



OPEN

# Targeted base editing in the plastid genome of *Arabidopsis thaliana*

Issei Nakazato<sup>1</sup>, Miki Okuno<sup>1,2,3</sup>, Hiroshi Yamamoto<sup>1,4</sup>, Yoshiko Tamura<sup>1</sup>, Takehiko Itoh<sup>1,2</sup>, Toshiharu Shikanai<sup>1,4</sup>, Hideki Takanashi<sup>1</sup>, Nobuhiro Tsutsumi<sup>1</sup> and Shin-ichi Arimura<sup>1</sup>✉

**Bacterial cytidine deaminase fused to the DNA binding domains of transcription activator-like effector nucleases was recently reported to transiently substitute a targeted C to a T in mitochondrial DNA of mammalian cultured cells<sup>1</sup>. We applied this system to targeted base editing in the *Arabidopsis thaliana* plastid genome. The targeted Cs were homoplasmically substituted to Ts in some plantlets of the T<sub>1</sub> generation and the mutations were inherited by their offspring independently of their nuclear-introduced vectors.**

Plastid genomes, which encode key genes for photosynthetic processes, including both light reactions and carbon assimilation, are potential targets of plant breeding. Plastid genetic transformation can now be used for only a limited number of species<sup>2</sup> and is difficult even in the model plant *Arabidopsis*<sup>3,4</sup>. In addition, it requires insertion of a marker gene into the plastid genome<sup>5</sup>, so the created plants are regarded as genetically modified organisms (GMOs). Recently, cytidine deaminase (CD), which converts C to U to change G/C pairs to A/T pairs in double-stranded DNA, was successfully used for in vitro targeted base editing of mitochondrial DNAs in mammalian cultured cells<sup>1</sup>. Here, we applied this technology to edit targeted bases in three genes in the plastid genome in *Arabidopsis* plantlets, without leaving any foreign genes in either the plastid or nuclear genomes. Targeted single nucleotide substitutions are expected to be the best way to make desired single nucleotide polymorphisms (SNPs) without disturbing any other genes or regulatory regions in the plastid genomes of common crops or elite lines. For the targets, we selected three genes whose modifications would be expected to lead to observable effects; *16S rRNA*, whose modification was expected to confer resistance to an antibiotic and two genes whose modifications would lead to poor growth; *rpoC1*, which encodes a part of the DNA-directed RNA polymerase subunit beta'; and *psbA*, which encodes photosystem II (PSII) protein D1.

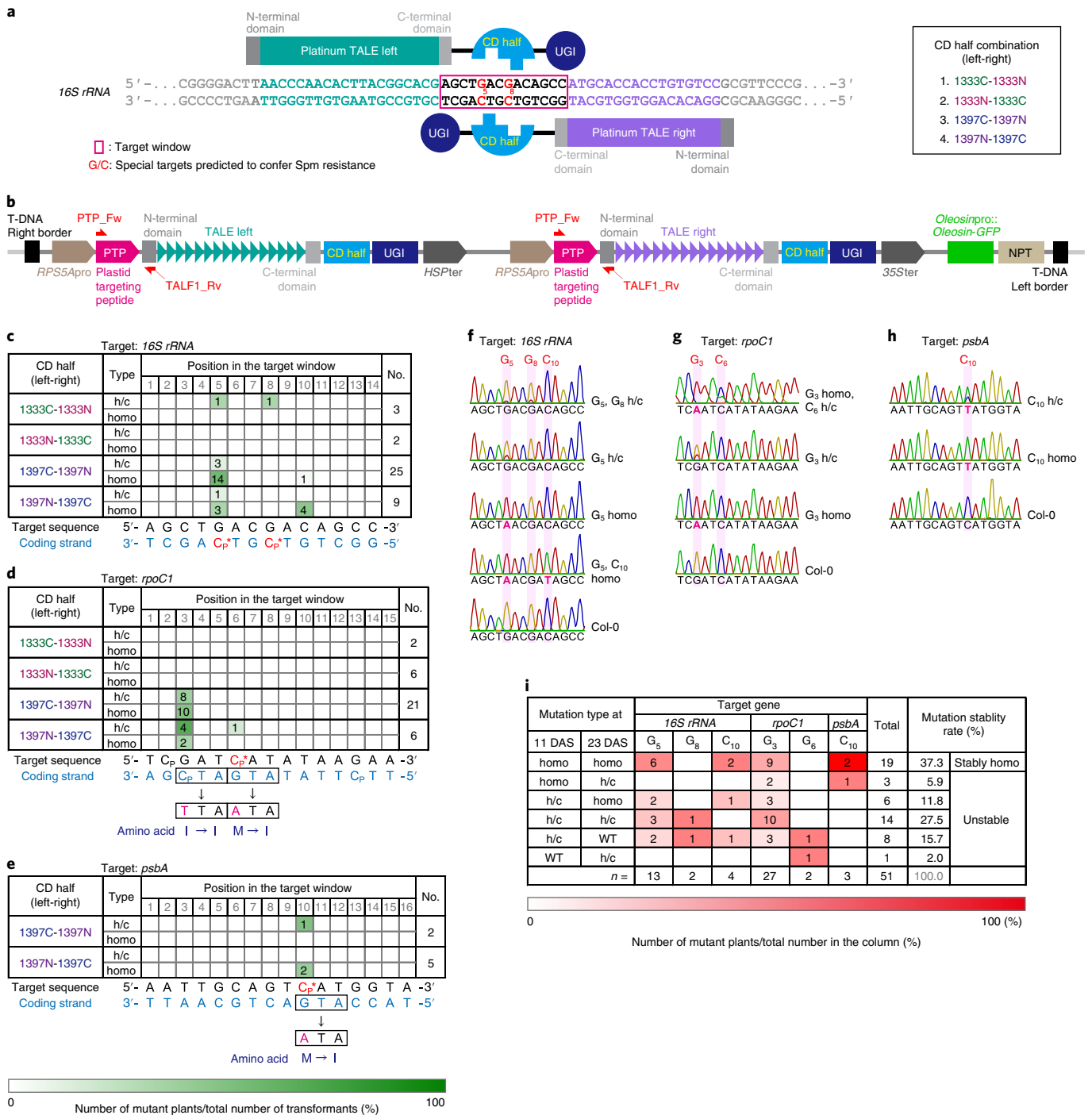
As in the previous study<sup>1</sup>, the CD domain (163 amino acids (aa)) of *Burkholderia cenocepacia* DddA toxin (1,427 aa) was split at the 1,333th or 1,397th amino acid. Each of the amino terminal or carboxy terminal halves of the CD was linked to the C terminus of the DNA binding domain of the platinum TALEN<sup>6</sup> (pTALECD; Fig. 1a). The N terminus of the pTALECD was linked to a plastid-targeting signal peptide (PTP) of *Arabidopsis thaliana* RecA1 protein (51 aa)<sup>7,8</sup> (Fig. 1b), while the C terminus was linked to an uracil glycosylase inhibitor (UGI)<sup>1,9</sup> to inhibit hydrolysis of the generated uracil (Fig. 1b). The nucleotide sequences of CD and UGI were optimized for *A. thaliana* codon usage. A pair of PTP-pTALECD-UGIs (ptpTALECDs) were expressed in a single plant transformation vector under control of efficient *RPS5A* promoters<sup>10</sup> (Fig. 1b).

We established a system to smoothly assemble the complicated tandem expression vectors of ptpTALECD for each target sequence on the Ti plasmid (Extended Data Fig. 1) by replacing the FokI in the vectors used in a previous study<sup>11</sup> with CD-UGI (Extended Data Fig. 2). We introduced the vectors into the nucleus of *A. thaliana* by floral dipping<sup>12</sup> and attempted to substitute C/G to T/A in *16S rRNA* (Fig. 1c), *rpoC1* (Fig. 1d) and *psbA* (Fig. 1e). Substitution of G<sub>5</sub> and/or G<sub>8</sub> in *16S rRNA* (highlighted in red in Fig. 1a) to A would confer spectinomycin (Spm) resistance (below)<sup>13,14</sup> while substitutions of C<sub>6</sub> in *rpoC1* and C<sub>10</sub> in *psbA* to Ts would lead to changes in initiation codons from ATG (methionine) to ATA (isoleucine). As a result, accumulation of their coding proteins would decrease and mutants would grow poorly<sup>15,16</sup> and/or be unable to grow photoautotrophically<sup>17,18</sup>. Other neutral mutations in some C/G pairs in the target windows, which are the regions between the sequences that the left and right transcription activator-like effector (TALE) domains recognize, would also be expected.

Twelve ptpTALECD expression vectors were constructed (four pairs of CD halves for each of three targets). Each vector was introduced into *A. thaliana* and, 23 days after stratification (DAS), the targeted regions of the T<sub>1</sub> plants were sequenced by Sanger sequencing. Only constructs from which T<sub>1</sub> plants were obtained are shown in Fig. 1c–e and Supplementary Table 1a–c. In all three target windows, C/G pairs were replaced with T/A pairs in multiple T<sub>1</sub> lines (Fig. 1c–h and Supplementary Table 1a–c). Surprisingly, in many lines, the targeted base(s) seemed to be homoplasmically substituted (homo), while in other lines, they seemed to be heteroplasmically or chimaerically substituted (h/c; Fig. 1c–h and Supplementary Table 1a–c). Such homoplasmic mutations might have occurred through stochastic sorting processes, such as selection of mutations in a small number of copies of plastid genomes (that is, plastid sorting)<sup>19</sup> or gene conversion<sup>20</sup>. Nevertheless, it is also conceivable that ptpTALECD mutated the C/G pairs in the target windows of all plastid genomes at the early stage of embryogenesis because the *RPS5A* promoter used for ptpTALECD expression was reported to highly drive gene expression in egg cells and early embryos<sup>21,22</sup>. Not all C/G pairs in the target windows were substituted and the positions of the substituted C/G pairs were biased for all the three target windows (Fig. 1c–e). Three homoplasmically substituted bases were C of (5') TC(3') (Cp in Fig. 1c–e), which was the preferential target<sup>1</sup> but a C of (5')AC(3') in *16S rRNA* gene was also substituted (Fig. 1c).

To investigate the stability of mutation rates during plant development, total DNAs extracted from an emerging leaf of T<sub>1</sub> plants at 11 and 23 DAS (or from a cotyledon of slowly growing plants at 11 DAS; Supplementary Table 1a–c) were sequenced. Among the

<sup>1</sup>Laboratory of Plant Molecular Genetics, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. <sup>2</sup>School of Life Science and Technology, Tokyo Institute of Technology, Tokyo, Japan. <sup>3</sup>Division of Microbiology, Department of Infectious Medicine, Kurume University School of Medicine, Kurume, Japan. <sup>4</sup>Department of Botany, Graduate School of Science, Kyoto University, Kyoto, Japan. ✉e-mail: [arimura@g.ecc.u-tokyo.ac.jp](mailto:arimura@g.ecc.u-tokyo.ac.jp)



**Fig. 1 | pTALECD for three plastid genes. a**, A pair of pTALECD proteins and its targeting window (shown in a red rectangle) in *16S rRNA* gene and CD half combination. Substitution of the fifth and/or eighth C/G pairs (shown in red) with T/A pairs was predicted to confer Spm resistance. **b**, T-DNA region of the tandem expression vectors for pTALECD. **c-e**, Base-edited plant numbers, editing efficiencies shown in colour (bottom) and predicted amino acid substitutions in the three target windows of 23 DAS T<sub>1</sub> plants (**c**, *16S rRNA*; **d**, *rpoC1*; **e**, *psbA*). **f-h**, Representative data from Sanger sequencing in the pTALECD targeting windows of 23 DAS T<sub>1</sub> plants (**f**, *16S rRNA*; **g**, *rpoC1*; **h**, *psbA*). **i**, Transition of substitution frequency states at the targeted bases between 11DAS and 23 DAS T<sub>1</sub> plants. Abbreviations: h/c, heteroplasmically or chimaerically substituted; homo, homoplasmically substituted; Cp, preferential target cytosines; and C\*, special target cytosines predicted to cause biological effects.

plants with base change(s) in the target window on either day, some had bases heteroplasmically or chimaerically (h/c) substituted on both days with their mutation rate increased or decreased (27.5%, 14/51; Fig. 1i) and others had bases at which the mutation rate differed between the two time points (for example, homoplasmism

(homo) to h/c, 5.9% (3/51); h/c to null, 15.7% (8/51); h/c to homo, 11.8% (6/51); or null to h/c, 2.0% (1/51); Fig. 1i). Many of the remaining plants had bases that were homoplasmically substituted on both days (37.3%, 19/51; Fig. 1i). Interestingly, one leaf of one T<sub>1</sub> plant (*16S rRNA* 1397NC3) showed an example of sector

formation<sup>19</sup>; that is, it had differently coloured sectors (wild-type-like green and pale) and the mutation rate at the Cp\* in *16S*rRNA differed between the sectors (Extended Data Fig. 3a,b). Remarkably, most of the bases homoplasmically substituted at 11 DAS were also homoplasmically substituted at 23 DAS (86.4%, 19/22; Fig. 1i), suggesting that the targeted bases of T<sub>1</sub> plants transformed by the ptpTALECD expression vector were homoplasmically substituted at a high frequency and that the homoplasmic mutations were stably fixed through development.

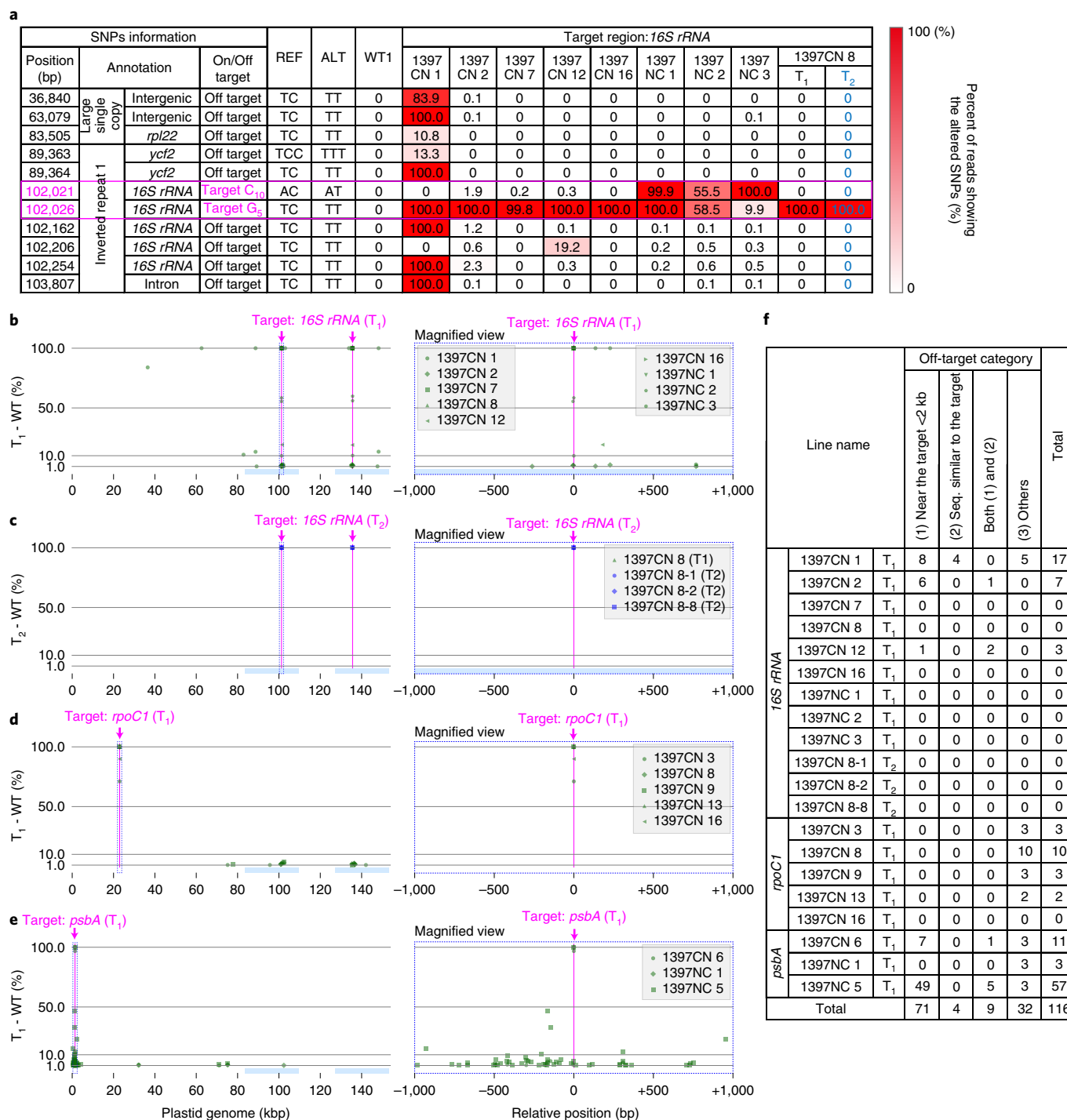
Next, we investigated the off-target effects of ptpTALECD on the plastid and mitochondrial genomes, because both organelle genomes are maternally inherited, so off-target mutations in these two organelle genomes cannot be segregated from the desired mutation by usual cross-breeding. The total genomes of 17 T<sub>1</sub> plants were sequenced (Novaseq, illumina) and Fig. 2a and Supplementary Table 2 show the results. As Fig. 2a and Supplementary Table 2 show, the targeted C appeared to be homoplasmically substituted to T in 16 of them (*16S*rRNA 1397C-1397N (1397CN) 1, 2, 7, 8, 12, 16; 1397N-1397C (1397NC) 1-3; *psbA* 1397CN 6; 1397NC 1, 5; and *rpoC1* 1397CN 8, 9, 13, 16), while it was either heteroplasmically or chimaerically substituted in the remaining one (*rpoC1* 1397CN 3). Each redundant mutation in the inverted repeats of the plastid genome was counted as one mutation. The targeted bases in these 16 lines were confirmed to be homoplasmically or dominantly substituted and the base in the remaining one line was confirmed to be heteroplasmically or chimaerically substituted (Fig. 2a and Supplementary Table 2). Dominant off-target point mutations (with substitution frequencies >50%) were detected at six places in the line *16S*rRNA 1397CN 1, while no dominant off-target point mutations were detected in the other lines (Fig. 2a and Supplementary Table 2). The *16S*rRNA 1397CN 1 did not have any true leaves (Extended Data Fig. 4) and died before 23 DAS. A total of 116 off-target mutations (allele frequencies ≥1%) were detected in the plastid genomes (Fig. 2b-f and Supplementary Table 2). Most (69.0%) were located within 2,000 base pairs (bp) of the target windows while only a few (11.2%) were located within 20 bp of sequences similar to those recognized by TALEs. The rest were found in other regions. No dominant off-target mutations were detected in the mitochondrial genomes of these 17 lines, including *16S*rRNA 1397CN 1 (Supplementary Table 2). These results indicate that ptpTALECD only infrequently introduced off-target point mutations into organelle genomes and can specifically and homoplasmically (or dominantly) substitute C/G to T/A in the target windows.

All but one of the T<sub>1</sub> plants that were transformed by the *16S*rRNA-targeting ptpTALECD vector and whose first Cp\* (G<sub>5</sub>) and/or C<sub>10</sub> was homoplasmically substituted were fertile (Supplementary Table 1a). The exception was *16S*rRNA 1397CN 1. To investigate whether the mutations were stably inherited by the offspring, the T<sub>2</sub> progenies of three of these lines (*16S*rRNA 1397CN 2, 8 and 1397NC 3) were genotyped (Fig. 3a and Extended Data Fig. 5a). Transgenic T<sub>2</sub> plants were identified by having seed-specific green fluorescent protein (GFP) fluorescence from *Ole1 pro::Ole1-GFP* (ref. <sup>23</sup>) on the transfer DNA (T-DNA; Fig. 1b) and/or a positive polymerase chain reaction (PCR) result showing the presence of the ptpTALECD reading frame (Fig. 1b and 3a). Both progeny stably inherited the homoplasmic mutations (Fig. 3a and Extended Data Fig. 5a). Interestingly, some T<sub>2</sub> plants had white, red or variegated cotyledons (Fig. 3b and Extended Data Fig. 5b), which were different from the phenotypes of their parents (Extended Data Fig. 4). All of these plants were GFP positive (Fig. 3a and Extended Data Fig. 5a) and many of them (8/9) had additional mutation(s) in or near the target window of *16S*rRNA (Extended Data Fig. 5a). Because, as mentioned above, the *RPS5A* promoter used for ptpTALECD expression was reported to highly drive gene expression in egg cells and early embryos<sup>21,22</sup>, de novo mutagenesis may have occurred during the early developmental stage in these

transgenic T<sub>2</sub> plants with abnormal cotyledons. In contrast, all T<sub>2</sub> plants of the T-DNA-free null segregants examined had the targeted mutations without any of the additional altered phenotypes described above. No major off-target mutations were detected in the three null-segregant T<sub>2</sub> plants (*16S*rRNA 1397CN 8 lines 1, 2 and 8), whose genomes were sequenced by next generation sequencing (Fig. 2a and Supplementary Table 2). Homoplasmic mutations in *rpoC1* (G<sub>3</sub>) and *psbA* (C<sub>10</sub>) in other lines were also inherited by their T<sub>2</sub> progeny (Extended Data Figs. 6a and 7a). These data indicate that plastid genomes with artificially introduced point mutations were stably inherited, independently of nuclear T-DNA inheritance and also suggest that transgene-free plants with targeted point mutations in the plastid genomes were successfully established.

The antibiotic Spm binds to a specific location in *Escherichia coli 16S*rRNA and inhibits translation<sup>24</sup>. Substitution of a specific G near this region to A confers Spm resistance (Spm<sup>r</sup>)<sup>13</sup>. The targeted G<sub>5</sub> in the *Arabidopsis* plastid *16S*rRNA gene is homologous to this G. Several mutations are known to confer Spm<sup>r</sup> to flowering plants<sup>25,26</sup> but none of them occur at the position of the targeted G<sub>5</sub>. T<sub>2</sub> seeds obtained from a T<sub>1</sub> plant in which G<sub>5</sub> was homoplasmically substituted to A (*16S*rRNA 1397CN 2; Supplementary Table 1a) were sown on plates containing Spm. Many of the seedlings that germinated from these seeds showed Spm<sup>r</sup>, regardless of the presence of seed GFP fluorescence (Fig. 3c and Extended Data Fig. 8b-d). However, some T<sub>2</sub> progenies from *16S*rRNA 1397CN 2 showed a Spm-sensitive (Spm<sup>s</sup>)-like phenotype (white plantlet with purple cotyledon; Fig. 3c and Extended Data Fig. 8a-d). All the Spm<sup>s</sup>-like plantlets germinated from GFP-positive seeds (Fig. 3c and Extended Data Fig. 8a-d) and many of them (5/5; Extended Data Fig. 9) had multiple de novo mutations (in addition to G<sub>5</sub>) in *16S*rRNA. This suggests that the de novo mutations caused dysfunction of *16S*rRNA, resulting in the Spm<sup>s</sup>-like phenotype. The existence of Spm<sup>s</sup>-like T<sub>2</sub> plants of *16S*rRNA 1397CN 2 on Spm-free medium (Extended Data Fig. 5a,b) with the de novo mutation(s) support this suggestion. Surprisingly, some of the progeny of a T<sub>1</sub> plant (*16S*rRNA 1397CN 15) that had the G<sub>5</sub> mutation at very low frequency at 11 DAS and no mutation at 23 DAS (Supplementary Table 1a) also showed Spm<sup>r</sup>. These progeny germinated from GFP-positive seeds (Fig. 3c). In five of them, the G<sub>5</sub> was homoplasmically substituted to A and in 13 others it was dominantly substituted to A (Extended Data Fig. 9). This suggests that the inherited nuclear T-DNA caused a major de novo mutation on the G<sub>5</sub>. These results suggest that homoplasmic substitution of G<sub>5</sub> to A confers Spm<sup>r</sup> to *A. thaliana*. Furthermore, the result that the GFP-negative T<sub>2</sub> progeny showed Spm<sup>r</sup> or Spm<sup>s</sup> phenotype that was predictable from SNPs at G<sub>5</sub> in the T<sub>1</sub> plants showed that the null-segregant T<sub>2</sub> plants were likely to inherit mutation(s) that their parent had and not likely to have additional mutations.

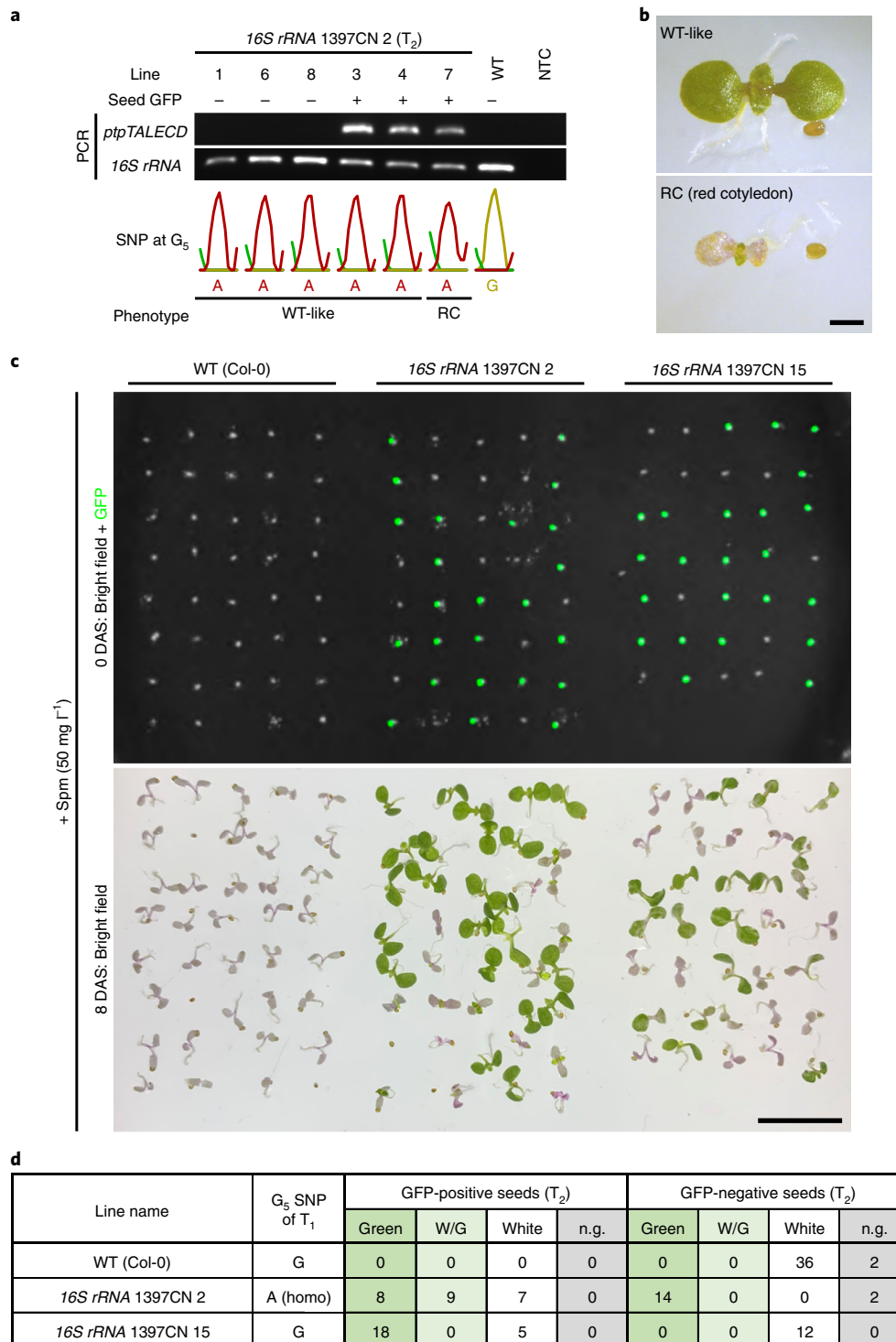
Previous studies showed that accumulation of D1 protein (encoded by *psbA*) and/or the maximum quantum yield of PSII (*Fv/Fm*) drastically decreased in mutants deficient in *psbA* expression<sup>17,18</sup>. Furthermore, these mutants looked pale<sup>17</sup> and could not grow photoautotrophically<sup>17,18</sup>. Surprisingly, *psbA* 1397NC 1, which had the homoplasmic mutation at the *psbA* initiation codon (C<sub>10</sub>) at both 11 and 23 DAS (Supplementary Table 1c), could grow photoautotrophically and set viable seeds. Thus, to investigate the effects of the homoplasmic mutation at the *psbA* initiation codon (C<sub>10</sub>) on its expression, we measured *Fv/Fm* and the accumulation of D1 protein in T-DNA-free null-segregant T<sub>2</sub> progeny of *psbA* 1397NC 1, which were confirmed to inherit the homoplasmic mutation. Unexpectedly, their growth (Extended Data Fig. 6b,c) and accumulation of D1 protein (Extended Data Fig. 6d-f) were comparable to those in wild-type plants, while *Fv/Fm* was only slightly decreased compared with wild-type plants (Extended Data Fig. 6g). One possibility is that another codon served as the initiation codon. It could be another AUG, or possibly a GUG or UUG, which can also serve



**Fig. 2 | Investigation of on- and off-target mutations and determinants of off-target mutations.** **a**, Mutated read frequencies in the plastid genomes of nine T<sub>1</sub> lines (11 DAS) and one T<sub>2</sub> line (49 DAS) targeted for 16S rRNA. SNPs where ≥10% of the reads were different from the reference genome (AP000423.1) in at least one plant are listed. Supplementary Table 2 shows all the on- and off-target mutations in both organelle genomes, where different reads frequencies were ≥1%. **b-e**, Positions and frequencies of on- and off-target mutations (shown as green dots) in T<sub>1</sub> plants targeting 16S rRNA (**b**), *rpoC1* (**d**) and *psbA* (**e**) and in null-segregant T<sub>2</sub> plants mutated in 16S rRNA (**c**). Magenta lines represent the target windows. The right panels are magnified views of the regions surrounded by the dotted rectangles in left panels within 1,000 bp from the target nucleotides in each gene (G<sub>5</sub> in 16S rRNA, G<sub>3</sub> in *rpoC1* and C<sub>10</sub> in *psbA*). **f**, List of the off-target mutations categorized by types. Off-target mutations within 2 kb of the target windows and/or within 20 bp of sequences similar to those recognized by TALEs and in other regions are shown. In **b-f**, off-target mutations were defined as SNPs in which C/G-to-T/A substitutions (allele frequencies ≥0.01) were detected only in the T<sub>1</sub> or T<sub>2</sub> plants but not in three wild-type plants used as controls.

as start codons in the chloroplast<sup>27</sup>. Upstream of the altered AUA, no such sites occur after the nearest stop codon. Downstream, the next potential start codons would shorten the protein by at least 10% but

they can be excluded because the recombinant protein was the same size as the wild-type protein (Extended Data Fig. 6d). These results suggest that the AUA codon does not greatly affect the initiation



**Fig. 3 | T<sub>2</sub> generation analysis. a,b**, Genotypes and phenotypes of T<sub>2</sub> progenies of 16S rRNA 1397CN 2. **a**, Gel images of bands of PCR products of 16S rRNA and *ptpTALECD*, presence of seed GFP fluorescence, genotypes of G<sub>5</sub> SNP and phenotypes of T<sub>2</sub> progenies of 16S rRNA 1397CN 2 are shown. Abbreviations: WT, wild type (Col-0); NTC, non-template control. **b**, Representative phenotype images of five WT-like plants (lines 1, 3, 4, 6, 8) and a plant with red cotyledons (line 7). Scale bar, 1 mm. **c,d**, Phenotypes of T<sub>2</sub> progenies of 16S rRNA 1397CN 2 and 15 in the presence of Spm. **c**, Images of seeds (0 DAS) and seedlings (8 DAS) of the T<sub>2</sub> progenies of the two lines and WT (Col-0) on 1/2 MS medium containing 50 mg l<sup>-1</sup> of Spm. Scale bar, 1 cm. **d**, A table that displays the relationship between the presence of seed GFP fluorescence and 8 DAS plants colour. Abbreviations: W/G, white or red cotyledons and green leaves; n.g., not germinated.

of translation of *psbA* or the D1 level but that the AUG codon is necessary for the full activity of PSII. Thus, a better way to knock out a plastid gene might be to create a premature stop codon in its reading frame rather than to change the initiation codon to AUA.

In *rpoC1*, none of the homoplasmic mutations that were obtained were at the initiation codon as expected. Instead, they were at the second codon where they caused a synonymous mutation (Ile to Ile; Fig. 1d and Supplementary Table 1b). Null-segregant T<sub>2</sub>

progeny of *rpoCl* 1397CN 8, which had the synonymous homoplasmic mutation at both 11 and 23 DAS, inherited the homoplasmic mutation and appeared to grow as well as wild-type plants (Extended Data Fig. 7b,c).

These experiments showed that ptpTALECD could specifically introduce homoplasmic C-to-T mutations in target windows in the *A. thaliana* plastid genome and that the mutations were stably (and probably maternally) inherited by the progeny seeds. Previous attempts to introduce homoplasmic mutations in mammalian mitochondrial genomes were unsuccessful<sup>1,28</sup>. The method was also successful in a region of inverted repeats, where mutations are thought to occur at a lower rate due to their greater potential for copy correction<sup>29</sup>; *16S rRNA* occurs in inverted repeats and targeted point mutations were successfully introduced in both copies. Compared to traditional methods for plastid transformation, such as biolistic methods, ptpTALECD technology has three advantages. First, it allows plastid-genome editing of *A. thaliana* without using specific mutants<sup>3,4</sup> or a specific ecotype<sup>30</sup> and without tissue culture, which is a major obstacle to plastid transformation. Second, it could probably be used to edit plastid genomes of other plant species that are recalcitrant to plastid transformation but amenable to nuclear transformation. And third, it could be used to create plastid-genome-edited plants without leaving any foreign gene in their genomes. Such plants are not regarded as GMOs in several countries. On the other hand, the ptpTALECD method has some problems with respect to accuracy. For example, unwanted substitutions in the target windows occurred ( $C_{10}$  in *16S rRNA* and  $G_3$  in *rpoCl*; Fig. 1c,d), while homoplasmic mutations at some special target C/G pairs in the target windows were not introduced ( $G_8$  in *16S rRNA* and  $C_6$  in *rpoCl*; Fig. 1c,d). These problems might be avoided by sliding the TALE recognition targets a few base pairs upstream or downstream or by using different sizes of target windows or by optimizing the sequences linking the TALE and CD<sup>31</sup>. In any case, only a few mutations in this study were off target. We also obtained null-segregant  $T_2$  plants that had the targeted homoplasmic mutation but had no off-target mutations (Fig. 2a and Supplementary Table 2).

This technology may also be useful for strengthening agronomic traits. For example, amino acid polymorphisms in the plastid-encoded ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit are expected to affect the carbon assimilation (and oxidation) rate<sup>32,33</sup> and some polymorphisms in *psbA* (not involving the  $C_{10}$  in this study) enhance herbicide resistance<sup>34</sup>. In addition, null-segregant plants are not regarded as GMOs in some countries and the introduced mutations would not leak out of the pollen<sup>2,5</sup>. Therefore, plants with their plastid genomes precisely edited by ptpTALECD might be more acceptable to the public. Also, this technology could be used for creating premature stop codons, substituting amino acids and modifying RNA editing sites. Thus, ptpTALECD technology has the potential to accelerate both plant breeding and basic research on plastid-encoded genes.

## Methods

**Plant material and growth conditions.** *A. thaliana* Columbia-0 (Col-0) and transgenic plants were grown at 22°C and under long-day conditions (16 h light, 8 h dark). Col-0 seeds were sown on 1/2 Murashige and Skoog (MS) medium (pH 5.7) containing 2.3 g l<sup>-1</sup> of MS Plant Salt Mixture (Wako), 500 mg l<sup>-1</sup> of MES, 10 g l<sup>-1</sup> of sucrose, 1 ml l<sup>-1</sup> of Plant Preservative Mixture (Plant Cell Technology), 1 ml l<sup>-1</sup> of Gamborg's Vitamin Solution (Sigma-Aldrich) and 8 g l<sup>-1</sup> of agar. Seedlings at 2–3-weeks-old were transferred to Jiffy-7 (Jiffy Products International) and thereafter subjected to *Agrobacterium* transfection. Most  $T_1$  plants were transplanted to Jiffy-7 but several growth-retarded plants were transplanted to plant boxes containing the 1/2 MS medium at 23 DAS.

**Designing the TALE binding sequence.** The TALE targeting sequence was designed to be on both sides of the CD targeting window, with Old TALEN Targeter (<https://tale-nt.cac.cornell.edu/node/add/talen-old>). The first recognized base was required to be adjacent to the 3' side of 'T' as far as possible. The minimum length of TALE targeting sequence was 15 bp so that TALE would specifically bind the

sequence. All the sequences that TALE binds and the target windows between the TALE binding sequences are shown in Supplementary Table 3 and Fig. 1c–e.

**Vector constructions.** A pair of left and right ptpTALECDs in Ti-plasmids (Extended Data Fig. 1b) for each target was constructed by using Platinum Gate assembling kit and multisite Gateway (Thermo Fisher) as described in our previous study of mitochondria-targeted TALEN<sup>11</sup>.

The DNA binding domains of ptpTALECD were assembled with the Platinum Gate TALEN system<sup>9</sup> on the basis of the same previous study<sup>11</sup> (Extended Data Fig. 1a). Each FokI coding sequence in the previous vectors of mitoTALENs used for assembly-step2 was replaced in advance by the CD half and UGI coding sequence with In-Fusion HD Cloning Kit (TaKaRa; Extended Data Fig. 2). The CD half and UGI coding sequences were designed to encode the same amino acids as those of Mok's experiment<sup>1</sup> and artificially synthesized by Eurofins Genomics with the codon usage optimized for *A. thaliana* (<https://www.eurofinsgenomics.jp/jp/orderpages/gsy/gene-synthesis-multiple/>; Supplementary Table 4). The reading frames in the assembled first and third entry vectors and the second entry vector (below) were transferred into the Ti plasmid<sup>10</sup> by a multi-LR reaction with LR Clonase II Plus enzyme (Thermo Fisher Scientific; Extended Data Fig. 1b). The second entry vector had an *Arabidopsis* heat-shock protein terminator<sup>35</sup>, an *Arabidopsis* RPS5A promoter and the N terminal (51 aa) PTP of *Arabidopsis* RECA1 (refs. 78; Extended Data Fig. 10a). This Ti plasmid was made from a Gateway destination Ti plasmid pK7WG2 (ref. 36) by replacing the CaMV 35S promoter with the *Arabidopsis* RPS5A promoter and inserting the PTP coding sequence and *Ole1* pro::*Ole1-GFP* derived from pFAST02 (ref. 23; Extended Data Fig. 10b; <http://www.inplanta.jp/pfast.html>). All primers used for vector construction are listed in Supplementary Table 5. All plasmids are deposited in Addgene and their sequences are also available in Addgene (ID 171723–171736).

**Plant transformation and screening transformants.** Col-0 plants were transformed by floral dipping<sup>12</sup> with *Agrobacterium tumefaciens* strain C58C1 that harboured one of the transformation vectors described above. Transgenic  $T_1$  seeds were selected at first by observing seed GFP fluorescence. GFP-positive seeds were sown on the 1/2 MS medium (section Plant material and growth conditions) further containing 125 mg l<sup>-1</sup> of claforan. In addition, GFP-negative seeds were sown on the 1/2 MS medium containing 50 mg l<sup>-1</sup> of kanamycin and 125 mg l<sup>-1</sup> of claforan.

**Sanger sequencing and next generation sequencing and their analyses.** Total DNAs were extracted from an emerging true leaf or a cotyledon of the selected seedlings with Maxwell RSC Plant DNA Kit (Promega). To genotype transgenic lines, plastid DNA sequences adjacent to the CD targeting windows were amplified with primer sets (Supplementary Table 6). Purified PCR products were subjected to Sanger sequencing (Eurofins Genomics) to detect substitution of the targeted bases. The data were analysed with Geneious prime (v.2020.2.2).

We called SNPs in the plastid and mitochondrial genomes using total DNA sequenced data. First, we ordered Macrogen Japan to prepare paired-end libraries using a Nextera XT DNA library Prep Kit (Illumina) and sequenced using Illumina NovaSeq 6000 platform. As preprocess for analysis, low-quality and adaptor sequences in the reads were trimmed using Platanus\_trim v.1.0.7 ([http://platanus.bio.titech.ac.jp/platanus\\_trim](http://platanus.bio.titech.ac.jp/platanus_trim)). Pair-end reads of each strain were mapped to reference sequences (AP000423.1 and BK010421.1) using BWA (v.0.7.12)<sup>37</sup> in single-ended mode. We filtered out inadequate mapped reads with mapping identities  $\leq 97\%$  or alignment cover rates  $\leq 80\%$ . SNPs were then called using samtools mpileup command (-uf -d 30000 -L 2000) and bcftools call command (-m -A -P 0.1)<sup>38</sup>. We finally listed positions in which variants with allele frequencies (AFs)  $\geq 0.1$  were detected in at least one strain including the WT (Fig. 2a). SNP calls with AFs  $\geq 0.01$  were also performed for positions with read depths  $\geq 500$  (Supplementary Table 2).

To evaluate whether closeness to target sites or similarity to TALE sequences influenced the locations of off-target mutations, we tallied off-target mutations that were either within 2,000 bp of the target site or within 20 bp of sequences  $\geq 70\%$  similar to those recognized by one of the TALEs.

**Genotyping  $T_2$  plants.**  $T_2$  seeds gained from several  $T_1$  lines were sown on the 1/2 MS medium (section Plant material and growth conditions). The genotypes of the target windows of a cotyledon of the 7DAS (for Fig. 3a and Extended Data Figs. 5a, 6a and 7a) or 13 DAS (for Extended Data Fig. 9) seedlings were determined in the same way as determining those of  $T_1$  plants (above). The ptpTALECD PCRs were performed with primers described in Supplementary Table 6.

**Screening Spm-resistant plants.**  $T_2$  seeds obtained from a  $T_1$  line of which  $G_3$  in *16S rRNA* was homoplasmically substituted at 11 and 23 DAS and control seeds were sown on the 1/2 MS medium (section Plant material and growth conditions) containing 0, 10, 50 or 100 mg l<sup>-1</sup> of Spm (without Plant Preservative Mixture for Extended Data Fig. 8a–d). Phenotypes of germinated seedlings were observed on 8 DAS.

**Measurement of chlorophyll fluorescence.** Chlorophyll fluorescence was measured using a MINI-pulse-amplitude modulation portable chlorophyll

fluorometer (MINI-PAM; Walz). Minimal fluorescence at open PSII centres in the dark-adapted state ( $F_0$ ) was excited by a weak measuring light (650 nm) at a photon flux density of 0.05 to 0.1  $\mu\text{mol}$  of photons  $\text{m}^{-2} \text{s}^{-1}$ . A saturating pulse of white light (800 ms, 8,000  $\mu\text{mol}$  of photons  $\text{m}^{-2} \text{s}^{-1}$ ) was applied to determine the maximal fluorescence at closed PSII centres in the dark-adapted state ( $F_m$ ). Maximum quantum yield of PSII was calculated as  $F_v/F_m$ . These procedures were done independently three times (experimental replicates = 3). In each replicate, four plants of each genotype (Col-0 and *psbA* 1397NC 1 T<sub>2</sub>) were analysed, average values and standard errors were calculated and  $F_v/F_m$  values of the two groups were tested by two-tailed Welch's test.

**SDS-polyacrylamide gel electrophoresis and immunoblot analyses.** Leaf extract was prepared by grinding the rosette leaves using mortar and pestle in an ice-cold buffer (20 mM Tricine (pH 8.4) containing 330 mM sorbitol, 10 mM NaHCO<sub>3</sub>, 5 mM EGTA and 5 mM EDTA). After filtration with two layers of Miracloth, intact chloroplasts were collected by centrifugation for 5 min at 4,800g. The purified chloroplasts were ruptured in a buffer (20 mM HEPES-KOH (pH 7.6), 5 mM MgCl<sub>2</sub>, 2.5 mM EDTA and complete ULTRA protease-inhibitor cocktail (Roche)). The insoluble fraction containing thylakoids and envelopes was separated from the soluble fraction by centrifugation for 2 min at 15,000g and resuspended in the above buffer. The concentration of chlorophyll was determined as described previously<sup>39</sup>. Chloroplast thylakoid and membrane proteins were solubilized in SDS-PAGE sample buffer. Proteins solubilized from the thylakoid membrane corresponding to 1–2  $\mu\text{g}$  of chlorophyll were separated by 12.5% (w/v) SDS-PAGE and electrotransferred onto polyvinylidene fluoride membranes. The antibodies were added and the protein-antibody complexes were labelled using the ECL Prime western-blotting detection system (GE Healthcare). The chemiluminescence was detected with a lumino-image analyser (LAS4000, GE Healthcare). Anti-PsbA and anti-AtpB were purchased from Agrisera. Anti-PetA and anti-PsbO were kindly provided by A. Makino (Tohoku University, Japan) and T. Endo (Kyoto University, Japan), respectively.

**Image processing.** Plant images were taken by iPhone Xs (Apple) and LEICA MC 170 HD (Leica). Gel images were taken by ChemiDoc MP Imaging System (BIORAD). These images were processed with Adobe Photoshop 2021 (Adobe). Figures and tables were made with Adobe Photoshop 2021 and Adobe Illustrator 2021 (Adobe).

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

## Data availability

All data are available in the main text or the Supplementary Information. Source data are provided with this paper.

Received: 9 January 2021; Accepted: 26 May 2021;  
Published online: 1 July 2021

## References

- Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
- Piatek, A. A., Lenaghan, S. C. & Stewart, C. N. Advanced editing of the nuclear and plastid genomes in plants. *Plant Sci.* **273**, 42–49 (2018).
- Yu, Q., Lutz, K. A. & Maliga, P. Efficient plastid transformation in *Arabidopsis*. *Plant Physiol.* **175**, 186–193 (2017).
- Ruf, S. et al. High-efficiency generation of fertile transplastomic *Arabidopsis* plants. *Nat. Plants* **5**, 282–289 (2019).
- Fuentes, P., Armarego-Marriott, T. & Bock, R. Plastid transformation and its application in metabolic engineering. *Curr. Opin. Biotechnol.* **49**, 10–15 (2018).
- Sakuma, T. et al. Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci. Rep.* **3**, 3379 (2013).
- Cerutti, H., Osman, M., Grandoni, P. & Jagendorf, A. T. A homolog of *Escherichia coli* RecA protein in plastids of higher plants. *Proc. Natl Acad. Sci. USA* **89**, 8068–8072 (1992).
- Cao, J., Combs, C. & Jagendorf, A. T. The chloroplast-located homolog of bacterial DNA recombinase. *Plant Cell Physiol.* **38**, 1319–1325 (1997).
- 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).
- Arimura, S. et al. Targeted gene disruption of ATP synthases 6-1 and 6-2 in the mitochondrial genome of *Arabidopsis thaliana* by mitoTALENs. *Plant J.* **104**, 1459–1471 (2020).
- Kazama, T. et al. Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. *Nat. Plants* **5**, 722–730 (2019).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).

- Miyazaki, K. & Kitahara, K. Functional metagenomic approach to identify overlooked antibiotic resistance mutations in bacterial rRNA. *Sci. Rep.* **8**, 5179 (2018).
- D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic resistome. *Science* **311**, 374–377 (2006).
- Allison, L. A., Simon, L. D. & Maliga, P. Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J.* **15**, 2802–2809 (1996).
- Chateigner-Boutin, A.-L. et al. OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J.* **65**, 532–542 (2011).
- Baena-González, E. et al. Deletion of the tobacco plastid *psbA* gene triggers an upregulation of the thylakoid-associated NAD(P)H dehydrogenase complex and the plastid terminal oxidase (PTOX). *Plant J.* **35**, 704–716 (2003).
- Schult, K. et al. The nuclear-encoded factor HCF173 is involved in the initiation of translation of the *psbA* mRNA in *Arabidopsis thaliana*. *Plant Cell* **19**, 1329–1346 (2007).
- Lutz, K. A. & Maliga, P. Plastid genomes in a regenerating tobacco shoot derive from a small number of copies selected through a stochastic process. *Plant J.* **56**, 975–983 (2008).
- Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **46**, 85–94 (2006).
- Weijers, D. et al. An *Arabidopsis* minute-like phenotype caused by a semi-dominant mutation in a *RIBOSOMAL PROTEIN S5* gene. *Development* **128**, 4289–4299 (2001).
- Tsutsui, H. & Higashiyama, T. pKAMA-ITACHI vectors for highly efficient CRISPR/Cas9-mediated gene knockout in *Arabidopsis thaliana*. *Plant Cell Physiol.* **58**, 46–56 (2017).
- Shimada, T. L., Shimada, T. & Hara-Nishimura, I. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *Plant J.* **61**, 519–528 (2010).
- Moazed, D. & Noller, H. F. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* **327**, 389–394 (1987).
- Svab, Z. & Maliga, P. Mutation proximal to the tRNA binding region of the *Nicotiana* plastid 16S rRNA confers resistance to spectinomycin. *Mol. Gen. Genet.* **228**, 316–319 (1991).
- Filipenko, E. A., Sidorchuk, Y. V. & Deineko, E. V. Spontaneous spectinomycin resistance mutations of the chloroplast *rrn16* gene in *Daucus carota* callus lines. *Russian J. Genet.* **47**, 35–40 (2011).
- Zoschke, R. & Bock, R. Chloroplast translation: structural and functional organization, operational control, and regulation. *Plant Cell* **30**, 745–770 (2018).
- Lee, H. et al. Mitochondrial DNA editing in mice with DddA-TALE fusion deaminases. *Nat. Commun.* **12**, 1190 (2021).
- Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **209**, 1747–1756 (2016).
- Sikdar, S. R., Serino, G., Chaudhuri, S. & Maliga, P. Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* **18**, 20–24 (1998).
- Tan, J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun.* **10**, 439 (2019).
- Orr, D. J. et al. Surveying Rubisco diversity and temperature response to improve crop photosynthetic efficiency. *Plant Physiol.* **172**, 707–717 (2016).
- Sharwood, R. E., Ghannoum, O., Kapralov, M. V., Gunn, L. H. & Whitney, S. M. Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C3 photosynthesis. *Nat. Plants* **2**, 16186 (2016).
- Oettmeier, W. Herbicide resistance and supersensitivity in photosystem II. *Cell. Mol. Life Sci.* **55**, 1255–1277 (1999).
- Nagaya, S., Kawamura, K., Shinmyo, A. & Kato, K. The HSP terminator of *Arabidopsis thaliana* increases gene expression in plant cells. *Plant Cell Physiol.* **51**, 328–332 (2010).
- Karimi, M., Inzé, D. & Depicker, A. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195 (2002).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Porra, R. J., Thompson, W. A. & Kriedemann, P. E. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta Bioenerg.* **975**, 384–394 (1989).

## Acknowledgements

This research was supported partly by the University of Tokyo GAP fund programme to S.A. and the Japan Society for the Promotion of Science (grant number 20H00417 to N.T. and 16H06279 (PAGS), 19H02927 and 19KK0391 to S.A.).

### Author contributions

I.N., N.T. and S.A. initiated and designed the project. I.N. constructed the vectors and performed the main part of this research. Y.T. carried out DNA isolations, PCRs and Sanger sequences. H.Y. and T.S. carried out phenotyping of *psbA* mutants. M.O., T.I. and S.A. analysed next-generation sequencing data. I.S., H.T. and S.A. wrote the manuscript.

### Competing interests

A patent related to the method described in this paper is pending in Japan (no. 2021-9001) and is in preparation for submission to countries in the Patent Cooperation Treaty. The patent is held by the University of Tokyo.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41477-021-00954-6>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41477-021-00954-6>.

**Correspondence and requests for materials** should be addressed to S.-i.A.

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

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

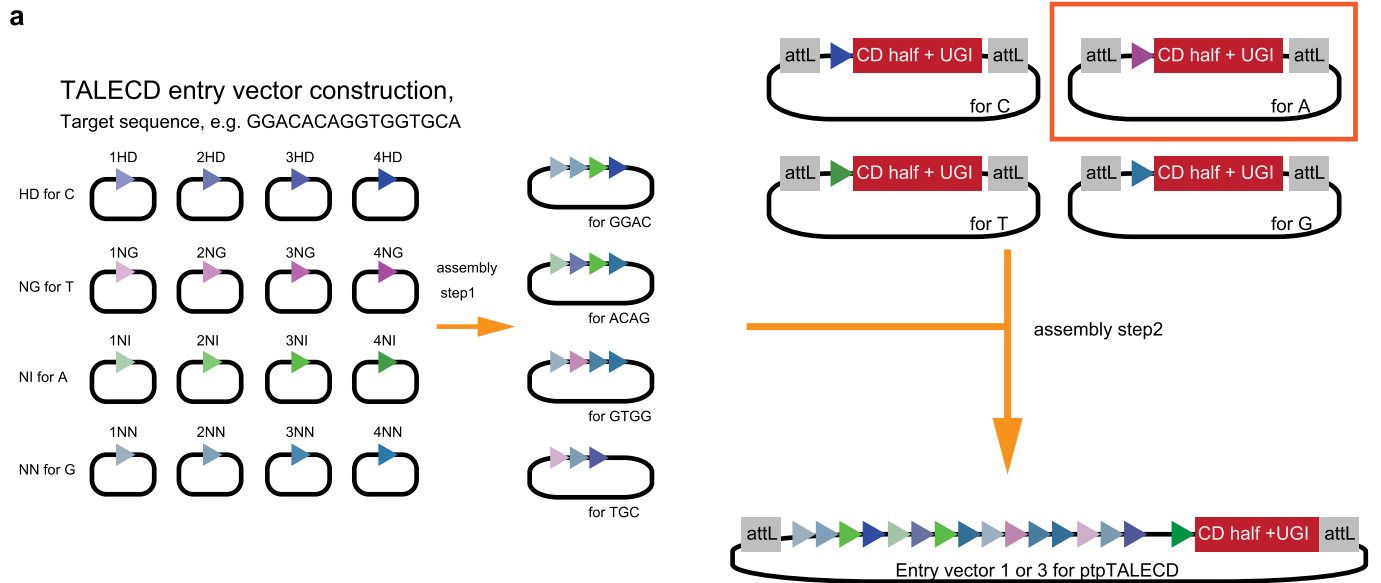
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



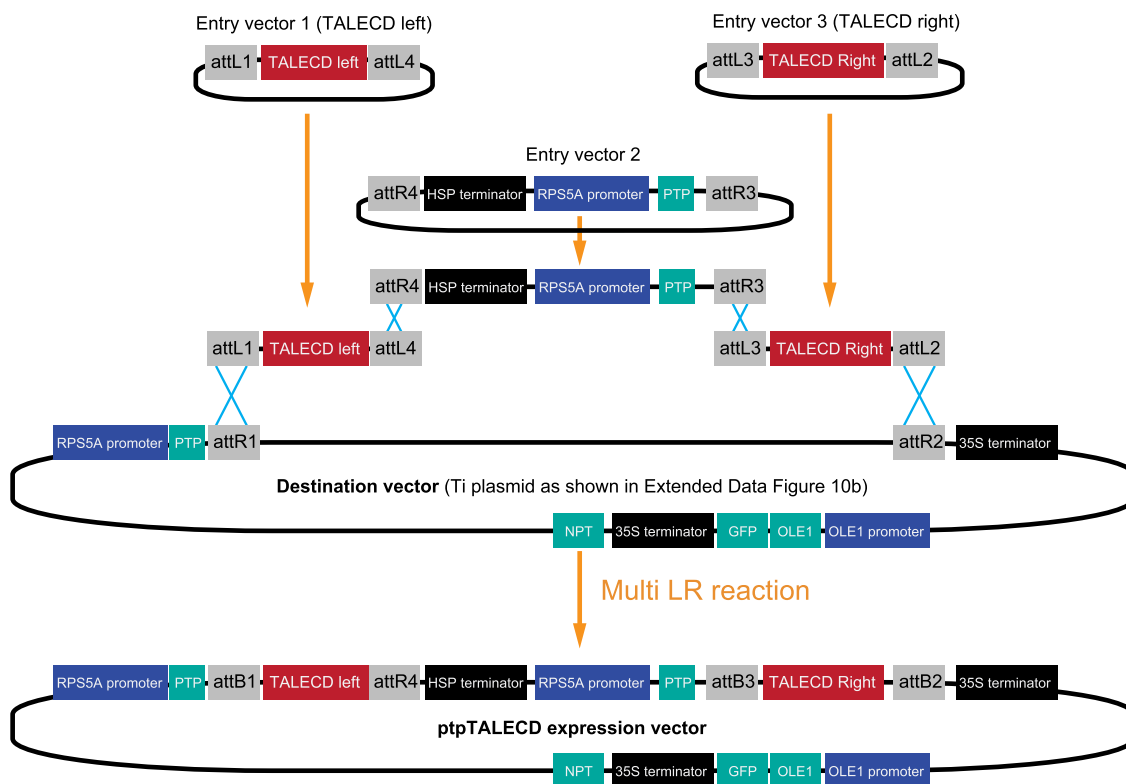
**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021





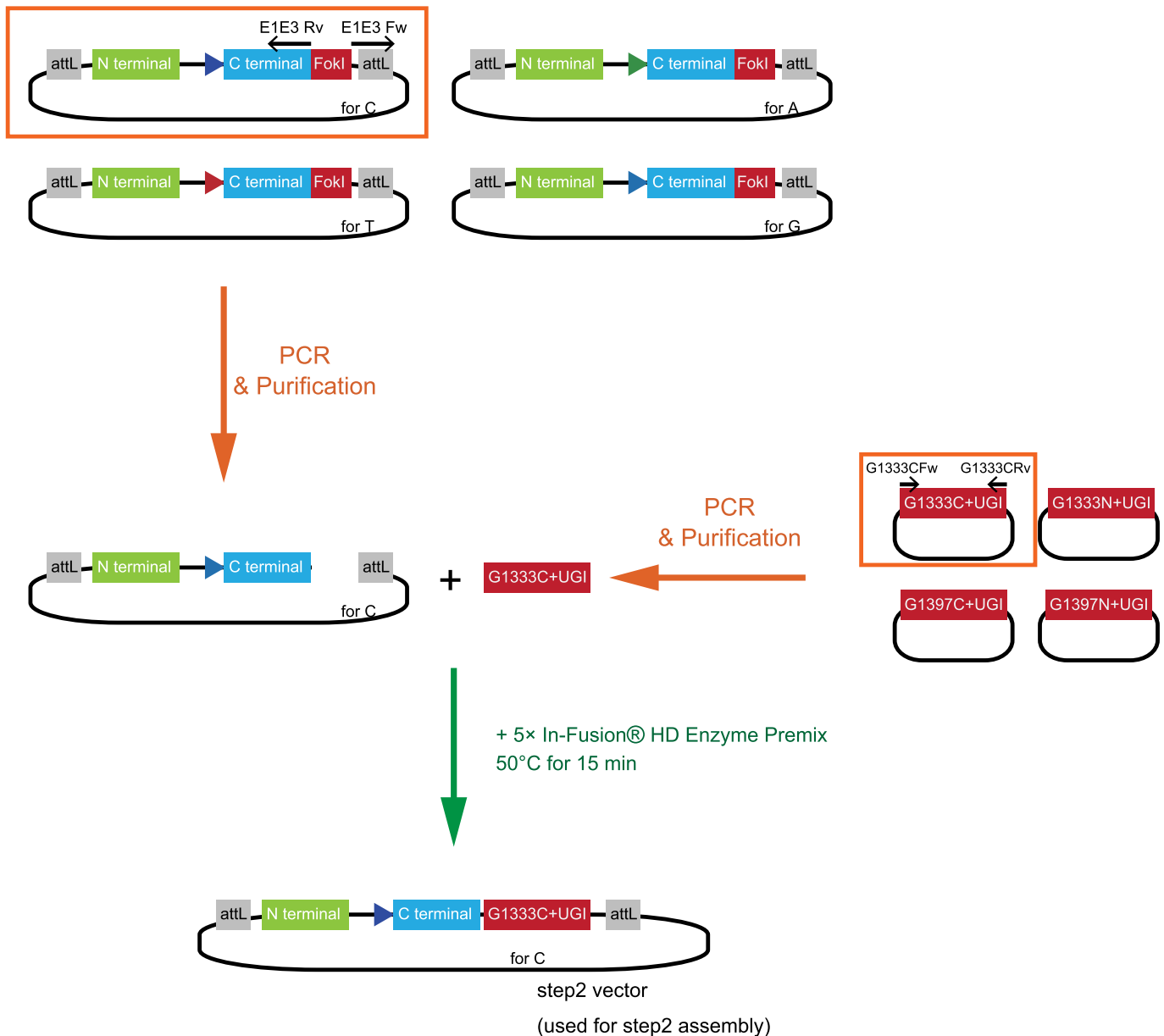
**b**



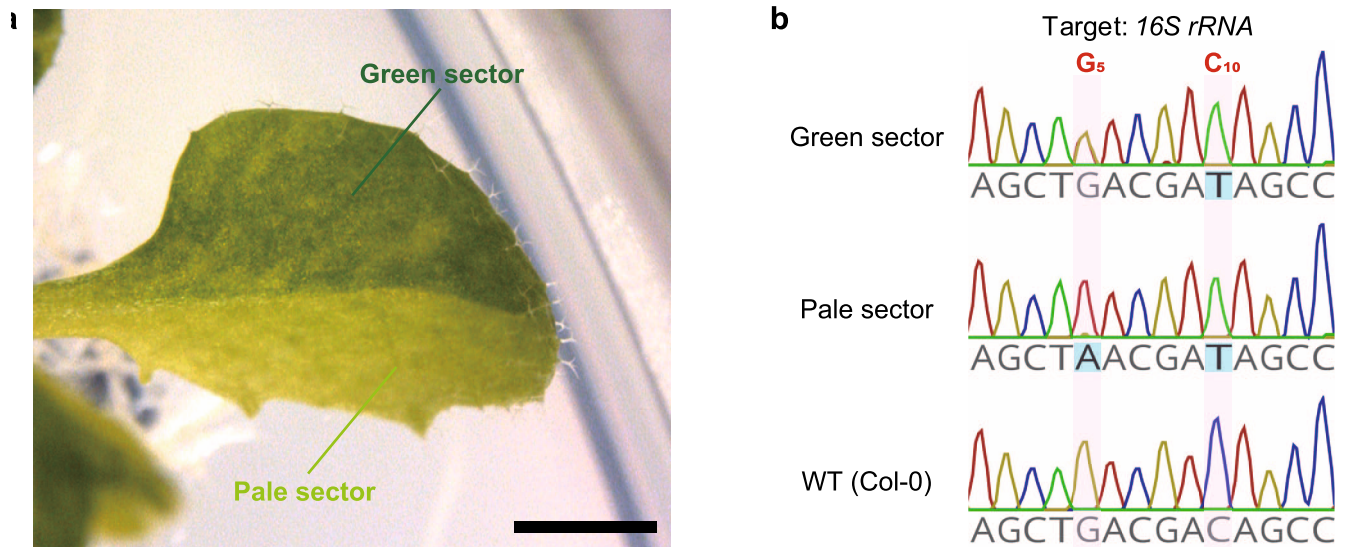
**Extended Data Fig. 1 | Schematic image of constructing plastid-targeting pTALECD (ptpTALECD) expression vectors. a**, Assembly steps for constructing pTALECD ORFs. Platinum TALEN Kit was used but entry vectors for step2 were those generated as Extended Data Fig. 10a shows. **b**, The ptpTALECD expression vectors were constructed with LR Clonase™ II Plus enzyme (Thermo Fisher Scientific). NPT: kanamycin resistance gene.

## Replace FokI by the CD half in step2 vectors

e.g. FokI to G1333C replacement in the entry vector for C



**Extended Data Fig. 2 | Replacement of the FokI coding sequence in entry vectors for step2 assembly by the cytidine deaminase (CD) half coding sequence.** Artificially synthesized CD half coding sequence (Supplementary Table 4) and entry vectors for step2 assembly containing FokI coding sequence were amplified with primes corresponding to the template (Supplementary Table 5). The purified PCR products were mixed with 5x In-Fusion HD Cloning Enzyme Premix (TaKaRa) and incubated at 50 °C for 15 minutes.



**Extended Data Fig. 3 | A chimerically base-edited leaf.** A differently coloured leaf of 23 DAS 16S rRNA 1397NC 3 (bar = 2 mm) plantlet (**a**) and its genotype at the ptpTALECD targeting window (**b**). The phenotype 'half-pale variegation' was observed in only one leaf on a T<sub>1</sub> plant of the line 16S rRNA 1397NC 3.



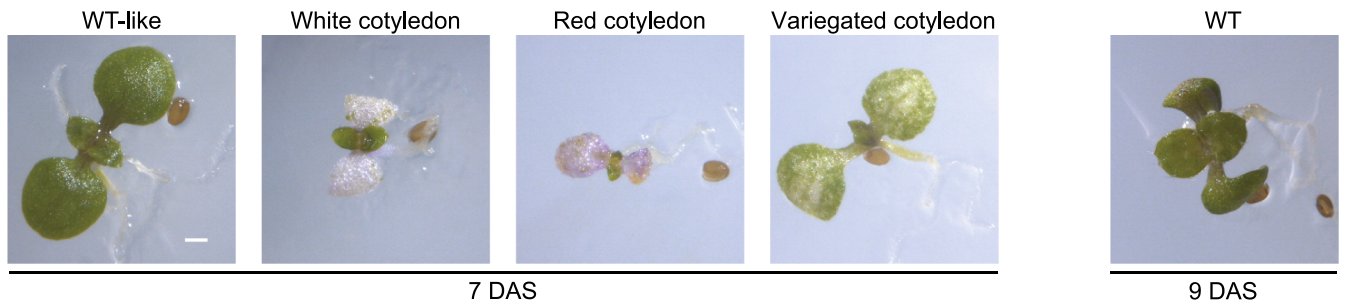
**Extended Data Fig. 4 | Appearance and genotypes of all T<sub>1</sub> plants at 11 and 23 DAS in which C/Gs were substituted on at least one of the two days.** The degrees of substitution at G<sub>5</sub>, G<sub>6</sub>, and C<sub>10</sub> (16S rRNA, **a**), G<sub>3</sub> and C<sub>6</sub> (rpoC1, **b**), and C<sub>10</sub> (psbA, **c**) are shown in magenta on the top of the images. All genotyped T<sub>1</sub> plants are shown. **d**, Col-0 representative of ≥ four plants. Genotypes are shown in Supplementary Table 1. Bars = 1 mm (white bars on 11 DAS), and 5 mm (a black bar on 11 DAS and all bars on 23 DAS).

**a** Genotypes and phenotypes of T<sub>2</sub> generation plants

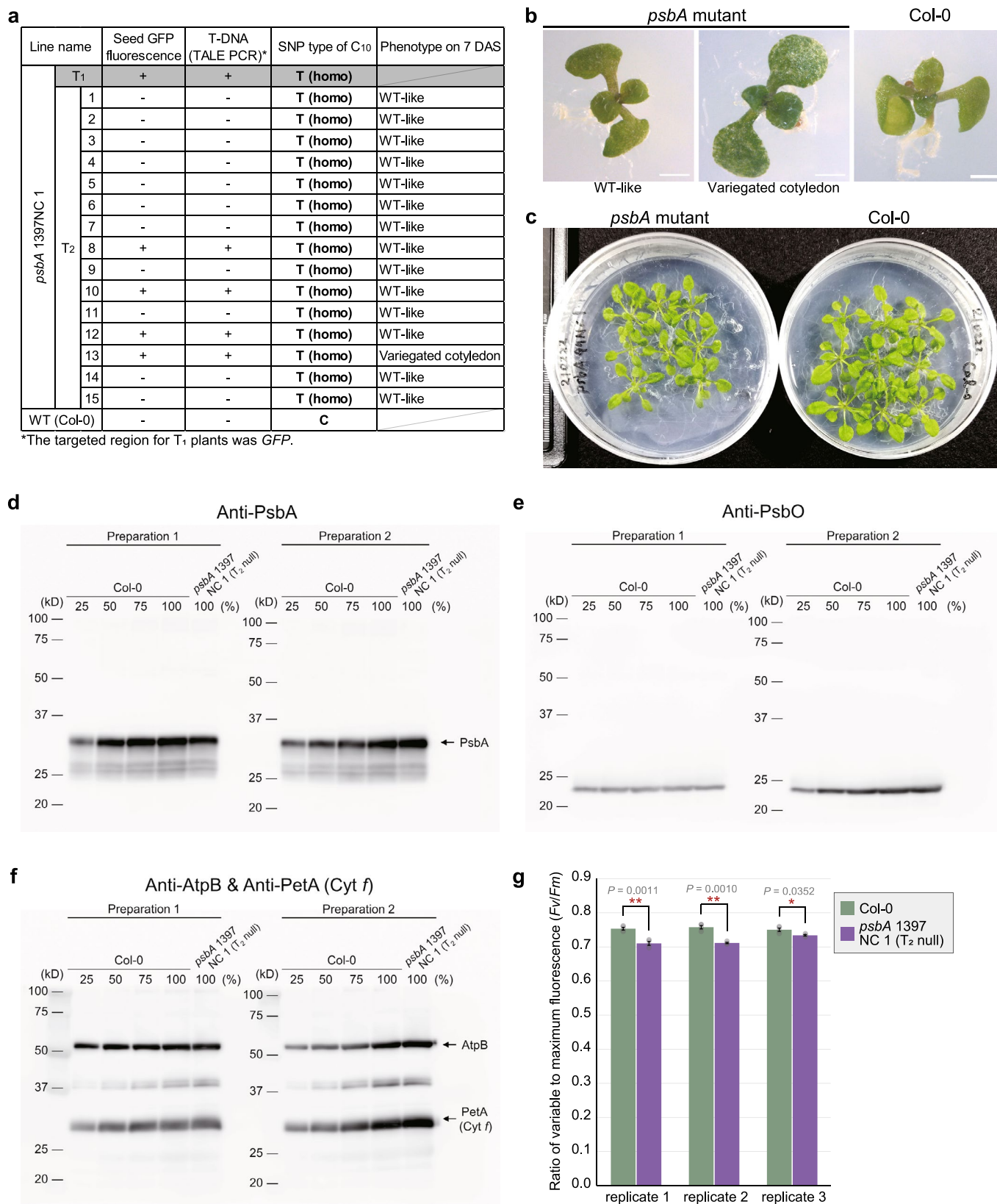
Line name	Seed GFP fluorescence	T-DNA* (TALE PCR)	SNP type of G <sub>5</sub>	G <sub>8</sub>	C <sub>10</sub>	Phenotype at 7 DAS	Note		
16S rRNA 1397CN 2	T <sub>1</sub>	+	+	A (homo)	G	C			
	T <sub>2</sub>	1	-	-	A (homo)	G	C	WT-like	
		2	+	+	A (homo)	A (h/c)	T (h/c)	White cotyledon	753 th G in 16S rRNA was h/c substituted to A
		3	+	+	A (homo)	G	C	WT-like	
		4	+	+	A (homo)	G	C	WT-like	
		5	-	-	A (homo)	G	C	WT-like	
		6	-	-	A (homo)	G	C	WT-like	
		7	+	+	A (homo)	A (h/c)	T (h/c)	Red cotyledon	753 th G in 16S rRNA was h/c substituted to A
		8	-	-	A (homo)	G	C	WT-like	
		9	+	+	A (homo)	A (h/c)	T (h/c)	Variegated cotyledon	753 th G in 16S rRNA was h/c substituted to A
10		+	+	A (homo)	G	C	Variegated cotyledon		
16S rRNA 1397CN 8	T <sub>1</sub>	+	+	A (homo)	G	C			
	T <sub>2</sub>	1	-	-	A (homo)	G	C	WT-like	
		2	-	-	A (homo)	G	C	WT-like	
		3	+	+	A (homo)	A (h/c)	T (h/c)	Variegated cotyledon	753 th G in 16S rRNA was h/c substituted to A
		4	+	+	A (homo)	A (h/c)	T (h/c)	Variegated cotyledon	753 th G in 16S rRNA was h/c substituted to A
		5	+	+	A (homo)	G	C	WT-like	
		6	+	+	A (homo)	A (h/c)	T (h/c)	Variegated cotyledon	753 th G in 16S rRNA was h/c substituted to A
		7	-	-	A (homo)	G	C	WT-like	
8		-	-	A (homo)	G	C	WT-like		
16S rRNA 1397NC 3	T <sub>1</sub>	+	+	h/c or A (homo)*	G	T (homo)		*mutation frequency at 5 th base was different between 10 and 22 DAS	
	T <sub>2</sub>	1	-	-	G	G	T (homo)	WT-like	
		2	+	+	A (homo)	G	T (homo)	Red cotyledon	
		3	-	-	G	G	T (homo)	WT-like	
		4	+	+	A (homo)	G	T (homo)	Red cotyledon	
		5	+	+	G	G	T (homo)	WT-like	
		6	-	-	G	G	T (homo)	WT-like	
		7	-	-	G	G	T (homo)	WT-like	
8		+	+	G	G	T (homo)	WT-like		
WT (Col-0)	-	-	G	G	C				

\*The targeted region for T<sub>1</sub> plants was *GFP*.

**b**



**Extended Data Fig. 5 | Genotypes and phenotypes of T<sub>2</sub> progeny of T<sub>1</sub> plants that had homoplasmic mutations in 16S rRNA. a**, Genotypes and phenotypes of T<sub>2</sub> plants obtained from 16S rRNA 1397CN 2, 8 and 1397NC 3 self-reproduction. **b**, Phenotype images of the T<sub>2</sub> plants listed in (a). Representative images of at least three plants, except for the white cotyledon for which only one plant was available (bar = 0.5 mm).



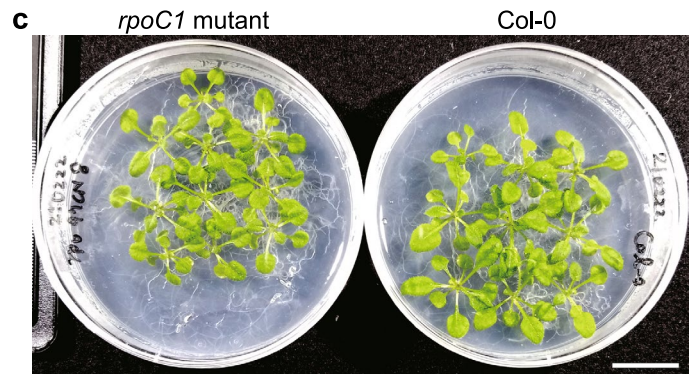
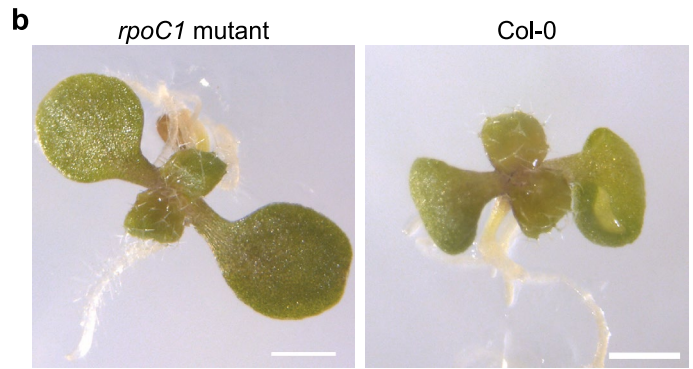
Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Genotypes and phenotypes of the progeny of *psbA* 1397NC 1.** **a**, Genotypes of T<sub>2</sub> progenies of *psbA* 1397NC 1. **b,c**, Representative phenotype images at 7 DAS of at least four plants except for variegated cotyledon for which only one plant was available (**b**) and 27 DAS (**c**). Bars = 1 mm in **b**, and diameter of plates = 9 cm in **c**. **d-f**, Immunoblot analysis of photosynthetic proteins in the thylakoid membrane of two independent experiments. Similar results were observed in each preparation. Labels on the lanes indicate the relative amounts of proteins loaded on the gel. 100 means proteins extracted from thylakoid membrane corresponding to 1 µg chlorophyll (**d**) or 2 µg of chlorophyll (**e** and **f**). **g**, Maximum quantum yield of PSII calculated as  $F_v/F_m$ . Chlorophyll fluorescence of four plants of each group (mutants and wild-type plants) were measured in each replicate. \* $P < 0.05$  and \*\* $P < 0.01$  (two-tailed Welch's test). Error bars represent standard errors.

**a**

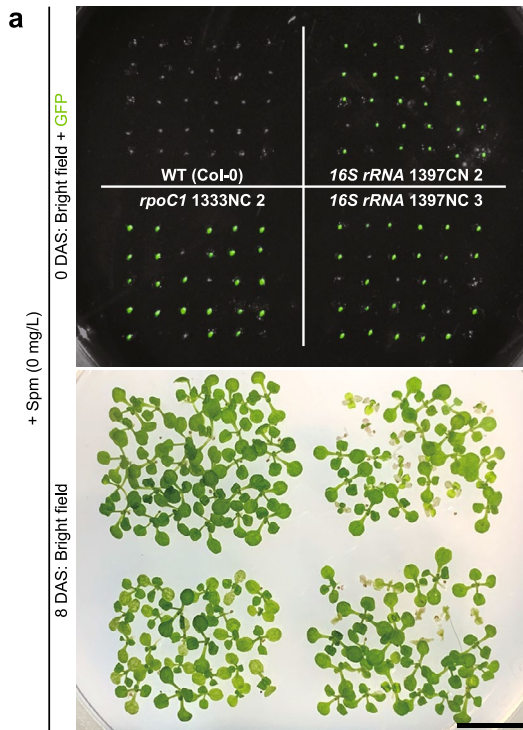
Line name	Seed GFP fluorescence	T-DNA (TALE PCR)*	SNP type of Gs	Phenotype on 7 DAS		
<i>rpoC1</i> 1397CN 8	T <sub>1</sub>	+	+	A (homo)		
	T <sub>2</sub>	1	+	+	A (homo)	WT-like
		2	+	+	A (homo)	WT-like
		3	-	-	A (homo)	WT-like
		4	+	+	A (homo)	WT-like
		5	-	-	A (homo)	WT-like
		6	-	-	A (homo)	WT-like
		7	+	+	A (homo)	WT-like
		8	-	-	A (homo)	WT-like
<i>rpoC1</i> 1397CN 9	T <sub>1</sub>	+	+	A (homo)		
	T <sub>2</sub>	1	+	+	A (homo)	WT-like
		2	+	+	A (homo)	WT-like
		3	+	+	A (homo)	WT-like
		4	+	+	A (homo)	WT-like
		5	+	+	A (homo)	WT-like
		6	+	+	A (homo)	WT-like
		7	+	+	A (homo)	WT-like
		8	+	+	A (homo)	WT-like
<i>rpoC1</i> 1397CN 13	T <sub>1</sub>	+	+	A (homo)		
	T <sub>2</sub>	1	+	+	A (homo)	WT-like
		2	+	+	A (homo)	WT-like
		3	+	+	A (homo)	WT-like
		4	+	+	A (homo)	WT-like
		5	+	+	A (homo)	WT-like
		6	+	+	A (homo)	WT-like
		7	+	+	A (homo)	WT-like
WT (Col-0)	-	-	G			

\*The targeted region for T<sub>1</sub> plants was *GFP*.



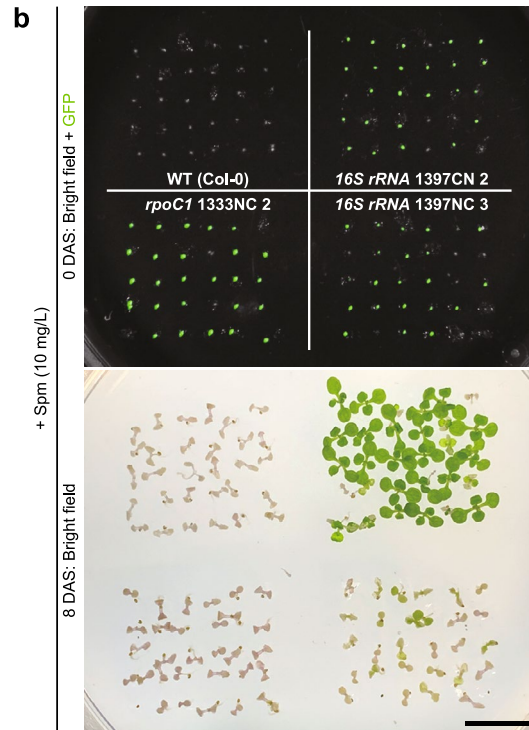
**Extended Data Fig. 7 | Genotypes and phenotypes of the progeny of *rpoC1* 1397CN 8, 9, and 13.** **a**, Genotypes of T<sub>2</sub> progenies of *rpoC1* 1397CN 8, 9, and 13. **b,c**, Representative phenotype images on 7 DAS (**b**, representative of ≥ four plants) and 27 DAS (**c**). Bars = 1 mm in **b**, and 2 cm in **c**.





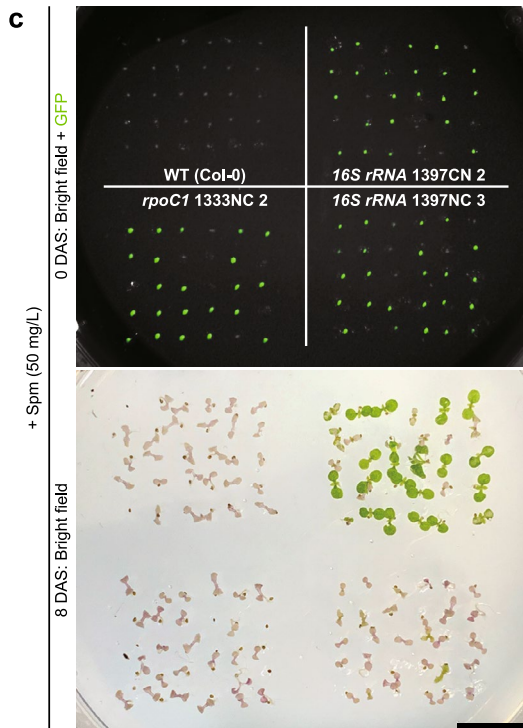
Spm: 0 mg/L

Line name	G <sub>s</sub> SNP of T <sub>1</sub>	GFP positive seeds (T <sub>2</sub> )			GFP negative seeds (T <sub>2</sub> )		
		Green	White	n.g.	Green	White	n.g.
WT (Col-0)	G	0	0	0	30	0	0
16S rRNA 1397CN 2	A (homo)	14	12	1	3	0	0
16S rRNA 1397NC 3	*	17	6	0	6	1	0
rpoC1 1333NC 2	G	24	0	0	3	0	2



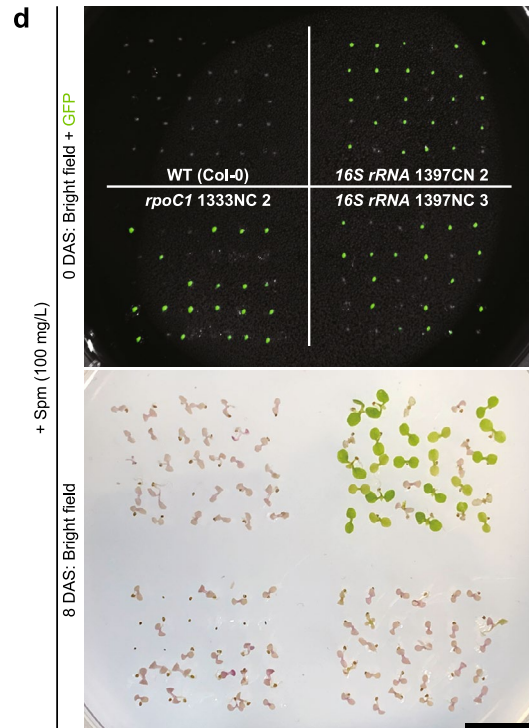
Spm: 10 mg/L

Line name	G <sub>s</sub> SNP of T <sub>1</sub>	GFP positive seeds (T <sub>2</sub> )			GFP negative seeds (T <sub>2</sub> )		
		Green	White	n.g.	Green	White	n.g.
WT (Col-0)	G	0	0	0	0	30	0
16S rRNA 1397CN 2	A (homo)	14	10	0	7	0	0
16S rRNA 1397NC 3	*	0	22	0	0	6	2
rpoC1 1333NC 2	G	0	27	0	0	3	1



Spm: 50 mg/L

Line name	G <sub>s</sub> SNP of T <sub>1</sub>	GFP positive seeds (T <sub>2</sub> )			GFP negative seeds (T <sub>2</sub> )		
		Green	White	n.g.	Green	White	n.g.
WT (Col-0)	G	0	0	0	0	28	2
16S rRNA 1397CN 2	A (homo)	11	11	0	8	0	0
16S rRNA 1397NC 3	*	0	25	0	0	5	0
rpoC1 1333NC 2	G	0	22	0	0	3	5


























Spm: 100 mg/L

Line name	G <sub>s</sub> SNP of T <sub>1</sub>	GFP positive seeds (T <sub>2</sub> )			GFP negative seeds (T <sub>2</sub> )		
		Green	White	n.g.	Green	White	n.g.
WT (Col-0)	G	0	0	0	0	29	1
16S rRNA 1397CN 2	A (homo)	13	9	0	7	0	0
16S rRNA 1397NC 3	*	0	17	0	0	13	0
rpoC1 1333NC 2	G	0	21	0	0	0	9

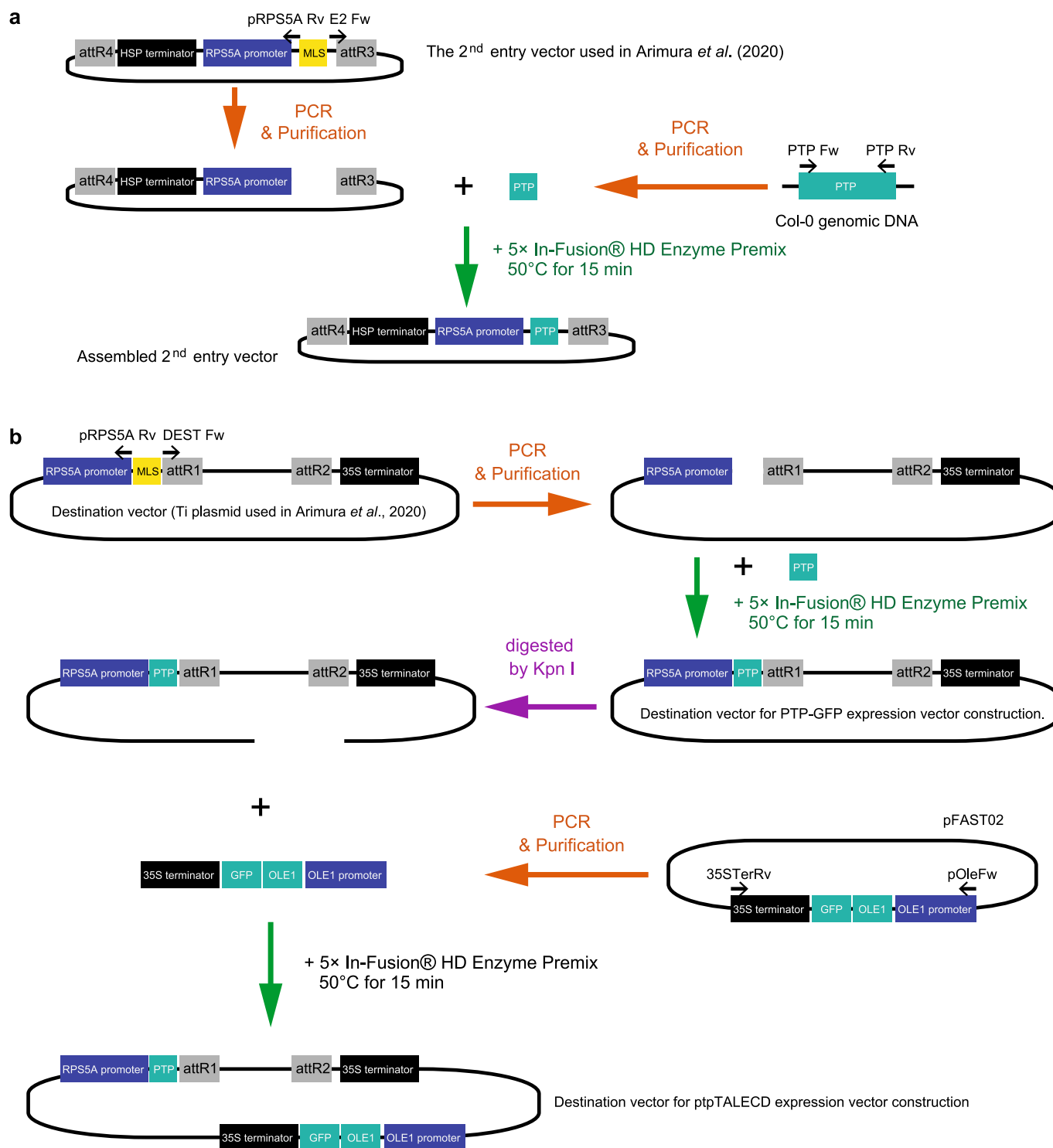
Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Spectinomycin (Spm) resistance of T<sub>2</sub> plants. a-d.** Seeds and 8 DAS seedlings images of Col-0 (upper left) and T<sub>2</sub> progenies of *16S rRNA* 1397CN 2 (upper right), 1397NC 3 (lower right), and *rpoC1* 1333NC 2 (lower left) on 1/2 MS medium containing 0 (**a**), 10 (**b**), 50 (**c**), or 100 (**d**) mg/L Spm are shown. We show merged images of a seeds image and a seeds GFP fluorescence image. Bars = 1 cm. Tables show the relationship between seed GFP and cotyledons colour on 8 DAS. n.g.: not germinated. \*means that T<sub>1</sub> G<sub>s</sub> SNP was different between 11 and 23 DAS (Supplementary Table 1a).

Line name	Seed GFP fluorescence	SNP at G <sub>5</sub>	Phenotype at 13 DAS	
16S rRNA 1397CN 15	1	+	A (homo) 	Spm <sup>r</sup>
	2	+	A (homo) 	Spm <sup>r</sup>
	3	+	A/G 	Spm <sup>r</sup>
	4	+	A/G 	Spm <sup>r</sup>
	5	+	A (homo) 	Spm <sup>r</sup>
	6	+	A (homo) 	Spm <sup>r</sup>
	7	+	A/G 	Spm <sup>r</sup>
	8	+	A/G 	Spm <sup>r</sup>
	9	+	A/G 	Spm <sup>r</sup>
	10	+	A/G 	Spm <sup>r</sup>
	11	+	A/G 	Spm <sup>r</sup>
	12	+	A/G 	Spm <sup>r</sup>
	13	+	A/G 	Spm <sup>r</sup>
	14	+	A/G 	Spm <sup>r</sup>
	15	+	A/G 	Spm <sup>r</sup>
	16	+	A/G 	Spm <sup>r</sup>
	17	+	A (homo) 	Spm <sup>r</sup>
	18	+	A/G 	Spm <sup>r</sup>
16S rRNA 1397CN 2*	21	+	A (homo)** 	W/G
	22	+	A (homo)** 	W/G
	23	+	A (homo)** 	W/G
	24	+	A (homo)** 	Spm <sup>s</sup> (died)
	26	+	A (homo)** 	Spm <sup>s</sup> (died)
WT (Col-0)	-	G (homo)	Spm <sup>s</sup> (died)	

\*16S rRNA 1397CN 2\_25 was lost when sampled. \*\*G<sub>8</sub>, C<sub>10</sub>, and other bases was also substituted in 16S rRNA. Spm<sup>r</sup>; Spm resistant, Spm<sup>s</sup>; Spm sensitive-like, W/G; white cotyledon with green leaves.

**Extended Data Fig. 9 | Cotyledon genotypes of 13 DAS Spm<sup>r</sup> and Spm<sup>s</sup>-like plants.** The presence of seed GFP, SNP at G<sub>5</sub>, and phenotypes of 13 DAS Spm<sup>r</sup> plants (T<sub>2</sub> of 16S rRNA 1397CN 15) and Spm<sup>s</sup>-like plants (T<sub>2</sub> of 16S rRNA 1397CN 2) in Fig. 3c are shown.



**Extended Data Fig. 10 | The 2<sup>nd</sup> entry vector and destination vector construction.** **a**, The 2<sup>nd</sup> entry vector construction. The 2<sup>nd</sup> entry vector used in Arimura *et al.*<sup>10</sup>, and RECA1 plastid transit peptide coding sequence (PTP; Supplementary Table 4) were amplified with primers corresponding to the DNA template (Supplementary Table 5). The purified PCR products (1: PTP + 2<sup>nd</sup> entry vector, 2: PTP + destination vector) were mixed with 5x In-Fusion HD Cloning Enzyme Premix (TaKaRa) and incubated at 50 °C for 15 minutes. **b**, The destination vector construction. The destination vector used in Arimura *et al.* (2020) was amplified with primers (Supplementary Table 5); purified PCR product was mixed with the PTP PCR product and 5x In-Fusion HD Cloning Enzyme Premix and incubated at 50 °C for 15 minutes to make the destination vector for PTP-GFP expression vector construction. The assembled destination vector was digested by KpnI and the purified product was incubated at 50 °C for 15 minutes after mixed with 5x In-Fusion HD Cloning Enzyme Premix and OLE1-GFP coding sequence amplified from pFAST02 (INPLANTA INNOVATIONS INC), to make the destination vector for ptpTALECD expression vector construction.

## 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 our [Editorial Policies](#) 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

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 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 Sanger sequencings were performed by Eurofins Genomics. Whole genome sequencing was performed by Macrogen Japan using an Illumina Novaseq platform. No software was used for data collection.

Data analysis Platanus\_trim v1.0.7, Geneious prime (v 2020.2.2), and BWA (v 0.7.12) were used to analyze sequencing data. Adobe Photoshop 2021 and Adobe Illustrator 2021 were used to prepare figures and tables.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

All data supporting the findings of this study are available in the article and the supplementary information. Vector sequences will be deposited in GenBank and addgene. Accession numbers of the reference sequences for organelle genome used in this study are AP000423.1 (plastid) and BK010421.1 (mitochondrion).

## 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](https://www.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	At least seven T1 plants for each targeted gene were analyzed. At least eight T2 progeny of three T1 lines were genotyped. To investigate spectinomycin resistance, at least 30 seeds per line were sown on one plate. The number of T1 plants used was determined by the maximum number that could be generated in our growth cabinets. The numbers of T2 seeds used was determined by the the maximum number of seeds (T2) available at the time of the experiment.
Data exclusions	No data exclusion.
Replication	For each targeted gene, at least seven T1 lines were genotyped and targeted bases were successfully substituted in at least three lines. Total DNA NGS was successfully performed on 17 T1 lines, 3 T2 plants and 3 wild-type plants. At least eight T2 progeny of the seven T1 lines were genotyped and all data are shown. Spectinomycin resistance was assayed four times, with the same results. Chlorophyll fluorescence measurements and Immunoblot experiments were successfully done in triplicate and duplicated respectively.
Randomization	All T1 plants were subjected to Sanger sequencing. The samples that were subjected to total DNA NGS were chosen randomly among the plants that appeared to have a homoplasmic mutation at the targeted base based on Sanger sequencing. T2 seeds used for genotyping were selected randomly among those whose T1 parents had a homoplasmic mutation in the target window. T2 seeds used for spectinomycin resistance assay were selected randomly among those whose T1 parents had a homoplasmic mutation at G5 in 16S rRNA. T2 progeny of psbA 1397NC 1 that were subjected to chlorophyll measurements and immunoblot experiments were selected randomly among the null segregants of T-DNA.
Blinding	Blinding was not required because all analyses including genotyping by sequencing and phenotyping by antibiotics and molecular methods could be carried out without making any subjective judgements.

## 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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	PsbA (AS11 1786, Agrisera); AtpB (AS05 085, Agrisera); PsbO (Murakami et al., FEBS Lett., 523: 138-142, 2002); Cyt f (Hidema et al., Plant Physiol., 97: 1287-1293, 1991).
Validation	Antibodies against PsbA, AtpB, PsbO or Cyt f can detect Arabidopsis proteins.