Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors

Chao Li[1,2,6], Rui Zhang[1,6], Xiangbing Meng[3,6], Sha Chen[1,2], Yuan Zong[1,2], Chunju Lu[1,2], Jin-Long Qiu [ID][2,4], Yu-Hang Chen[2,5], Jiayang Li [ID][2,3]* and Caixia Gao [ID][1,2]*

Targeted saturation mutagenesis of crop genes could be applied to produce genetic variants with improved agronomic performance. However, tools for directed evolution of plant genes, such as error-prone PCR or DNA shuffling, are limited[1]. We engineered five saturated targeted endogenous mutagenesis editors (STEMEs) that can generate de novo mutations and facilitate directed evolution of plant genes. In rice protoplasts, STEME-1 edited cytosine and adenine at the same target site with C > T efficiency up to 61.61% and simultaneous C > T and A > G efficiency up to 15.10%. STEME-NG, which incorporates the nickase Cas9-NG protospacer-adjacent motif variant, was used with 20 individual single guide RNAs in rice protoplasts to produce near-saturated mutagenesis (73.21%) for a 56-amino-acid portion of the rice acetyl-coenzyme A carboxylase (OsACC). We also applied STEME-1 and STEME-NG for directed evolution of the *OsACC* gene in rice and obtained herbicide resistance mutations. This set of two STEMEs will accelerate trait development and should work in any plants amenable to CRISPR-based editing.

Heredity and variation are the basis of organismic evolution. Random mutagenesis by physical (for example, ultraviolet) or chemical (for example, ethylmethane sulfonate) methods has long been applied to improve traits in plants, but is labor-intensive and time-consuming[2,3]. Although directed evolution is an effective strategy for obtaining desired genetic or protein variants[1,4], most approaches to library diversification involve error-prone PCR, DNA synthesis, recombination or mutator strains[1,4]. In higher organisms, especially in plants, a target gene is usually transferred into a bacterial or yeast cell to generate the required diversity for selection[1,5–8], but once a target gene is not in situ, the functional consequences of changes may not be the same as in the native context[1].

CRISPR–Cas has been used for targeted mutations in various organisms[9,10] including rice, in which CRISPR–Cas9 was applied to produce several in-frame mutations resistant to splicing inhibitors[11]. However, most mutations generated by CRISPR–Cas9 are insertions or deletions (indels). These types of mutations typically produce nonfunctional, out-of-frame mutations or premature terminations, whereas directed evolution can be further empowered by making nucleotide changes to generate gain-of-function mutations[12].

CRISPR-based cytosine base editors (CBEs) and adenine base editors (ABEs) make C:G > T:A and A:T > G:C substitutions,

respectively, in single guide RNA-specified target sites[13–17]. CBEs and ABEs are fusions of Cas9 variants (nCas9 and dCas9) with corresponding deaminase domains, and their functions have led to suggestions that they could be applied for saturation base editing (C/G or A/T bases) of protein-coding sequences[13,17]. Although directed evolution in mammalian cells[15,16] and mice[18] using CBEs has been reported, increasing the mutation potential beyond CBEs would enable greater diversity in directed evolution strategies.

We hypothesized that it might be possible to engineer a single protein that used a single sgRNA to achieve C:G > T:A and A:T > G:C substitutions. To engineer dual base editors, we fused cytidine deaminase with adenosine deaminase to obtain editors that we named STEMEs. In addition to the fusion deaminase, STEMEs contain nCas9 (D10A) and uracil DNA glycosylase inhibitor (UGI)[19,20] (Fig. 1a).

We previously developed a high-efficiency CBE for plants, A3A-PBE, which has an enlarged base-editing window[21,22], and an ABE, PABE-7, which contains an evolved tRNA adenosine deaminase (ecTadA-ecTadA7.10)[23]. To enable both C:G > T:A and A:T > G:C substitutions in the same target sequence with a single editor, we fused APOBEC3A-ecTadA-ecTadA7.10 or ecTadA-ecTadA7.10-APOBEC3A, to the N terminus of nCas9 (D10A), together with either UGI or two copies of free UGI at the C terminus of nCas9 (D10A), to produce STEME-1, STEME-2, STEME-3 and STEME-4, respectively (Fig. 1b and Supplementary Sequences 1). All four STEMEs were codon optimized for crop plants and transcribed from the maize *Ubi-1* promoter.

To examine their base-editing activities on endogenous genes, six sgRNAs targeting different rice genes were designed (Supplementary Table 1) and cloned into pOsU3-esgRNA (ref. [23]). Each sgRNA was cotransfected into rice protoplasts along with each of the four STEMEs. A3A-PBE, PABE-7 and wild-type Cas9 were used as controls. Amplicon deep sequencing showed that all four STEMEs produced C > T and A > G conversions efficiently (Fig. 1c,d; BioProject accession PRJNA590652). The C > T base-editing windows were equivalent to that of A3A-PBE (ref. [21]) and editing efficiencies ranged from 0.10% to 61.61%, with STEME-1 the most efficient (Fig. 1c and Supplementary Fig. 1). Within the primary editing window of A3A-PBE (C_1–C_{17}; counting the end distal to the protospacer-adjacent motif (PAM) as position 1), STEME-1 had a C > T editing efficiency averaging 25.14% in *OsAAT*, *OsACC*,

OsCDC48 and *OsDEP1* that was 1.5-fold higher than A3A-PBE (average 17.25%) (Fig. 1c and Supplementary Fig. 1).

STEME-1 also had the highest A>G base-editing efficiency (0.69–15.50%) among the four STEMEs and the expected A>G base-editing window of A_4 to A_8 (Fig. 1d and Supplementary Fig. 1). Although this was lower than PABE-7 (1.74–21.54%), the STEME-1 A>G editing efficiency was still sufficient to provide the desired mutation diversity for an improved directed evolution strategy (Fig. 1e and Supplementary Fig. 2a). Sequencing reads indicated that the products of STEME-1 were mainly C>T conversions (16.50–50.78%) and simultaneous C>T and A>G conversions (0.49–15.10%) in rice protoplasts (Fig. 1f), while A3A-PBE or PABE-7 produced only C>T or only A>G conversions, respectively (Fig. 1g). No undesired editing at any of the sgRNA targets was detected (<0.05%) (Supplementary Fig. 1). Indel frequencies with STEMEs (0.04–0.63%) were also equivalent to that in untreated control protoplasts (0.04–0.51%), much lower than with Cas9 (6.30–15.61%) (Supplementary Fig. 2b).

We next codelivered A3A-PBE and PABE-7 with three sgRNAs to evaluate whether individually delivered base editors could perform as well as our dual base editors. Amplicon deep sequencing showed that all A>G substitutions were accompanied by simultaneous C>T editing, but that the efficiency was about fourfold lower than that of STEME-1 (Fig. 1g). This is most likely because only one base-editor protein can occupy a single target site in the genome and will therefore compete with the other base editor. A3A-PBE is more active than PABE-7, so it seems reasonable to conclude that A3A-PBE edits sequences before PABE-7 can access the target site. By the time PABE-7 is present, the target will have been edited and will be a mismatch. Taken together, these results indicate that STEME-1 can increase the mutation diversity at a target site using only a single sgRNA.

We sought to expand the targeting scope of STEME-1 to increase its utility. We replaced the nCas9 (D10A) in STEME-1 with codon-optimized nCas9-NG (D10A)[24] to produce STEME-NG (Fig. 2a and Supplementary Sequence 1). We also generated A3A-PBE-NG, PABE7-NG and pCas9-NG constructs by replacing the corresponding portions of A3A-PBE (ref. [21]), PABE-7 (ref. [23]) and pCas9 (ref. [25]) with codon-optimized nCas9-NG (D10A) or Cas9-NG (Supplementary Fig. 3a and Supplementary Sequence 1). We designed 16 20-nucleotide spacers with NG PAMs from four different rice loci (Supplementary Fig. 3b,c). STEME-NG along with each of these 16 sgRNAs was then cotransfected into rice protoplasts. We found that STEME-NG had a broad capacity for editing C>T and A>G in NG PAM sequences, but preferred NGD (D=A, T or G) PAMs (Fig. 2b and Supplementary Fig. 4). Similar to Cas9-NG (ref. [24]), STEME-NG exhibited compromised activity (average C>T 7.92%, A>G 1.84%) at canonical NGG PAM sequences compared with STEME-1 (average C>T 17.89%, A>G 3.80%) (Fig. 2b and Supplementary Fig. 4). STEME-NG edited cytosines in a window of C_1 to C_{17} and adenines in a window of A_4 to A_8, which was the same as observed for the individual A3A-PBE-NG and PABE7-NG,

respectively (Supplementary Fig. 5a,b). In addition, STEME-NG, A3A-PBE-NG and PABE7-NG generated indels at much lower frequencies (<0.10%) than pCas9-NG (0.16–13.24%) in rice protoplasts (Supplementary Fig. 6). Taken together, these data show that the editing activities of STEME-NG, A3A-PBE-NG and PABE7-NG at NG PAMs depend mainly on the nature of the Cas9-NG. All prefer NGD PAMs to NGC PAMs, which is consistent with a previous report[24]. As the Cas9-NG recognizes a relaxed PAM, its kinetics of nuclease activity are slower than that of wild-type Cas9 toward canonical NGG PAM, which could be improved by further molecular engineering[24]. Although the editing efficiency of STEME-NG was on average 2.2-fold lower than that of STEME-1 on NGG PAM, the above data suggest that STEME-NG is able to expand the scope of C>T and A>G base editing and facilitate the application of directed evolution in plants.

To test whether STEME-NG can be used for saturated de novo mutagenesis in rice protoplasts, we targeted *OsACC*. ACC is an enzyme in fatty acid biosynthesis and its carboxyltransferase (CT) domain is the target of herbicides[26] (Fig. 2c and Supplementary Sequences 2). Amino acid substitutions in the CT domain can confer herbicide resistance in grass[26]. We designed 20 sgRNAs, including 11 sgRNAs with forward direction NGD-3′ PAMs and 9 sgRNAs with reverse complement 5′-HCN (H=A, T or C) PAMs spanning a 168-base pair DNA sequence that encodes 56 amino acids of the CT domain (Fig. 2c and Supplementary Table 2). These 20 sgRNAs covered 90.32% of the cytosines, 40.43% of the adenines, 77.78% of the guanines and 38.89% of the thymines in the editing windows, corresponding in all to 61.31% of the bases of the coding strand (Supplementary Table 3). We cotransfected individual sgRNAs together with STEME-NG into rice protoplasts. A3A-PBE-NG and pCas9-NG served as controls. Amplicon deep sequencing showed that STEME-NG converted 96.43% of the Cs to Ts, 63.16% of the As to Gs, 92.86% of the Gs to As and 42.86% of the Ts to Cs in the covered bases on the coding strand; average base-editing efficiencies were 11.50%, 0.35%, 13.33% and 0.45%, respectively (Fig. 2d and Supplementary Table 3). Meanwhile, A3A-PBE-NG edited 89.29% of Cs to Ts and 92.86% of Gs to As on the coding strand, and no A>G or T>C substitutions were found (Supplementary Fig. 7a and Supplementary Table 3). No base conversions were detected in the untreated control (Supplementary Fig. 7b). The diversity of mutations induced by these 20 sgRNAs using STEME-NG was about twofold greater than that observed using A3A-PBE-NG (Fig. 2e and Supplementary Dataset 1). Simultaneous C:G>T:A and A:T>G:C events contributed to 18.4% of the observed STEME-NG diversity (Supplementary Dataset 1), efficiency up to 2.71% (Supplementary Fig. 8a). Consistent with the above experiments, STEME-NG showed very few indels (<0.02%) with this different target set, similar to A3A-PBE-NG (<0.01%) and untreated control protoplasts (<0.02%), but much less than Cas9-NG (0.32–39.72%) (Supplementary Fig. 8b).

We also analyzed the amino acid substitutions generated by STEME-NG in the targeted 56 amino acids. We found that 41 of the

Fig. 1 | Base editing of STEMEs via fused cytidine and adenosine deaminases. a, The STEME-mediated C:G>T:A and A:T>G:C base-editing strategy. nCas9 (D10A) stimulates cellular mismatch repair, UGI blocks the activity of uracil DNA glycosylase (UDG), and the damaged strand is repaired using the deaminated strand as template. **b,** Architectures of STEME-1, STEME-2, STEME-3 and STEME-4. ecTadA7.10, evolved *Escherichia coli* TadA; aa, amino acid; XTEN, a 16-aa linker; NLS, nuclear localization signal; CaMV, cauliflower mosaic virus; term, terminator. **c,** Comparison of the C>T editing frequencies of A3A-PBE and the four STEME constructs ($n=3$). Values and error bars indicate mean±s.e.m. of three independent experiments. **d,** Comparison of the A>G editing frequencies of PABE-7 and the four STEME constructs ($n=3$). An untreated protoplast sample served as control. Values and error bars indicate mean±s.e.m. of three independent experiments. **e,** The allelic outcome of the *OsDEP1* site edited by STEME-1 ($n=1$) in rice protoplasts. **f,** The product distribution among edited DNA sequencing reads in rice protoplasts is shown for STEME-1. The resulting efficiencies were log-scaled by base 10. Values and error bars indicate mean±s.e.m. of three independent experiments. **g,** Comparison of the product distribution among edited DNA sequencing reads of A3A-PBE, PABE-7, codelivery A3A-PBE and PABE-7, and STEME-1 in rice protoplasts. The individual data points are shown as light blue (only C>T), light green (only A>G) and light red (simultaneous C>T and A>G) dots. Values and error bars indicate mean±s.e.m. of three independent experiments.

amino acids were substituted (including silent, missense and non-sense mutations) (Fig. 2f and Supplementary Dataset 1). Of these, 24, 12 and 5 amino acids had 1, 2 or 3 kinds of amino acid substitutions, respectively (Fig. 2f and Supplementary Dataset 1). We achieved almost-saturated mutagenesis (73.21%) for these 56 amino acids

using STEME-NG and only 20 sgRNAs. Similarly, A3A-PBE-NG mutated 33 amino acids, of which 26, 6 and 1 contained 1, 2 or 3 kinds of amino acid substitutions, respectively (Supplementary Fig. 9 and Supplementary Dataset 1). These results collectively show that STEME-NG is more than just a C > T editor; it is indeed a dual
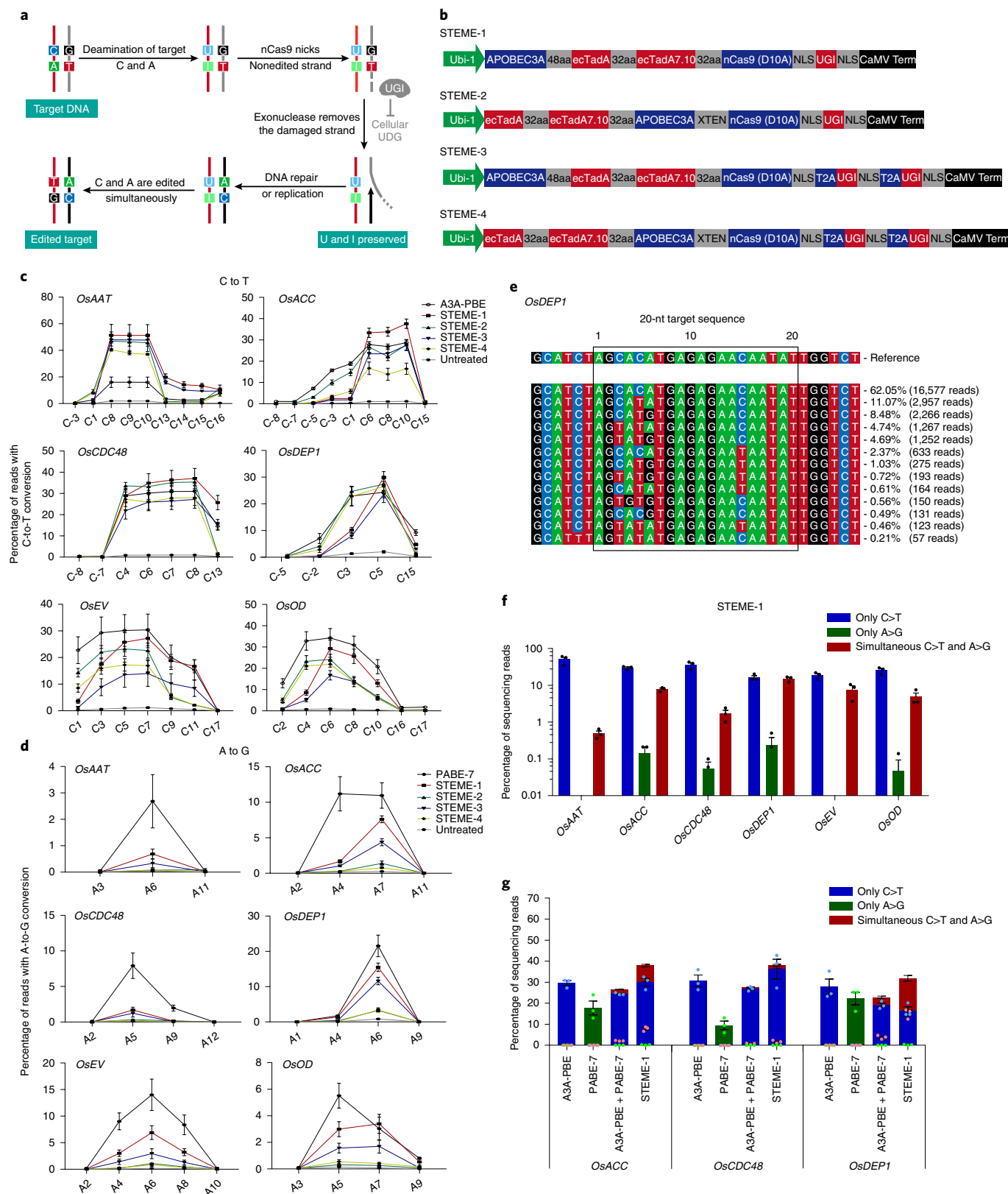
base-editing system that provides an improved saturated mutagenesis outcome compared with A3A-PBE-NG. Thus, it promises to be a powerful tool for directed evolution of endogenous genes by saturated de novo mutagenesis in situ.

In a proof-of-concept experiment, we used STEMEs for directed evolution of *ACC* in rice plants (Fig. 3a). We chose a 1,200-nucleotide region encoding 400 amino acids of the CT domain as the mutagenesis target[26]. A total of 200 sgRNAs were designed, including 118 forward direction NGD-3′ and 82 reverse complement 5′-HCN PAM sgRNAs (Supplementary Fig. 10a and Supplementary Table 4). STEME-1 was chosen for 102 sgRNAs with NGG-3′ or 5′-CCN PAMs, while STEME-NG was used for the remaining sgRNAs, which had NGW-3′ or 5′-WCN (W = A or T) PAMs (Supplementary Fig. 10a). These sgRNAs covered 94.61% of the Cs, 48.26% of the As, 83.39% of the Gs and 37.46% of the Ts in the editing windows, representing in all 63.95% of the bases on the coding strand (Supplementary Table 5). We inserted these sgRNAs separately into the binary vector pH-STEME-1-esgRNA or pH-STEME-NG-esgRNA (Supplementary Fig. 10b). To perform plant transformation and genotyping efficiently, the 200 sgRNAs were divided into 27 groups (Groups 1–27). In each group, equal amounts of 4–11 sgRNA plasmids covering 80–142 nucleotides in *OsACC* were pooled (Fig. 3a and Supplementary Table 4).

To evaluate the transformation coverage, we amplified the guide RNA sequences from genomic DNA extracted from each group of regenerated seedlings for amplicon deep sequencing (Supplementary Fig. 10b). We found that 72.73–100% of the sgRNAs had been transformed into the plants in each group, and in total 92.50% (185/200) of the sgRNAs had been successfully introduced (Fig. 3b). We then characterized the mutational coverage by deep sequencing and observed 377 nucleotide substitutions among the 768 nucleotides covered, involving 168 Cs (73.68%), 23 As (15.03%), 164 Gs (61.65%) and 22 Ts (18.18%) (Supplementary Table 5). The average editing efficiency in each group was 13.18% (Supplementary Fig. 11a and Supplementary Dataset 2). Moreover, unlike the uniform substitutions seen in protoplasts, the STEMEs induced C > G/A and G > C/T conversions and in-frame indels in addition to the canonical C:G > T:A and A:T > G:C base conversions (Fig. 3c, Supplementary Fig. 11b and Supplementary Dataset 2). The product distributions among the edited bases were 81.86% C > T, 13.73% C > G, 4.41% C > A, 76.63% G > A, 19.02% G > C, 4.35% G > T, 100% A > G and 100% T > C (Fig. 3c); this somewhat altered distribution may be due to differences in base excision repair mechanisms in protoplasts and plants. Thus, STEMEs can also be used to generate C:G > G:C or C:G > A:T substitutions in addition to the canonical edits, which should further improve the diversity possible for protein-directed evolution in plants.

We analyzed the details of the mutational reads created in rice plants. Of the 495 types of mutational reads induced by the 185 sgRNAs, 76.36%, 19.80%, 3.64% and 0.20% involved 1, 2, 3 and 4 amino acid substitutions, respectively (Supplementary Dataset 2).

In addition, 2.83% of the mutated sequences involved adenine editing and 3.84% involved simultaneous adenine and cytosine editing (Supplementary Dataset 2). Of the 400 amino acids targeted, 209 (52.25%) were altered, including silent, missense and nonsense mutations (Fig. 3d). Of these, 116, 66, 19, 7 and 1 had 1, 2, 3, 4 and 6 kinds of amino acid substitutions, respectively (Fig. 3d). Taken together, these data demonstrate that STEMEs are able to generate large numbers of mutations to serve as the basis for directed evolution of endogenous genes in the rice genome.

To identify the desired mutants, we sprayed a commonly used ACC inhibitor, haloxyfop, to select for herbicide resistance seedlings in Groups 1–27. Three weeks later, a few normal-looking seedlings appeared and were clearly herbicide resistant (Fig. 3e). Sanger sequencing showed that ten in Group 6 carried mutations: seven were P1927F homozygotes and two were heterozygotes, and the remaining seedling was a Q1926*/P1927F and P1927F biallelic mutant; two seedlings in Group 20 carried mutations: one was a W2125C homozygote, and the other a A2123T/W2125C heterozygote (Fig. 3e, Supplementary Fig. 12a,b and Supplementary Table 6). We also observed other seedlings with a slightly weaker haloxyfop resistance than that observed above (Fig. 3f), suggesting that these may represent different alleles. Sanger sequencing showed that two of the seedlings, in Group 2, were S1866F heterozygotes, whereas three seedlings in Group 3 were A1884P heterozygotes (Supplementary Fig. 12c,d and Supplementary Table 6). In all plants containing the S1866F, P1927F or A2123T substitutions, these were the result of C:G > T:A transitions, whereas a C:G > G:C transversion was responsible for all observed W2125C substitutions (Supplementary Fig. 12 and Supplementary Table 6). This was consistent with the amplicon data of STEMEs in rice plants, showing the occurrence of C:G > G:C transversions (Fig. 3c and Supplementary Table 5). In contrast, the A1884P substitutions observed were caused by different activities; two plants contained a single C:G > G:C transversion, whereas the third plant contained both a C:G > G:C transversion and an A:T > G:C transition within the A1884P codon indicative of simultaneous deaminase activities from STEME-NG (Supplementary Fig. 12d and Supplementary Table 6). W2125C is a herbicide resistance mutation, which has been reported in grasses[26], indicating that mutagenesis of ACC by STEMEs is able to generate known mutations. Importantly, these results also confirmed that STEMEs can generate multiple mutations, such as P1927F, S1866F and A1884P, which have not been reported previously[26].

Finally, we tested a strategy for using STEMEs for targeted mutagenesis under concurrent selection pressure. Based on the above results, Group 6 (P1927) and Group 20 (W2125) were selected as representative targets and used to transform rice with a modified protocol in which the herbicide selection pressure was applied during callus induction and regeneration (Supplementary Fig. 13a). Vigorous growth of calli was observed in the target transformations and the mutations conferring herbicide resistance were enriched

Fig. 2 | STEME-NG performs saturated mutagenesis in rice protoplasts. a, Structure of STEME-NG. **b**, STEME-NG edits the targets of NGA, NGT, NGC and NGG PAMs. Cytosines in editing windows of C_1 to C_{17} and adenines in editing windows of A_4 to A_8 are shown. STEME-1 with the NGG PAM served as a positive control, and an untreated protoplast sample served as a negative control. Values and error bars indicate the mean ± s.e.m. of three independent experiments. **c**, An overview of the OsACC protein domains generated by Pfam. Design of sgRNAs with forward direction NGD-3′ (D = A, T or G) PAMs, and reverse complement 5′-HCN (H = A, T or C) PAMs. BC-N, biotin carboxylase (N-terminal domain); CPSase_L_D2, carbamoyl-phosphate synthase L chain (ATP-binding domain); BC-C, biotin carboxylase (C-terminal domain); BA, biotin-requiring enzyme (the attachment domain binds biotin); ACC_central, acetyl-coenzyme A carboxylase (central region). **d**, Saturated mutagenesis of 168 base pairs of the coding strand of the OsACC CT domain in rice protoplasts. Maximum values were chosen when different sgRNAs converted the same cytosine, adenine, thymine and guanine. Values and error bars indicate the mean ± s.e.m. of three independent experiments. **e**, The higher mutation diversity within the 20-nt target site generated by STEME-NG compared with that generated by A3A-PBE-NG using 20 sgRNAs. Data are presented as box plots (center line, median; box limits, 25th and 75th percentiles of the data; lower and upper whiskers, smallest and largest values). Data in each box plot included three independent experiments ($n = 60$). P value was obtained by two-tailed Student's t-test. **P < 0.01. **f**, Heatmap showing conversion saturation by STEME-NG of the 56 amino acids associated with **d** in rice protoplasts. Silent, missense and nonsense mutations were all counted.

under concurrent selection pressure, whereas calli transformed with the control vector died (Fig. 3g, Supplementary Fig. 13a–c and Supplementary Dataset 3). Twenty plants each from the Group 6 and Group 20 transformations were selected for further analysis and all carried P1927F or W2125C mutations, respectively (Supplementary Fig. 13d,e). Three of 20 mutants carrying W2125C also contained
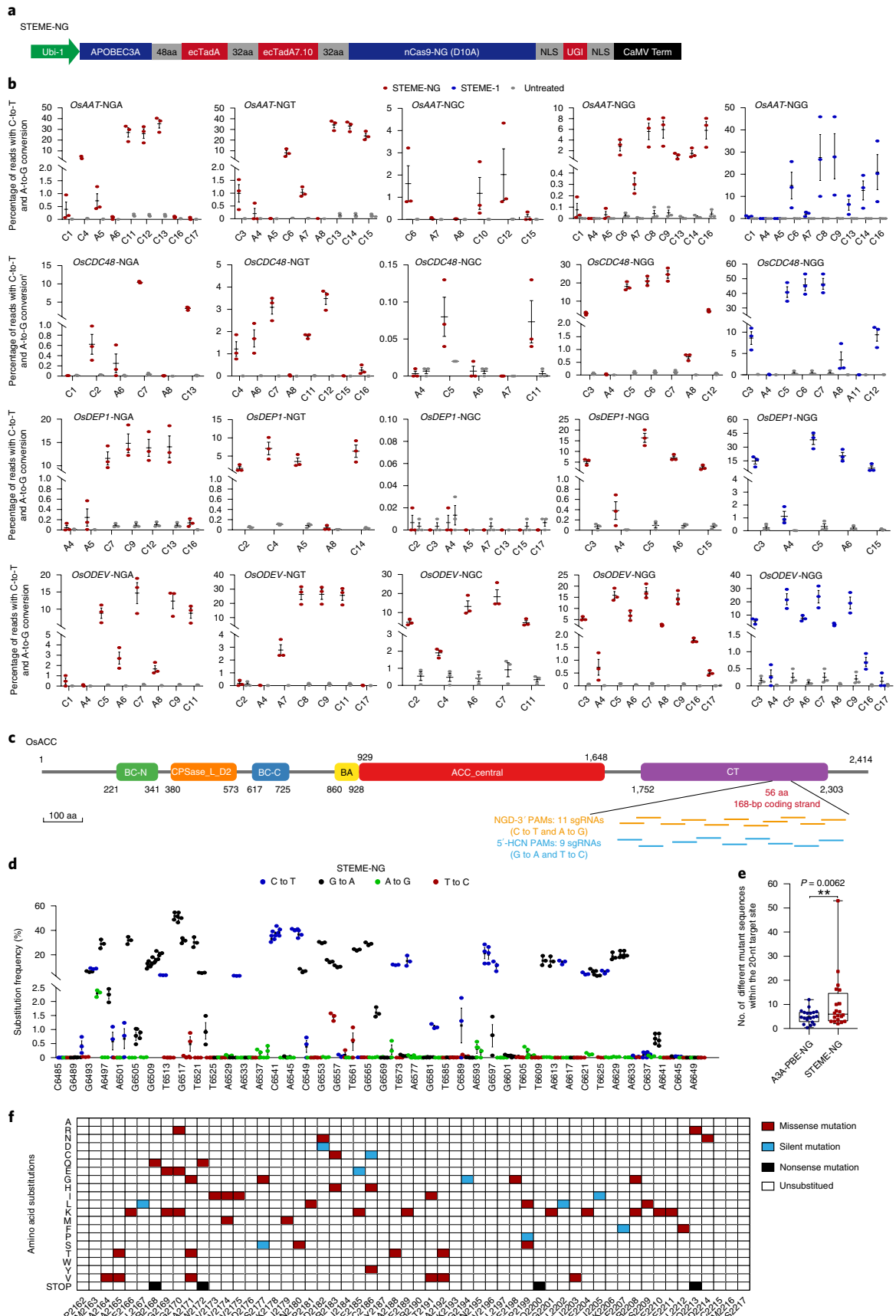
A2123T mutations with nucleotide changes resulting from simultaneous adenosine and cytidine deaminase activity within the A2123 codon (Supplementary Fig. 13e). In addition, we also sequenced the *OsACC* gene of representative resistance seedlings harboring P1927F, W2125C, S1866F or A1884P and found no other mutational changes (Supplementary Fig. 14). Therefore, mutagenesis of *OsACC* by STEMEs can reveal a range of functional herbicide resistance mutations, demonstrating their potential value in carrying
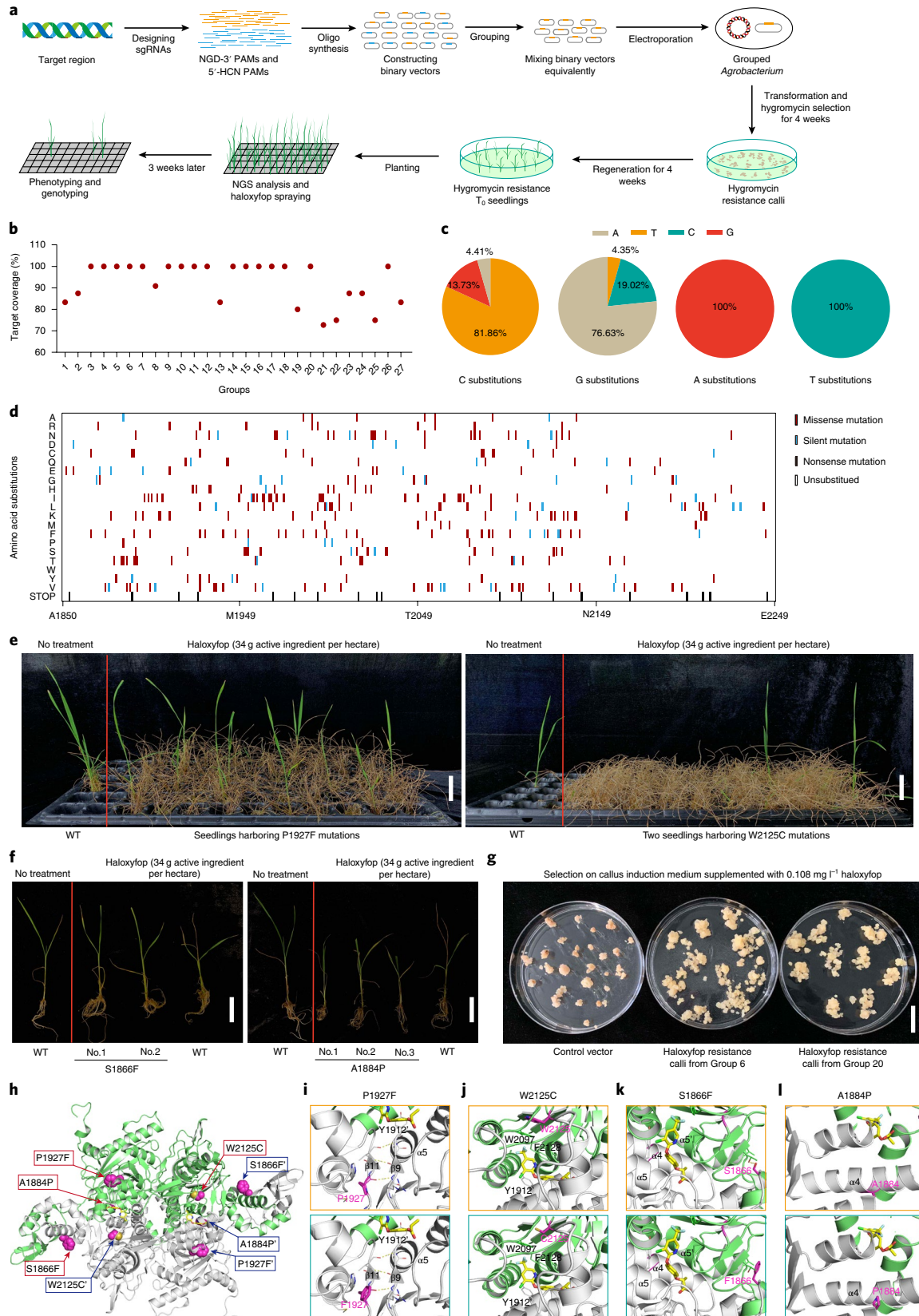
Fig. 3 | Directed evolution by saturated mutagenesis generates herbicide-tolerant mutants. a, Schematic of the procedure for mutating the OsACC CT domain via STEMEs using groups of individual sgRNAs. **b**, Transformation coverage by the binary vectors in each regenerated seedling group. Percentages were based on the presence of protospacers in the deep sequencing reads. **c**, The product distributions among edited bases are shown for STEMEs in rice plants. Reads in which the target C, G, T or A was mutated. **d**, Heatmap showing conversion saturation by STEMEs of 400 amino acids of OsACC CT in regenerated rice plants. Silent, missense and nonsense mutations were all enumerated. **e**, Mutants of STEMEs-induced amino acid substitutions in OsACC confer resistance to herbicide. The T_0 seedlings harboring P1927F (Group 6) or W2125C (Group 20) mutations had a normal morphological phenotype 3 weeks after haloxyfop treatment. One biological experiment was performed. Scale bar, 5 cm. **f**, The mutations of S1866F (Group 2) and A1884P (Group 3) of *OsACC* conferred weak herbicide resistance. Two S1866F heterozygous mutants and three A1884P heterozygous mutants survived 2 weeks after haloxyfop treatment, while the newest grown leaves of the wild type died. One biological experiment was performed. Scale bar, 4 cm. **g**, Herbicide-resistant calli in Group 6 and Group 20. One biological experiment was performed. Scale bar, 2 cm. **h**, The structural model of the OsACC CT domain was based on the structure of the yeast ACC CT domain. The mutations (P1927F, W2125C, S1866F and A1884P) related to haloxyfop resistance are shown as spheres and are colored in purple. **i–l**, Molecular basis of herbicide resistance in the rice *OsACC* mutants. The mutations P1927F (**i**), W2125C (**j**), S1866F (**k**) and A1884P (**l**) related to haloxyfop resistance are shown as thick sticks and are colored in purple. Key residues involved in herbicide binding are shown as thin sticks: Y1912, W2097 and F2128 correspond to residues Y1738, W1924 and F1956 in yeast ACC. The monomers of the CT domain dimer are colored gray and green, respectively. Haloxyfop (yellow) is superimposed on the model structure. Haloxyfop binds to a pocket at the dimer interface, and blocks access by acetyl-coenzyme A. WT, wild type.

out directed protein evolution. Moreover, mutants harboring P1927F or W2125C could potentially be used for weed control in the field.

To assess the potential for off-target effects, we scanned the genomic sequence for all similar target sites that contained up to a 3-nucleotide mismatch and sequenced these sites in the respective mutants. From this analysis, a single off-target mutation was found in only one of the mutants (the biallelic mutant harboring A2123T and A2123T/W2125C), whereas no off-target mutations were found in any of the other mutants (Supplementary Table 7). As STEMEs are composed of two deaminases there remains a potential for unpredictable DNA and RNA off-target mutations to occur over a wide spectrum[27,28], but this problem could be rapidly solved with the development of new deaminase variants with high efficiency and specificity in plants[28,29].

To understand the molecular basis of haloxyfop resistance mutations, we developed a homology model (Fig. 3h and Supplementary Sequence 3) based on the structure of the CT domain of yeast ACC[30]. The mutation P1927F is predicted to alter the interactions of the β-sheet (β11–β9) and affect the conformation of helix α5 (Fig. 3i). The mutation W2125C shortens the side-chain of C2125, abolishing the indole ring π–π stacked interaction with W2097. Both W2097 and the neighboring F2128 directly interact with haloxyfop (Fig. 3j). The mutation S1866F abolishes a hydrogen bond with helix α5′ at the dimer interface. Both the α5 and α5′ helices form direct interactions with haloxyfop (Fig. 3k). The mutation A1884P generates a kink in the middle of helix α4 at the bottom of the haloxyfop-binding pocket (Fig. 3l). All of the above mutations are expected to cause conformational changes within the haloxyfop-binding site, thus conferring herbicide resistance. Another mutation, A2123T, is not predicted to have a marked effect on herbicide binding, which is consistent with the fact that there were many A2123T reads in Group 20, but no A2123T mutants survived haloxyfop spraying (Supplementary Fig. 15 and Supplementary Dataset 2).

In conclusion, this report shows that STEME-editing is an effective tool for mutagenesis in plants. STEMEs can generate diverse mutations, including base substitutions and in-frame indels, facilitating analysis of protein function and development of agronomic traits. Meanwhile, the high product purity and low indel numbers obtained by editing protoplasts point to the importance of transient expression of CRISPR. Our STEME system could be applied for directed evolution of protein-coding genes for which a new or alternative functional activity is desired. For functional genes that lack easily selectable phenotypes, it may be possible to screen for mutants using high-throughput PCR of pooled template DNA samples from regenerated seedlings, and use CEL1 to cleave the resulting

amplicons, followed by further screening of the individuals in the positive pools with CEL1 and Sanger sequencing. Any confirmed positive mutants would need to be phenotyped at suitable developmental stages.

We envisage that this STEME system might also be applicable beyond plants; for example, for screening drug resistance mutants, altering *cis*-elements on noncoding regions and correcting pathogenic single-nucleotide variants in cell lines, yeast or animals.

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/s41587-019-0393-7.

References

1. Engqvist, M. K. M. & Rabe, K. S. Applications of protein engineering and directed evolution in plant research. *Plant Physiol.* **179**, 907–917 (2019).
2. Henikoff, S., Till, B. J. & Comai, L. TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.* **135**, 630–636 (2004).
3. Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N. & Facciotti, D. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* **23**, 75–81 (2005).
4. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
5. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
6. Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2016).
7. Halperin, S. O. et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).
8. Bao, Z. et al. Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat. Biotechnol.* **36**, 505–508 (2018).
9. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
10. Ran, Y., Liang, Z. & Gao, C. Current and future editing reagent delivery systems for plant genome editing. *Sci. China Life Sci.* **60**, 490–505 (2017).
11. Butt, H. et al. CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors. *Genome Biol.* **20**, 73 (2019).
12. Zhang, Y. & Qi, Y. CRISPR enables directed evolution in plants. *Genome Biol.* **20**, 83 (2019).
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. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, 1248 (2016).

15. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).

16. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).

17. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

18. Li, Q. et al. CRISPR–Cas9-mediated base-editing screening in mice identifies DND1 amino acids that are critical for primordial germ cell development. *Nat. Cell Biol.* **20**, 1315–1325 (2018).

19. Mol, C. D. et al. Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701–708 (1995).

20. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).

21. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).

22. Li, Z., Xiong, X. & Li, J.-F. New cytosine base editor for plant genome editing. *Sci. China Life Sci.* **61**, 1602–1603 (2018).

23. Li, C. et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **19**, 59 (2018).

24. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).

25. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).

26. Powles, S. B. & Yu, Q. Evolution in action: plants resistant to herbicides. *Annu. Rev. Plant Biol.* **61**, 317–347 (2010).

27. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).

28. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).

29. Grünewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).

30. Zhang, H., Yang, Z., Shen, Y. & Tong, L. Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase. *Science* **299**, 2064–2067 (2003).

Methods

Plasmid construction. The cytidine deaminase, adenosine deaminase, nCas9 (D10A) and UGI portions of STEME-1, STEME-2, STEME-3 and STEME-4 were amplified from A3A-PBE (ref. [21]) or PABE-7 (ref. [23]), and assembled into the pJIT163 backbone by One Step Cloning (ClonExpress II One Step Cloning Kit, Vazyme). PCR was performed using TransStart FastPfu DNA Polymerase (TransGen Biotech). The Cas9 variant nCas9-NG (D10A)[24] containing R1335V/ L1111R/D1135V/G1218R/E1219F/A1322R/T1337R substitutions was synthesized commercially (GENEWIZ). The sgRNA construct pOsU3-esgRNA was previously described[23]. Annealed oligos were inserted into *Bsa*I (New England BioLabs)-digested pOsU3-esgRNA. To construct the pH-STEME-1-esgRNA and pH-STEME-NG-esgRNA binary vectors, STEME-1 and STEME-NG along with the OsU3-esgRNA expression cassette were cloned into the pHUE411 backbone[31]. All of the primer sets used in this work are listed in Supplementary Table 8 and were synthesized by Beijing Genomics Institute.

Protoplast transfection. We used the *Japonica* rice variety Nipponbare to prepare protoplasts. Protoplast isolation and transformation were performed as described[32]. First, 10 μg each of nuclease and sgRNA plasmid DNA were introduced into the protoplasts by PEG-mediated transfection, with a mean transformation efficiency of 40–55% as measured by hemocytometer. The transfected protoplasts were incubated at 23 °C, and at 60 h post-transfection they were collected and genomic DNA was extracted for amplicon deep sequencing.

***Agrobacterium*-mediated transformation of rice callus cells.** The binary vectors for each group were pooled in equimolar ratios and transformed into *Agrobacterium tumefaciens* AGL1 by electroporation and used to transform about 240 rice calli. *Agrobacterium*-mediated transformation of callus cells of the *Japonica* rice variety Zhonghua11 was conducted as reported[32,33]. Hygromycin (50 μg ml⁻¹) was used to select transgenic plants.

Screening for herbicide tolerance. The regenerated rice seedlings were transferred to water, grown in a growth chamber (25 °C, 16 h light and 8 h dark) for 10 d and sprayed with haloxyfop (34 g active ingredient per hectare). The herbicide was applied with pressurized equipment at 0.2 MPa and a spray volume of 450 l ha⁻¹. Surviving seedlings were identified 3 weeks later.

Selection of haloxyfop-resistant seedlings in the medium. After transformation, the calli were selected on callus induction medium supplemented with hygromycin (50 μg ml⁻¹) for 4 weeks. Then, the hygromycin-resistant calli were transferred to callus induction medium supplemented with haloxyfop (0.108 mg l⁻¹). After 6 weeks of selection, the fresh and bright calli were transferred to regeneration medium supplemented with haloxyfop (0.108 mg l⁻¹) for regeneration (Supplementary Fig. 13a).

DNA extraction. The genomic DNA of protoplasts was extracted with a DNA-Quick Plant System (Tiangen Biotech). Genomic DNA of regenerated rice seedlings was extracted with CTAB, and all the seedlings in each group were sampled together. The targeted site was amplified with specific primers, and the amplicons were purified with an EasyPure PCR Purification Kit (TransGen Biotech), and quantified with a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific).

Amplicon deep sequencing and data analysis. To assess the coverage of sgRNAs, PCR was performed on transgenic plant DNA to amplify the protospacer region of the binary vectors with primers OsU3-F and esgRNA-R (Supplementary Table 8). Barcodes were added to the ends of the PCR products using primers (Supplementary Table 8). Two rounds of PCR were performed for the amplicons of protoplasts and grouped regenerated seedlings. In the first round of PCR, the target region was amplified using site-specific primers (Supplementary Table 8). In the second, both forward and reverse barcodes were added to the ends of the PCR products for library construction (Supplementary Table 8). Equal amounts of the PCR products were pooled and purified by Gel DNA Extraction, and samples were sequenced commercially (GENEWIZ) using the Illumina NextSeq 500 platform. The protospacer sequences in the reads were examined to identify C > T and A > G substitutions and indels. The amplicon sequencing was repeated three times for each target site, when using genomic DNA extracted from three independent protoplast samples. Amplicon reads with a quality score < 30 were filtered out. The distribution of identified alleles around edited targets for each sgRNA was analyzed using the online tool CRISPResso2 (ref. [34]). When analyzing the details of the amplicon reads in rice plants, we used 0.10% as a threshold to filter out unreliable sequencing reads, based on the control amplicon reads from wild-type rice plants. Analyses of base-editing processivity were performed as previously described[13].

Modeling of the 3D structure of the OsACC CT domain. Structural studies of the CT domains of yeast ACC in complex with CoA[30] and various herbicide inhibitors[35–37] revealed that the herbicide inhibitors all bind to the active site at the dimer interface. Pairwise sequence alignment of rice and yeast ACCs

was performed by the server (https://www.ebi.ac.uk/Tools/msa/clustalo/). The homology model of the rice OsACC CT domain was based on sequence alignment and the structures of the yeast ACC CT domain (1UYS, 1OD2) using the MODELLER[38] (v9.12) program. In the wild-type ACC, the indole rings of W2097 and W2125 form a π–π stacking pair near the herbicide-binding pocket. The pyridyl ring of haloxyfop is sandwiched between the side-chains of F2128 and Y1912′ (from the other monomer).

Detection of likely off-target sites. The potential off-target sites were predicted using the online tool Cas-OFFinder[39]. The PAM type parameter was chosen based on the PAM type of the on-target sgRNA. The off-target sites containing up to 3-nucleotide mismatches were examined in this study.

Statistical analysis. All of the statistical data were calculated using GraphPad Prism v.7.0. All numerical values of rice protoplasts were presented as means ± s.e.m. A statistical comparison adjustment was performed using two-tailed Student's *t*-test.

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 this study are available in the article and its supplementary figures and tables or are available from the corresponding author upon request. For sequence data, rice LOC_Os IDs LOC_Os01g55540 (*OsAAT*), LOC_Os05g22940 (*OsACC*), LOC_Os03g05730 (*OsCDC48*), LOC_Os09g26999 (*OsDEP1*), LOC_Os02g11010 (*OsOD*, *OsEV*) are available on the Rice Genome Annotation Project (http://rice.plantbiology.msu. edu/). The deep sequencing data have been deposited with the NCBI BioProject database under accession code numbers PRJNA590652 and PRJNA590653. Plasmids encoding STEME-1, STEME-2, STEME-3, STEME-4, STEME-NG, A3A-PBE-NG, PABE-7-NG, pH-STEME-1-esgRNA and pH-STEME-NG-esgRNA will be available from Addgene.

Code availability

The custom Python script to analyze types of mutational reads and amino acid substitutions is in the Supplementary Code file.

References

31. Xing, H. L. et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327 (2014).
32. Shan, Q. et al. Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Mol. Plant* **6**, 1365–1368 (2013).
33. Meng, X. et al. Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant* **10**, 1238–1241 (2017).
34. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
35. Zhang, H., Tweel, B. & Tong, L. Molecular basis for the inhibition of the carboxyltransferase domain of acetyl-coenzyme-A carboxylase by haloxyfop and diclofop. *Proc. Natl Acad. Sci. USA* **101**, 5910–5915 (2004).
36. Xiang, S., Callaghan, M. M., Watson, K. G. & Tong, L. A different mechanism for the inhibition of the carboxyltransferase domain of acetyl-coenzyme A carboxylase by tepraloxydim. *Proc. Natl Acad. Sci. USA* **106**, 20723–20727 (2009).
37. Yu, L. P., Kim, Y. S. & Tong, L. Mechanism for the inhibition of the carboxyltransferase domain of acetyl-coenzyme A carboxylase by pinoxaden. *Proc. Natl Acad. Sci. USA* **107**, 22072–22077 (2010).
38. Eswar, N. et al. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.* **31**, 3375–3380 (2003).
39. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).

Author contributions

C. Li, R.Z., X.M., J.L. and C.G. designed the project. C. Li, R.Z., S.C., Y.Z. and C. Lu performed most of the experiments. X.M. performed transformation. Y.-H.C. analyzed

the three-dimensional structure of the CT domain. J.L. and C.G. supervised the project. C. Li, R.Z., J.-L.Q., Y.-H.C., J.L. and C.G. wrote the manuscript.

Competing interests

The authors have submitted a patent application based on the results reported in this paper.

Additional information

Corresponding author(s): Caixia Gao

Last updated by author(s): Dec 29, 2019

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 NextSeq 500 platform was used to collect the amplicon deep sequencing data.
Data analysis	Graphpad prism 7 was used to analyze the data. Amplicon sequencing data of base-editing processivity was analyzed using the published code as previously described in reference 13. The distribution of identified alleles around edited targets for each sgRNA were analyzed using the online tool CRISPResso2 (http://http://crispresso2.pinellolab.org/). The potential off-target sites were predicted using the online tool Cas-OFFinder (http://http://www.rgenome.net/cas-offinder/). The custom Python script to analyze types of mutational reads and amino acid substitutions is available in Supplementary Code.

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 have been deposited with the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject ID PRJNA590652 and PRJNA590653.

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 regenerated seedlings was measured by regenerated seedlings from the same clones, totally about 6000. The herbicide resistance 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