzed 95 calli clones from each treatment to compare editing efficiencies and precision. PrimeRoot generated two precise Act1P insertions and two precise Act1P-*pigmR* insertions, whereas NHEJ generated three imprecise insertions of Act1P and one imprecise insertion of Act1P-*pigmR* (Extended Data Fig. 4c,d). These results show that PrimeRoot is an effective editing tool for creating large, targeted precise DNA insertions in contrast to NHEJ, which relies heavily upon double-strand DNA breaks as intermediates.

Precise targeted insertion of an actin promoter into the 5′ UTR of *OsHPPD*

Many desirable agronomic traits are quantitative, depending on the upregulation or downregulation of some gene or depending on tissue-specific expression. To see if PrimeRoot could insert favorable promoters accurately upstream of targeted genes, we used PrimeRoot to knock-in a strong promoter into the 5′ UTR of *OsHPPD* (Fig. 3c). We designed 16 pairs of pegRNAs in the 5′ UTR and compared their RS insertion editing efficiency in rice protoplasts. We identified T2+T7 as the optimal pair of pegRNAs, with a 30% RS insertion frequency (Fig. 3d). We next used PrimeRoot and T2+T7 to insert the rice Actin1 promoter (Act1P) into rice calli by particle bombardment. We identified edited plants by amplifying the junction between the genome and the inserted donor sequence and assessed insertion precision by Sanger sequencing. We detected a total of 12 precise Act1P insertion events among 507 regenerated rice plants (2.4%) (Fig. 3e). These results establish PrimeRoot as an effective genome insertion tool for introducing new genetic regulatory elements into plant genomes for breeding purposes.

Targeted insertion of gene cassettes into GSH regions in rice

To permit safe insertion of transgenes into plant genomes, we predicted GSH regions over the entire Kitaake rice genome. Based on previous studies on GSHs[51,52], we used a variety of algorithms to identify regions some distance from elements, such as gene-coding regions, small RNAs, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), transfer RNAs (tRNAs), promoters, enhancers, long terminal repeats (LTRs) and more. In this way, we generated a new set of GSH regions (Fig. 4a) comprising 30 regions and totaling 40 kb (Fig. 4b).

We selected GSH1 as a proof-of-concept region and designed four pairs of pegRNAs for inserting an RS in this region. When comparing RS insertion efficiencies using dual-ePPE at GSH1, we found that the T1+T2 pair was the most efficient, giving RS insertion efficiencies of more than 40% (Fig. 4c). We then chose to examine the insertion of *pigmR* into this GSH1 region[53]. We constructed a 4.9-kb DNA donor cassette comprising an Act1P driving *pigmR* expression (Extended Data Fig. 4a), and we co-delivered plasmid constructs expressing PrimeRoot. v2C-Cre using two pegRNAs (T1+T2) driving Lox66 insertion and the *pigmR* donor cassette into rice calli by particle bombardment (Fig. 5a). After plant regeneration, we used specific F and R primers to amplify the junctions between the rice genome and the Act1P-*pigmR* expression cassette to identify edited mutants (Fig. 5b). Gel electrophoresis and Sanger sequencing identified 19 Act1-*pigmR* insertion events out of 744 regenerated plants (2.6%) (Fig. 5c,f). Notably, all 19 junctions yielded amplified products of the same size and were shown by sequencing to be the result of precise insertion events in which the ends of the donor cassette were exactly as predicted. We further analyzed the insertion events of the T1 generation in subsequent experiments, and we found that three out of 24 T1 precise insertion plants were identified as isolation of the Cre component (Supplementary Fig. 1). Indeed, although the precise targeted insertion edits were heritable, the isolation of other transgenic components when using particle bombardment is limited, so we next sought to develop a delivery method to perform *Agrobacterium*-mediated PrimeRoot.

To compare results obtained with PrimeRoot with those yielded by another commonly used plant transgenesis method, we transformed *Agrobacterium* carrying the Act1P-*pigmR* expression cassette vector into rice calli and selected at random 30 positive transformation events identified by PCR of regenerated plants. We performed whole-genome sequencing of these 30 plants and identified 62 different Act1P-*pigmR* cassette insertion events (Fig. 4e and Extended Data Fig. 5b,c). Subsequent analyses showed that these insertion events were located randomly throughout the genome, with most in coding regions or other conserved elements such as small RNAs; notably, none was in any of the 30 predicted GSH regions of the Kitaake rice genome (Fig. 4e and Extended Data Fig. 5d,e).

We also performed whole-genome sequencing of 12 PrimeRoot-edited plants to evaluate PrimeRoot's specificity. We first identified sites in the Kitaake genome with up to five mismatches with the pegRNA targeting sequences or up to 10 mismatches with the Lox66 recombination site (Fig. 4d). We identified 59 sites for pegRNA-1, 61 sites for pegRNA-2 and 40 sites for Lox66. When we then examined each of these sites for all 12 sequenced plants, we found that they were all of wild-type sequences and did not contain any edits as a result of undesired off-target editing (Fig. 4d).

We also used Sanger sequencing to examine the insertion junctions of individual insertion events. While the *Agrobacterium*-mediated insertion events were extremely imprecise and contained indels, all the junctions between donor and host genome created by PrimeRoot were precise (Fig. 4f). These results show that *Agrobacterium*-mediated insertions are random and imprecise, whereas PrimeRoot is capable of specifically and precisely integrating a donor DNA segment of interest into a defined GSH region.

Lastly, we evaluated the disease resistance of the PrimeRoot-induced mutants with Act1P-*pigmR* inserted into the GSH1 region. We inoculated rice blast race Guy11 onto the leaves of wild-type and PrimeRoot-edited rice plants by dot-joining. After 6 days, we measured the lengths of the lesions formed by bacterial growth. Whereas the lesions on the control plants averaged 1.2 cm, those on the edited plants averaged only 0.53 cm, a 2.3-fold improvement in resistance (Fig. 5d,e). These results demonstrate PrimeRoot to be an effective tool for generating gene insertions and performing reliable molecular plant breeding.

Efficient targeted gene insertion by sequential transformation of PrimeRoot and donor components into rice plants

To further improve the editing efficiency of PrimeRoot in plants, we speculated that the sequential transformation of PrimeRoot and donor components might enhance insertion efficiencies during plant regeneration. We used Lox66 and the FRT1 variant F1m2 as the landing site to test whether sequential transformations of PrimeRoot and the donor components into rice calli (PrimeRoot. v3) would improve overall edited plant recovery efficiency. We first evaluated dual-ePPE-mediated RS insertion in rice calli and achieved editing efficiencies up to 84.7% (Fig. 6b). We transformed the PrimeRoot reagents into calli by *Agrobacterium*-mediated T-DNA insertion and, after 1 month of selection with hygromycin, we enriched for edited calli containing the desired RS insertion (Fig. 6c). These calli were then used as substrates for a second round of transformation containing the donor vector delivered by either particle bombardment or *Agrobacterium*. After selection by G418 and regeneration, we examined regenerated plants and measured editing frequencies of desired insertion events (Fig. 6a). When we performed the donor delivery by particle bombardment, we found that the editing efficiency of precise insertions of Act1P into the 5′ UTR of *OsHPPD* by Cre-Lox66 was 7.1% and by FLP-F1m2 was 8.3%, which is threefold and 3.5-fold higher than when performing all-in-one plant transformations, respectively. When evaluating the editing efficiency of precise insertions of Act1P-*pigmR* into GSH1, we obtained efficiencies

https://doi.org/10.1038/s41587-023-01769-w

a

GSH criteria	Theoretical basis	Candidates
5 kb away from gene-encoding regions	Avoiding disrupting functional regions associated with gene expression	30 regions ~40.12 kb
30 kb away from miRNAs 20 kb away from lncRNAs 20 kb away from tRNAs	Avoiding disruption of RNAs involved in differentiation and development programs determining cell fate	
5 kb away from promoters and enhancers	Reducing the chance of altering physiological gene expression	
10 kb away from centromeres	Prevents alterations in DNA replication, cell division and normal ageing	

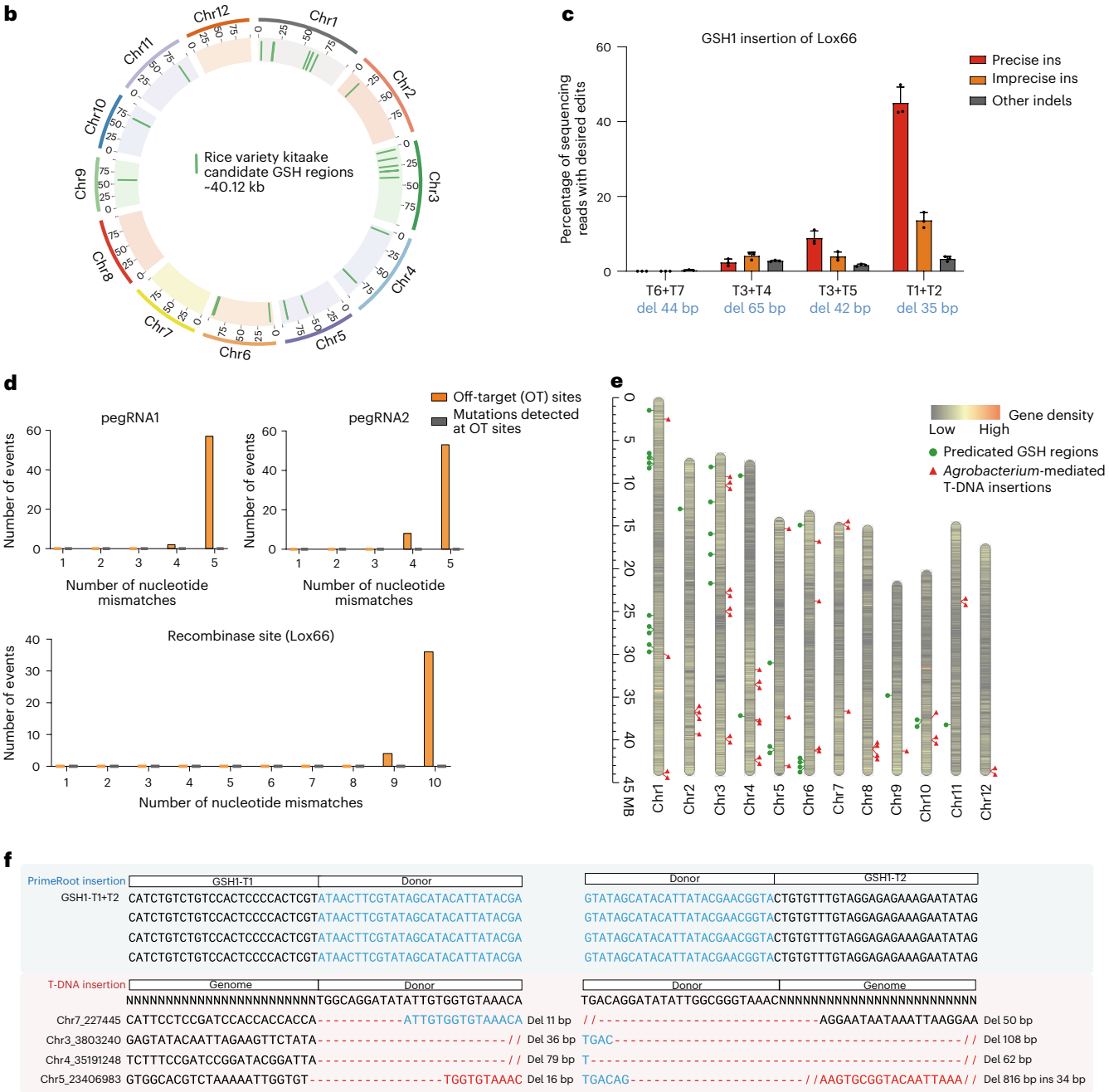
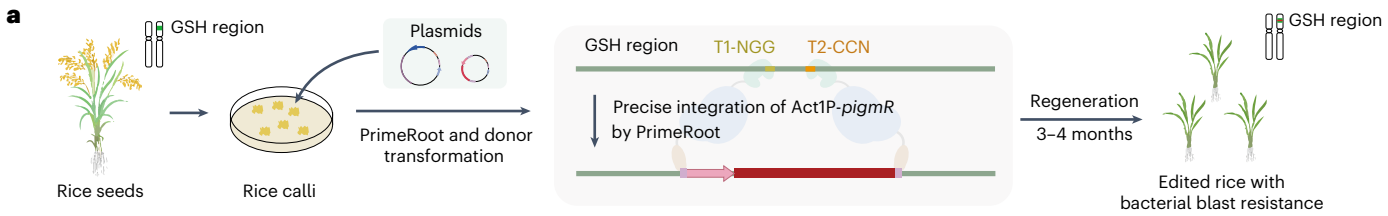
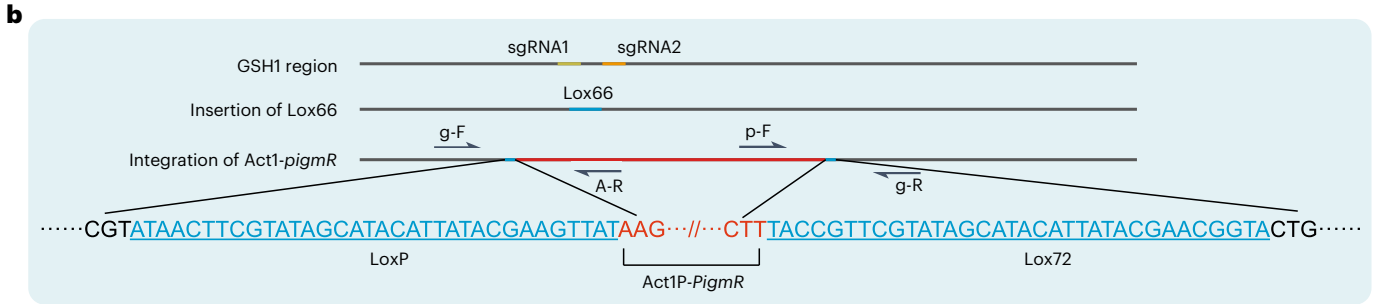
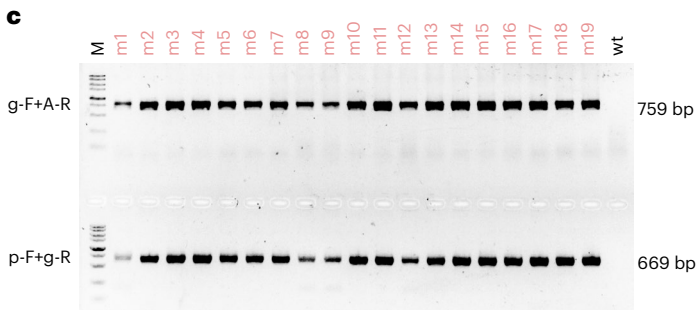


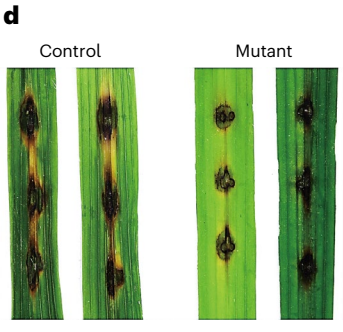
Fig. 4 | Prediction of GSH regions and specificity analysis. a, Criteria for predicting candidate GSH regions. **b**, Graphical summary of the 30 predicted GSH regions in the Kitaake genome. **c**, Recombinase site (Lox66) insertion efficiencies of dual-ePPE in the GSH1 region of Kitaake protoplasts as measured by high-throughput sequencing. Values and error bars represent means and standard errors of the means of three independent biological replicates. **d**, Numbers of pegRNA and RS off-target sites predicted based on numbers of mismatches and the corresponding number of sites with mutations as evaluated by whole-genome sequencing at each site. **e**, A graphical overview of insertion events mediated by *Agrobacterium* transformation ($n = 30$) in rice plants. The *y* axis represents chromosome size; green circles represent the 30 predicted GSH regions in the Kitaake rice genome; red arrows represent *Agrobacterium*-mediated insertions; and weak and strong shading of the chromosome indicates density of coding genes. **f**, Sanger sequencing traces of junctions of insertions mediated by *Agrobacterium* T-DNA insertion and PrimeRoot. Blue bases represent donor sequences, and red bases represent sporadic DNA insertions or deletions. del, deletion; ins, insertion.

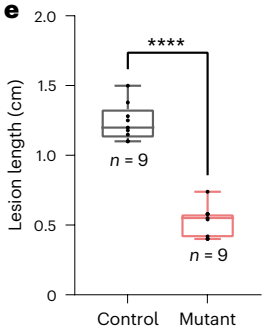
https://doi.org/10.1038/s41587-023-01769-w











f

Target region	Donor inserted	Number of regenerated plants	Precise insertion events (%)
GSH1	Act1P-*pigmR* (4.9 kb)	744	19 (2.6%)

Fig. 5 | Targeted integration of the Act1P-*pigmR* gene cassette into GSH1 to confer bacterial blast disease resistance in rice plants. a, Schematic overview of the use of PrimeRoot to insert the Act1P-*pigmR* gene cassette into the GSH1 region in rice plants. **b**, Schematic overview and Sanger sequencing primers for evaluating Act1P-*pigmR* integration into the GSH1 region. sgRNA1 and sgRNA2 are the paired pegRNAs for PrimeRoot insertion, and g-F, A-R, A-F and g-R are the primers used for PCR and Sanger sequencing. **c**, Gel electrophoresis of PCR products of the insertion junctions between the donor cassette and endogenous genome. The 759-bp band derives from the 5′ insertion junction (g-F+A-R) and the 669-bp band from the 3′ junction (A-F+g-R); wt is an untreated negative control. **d**, Visual portrayal of levels of rice blast disease resistance of untreated control rice plants and the edits resulting from insertion of the Act1P-*pigmR* gene cassette into GSH1; 20 days post-inoculation (dpi) with *M. oryzae* strain Guy11. **e**, Lengths of bacterial growth lesions of three edited mutants and three untreated controls measured 20 dpi with *M. oryzae* strain Guy11. Bounds of whiskers represent the values of Min to Max ($n = 9$ biologically independent samples). *P* values were obtained using two-tailed Student's *t*-tests: ****$P < 0.0001$. **f**, Statistics of PrimeRoot-mediated Act1P-*pigmR* targeted insertion into the GSH1 region of rice plants.

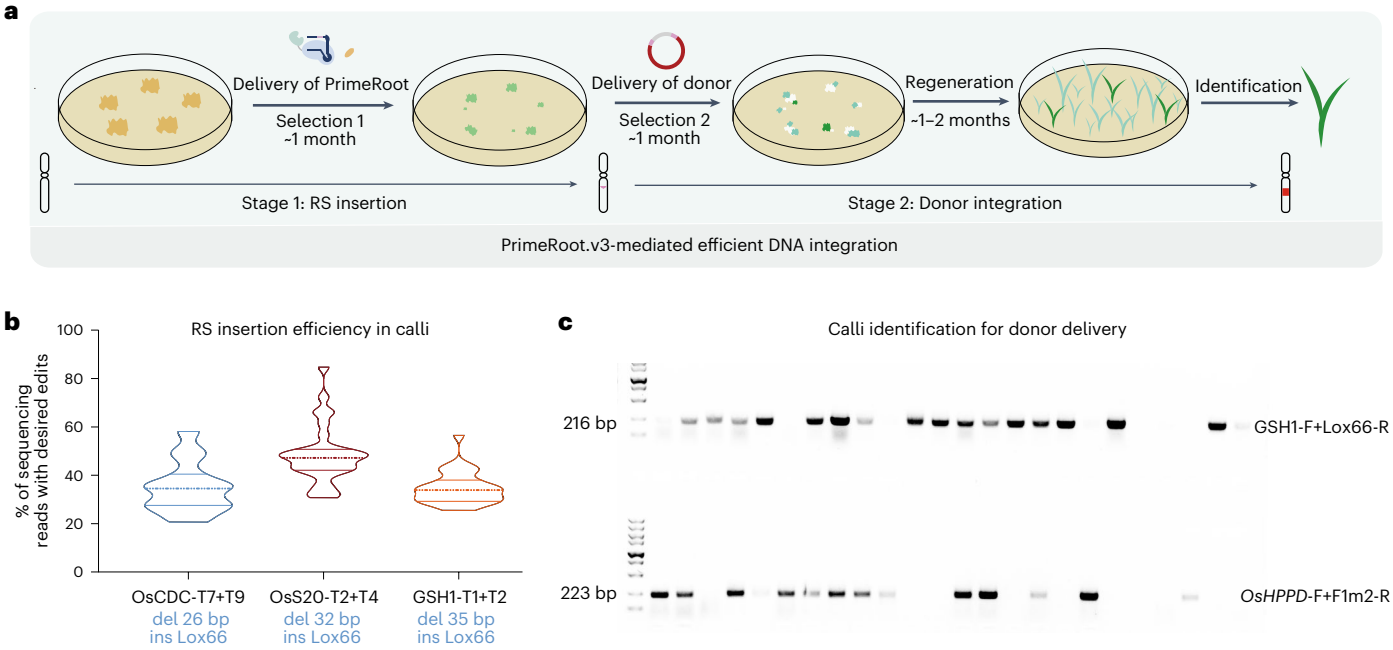
of 4.2% by Cre-Lox66 and 6.3% by FLP-F1m2, which are 1.6-fold and 2.4-fold higher than when performing all-in-one plant transformations, respectively (Fig. 6d). When we delivered the donor by *Agrobacterium* transformations, we obtained an efficiency of 3.9% of precise insertion events comprising Act1P-*pigmR* inserted into the GSH1 site (Fig. 6d). These results highlight that PrimeRoot. v3 can be performed using different delivery methods, which further improves the editing efficiency of precise targeted gene insertions in plants.

Discussion

The ability to insert novel DNA sequences specifically and precisely into plant genomes is a major step toward realizing precision plant breeding. Here we describe PrimeRoot, a new genome insertion tool capable of inserting large genetic cargos into plant genomes. We found that the combination of an optimized epegRNA and our recently engineered ePPE was capable of efficiently inducing recombinase site insertion events in rice at frequencies approaching 50%. Although PE in its current state performs satisfactorily at some sites, there remains a need to develop better pegRNA design methods to permit robust editing genome wide. Furthermore, continued work on engineering prime editor proteins is needed to increase editing efficiency.

We developed a fluorescence recombinase reporter system to identify the best recombinase for use in rice. Although we identified Cre and FLP as the most active recombinases currently available, there is a great need to identify other recombinases for use in plant cells. It is also intriguing that different recombinase site sequences influence recombination activity. We expect that future efforts to optimize the recombinase and the corresponding recombinase sites will expand the utility of PrimeRoot. In this study, we used PrimeRoot to precisely insert DNA segments of up to 11.1 kb into the rice genome. Future

https://doi.org/10.1038/s41587-023-01769-w



d

Donor delivery method	Target region	RS inserted	Donor inserted	Number of regenerated plants	Precise insertion events (%)
Particle bombardment	*OsHPPD*	Lox66	Act1P (1.4 kb)	84	6 (7.1%)
Particle bombardment	*OsHPPD*	F1m2	Act1P (1.4 kb)	36	3 (8.3%)
Particle bombardment	GSH1	Lox66	Act1P-*pigmR* (4.9 kb)	213	9 (4.2%)
Particle bombardment	GSH1	F1m2	Act1P-*pigmR* (4.9 kb)	348	22 (6.3%)
Agrobacterium	GSH1	F1m2	Act1P-*pigmR* (4.9 kb)	416	16 (3.9%)

Fig. 6 | PrimeRoot.v3 for efficient precise targeted gene insertion by sequential transformation. a, Schematic overview of PrimeRoot.v3. **b**, Recombinase site insertion frequencies by dual-ePPE at three endogenous sites in rice calli. Bounds of whiskers represent the values of Min to Max ($n = 30$ biologically independent samples). **c**, Gel electrophoresis of PCR products in PrimeRoot transformation calli used to rapidly identify donor calli for subsequent transformation. **d**, Statistics of PrimeRoot-mediated precise insertion events in rice plants by PrimeRoot.v3. del, deletion; ins, insertion.

advances in this methodology should improve plant PE efficiency and recombination and expand the limit of insertion fragment size; this will enable stacking of more complex traits and other plant synthetic biology applications.

We highlighted PrimeRoot's superior efficiency and precision over current NHEJ methods for inserting large DNA donors, and we also demonstrated PrimeRoot's superior accuracy and programmability compared to *Agrobacterium*-mediated T-DNA insertion, which is one of the most common ways to perform transgenesis in plants. New genes of interest affecting agronomic plant traits are constantly being discovered, but there remains a need to quickly adapt these genes for breeding new crops. The ability to generate specific insertions at GSH sites will rapidly accelerate breeding because it allows one to generate insertions at any desired site and in any number of crop varieties in just one breeding cycle.

We identified 30 GSH regions in the rice Kitaake genome as proof of concept for selection criteria. To expand the utility of GSH regions, we performed genome annotations and comparative genomics across 33 rice species and identified one shared GSH region of interest (Extended Data Fig. 5a). These methods suggest the possibility of identifying species-specific GSH sites by evaluating many genomes of a particular plant species. Future studies should evaluate the potential of GSH sites for use in transgenic crop breeding.

With rapidly changing climates and a growing world population, there exists an urgent need to breed new crop varieties[54]. PrimeRoot offers many opportunities to engineer quantitative trait changes, trait stacking and more, all of which are useful for generating critical agronomic traits. Future applications, such as insertion of regulatory elements, tagging of endogenous genes and the introduction of new transgenes, will rely upon precise genome editing technologies to achieve precise molecular crop breeding and advance plant synthetic biology.

Online content

Any methods, additional references, Nature Portfolio 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/s41587-023-01769-w.

References

1. Fuchs, M. Pyramiding resistance-conferring gene sequences in crops. *Curr. Opin. Virol.* **26**, 36–42 (2017).
2. Shukla, V. K. et al. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* **459**, 437–441 (2009).
3. Shehryar, K. et al. Transgene stacking as effective tool for enhanced disease resistance in plants. *Mol. Biotechnol.* **62**, 1–7 (2020).
4. Zhang, B. Transgenic cotton: from biotransformation methods to agricultural application. *Methods Mol. Biol.* **958**, 3–15 (2013).

5. Guo, M., Ye, J., Gao, D., Xu, N. & Yang, J. *Agrobacterium*-mediated horizontal gene transfer: mechanism, biotechnological application, potential risk and forestalling strategy. *Biotechnol. Adv.* **37**, 259–270 (2019).
6. Stam, M. Review article: the silence of genes in transgenic plants. *Ann. Bot.* **79**, 3–12 (1997).
7. Iyer, L. M., Kumpatla, S. P., Chandrasekharan, M. B. & Hall, T. C. Transgene silencing in monocots. *Plant Mol. Biol.* **43**, 323–346 (2000).
8. Rajeevkumar, S., Anunanthini, P. & Sathishkumar, R. Epigenetic silencing in transgenic plants. *Front. Plant Sci.* **6**, 693 (2015).
9. Mumm, R. H. & Walters, D. S. Quality control in the development of transgenic crop seed products. *Crop Sci.* **41**, 1381–1389 (2001).
10. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
11. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
12. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
13. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
14. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
15. Gao, C. Genome engineering for crop improvement and future agriculture. *Cell* **184**, 1621–1635 (2021).
16. Dong, O. X. & Ronald, P. C. Targeted DNA insertion in plants. *Proc. Natl Acad. Sci. USA* **118**, e2004834117 (2021).
17. Lu, Y. et al. Targeted, efficient sequence insertion and replacement in rice. *Nat. Biotechnol.* **38**, 1402–1407 (2020).
18. Terada, R., Urawa, H., Inagaki, Y., Tsugane, K. & Iida, S. Efficient gene targeting by homologous recombination in rice. *Nat. Biotechnol.* **20**, 1030–1034 (2002).
19. Leibowitz, M. L. et al. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* **53**, 895 (2021). -+.
20. Alanis-Lobato, G. et al. Frequent loss of heterozygosity in CRISPR–Cas9-edited early human embryos. *Proc. Natl Acad. Sci. USA* **118**, e2004832117 (2021).
21. Cullot, G. et al. CRISPR–Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.* **10**, 1136 (2019).
22. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
23. Jiang, Y. Y. et al. Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol.* **21**, 257 (2020).
24. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell Discov.* **6**, 27 (2020).
25. Lin, Q. et al. High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat. Biotechnol.* **39**, 923–927 (2021).
26. Grindley, N. D., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75**, 567–605 (2006).
27. Van Duyne, G. D. Cre recombinase. *Microbiol. Spectr.* **3**, MDNA3-0014-2014 (2015).
28. Argos, P. et al. The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**, 433–440 (1986).
29. Albert, H., Dale, E. C., Lee, E. & Ow, D. W. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J.* **7**, 649–659 (1995).
30. Ahmadi, M., Damavandi, N., Akbari Eidgahi, M. R. & Davami, F. Utilization of site-specific recombination in biopharmaceutical production. *Iran. Biomed. J.* **20**, 68–76 (2016).
31. Jiang, L. et al. Target lines for recombinase-mediated gene stacking in soybean. *Theor. Appl. Genet.* **135**, 1163–1175 (2022).
32. Hirano, N., Muroi, T., Takahashi, H. & Haruki, M. Site-specific recombinases as tools for heterologous gene integration. *Appl. Microbiol. Biotechnol.* **92**, 227–239 (2011).
33. Ow, D. W. The long road to recombinase-mediated plant transformation. *Plant Biotechnol. J.* **14**, 441–447 (2016).
34. Gao, H. et al. Complex trait loci in maize enabled by CRISPR–Cas9 mediated gene insertion. *Front. Plant Sci.* **11**, 535 (2020).
35. Wang, Y., Yau, Y. Y., Perkins-Balding, D. & Thomson, J. G. Recombinase technology: applications and possibilities. *Plant Cell Rep.* **30**, 267–285 (2011).
36. Ioannidi, E. I. et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01527-4 (2022).
37. Anzalone, A. V. et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).
38. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
39. Vo, P. L. H. et al. CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* **39**, 480–489 (2021).
40. 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).
41. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9–cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
42. Manghwar, H., Lindsey, K., Zhang, X. & Jin, S. CRISPR/Cas system: recent advances and future prospects for genome editing. *Trends Plant Sci.* **24**, 1102–1125 (2019).
43. Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T. & Buchholz, F. Cre recombinase and other tyrosine recombinases. *Chem. Rev.* **116**, 12785–12820 (2016).
44. Wang, J. et al. Efficient targeted insertion of large DNA fragments without DNA donors. *Nat. Methods* **19**, 331–340 (2022).
45. Zong, Y. et al. An engineered prime editor with enhanced editing efficiency in plants. *Nat. Biotechnol.* **40**, 1394–1402 (2022).
46. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
47. Bruckner, R. C. & Cox, M. M. Specific contacts between the flp protein of the yeast 2-micron plasmid and its recombination site. *J. Biol. Chem.* **261**, 1798–1807 (1986).
48. Senecoff, J. F., Rossmeissl, P. J. & Cox, M. M. DNA recognition by the FLP recombinase of the yeast 2 μ plasmid: a mutational analysis of the FLP binding site. *J. Mol. Biol.* **201**, 405–421 (1988).
49. Li, J. et al. Gene replacements and insertions in rice by intron targeting using CRISPR–Cas9. *Nat. Plants* **2**, 16139 (2016).
50. Dong, O. X. O. et al. Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR–Cas9. *Nat. Commun.* **11**, 1178 (2020).
51. Aznauryan, E. et al. Discovery and validation of human genomic safe harbor sites for gene and cell therapies. *Cell Rep. Methods* **2**, 100154 (2022).
52. Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2011).

53. Deng, Y. et al. Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* **355**, 962–965 (2017).
54. Springmann, M. et al. Options for keeping the food system within environmental limits. *Nature* **562**, 519–525 (2018).

Methods

Plasmid construction

Plasmids expressing epegRNAs to produce short insertions were cloned as previously described[46,55]. Plasmids expressing epegRNAs for generating long insertions were fused with three fragments using a Uniclone One Step Seamless Cloning Kit (Genesand): fragment 1 contained an OsU3, TaU3 or pGSPE backbone[23]; fragment 2 contained a spacer and guide RNA scaffold; and fragment 3 contained the RT template, PBS and tevopreQ1. For recombinase screening, Cre, FLP, phiC31, Bxb1, B2R, Dre, KD and PSR1 were codon optimized for cereal plants and synthesized commercially by GeneScript (Supplementary Sequences). All of the donor vectors for recombinase screening fused with GFP-N or GFP-C sequence, intron and recombinase recognition sites were generated by Gibson assembly.

To construct rice transformation components for the all-in-one strategy, ePPE, recombinase and epegRNA expression cassettes were cloned into pH-ePPE[45]. Genes of interest were amplified using primer sets containing recombinase sites at their 5′ ends and cloned into the backbone with ampicillin resistance. To avoid recombination between the two recombination sites during Gibson assembly, we introduced a recombination site at one end at a time. For sequential transformation, we constructed donor vectors with nptIIgene and kanamycin resistance.

Isolation and transformation of rice protoplasts

Japonica rice cultivar Zhonghua11 was used to isolate protoplasts[56]. For transformation, high-quality plasmids were purified through the Wizard Plus Midipreps DNA Purification System (Promega). Then, 5 µg of each plasmid was mixed and introduced into rice protoplasts by PEG-mediated transformation. Transformation efficiency was checked after 24 h by fluorescence intensity. The mean efficiency was 30–50% quantified by flow cytometry. After incubation at 26 °C for 72 h, transformed protoplasts were collected by centrifugation for genomic DNA extraction or used in flow cytometry analysis.

Flow cytometry analysis

Samples were sorted and counted for GFP^+ cells. A FACSAria III (BD Biosciences) was used for flow cytometry, and FACSDiva version 6.1.3 and FlowJo version 7.6 software were used for analyzing results.

Next-generation sequencing and analysis of the results

Protoplast DNA was extracted by the DNA Quick Plant System (Tiangen Biotech). Specific primers with barcodes at their 5′ ends were designed to amplify the targeted sequence using 2× Phanta Max Master Mix (Vazyme) (Supplementary Table 3). PCR reactions were carried out as follows: 95 °C for 5 min and then 30 cycles of 95 °C for 15 s, 55 °C for 15 s and 72 °C for 15 s, followed by a final 72 °C extension for 5 min. PCR products were checked by electrophoresis in a 2.0% agarose gel. Roughly the same amount of each sample was mixed and purified using a Thermo Fisher Scientific GeneJET Gel Recovery Kit. DNA concentrations were measured with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), and pooled PCR products were sequenced commercially (Novogene) using the NovaSeq platform. Analyses of editing efficiencies were performed as previously described[13] with custom shell scripts to analyze the different insertion outcome types.

Frequency of precise insertion was calculated as: percentage (number of reads with seamless insertion without byproducts) / (number of total reads). Frequency of imprecise insertion was calculated as: percentage (number of reads with insertions containing at least half of the continuous donor sequence but not the precise insertion) / (number of total reads). Frequency of other indels was calculated as: percentage (number of reads with indels but neither precise or imprecise insertions) / (number of total reads).

Measurement of insertion efficiency by ddPCR

Primers and TaqMan probes for genome–donor junctions and *OsCDC48* were designed (Supplementary Table 2). Reaction mixtures contain 50 ng of rice genomic DNA; 1.8 µl of each primer; 0.5 µl of each probe; 10 µl of ddPCR Supermix for Probes (Bio-Rad, 1863026); and water to 20 µl. The reaction mixtures were transferred into DG8 cartridges and turned into droplets with 70 µl of DG Oil using a QX200 Manual Droplet Generator (Bio-Rad, 186-4002). ddPCR was performed under the following conditions: 94 °C for 10 min and 50 cycles of 95 °C for 30 s and 58 °C for 2 min, with a ramp rate of 1 °C s^{-1}. Final incubation was at 98 °C for 10 min. The droplets were read on a QX200 Droplet Reader (Bio-Rad, 1864001), and data were analyzed with QuantaSoft (version 1.6).

Transformation of rice calli by particle bombardment

Plasmids for the ePPE, recombinase and epegRNA expression cassette and donor plasmid were simultaneously delivered into embryonic calli of *Oryza sativa* L. (cv. Kitaake) as previously described[57]. Then, 50 µg ml^{-1} hygromycin was used to select transgenic calli, and transgenic plantlets had regenerated on the selection medium 10–12 weeks later.

Agrobacterium-mediated transformation of Act1P-*pigmR* in rice

The Act1P-*pigmR* expression cassette was transformed into *A. tumefaciens* strain EHA105 by electroporation. *Agrobacterium*-mediated transformation of *O. sativa* L. (cv. Kitaake) was conducted as previously described[58]. Then, 50 µg ml^{-1} hygromycin was used to select transgenic calli. Transgenic plantlets had regenerated on the selection medium 10–12 weeks later.

Sequential transformation of PrimeRoot in rice

In the first step, binary plasmid containing ePPE, pegRNA and recombinase expression cassette was transformed into *A. tumefaciens* strain EHA105 by electroporation. *Agrobacterium*-mediated transformation of *O. sativa* L. (cv. Kitaake) was conducted as previously described[58]. Then, 50 µg ml^{-1} hygromycin was used to select the transgenic callus for 3–4 weeks. In the second step, either particle bombardment or *Agrobacterium* can be used to transform the RS inserted calli obtained from the first step. Donor cassette was cloned into a vector used for *Agrobacterium* transformation and subsequently used for the particle bombardment or *Agrobacterium* transformation. Then, 150 mg L^{-1} G418 was used to select the transgenic callus. Transgenic plantlets were regenerated on the selection medium after 10–12 weeks.

Analysis of the precision of long DNA fragments insertion

To analyze the precision of long fragment insertion, we amplified the 5′ and 3′ genome–donor junction sequences. The PCR products were Sanger sequenced by the Beijing Genomics Institute, and alignments were analyzed. Individual clones were isolated and cloned into the commercial T vector using a 5-min TA/Blunt-Zero Cloning Kit (Vazyme). Individual clones were then sequenced through Sangar sequencing. Primers for PCR reactions are listed in Supplementary Table 3.

Identification of mutant plants

Two or three rice plantlets were sampled and mixed into individual wells and extracted by the SDS-based DNA extraction method. Next, DNA samples were amplified using primer sets that could detect genome–donor junctions. Plantlets in positive wells were sampled and identified individually. The PCR products of candidate mutants were checked by electrophoresis in 2.0% agarose gels and validated by Sanger sequencing. Each mutant DNA was extracted from independent leaves at least three times and amplified by at least two primer sets for verification. Primers for mutant identification are listed in Supplementary Table 3.

Prediction of GSH regions of rice kitaake genome

Annotations of rice Kitaake genome and coding genes are from phytozome (https://phytozome-next.jgi.doe.gov/info/OsativaKitaake_v3_1)[59].

Annotation of tRNA regions was performed by tRNAscan-SE version 2.0 software using default parameters to exclude the tRNA coding regions and their surrounding 20-kb-long regions[60]; miRNA regions were annotated by aligning cmscan of tRNAscan-SE version 2.0 with the Rfam database to avoid the miRNAs and their surrounding 30-kb-long regions; cpc2 (CPC2 standalone-1.0.1), PlncPro version 1.2.2 and pfam were all used to annotate lncRNAs and their intersection, including the surrounding 20-kb-long regions[61–63].

The RNA sequencing data of four samples of Kitaake (GenBank accession numbers SRP182736, SRP182738, SRP182741 and SRP182741) were alignd to the Kitaake genome through Hisat2-2.2.1 and stringTie version 2.2.1 to the assembled transcripts, and then the transcripts whose transcripts per million (TPM) value is greater than 1 in each tissue are merged, and the transcripts with length greater than 200 bp are retained. The cuffcompare function of cufflinks-2.2.1 is used to compare the genome coordinates with the existing coding genes of Kitaake, and the transcripts of tag type 'u, x, i, j, o' are retained to analyze the coding ability. The filtered transcripts were predicted by cpc2, PlncPro and pfam, respectively. In total, 4,839 transcripts of the type 'non-coding' of cpc2 were obtained. A training set of PlncPro was constructed from known rice lncRNAs and mRNAs. The dataset of known lncRNAs in rice was from the following three databases: NONCODE V6 (http://www.noncode.org/datadownload), RiceLncPedia (http://3dgenome.hzau.edu.cn/RiceLncPedia#/Data) and cantata (http://cantata.amu.edu.pl/DOWNLOADS). The data of known mRNA regions were derived from the Kitaake genome annotation file, and the 5,298 candidate lncRNA regions were predicted by PlncRNA. The candidate transcripts assembled of stringTie by PfamScan were compared with Pfam-A, and 5,429 candidate lncRNA regions were obtained. Finally, 3,337 candidate lncRNAs remained by taking the intersection of the three results.

The LTRs, such as centromeric regions, were annotated by Tandem Repeats Finder version 4.04 (parameter, 112 80 5 200 2000 -d -h) based on the tandem repeat features to avoid the surrounding 20-kb-long regions[64]. Lastly, promoters and enhancers and the surrounding 5-kb-long regions were annotated to avoid affecting gene expression and distal gene–enhancer interactions[65–68].

Combine all the annotated element intervals and files with specific upstream and downstream distances, and then use BEDTools to take the complement of the genome and perform sliding window statistics on the complement genome segments (the window size is 10 kb, and the window step size is 1 kb). The GC ratio of the sliding window segment, the gap ratio and the specificity of the genome were counted to filter the results to select the regions of better specificity (the number of blast alignments in the whole genome is less than or equal to 1) and with a minimal fragment length of 1 kb. Following these criteria, we identified 30 regions in the Kitaake rice genome spanning a total of 40 kb (Supplementary Table 6).

Prediction of common GSH regions of rice variety genomes

GSH regions of 32 other rice varieties were also predicted using the same method as for the Kitaake rice mentioned above. A common GSH region was obtained by bwa comparison (Supplementary Table 7).

Whole-genome sequencing and data analysis

A total of 44 plants, including 30 *Agrobacterium*-mediated T-DNA insertion mutants, 12 PrimeRoot mediated insertion mutants and two wild-types, were used to analyze insertions. They were sequenced using the NovaSeq platform (Novogene). An average of 22 Gb of data (~50×) were generated per plant. All the raw data reads were mapped to the reference genome (OsativaKitaake_499_v3.0) to perform bwa comparison, extract all the paired-end sequences that had aligned to the vector and then use SOAPdenovo to assemble under different parameters of 31-mer, 41-mer, 51-mer and 61-mer. All assembled Contig sequences were aligned to the vector sequence and the reference genome by bwa. The insertion sequence interface was obtained by Contig alignment analysis, whereby half of the alignment appeared on the genome and the other half appeared on the vector. Background filtering was performed by comparison with a wild-type control to finally obtain all insertion positions of the materials. The position where the sequence was aligned to the reference genome was the insertion site, and the position where the sequence was aligned to the vector was the start and end positions of the vector (Supplementary Table 11).

Predication of off-target sites and identification of mutations at these sites of whole-genome sequencing samples

Off-target sites were predicted using an offline version of Cas-OFFinder. The filtered clean data were aligned to the reference genome with BWA-MEM. The resulting SAM files were converted to source BAM files by SAMtools. Reads whose mapQ value was more than 30 were selected for subsequent mutation detection analysis. Single-nucleotide polymorphism (SNP) detection was performed using the UnifiedGenotyper of the GATK (GATK3.5). The predicted off-target sites and all mutation (SNPs and indels) sites were cross-aligned through BEDTools to obtain all mutations that finally fall within the predictable off-target sites (Supplementary Table 10).

Inoculation of bacterial rice blast fungus

The *oryzae* isolate Guy11 was grown on oatmeal medium for 2 weeks to produce conidial spores. The spores were collected in sterile water with 0.02% (v/v) Tween 20 and adjusted to 5×10^5 spores per milliliter before punch inoculation. Four-leaf stage T0 regenerated mutants were used as previously described[69], and a 20-µl volume of the spore suspension was applied. Inoculated leaves were kept in a growth chamber at 28 °C and 90% humidity in the dark for the first 24 h, followed by a photoperiod of 16-h light and 8-h dark. Photographs were taken 5–7 days after inoculation.

Statistical analysis

GraphPad Prism 9 software was used to analyze the data. All numerical values are presented as mean ± s.e.m. Differences between control and treatments were tested using two-tailed Student's *t*-tests.

Reporting summary

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

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 upon reasonable request. All sequencing data were deposited in the National Center for Biotechnology Information BioProject under accession code PRJNA879048 (ref. 70). For sequence data, rice OsKitaake identifiers (https://phytozome-next.jgi.doe.gov/) are: OsKitaake03g041600 (*OsCDC48*), OsKitaake08g207700 (*OsIPA1*), OsKitaake02g183100 (*OsALS*) and OsKitaake08g018600 (*OsS20*). Source data are provided with this paper.

References

55. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
56. Shan, Q. W. et al. Targeted genome modification of crop plants using a CRISPR–Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
57. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **9**, 2395–2410 (2014).
58. Shan, Q. W., Zhang, Y., Chen, K. L., Zhang, K. & Gao, C. X. Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. *Plant Biotechnol. J.* **13**, 791–800 (2015).
59. Jain, R. et al. Genome sequence of the model rice variety KitaakeX. *BMC Genomics* **20**, 905 (2019).

60. Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
61. Zhang, Z. et al. PMRD: plant microRNA database. *Nucleic Acids Res.* **38**, D806–D813 (2010).
62. Yi, X., Zhang, Z., Ling, Y., Xu, W. & Su, Z. PNRD: a plant non-coding RNA database. *Nucleic Acids Res.* **43**, D982–D989 (2015).
63. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).
64. Villasante, A., Abad, J. P. & Mendez-Lago, M. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc. Natl Acad. Sci. USA* **104**, 10542–10547 (2007).
65. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
66. Sun, J. et al. Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics Proteomics Bioinformatics* **17**, 140–153 (2019).
67. Vangala, P. et al. High-resolution mapping of multiway enhancer-promoter interactions regulating pathogen detection. *Mol. Cell* **80**, 359–373 (2020).
68. Gao, Y., Chen, Y., Feng, H., Zhang, Y. & Yue, Z. RicENN: prediction of rice enhancers with neural network based on DNA sequences. *Interdiscip. Sci.* **14**, 555–565 (2022).
69. Yang, C. et al. Poaceae-specific cell wall-derived oligosaccharides activate plant immunity via OsCERK1 during *Magnaporthe oryzae* infection in rice. *Nat. Commun.* **12**, 2178 (2021).
70. Sun, C. et al. Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *National Center for Biotechnology Information (NCBI)*. https://dataview.ncbi.nlm.nih.gov/object/PRJNA879048 (2022).

Acknowledgements

We thank Z. He (Institute of Plant Physiology & Ecology, Chinese Academy of Sciences) for providing a *pigmR* gene coding plasmid. We also thank Y. Liu, Q. Lin, S. Jin and Z. He for helpful advice.

This work was supported by grants from the National Key Research and Development Program (2022YFF1002802 to C.G.), the Strategic Priority Research Program of the Chinese Academy of Sciences (Precision Seed Design and Breeding, XDA24020102, to C.G. and XDA24020310 to Y.W.), the Ministry of Agriculture and Rural Affairs of China, the National Natural Science Foundation of China (32122051 to Y.W.) and the Schmidt Science Fellows to K.T.Z.

Author contributions

K.T.Z. and C.G. designed the experiments and supervised the project. C.S., Y. Lei and H.L performed experiments. B.L. performed rice transformation experiments. Q.G., Yunjia Li, Z.W. and Yan Li performed the GSH prediction and whole-genome sequencing analysis. W.C. and C.Y. performed blast resistance experiments. Y.W. and J.L. consulted and designed experiments. C.S., Y. Lei, B.L., K.T.Z. and C.G. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

The authors have submitted a patent application based on the results reported in this article. K.T.Z. is a founder and employee of Qi Biodesign. Q.G., Z.W. and Y.L. are employees of Qi Biodesign.

Additional information

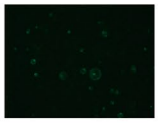
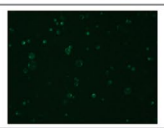
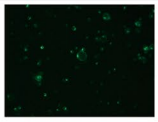
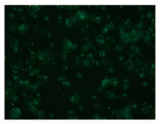
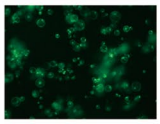
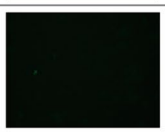
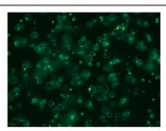
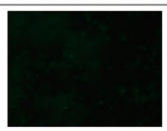
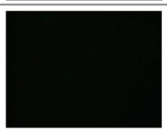
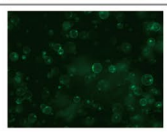
Extended data is available for this paper at https://doi.org/10.1038/s41587-023-01769-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-023-01769-w.

Correspondence and requests for materials should be addressed to Kevin Tianmeng Zhao or Caixia Gao.

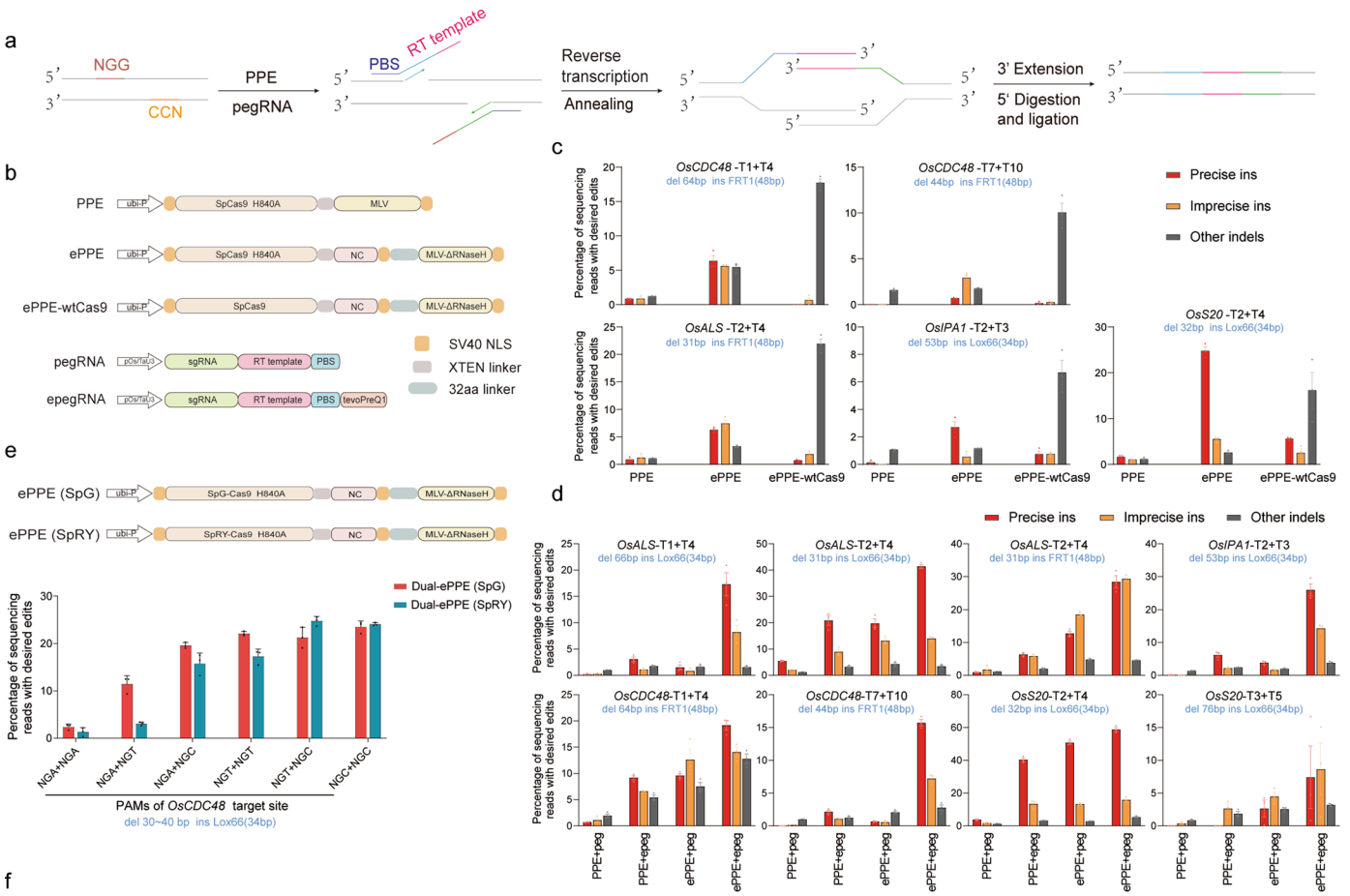
Peer review information *Nature Biotechnology* thanks Wendy Harwood and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41587-023-01769-w

Recombinase	Recombinase site	Cell fluorescence -/Recombinase	Cell fluorescence +/Recombinase
Dre	TAACTTTAAATAATGCCAATTATTTAAAGTTA TAACTTTAAATAATGTCCATTATTTAAAGTTA		
pSR1	TTGATGAAAGAATAACGTATTCTTTCATCAA		
KD	AAACGATATCAGACATTTGTCTGATAATGCTTCATTATCAGACAAATGTCTGATATCGTTT		
B2	GAGTTTCATTAAGGAATAACTAATTCCCTAATGAAACTC		
Cre	ATAACTTCGTATAGCATACATTATACGAAGTTAT		
Cre	ATAACTTCGTATAGCATACATTATACGAACGGTA TACCGTTCGTATAGCATACATTATACGAAGTTAT		
FLP	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC		
phiC31	CGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCAC GTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAG		
Bxb1	GGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGG GGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACC		

Extended Data Fig. 1 | Evaluating different site-specific recombinases using a fluorescent reporter directly in rice protoplasts. Eight recombinases are evaluated with each corresponding recombinase site sequence listed. Microscopy images are of rice protoplasts with or without the corresponding recombinase transformed.

https://doi.org/10.1038/s41587-023-01769-w

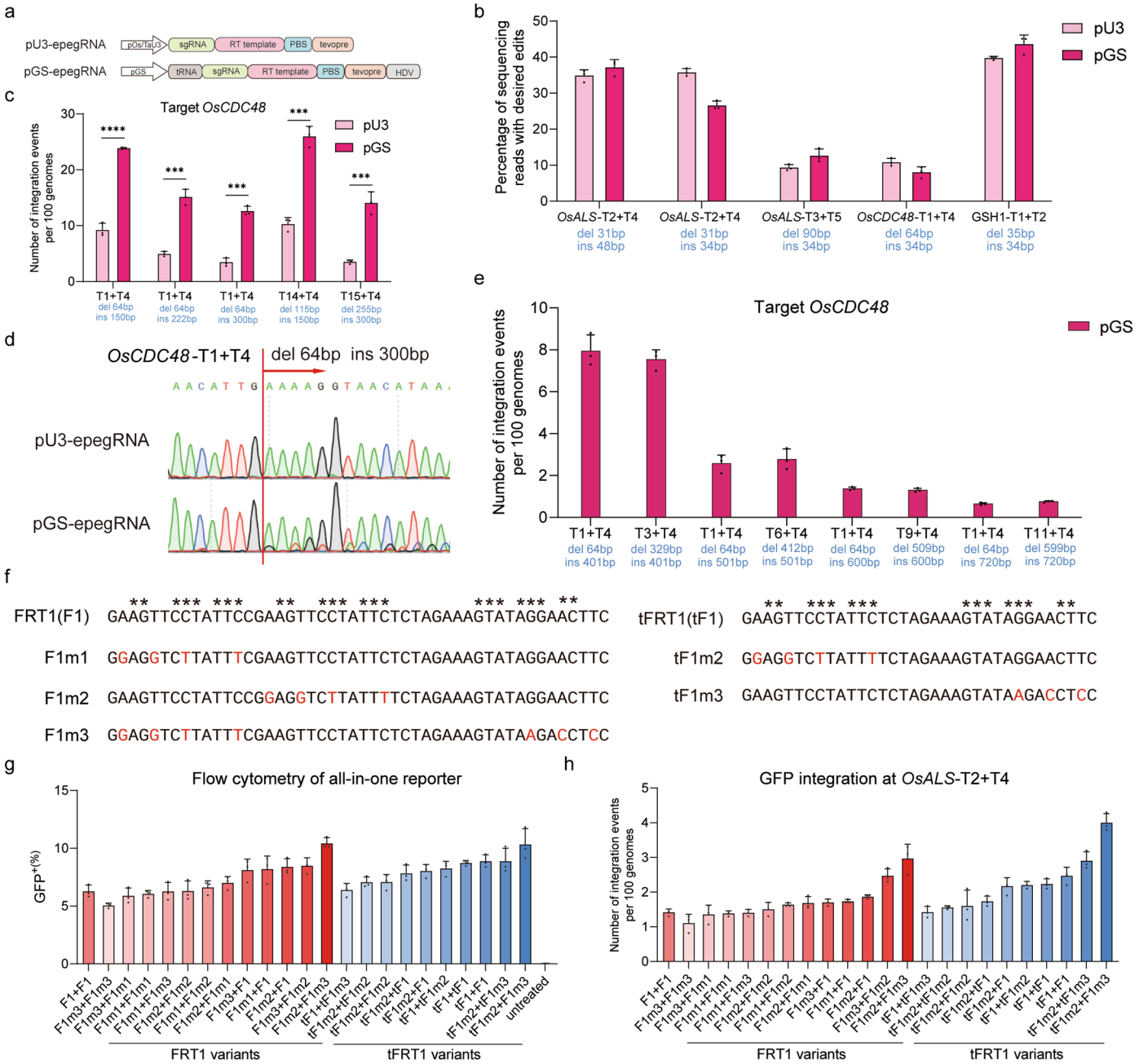


f

Target site	DNA fragment inserted	Number of transgenetic rice plants	Number of mutants/mutagenesis (%) of precise insertion
OsCDC48-T1+T4	FRT1	95	13 (13.7)
*OsCDC*48-T2+T4	FRT1	95	24 (25.3)
*OsCDC*48-T7+T9	Lox66	95	39 (41.0)
GSH1-T1+T2	Lox66	95	44 (46.3)
OsS20-T2+T4	Lox66	95	44 (46.3)

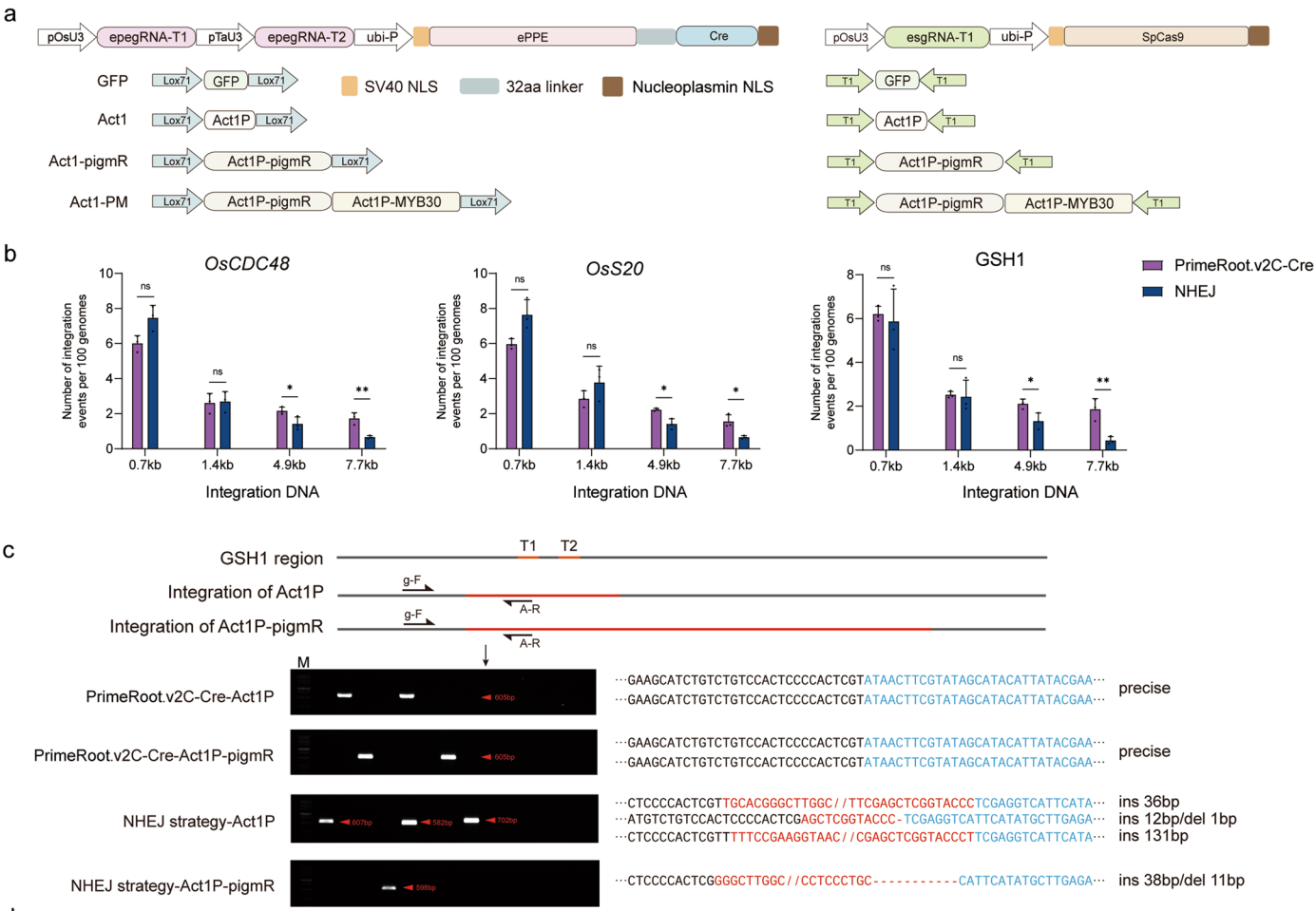
Extended Data Fig. 2 | Development and optimization of dual-ePPE system in rice. a, Schematic overview of dual-PPE system-mediated targeted DNA insertions. NGG and CCN represent the PAMs of two pegRNAs targeting opposing DNA strands; the blue and green lines represent the corresponding PBS/RT template on each DNA strand in each pegRNA and the red line represents the complementary sequence between the two pegRNAs; the blue and green arrows show the directions of reverse transcription. PBS: primer binding site; RT template: reverse transcription template. **b**, Overview of PPE, ePPE, ePPE-wtCas9, pegRNA, and epegRNA construct architectures. **c**, Recombinase site insertion (ins) efficiencies mediated by PPE, ePPE, and ePPE-wtCas9 with pegRNA across five endogenous genomic sites in rice protoplasts as measured using high-throughput sequencing; precise ins represent precise insertions; imprecise ins represent insertions comprised of more than half of the insertion sequence inserted but not the complete sequence; other indels represent all other edits; Values and error bars represent the mean and standard error of mean for three independent biological replicates. **d**, Recombinase site insertion efficiencies mediated by PPE + peg, PPE + epeg, ePPE+peg, and ePPE+epeg across eight endogenous genomic sites in rice protoplasts as measured using high-throughput sequencing; Values and error bars represent the mean and standard error of mean for three independent biological replicates. **e**, Overview of the ePPE (SpG) and the ePPE (SpRY) construct architectures and their corresponding insertion frequencies with six pairs of epegRNAs at the *OsCDC48* site in rice protoplasts as measured using high-throughput sequencing; Values and error bars represent the mean and standard error of mean for three independent biological replicates. **f**, Statistical overview of dual-ePPE-mediated insertions of recombinase sites in rice plants.

https://doi.org/10.1038/s41587-023-01769-w



Extended Data Fig. 3 | epegRNA or RS sequence optimizations to improve editing efficiency. a, Overview of pU3-epegRNA and pGS-epegRNA construct architectures. **b**, Dual-ePPE editing efficiencies mediated by a pU3 or pGS promoter driving epegRNA expression across five endogenous genomic sites in rice protoplasts as measured using high-throughput sequencing; Values and error bars represent the mean and standard error of mean for three independent biological replicates. **c**, Dual-ePPE editing efficiencies of varying insertion or deletion sizes mediated by a pU3 or pGS promoter driving epegRNA expression at the *OsCDC48* genomic site in rice protoplasts as measured using high-throughput sequencing; Values and error bars represent the mean and standard error of mean for three independent biological replicates; P values were obtained using the two-tailed Student's t-test: ***P < 0.001, ****P < 0.0001. **d**, Sanger sequencing traces of dual-ePPE editing at *OsCDC48* mediated by a pU3 or pGS promoter to drive epegRNA expression; The red line and arrow represent the point of insertion and insertion direction, respectively. **e**, Dual-ePPE editing efficiencies to generate larger DNA donor insertions mediated by the pGS promoter driving epegRNA expression at the *OsCDC48* genomic site in rice protoplasts as measured using high-throughput sequencing; Values and error bars represent the mean and standard error of mean for three independent biological replicates. **f**, FRT recombinase site truncation and engineered variants. tFRT1 (tF1) represents a truncated form of FRT1 (F1), * identifies key residues recognized by FLP; the red bases represent mutated residues in each variant. **g**, Percent GFP positive plant protoplast cells reflective of overall insertion efficiencies as evaluated using the all-in-one reporter and measured using flow cytometry; Each bar represents a unique pair of recombinase sites evaluated using the PrimeRoot; Values and error bars represent the mean and standard error of mean for three independent biological replicates. **h**, GFP insertion efficiencies at *OsALS* in rice protoplasts as measured using ddPCR; Each bar represents a unique pair of recombinase sites evaluated using the PrimeRoot; Values and error bars represent the mean and standard error of mean for three independent biological replicates.

https://doi.org/10.1038/s41587-023-01769-w

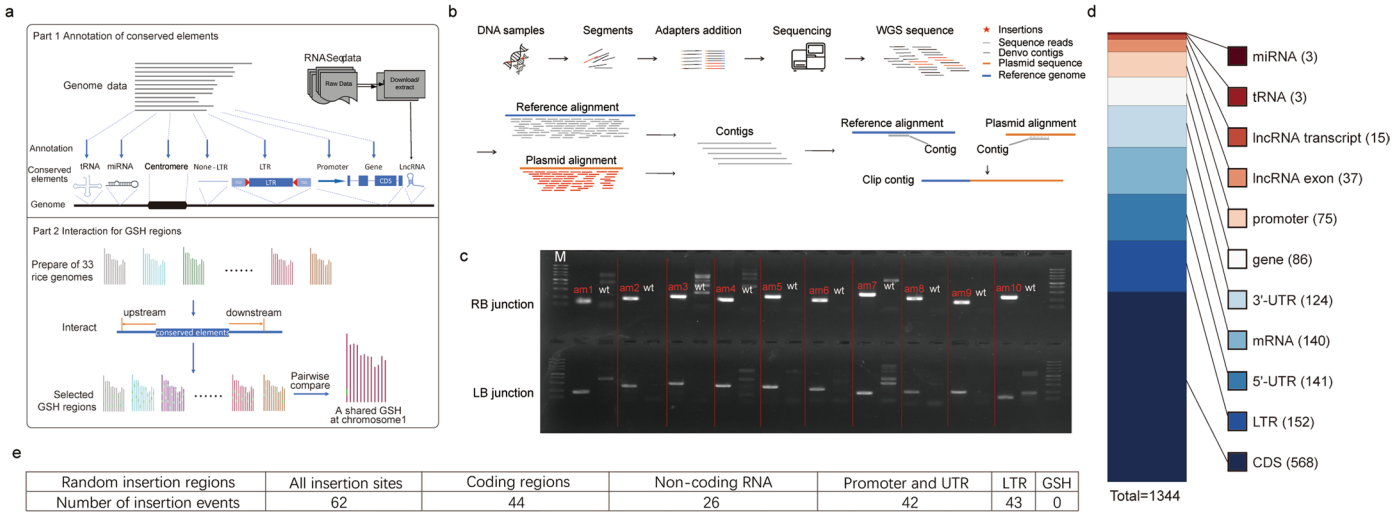


d

Method	Donor inserted	Clones of callis	All insertion events	Precise insertion events
PrimeRoot.v2C-Cre	Act1P (1.4kb)	95	2	2
NHEJ	Act1P (1.4kb)	95	3	0
PrimeRoot.v2C-Cre	Act1P-pigmR (4.9kb)	95	2	2
NHEJ	Act1P-pigmR (4.9kb)	95	1	0

Extended Data Fig. 4 | Comparison of targeted insertions mediated by PrimeRoot and NHEJ. a, Overview of PrimeRoot construct architectures and NHEJ donor constructs for inserting GFP (720 bp), Act1P (1.4 kb), Act1P-*pigmR* (4.9 kb) and Act1P-PM (7.7kb). **b**, Comparison of PrimeRoot and NHEJ editing efficiencies for targeted insertions of the three donors at the *OsCDC48*, *OsS20*, and GSH1 sites in rice protoplasts as measured by ddPCR; Values and error bars represent the mean and standard error of mean for three independent biological replicates; P values were obtained using the two-tailed Student's t-test: $^{ns}P > 0.05$, $^{*}P < 0.05$, $^{**}P < 0.01$. **c**, Gel electrophoresis of PCR outcomes of the insertion junction between the donor cassette and endogenous genome. sgRNA1 and sgRNA2 are the target sites used in PrimeRoot; sgRNA1 is the Cas9 target site used for NHEJ -F, A-R are the primers for PCR and Sanger sequencing; Blue bases represent sequences from the donor and red bases represent sporadic DNA insertions or deletions; M represents the marker; the numbers behind the red arrow represent the size of PCR outcomes. **d**, Overview of insertion statistics mediated by PrimeRoot.v2C-Cre or NHEJ.

https://doi.org/10.1038/s41587-023-01769-w



e

Random insertion regions	All insertion sites	Coding regions	Non-coding RNA	Promoter and UTR	LTR	GSH
Number of insertion events	62	44	26	42	43	0

Extended Data Fig. 5 | Predicting GSH regions and analyzing Agrobacterium insertion events. a, Schematic overview on annotating rice genomes and predicting conserved GSH regions across 33 rice varieties. **b**, Schematic overview on using whole genome sequencing to identify insertion events. **c**, PCR validation of Agrobacteria insertion events as identified by whole-genome sequencing. RB and LB junctions are amplified using primers designed based on surrounding contig sequences. am1 to am10 represent 10 different Agrobacteria-derived mutants. wt represent wild-type rice plants amplified using the junction-spanning primer pairs. **d**, Overview of Agrobacteria-mediated insertion events and their genome insertion locations.

nature research

Corresponding author(s): Caixia Gao, Kevin Tianmeng Zhao

Last updated by author(s): Feb 26, 2022

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 NovaSeq platform was used to collect the amplicon deep sequencing data. BD FACSAriaIII was used to do flow cytometry.
Data analysis	Amplicon sequencing data of prime-editing processivity was analyzed using the published code as previously described in reference 13. The custom Python script to analyze types of mutational reads and amino acid substitutions will be made available upon request. Graphpad prism 7 was used to analyze the data. FACSDiva Version 6.1.3 software and FlowJo Version 7.6 software was used for flow cytometry result analysis.

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

The authors declare that all data supporting the findings of this study are available in the article and its Supplementary Information files or are available from the corresponding author on request. Datasets of high-throughput sequencing experiments can be deposited with the National Center for Biotechnology Information (NCBI) under accession code PRJNA879048 before publication.

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	The experiments of protoplasts were performed with three biological repeats. About 500,000 protoplasts were used for each transfection. The number of protoplasts in each transfection was measured by thrombocytometry. The experiment in rice regenerated plants was performed once, all the regenerated seedlings were sampled, the number of mutants were confirmed by Sanger sequencing.
Data exclusions	No data exclusion.
Replication	All attempts for replication were successful. For the experiments in rice protolasts, a minimum of three independent experiments were included.
Randomization	Rice protoplasts were isolated and randomly separated to each transformation.
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

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Rice protoplasts were isolated from the stem of rice seedlings, transfected as described in the Mehtods and incubated in 1 ml WI solution for 2 days.
Instrument	BD FACSAriaIII
Software	FACSDiva Version 6.1.3 software and FlowJo Version 7.6 software was used for analysis.
Cell population abundance	The abundance of cells for flow cytometry analysis was 10,000 for each sample.

Gating strategy

Negative control (untreated) and fluorophore-positive cells were used to establish gates for each cell type. Gates were drawn to collect cells expressing either fluorophore. See the provided examples for gates used.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.