



Genome editing in plants using the compact editor CasΦ

Zheng Li^a, Zhenhui Zhong^a, Zhongshou Wu^a, Patrick Pausch^b, Basem Al-Shayeb^{c,d}, Jasmine Amerasekera^e, Jennifer A. Doudna^{c,f,g,h,i,j,k} and Steven E. Jacobsen^{a,l}

Contributed by Steven E. Jacobsen; received October 1, 2022; accepted December 21, 2022; reviewed by Xiaoqi Feng and Feng Zhang

Clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins (CRISPR-Cas) systems have been developed as important tools for plant genome engineering. Here, we demonstrate that the hypercompact CasΦ nuclease is able to generate stably inherited gene edits in *Arabidopsis*, and that CasΦ guide RNAs can be expressed with either the Pol-III U6 promoter or a Pol-II promoter together with ribozyme mediated RNA processing. Using the *Arabidopsis fwa* epiallele, we show that CasΦ displays higher editing efficiency when the target locus is not DNA methylated, suggesting that CasΦ is sensitive to chromatin environment. Importantly, two CasΦ protein variants, vCasΦ and nCasΦ, both showed much higher editing efficiency relative to the wild-type CasΦ enzyme. Consistently, vCasΦ and nCasΦ yielded offspring plants with inherited edits at much higher rates compared to WTCasΦ. Extensive genomic analysis of gene edited plants showed no off-target editing, suggesting that CasΦ is highly specific. The hypercompact size, T-rich minimal protospacer adjacent motif (PAM), and wide range of working temperatures make CasΦ an excellent supplement to existing plant genome editing systems.

CRISPR-CasΦ | plant genome editing | DNA methylation | CasΦ variants | off-target editing

CRISPR-Cas systems, originally discovered as an adaptive immune system in bacteria and archaea (1–3), have been widely used as genome engineering tools (4–6). Class 2 CRISPR-Cas systems are of particular interest as the guide RNA binding and the cleavage of target nucleic acids are accomplished by a single effector protein (7). For plant genome engineering, Cas9 (2, 8–12) and Cas12a (13–17) have been routinely applied in multiple plant species. With the discovery of novel CRISPR-Cas systems, additional DNA and RNA targeting Cas proteins serve as potential plant genome engineering tools. CRISPR-CasΦ, discovered from huge bacteriophages (18), targets double-stranded DNA and generates staggered cuts (19). Importantly, CasΦ proteins are only 700 to 800 amino acids (aa) (19), which are much smaller in size compared to Cas9 (1,000 to 1,400 aa) (2, 20, 21) and Cas12a (1,100 to 1,300 aa) (13). The compact sizes of the CRISPR-CasΦ systems may allow for approaches where protein or nucleic acid size is a limiting factor. CasΦ systems also have T-rich minimal PAM sequences (19). Huge bacteriophages encoding CasΦ systems are from diverse ecosystems potentially providing a range of optimum temperatures for CasΦ activity (18). CasΦ proteins are therefore interesting candidates for novel plant genome engineering tools.

The CasΦ-2 ortholog was previously shown to be capable of target DNA editing in both human and plant cells (19). It recognizes a 5'-TBN-3' PAM sequence (where B is G, T or C) and generates staggered 5' overhangs which usually yield multiple base pair deletions after the action of cellular DNA repair machineries (19). Similar to Cas12a, the CasΦ-2 protein is also able to process pre-crRNA into mature crRNAs (19). The CasΦ-2 protein employs a single RuvC active site for DNA cleavage, which is also used for pre-crRNA processing (19, 22). A structural study of CasΦ-2 revealed that helix α7 blocks the path of the target strand of the substrate DNA toward a PAM-proximal position (22). Mutations in the negatively charged tip of helix α7 either by substitution of the whole negatively charged tip to a linker sequence glycine-serine-serine-glycine (GSSG) (vCasΦ) or substitution of residues in the negatively charged tip to alanine (E159A, S160A, S164A, D167A, E168A) (nCasΦ) resulted in significantly faster substrate DNA cleavage compared to the wild-type (WT) CasΦ (22). This finding provides potentially more efficient CasΦ variants for genome engineering purposes.

In this study, we provide evidence that CasΦ can be utilized as a novel plant genome editing tool to generate heritable mutations. The CasΦ genome editing system is compatible with Pol-II promoter-driven guide RNA transcription and ribozyme mediated guide RNA processing machineries. We also found that CasΦ editing is more efficient at unmethylated DNA than at methylated DNA. Finally, the hyperactive CasΦ variants vCasΦ and nCasΦ exhibited much higher editing efficiency with all of the gRNAs tested, and off-target editing of the CasΦ variants vCasΦ and nCasΦ was not

Significance

Plant genome engineering with Clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins (CRISPR-Cas) systems is frequently used in both research and agriculture. Here, we demonstrate that the hypercompact CasΦ-2 nuclease is able to generate heritable gene edits in *Arabidopsis*. Two CasΦ protein variants vCasΦ and nCasΦ increased the editing efficiency in plants and yielded more offspring plants with heritable edits. CasΦ also has a wide range of working temperatures, and the editing by CasΦ is highly specific. We also observed that editing by CasΦ is sensitive to chromatin environment. The hypercompact size, T-rich minimal protospacer adjacent motif (PAM) and wide range of working temperatures make CasΦ an excellent supplement to existing plant genome editing systems.

Author contributions: Z.L., P.P., B.A.-S., J.A.D., and S.E.J. designed research; Z.L., Z.W., and J.A. performed research; Z.L. and P.P. contributed new reagents/analytical tools; Z.L. and Z.Z. analyzed data; and Z.L., Z.Z., Z.W., P.P., and S.E.J. wrote the paper.

Reviewers: X.F., John Innes Centre; and F.Z., University of Minnesota Twin Cities.

Competing interest statement: The authors declare competing interest. The authors have patent filings to disclose, S.E.J., Z.L., J.A.D., P.P., and B.A.-S. are inventors on a published patent WO2021216512A1 from The Regents of the University of California, which includes applications of CasΦ systems in plants in this study. J.A.D., P.P., S.E.J., and Z.L. are inventors on International Application Number PCT/US22/13539, by The Regents of the University of California, which includes the applications of CasΦ variants in plants in this study.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹To whom correspondence may be addressed. Email: jacobsen@ucla.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2216822120/-DCSupplemental>.

Published January 18, 2023.

observed, suggesting that these variants can be utilized for highly specific genome editing in plants.

Results

CasΦ-2 Is Capable of Generating Heritable Mutations in *Arabidopsis*. Previously, it was shown that CasΦ-2 ribonucleoproteins (RNPs) were capable of editing the *AtPDS3* gene in *Arabidopsis* mesophyll protoplasts (19). To investigate if CasΦ-2 is able to edit a target gene in transgenic plants, we used the *Arabidopsis* *UBQ10* promoter to drive the expression of an *Arabidopsis* codon optimized CasΦ-2 and the U6 promoter to drive transcription of guide RNAs (Fig. 1A). The ubiquitous expression by the *UBQ10* promoter allows for assessment of editing efficiency in somatic tissues of transgenic plants as well as using mesophyll protoplasts for further optimization. We utilized *AtPDS3* gRNA10, which had the highest editing efficiency of all *AtPDS3* gRNAs tested in protoplasts (19). Version 1 and version 2 constructs were made with nuclear localization signal peptides either flanking N and C termini or only at the C terminus of the CasΦ-2 protein, respectively (Fig. 1A). Seventy-eight T1 transgenic plants of the version 1 construct were screened by Sanger sequencing, and a plant heterozygous for mutation of the gRNA10 targeted region was identified (T1 #33) (Fig. 1B). Further amplicon sequencing of different parts of this plant indicated that it was mosaic for the heterozygous mutation, with leaf a and leaf b showing about 50% of reads carrying mutation, but other leaves showing little to no editing (*SI Appendix, Table S1*). The dominant form of mutation detected in this plant by amplicon sequencing was a 6 bp deletion in the *AtPDS3* gRNA10 targeted region, although a small number of reads with other forms of deletion were also

detected (*SI Appendix, Table S1*). The *AtPDS3* gene encodes a phytoene desaturase enzyme that is essential for chloroplast development. Disruption of this gene function results in albino and dwarfed seedlings (23). Indeed, albino and dwarfed seedlings were observed from the offspring (T2) population of plant #33 (Fig. 1C). Twenty albino/dwarf seedlings were tested individually for the *AtPDS3* gRNA10 target region, and all of the tested seedlings were homozygous for a 6 bp deletion at the gRNA10 target region (Fig. 1D). This 6 bp deletion leads to the loss of two amino acids which are highly conserved among orthologous proteins (*SI Appendix, Fig. S1A*). In addition, two of these 20 albino/dwarf T2 seedlings had segregated away the transgene and were thus CasΦ-2 transgene-free (Fig. 1E), confirming the germline transmission (heritability) of the CasΦ-2 generated mutation in the *AtPDS3* gene from the T1 to the T2 plants.

For the version 2 construct, six T1 transgenic plants were screened, this time directly by amplicon sequencing so that we could more sensitively detect plants with low editing efficiency. T1 plant #6 showed the highest number of edits, with 13.5% of reads carrying mutations, while the other five plants showed much lower or no editing (*SI Appendix, Fig. S1B*). Ninety-six offspring T2 plants of T1 plant #6 were analyzed, and six of them were heterozygous for mutations at the *AtPDS3* gRNA10 targeted region (*SI Appendix, Fig. S1C*). As expected, albino seedlings were identified in the T3 offspring populations of these heterozygous T2 plants (*SI Appendix, Fig. S1D*). DNA sequencing of T3 albino seedlings derived from different T2 lineages showed deletions of 3 bp, 5 bp, 9 bp (two plants), or a 42 bp deletion plus a 1 bp deletion that was 16 bp upstream of the larger deletion (two plants) (*SI Appendix, Fig. S2A*). CasΦ-2 transgene-free albino seedlings were also identified from these T3 populations, further supporting the heritability of the

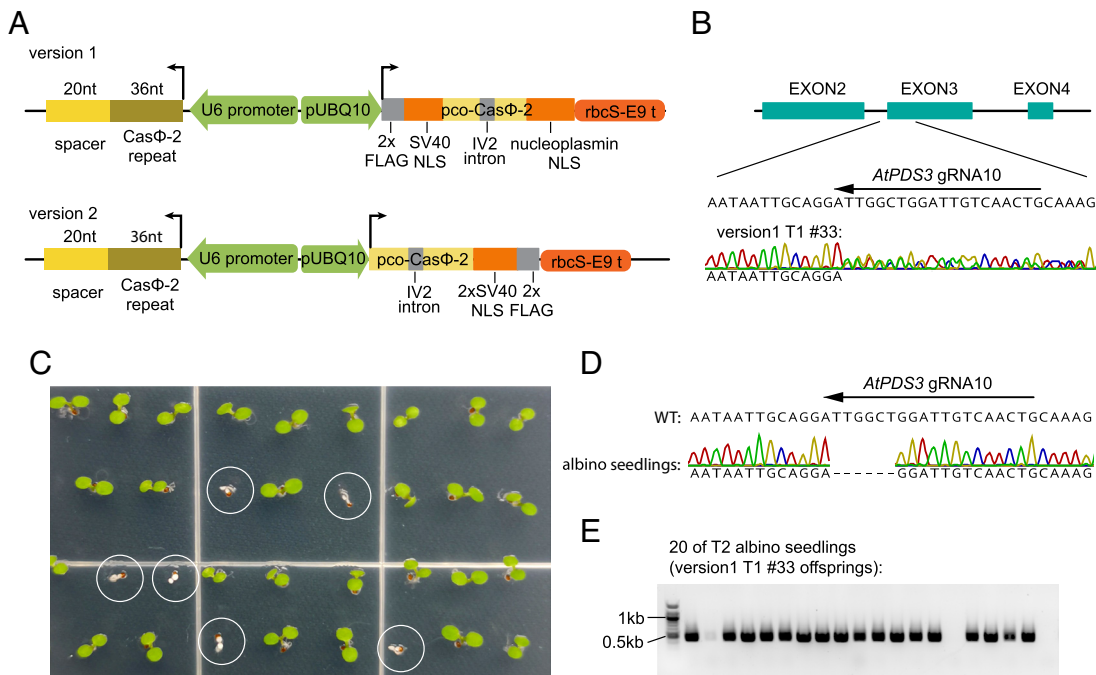


Fig. 1. Editing of the *AtPDS3* gene by CasΦ-2 in transgenic plants. (A) Schematics of the version 1 and version 2 constructs to generate *Arabidopsis* transgenic plants expressing the CasΦ-2 protein and the CasΦ-2 guide RNA. NLS, nuclear localization signal. pUBQ10, *Arabidopsis* *UBQ10* gene promoter. pco-CasΦ-2, *Arabidopsis* codon optimized CasΦ-2. rbcS-E9 t, rbcS-E9 terminator. U6 promoter, *AtU6-26* gene promoter. (B) Sanger sequencing result of T1 transgenic plant #33 of version 1 construct with *AtPDS3* gRNA10 at the target region. (C) Picture of seedlings from T2 populations of T1 plant #33, with albino seedlings circled. (D) Sanger sequencing result of albino seedlings from the T2 populations of T1 plant #33 at the *AtPDS3* gRNA10 targeted region. (E) PCR amplification of the DNA of 20 randomly selected albino seedlings from the T2 population of T1 plant #33 for a fragment of the CasΦ-2 transgene. Successful DNA extraction was suggested by successful PCR amplification of the *AtPDS3* gRNA10 targeted region followed by Sanger sequencing with the same DNA samples from these 20 albino seedlings, with all samples displaying mutation pattern as shown in (D).

CasΦ-2 generated mutation in the *AtPDS3* gene (SI Appendix, Fig. S2B). These results suggest that the initial T1 plant #6 was chimeric for different deletions which were inherited into the different T2 and T3 lines.

RDR6 Mediated Transgene Silencing Attenuates CasΦ-2 Mediated Editing in Transgenic Plants. In *Arabidopsis*, RNA-dependent RNA polymerase 6 (RDR6) plays an important role in the initiation of transgene silencing (24, 25). To evaluate if the CasΦ-2 transgene is also affected by transgene silencing, the editing efficiencies of CasΦ-2 transgenic T1 plants in the WT background or in the *rdr6-15* mutant (26) background were compared. For both the version 1 and the version 2 constructs with *AtPDS3* gRNA10, we detected significantly higher editing efficiency in the population of T1 transgenic plants in the *rdr6-15* mutant background compared to the WT background (SI Appendix, Fig. S3A), suggesting that RDR6 mediated silencing is limiting the editing efficiency in CasΦ-2 transgenic plants. No significant difference was detected between the version 1 and version 2 constructs in the same genetic background, suggesting that the two configurations of nuclear localization signal work similarly (SI Appendix, Fig. S3A).

CasΦ-2 Shows Similar Editing Efficiency at 23 °C and 28 °C.

We previously observed that CasΦ-2 RNPs with gRNA10 produced around 0.85% edits in *Arabidopsis* protoplasts using a 2-d room temperature incubation with a 2.5-h 37 °C pulse (79,768 edited reads/9,404,589 total reads) (19). We performed a similar experiment in the absence of the 37 °C pulse and found similar (1.1%) efficiency (91,950 edited reads/8,572,470 total reads), suggesting that CasΦ-2 is functional at lower temperatures. *Arabidopsis* grows ideally at around 23 °C but can also grow at temperatures up to about 28 °C (27), so we also performed an experiment with the version 2 plasmid comparing editing efficiencies in protoplasts at 23 °C versus 28 °C and found a roughly similar editing efficiency in both with gRNA10 (roughly 1.1%) (SI Appendix, Fig. S3B). To test the effect of temperature with a longer incubation time and in plants stably expressing CasΦ-2, T1 transgenic plants of version 1 or version 2 constructs either in the WT Col-0 background or the *rdr6-15* background were either continuously grown at room temperature (23 °C) or initially grown at 28 °C for 2 wk then transferred to room temperature. No significant difference was detected between the editing efficiencies of the T1 plants grown under these two temperature regimes (SI Appendix, Fig. S3 C and D), suggesting that CasΦ-2 expressing transgenes function similarly between 23 °C and 28 °C. Previously, CasΦ-2 has also been shown to be functional at 37 °C in human cells (19). Altogether, these data suggest that the CasΦ-2 protein has a wide functional temperature range, which should make it useful for editing in a wide variety of eukaryotic systems.

CasΦ-2 Mediated Editing Is Sensitive to DNA Methylation Status.

Previously, it was observed that DNA methylated chromatin showed a lower efficiency of editing by Cas9 (28, 29). To examine this for CasΦ-2 editing, we took advantage of the *FWA* gene and the *fwa-4* epiallele. In WT *Arabidopsis* plants, the *FWA* gene is silent in all adult plant tissues owing to DNA methylation in the promoter. *FWA* is normally only expressed by the maternal allele in the developing endosperm where it is imprinted and demethylated (30). In the epiallele *fwa-4*, the promoter is heritably unmethylated, and thus, the *FWA* gene is expressed ectopically (31). Ten guide RNAs were designed targeting the promoter region of the *FWA* gene, immediately upstream of

the transcription start site (Fig. 2A). In WT plants, this region contains DNA methylation and is targeted by the RNA directed DNA methylation machinery (32, 33), as evidenced by Pol V occupancy (Fig. 2A). Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) signals are also significantly lower in the WT plant compared to the *fwa-4* epi-mutant (34) (Fig. 2A), suggesting that the accessibility of the *FWA* promoter region is lower in the WT plant. An in vitro cleavage assay with CasΦ-2 RNPs showed that all 10 *FWA* guide RNAs led to cleavage of the PCR amplified *FWA* gene fragment, with the majority of the substrate DNA cleaved by RNPs of gRNA1, gRNA4, gRNA5, gRNA6, and gRNA7 (SI Appendix, Fig. S4). When CasΦ-2 RNPs were transfected into *fwa-4* epi-mutant protoplasts, gene editing events were detected with gRNA1, gRNA4, gRNA5, and gRNA6 (Fig. 2B). To compare the editing efficiency at *FWA* in different chromatin states, in an independent experiment, protoplasts of WT and *fwa-4* epi-mutant plants were prepared and transfected with CasΦ-2 RNPs with *FWA* gRNA1, gRNA4, gRNA5, and gRNA6. Much higher editing efficiency was observed for each of the four gRNAs in the *fwa-4* protoplasts compared to the WT protoplasts (Fig. 2C), suggesting that the CasΦ-2 mediated editing is more efficient when the target DNA has a more open chromatin state compared to a repressive chromatin state.

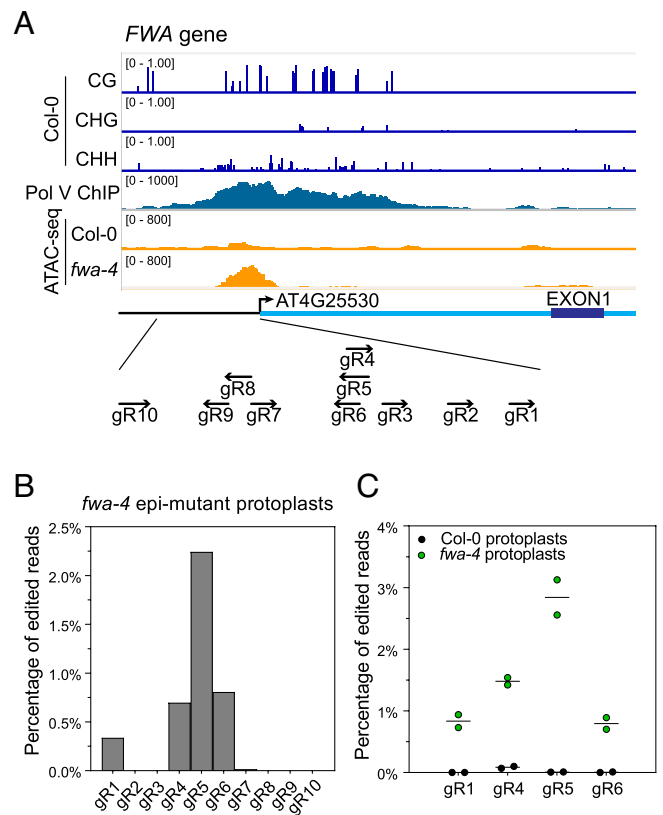


Fig. 2. Effect of the DNA methylation status at the *FWA* gene promoter on editing efficiency by CasΦ-2. (A) DNA methylation level (33) and normalized Pol V ChIP-seq signal (normalized by RPKM) (32) in Col-0 plants, as well as normalized ATAC-seq signal (normalized by RPKM) (34) in Col-0 plants versus the *fwa-4* epi-mutant plants at the *FWA* gene (AT4G25530) promoter region. The relative positions of the CasΦ-2 guide RNAs are illustrated in the schematic below. (B) CasΦ-2 editing efficiencies at the target loci in the *FWA* promoter region with gRNA 1 to gRNA 10 in the *fwa-4* epi-mutant protoplasts. (C) Comparison of CasΦ-2 editing efficiencies in the *FWA* promoter region with *FWA* gRNA1, gRNA4, gRNA5, and gRNA6 in Col-0 protoplasts versus in *fwa-4* epi-mutant protoplasts. Two replicates were performed for each gRNA in each type of protoplast. Dots indicate the editing efficiencies of individual replicates, and the line indicates the mean of the two replicates.

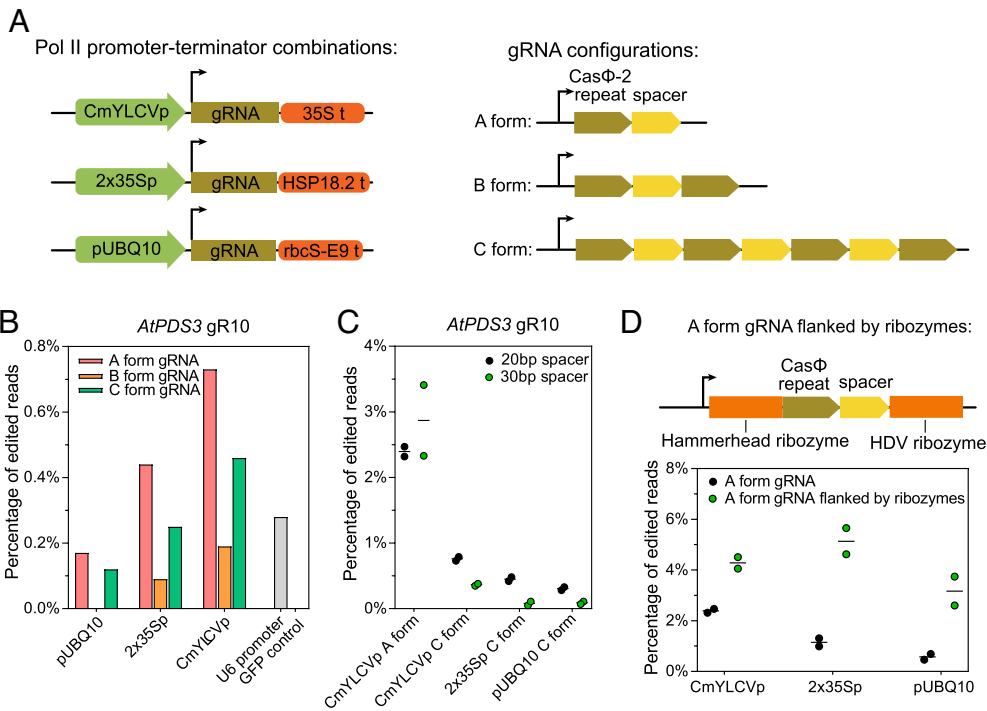


Fig. 3. Pol-II promoter-terminator combinations and ribozymes for transcription and processing of CasΦ-2 guide RNA. (A) Schematics of the tested Pol-II promoter-terminator combinations (Left) and the configurations of CasΦ-2 guide RNA arrays (Right). CmYLCVp, cestrum yellow leaf curling virus promoter. 35S t, 35S terminator. 2x35Sp, two times 35S promoter. HSP18.2 t, *Arabidopsis HSP18.2* gene terminator. For gRNA configurations: A form, a single CasΦ-2 repeat followed by *AtPDS3* gRNA10 spacer; B form, a CasΦ-2 repeat followed by *AtPDS3* gRNA10 spacer followed by a second CasΦ-2 repeat; C form, a triple array of CasΦ-2 repeats/*AtPDS3* gRNA10 spacers followed by a fourth CasΦ-2 repeat. (B) Summary of the target (*AtPDS3*) editing efficiencies in protoplasts, comparing promoter-terminator combinations and gRNA configurations, with the U6 promoter driving *AtPDS3* gRNA10 as a control. (C) CasΦ-2 editing efficiencies of *AtPDS3* gRNA10 with 20 bp spacer and 30 bp spacer were compared. (D) CasΦ-2 editing efficiencies of *AtPDS3* gRNA10 with or without the ribozyme processing machineries were compared. For (C) and (D), individual replicate values and mean of the two replicates of each test were plotted.

CasΦ-2 Guide RNAs Can Be Expressed with Pol-II Promoters and Ribozyme Processing Machineries.

As Pol-II promoters have been successfully used to drive the gRNA transcription for the Cas9 and Cas12 systems (14, 35), we wanted to test if Pol-II promoters are able to drive CasΦ-2 guide RNA transcription. Three Pol-II promoter and terminator sets were tested, the CmYLCV promoter and 35S terminator, a 2x35S promoter and HSP18.2 terminator, and a UBQ10 promoter with a rbcS-E9 terminator (Fig. 3A). Since CasΦ-2 has intrinsic pre-crRNA processing activity (19), *AtPDS3* gRNA10 without additional RNA processing machinery in three configurations was tested, a single CasΦ-2 repeat followed by *AtPDS3* gRNA10 spacer, a CasΦ-2 repeat followed by *AtPDS3* gRNA10 spacer followed by a second CasΦ-2 repeat, and a triple array of CasΦ-2 repeats/*AtPDS3* gRNA10 spacers followed by a fourth CasΦ-2 repeat (Fig. 3A). Among the three combinations of Pol-II promoters and terminators, the CmYLCV promoter with the 35S terminator led to the highest editing efficiency, while the UBQ10 promoter with the rbcS-E9 terminator led to the lowest editing efficiency in protoplasts (Fig. 3B). Out of the three different gRNA configurations, the single CasΦ-2 repeat followed by the *AtPDS3* gRNA10 exhibited the highest editing efficiency, while the CasΦ-2 repeat followed by the *AtPDS3* gRNA10 with another CasΦ-2 repeat at the end exhibited the lowest editing efficiency (Fig. 3B). When combining the CmYLCV promoter/35S terminator with the single CasΦ-2 repeat followed by the *AtPDS3* gRNA10, the target gene editing efficiency was much higher than that of the AtU6-26 *AtPDS3* gRNA10 cassette in protoplasts (Fig. 3B). Consistent with the higher levels of editing observed, a higher level of *AtPDS3* gRNA10 was detected in protoplasts transfected with plasmid carrying the cassette with the CmYLCV promoter *AtPDS3* gRNA10 construct (SI Appendix, Fig. S5). This

data suggests that boosting the levels of gRNAs using a strong promoter can increase the efficiency of gene editing by CasΦ-2.

The fact that the single *AtPDS3* gRNA10 without another CasΦ-2 repeat at the end exhibited the highest editing efficiency among the three gRNA configurations suggests that the CasΦ-2 processing of the transcript was not able to provide enough mature gRNAs. The 20 bp spacer could be too short for CasΦ-2 to bind simultaneously to both of the repeats flanking the spacer for pre-crRNA processing. To examine this further, *AtPDS3* gRNA10 with a 30 bp spacer was used to test if a longer spacer might assist the self-processing of pre-crRNA by CasΦ-2. However, we observed no difference between the editing efficiencies of the 30 bp and the 20 bp *AtPDS3* gRNA10 spacers when constructed in the A form (Fig. 3C). We also observe a slight decrease of target gene editing when the 30 bp spacer was constructed in the C form compared to the 20 bp spacer (Fig. 3C). These results suggest that a longer 30 bp spacer did not promote more efficient processing of pre-crRNA by CasΦ-2.

We tested whether adding a secondary gRNA processing mechanism might facilitate the release of mature gRNA. A ribozyme gRNA processing system, previously proven to be able to facilitate Cas9 and Cas12a guide RNA processing (14, 36), was cloned to flank the CasΦ-2 repeat and the *AtPDS3* gRNA10 spacer sequence. The hammerhead (HH) type ribozyme was added to the 5' end of the CasΦ-2 *AtPDS3* gRNA10, and a hepatitis delta virus (HDV) ribozyme was added to the 3' end (Fig. 3D). Constructs with ribozymes led to significantly higher editing efficiency, with all three promoter-terminator combinations tested (Fig. 3D). These results suggest that ribozymes were able to promote the processing of gRNA from the Pol-II transcripts, leading to a higher editing efficiency.

The Engineered CasΦ-2 Variants vCasΦ and nCasΦ Have Higher Target Gene Editing Efficiency in Protoplasts. It has been previously reported that the engineered CasΦ-2 variants vCasΦ and nCasΦ cleaved substrate DNA faster than the WTCasΦ in vitro (22). To test if these engineered CasΦ-2 variants are able to edit a target gene with higher efficiency in *Arabidopsis*, mesophyll protoplasts were transfected with plasmids expressing WTCasΦ, vCasΦ, or nCasΦ, as well as the desired guide RNAs. Version 2 constructs with U6 promoter-driven *AtPDS3* gRNA8 (19) and gRNA10 were transfected into WT protoplasts, while version 2 constructs with U6 promoter-driven *FWA* gRNA1, gRNA4, gRNA5, and gRNA6 were transfected into *fwa-4* protoplasts. Higher editing efficiencies were observed with vCasΦ and nCasΦ variants for all guide RNAs tested (Fig. 4A). To statistically evaluate the differences between the editing efficiencies, for the six gRNAs tested, normalized editing efficiencies were calculated (ratio over WTCasΦ efficiency) and pooled for statistical tests. Both the vCasΦ and nCasΦ variants yielded significantly higher editing efficiencies than the WTCasΦ, with ~17-fold of increase in editing efficiency for vCasΦ and ~ten-fold of increase in editing efficiency for nCasΦ (Fig. 4B). Similar tests were performed with RNPs reconstituted with WTCasΦ, vCasΦ, or nCasΦ and the same guide RNAs (SI Appendix, Fig. S6A). A significant increase in the editing efficiency was detected with both the vCasΦ and nCasΦ variants when pooled analysis was performed (SI Appendix, Fig. S6B), although the overall fold increase in editing efficiency was lower compared to the plasmid transfection experiments (Fig. 4B and SI Appendix, Fig. S6B).

We also tested if the vCasΦ and nCasΦ variants were able to enhance the editing efficiency at more compact chromatin utilizing the *FWA* promoter region as the target. To test this, protoplasts from WT plants were used for plasmid transfections with *FWA* gRNA1, gRNA4, gRNA5, and gRNA6. Although zero to very low editing was observed for WTCasΦ, both vCasΦ and nCasΦ variants yielded readily detectable and higher editing frequencies for all four *FWA* guide RNAs tested (Fig. 4C). Thus, in compact chromatin environments, the vCasΦ and nCasΦ variants appear to dramatically enhance the editing efficiency compared to WTCasΦ.

To confirm that the vCasΦ and nCasΦ variants, like WTCasΦ, are compatible with guide RNAs driven by Pol-II promoters, the CmYLCV promoter and the UBQ10 promoter were used to drive transcription of *AtPDS3* gRNA10 flanked by ribozymes. For both the CmYLCV promoter and the UBQ10 promoter, the editing efficiencies of the vCasΦ and nCasΦ variants were higher than that of the WTCasΦ (Fig. 4D).

Despite the differences in editing efficiency, the editing profiles of WTCasΦ, vCasΦ, and nCasΦ were similar. Consistent with previous data for WTCasΦ (19), the vCasΦ and nCasΦ variants also generate deletions primarily around 8 to 10 bp in size (SI Appendix, Fig. S6C).

vCasΦ and nCasΦ Show Higher Editing Efficiency in Transgenic Plants. Like in protoplasts, we observed significantly higher editing efficiencies in T1 transgenic plants expressing the vCasΦ and nCasΦ variants compared to WTCasΦ for *AtPDS3* gRNA10 driven by the U6 promoter (Fig. 5A). Interestingly, some T1 plants

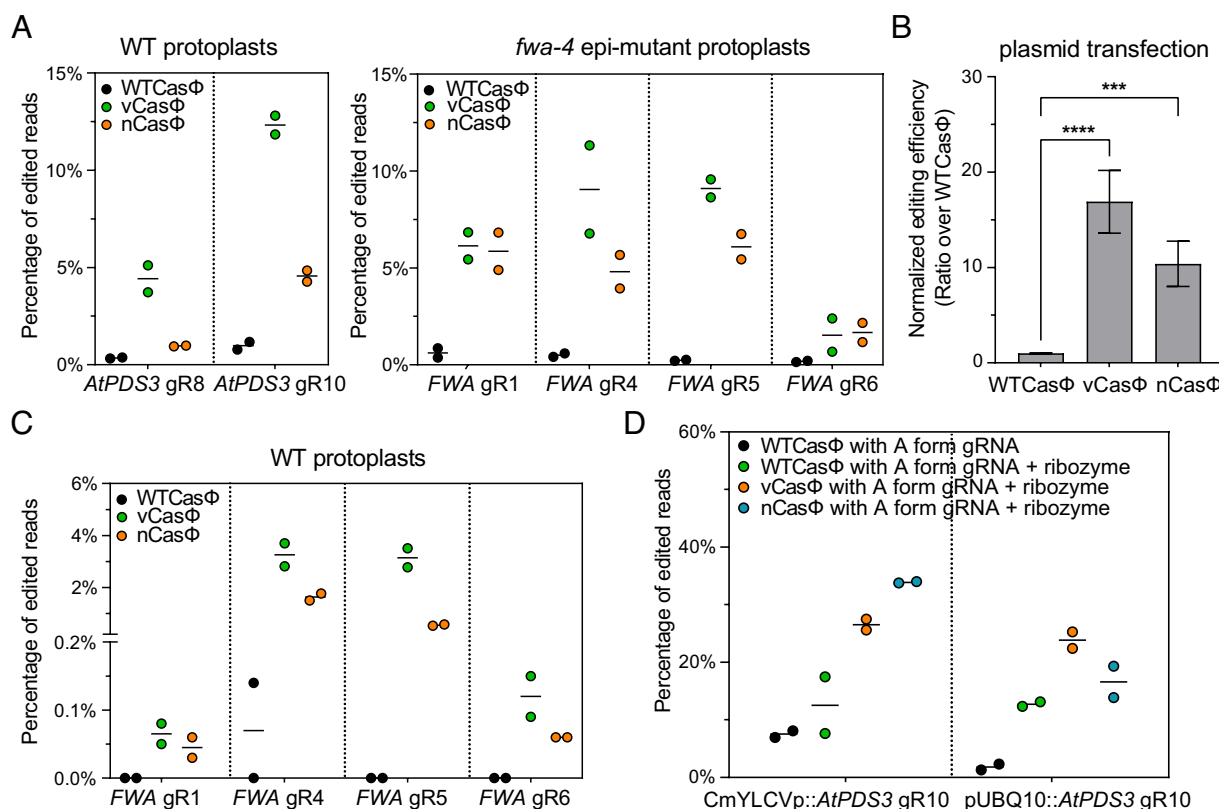


Fig. 4. Comparison of editing efficiency by the vCasΦ and nCasΦ variants and WTCasΦ in protoplasts. (A) Plasmids with WTCasΦ, vCasΦ, and nCasΦ expression cassettes and with indicated guide RNAs were transfected into protoplasts prepared from Col-0 plants (WT) (Left) and from *fwa-4* epi-mutant plants (Right). (B) Target gene editing efficiencies in (A) were normalized by calculating the ratio of editing efficiencies over that of mean editing efficiency by WTCasΦ for each guide RNA. Mean and SE of the normalized editing efficiencies for all gRNAs were plotted. Unpaired *t* test was used to calculate *P* value of indicated comparisons. ****0.0001 < *P* < 0.001, ******P* < 0.0001. (C) Plasmids with WTCasΦ, vCasΦ, and nCasΦ expression cassettes and guide RNAs targeting the *FWA* promoter region were transfected into WT protoplasts. (D) Plasmids with WTCasΦ, vCasΦ, and nCasΦ expression cassettes and *AtPDS3* gRNA10 with or without ribozymes driven by CmYLCVp and pUBQ10 were transfected into protoplasts. For (A), (C) and (D), individual replicate values and mean of the two replicates of each test were plotted.

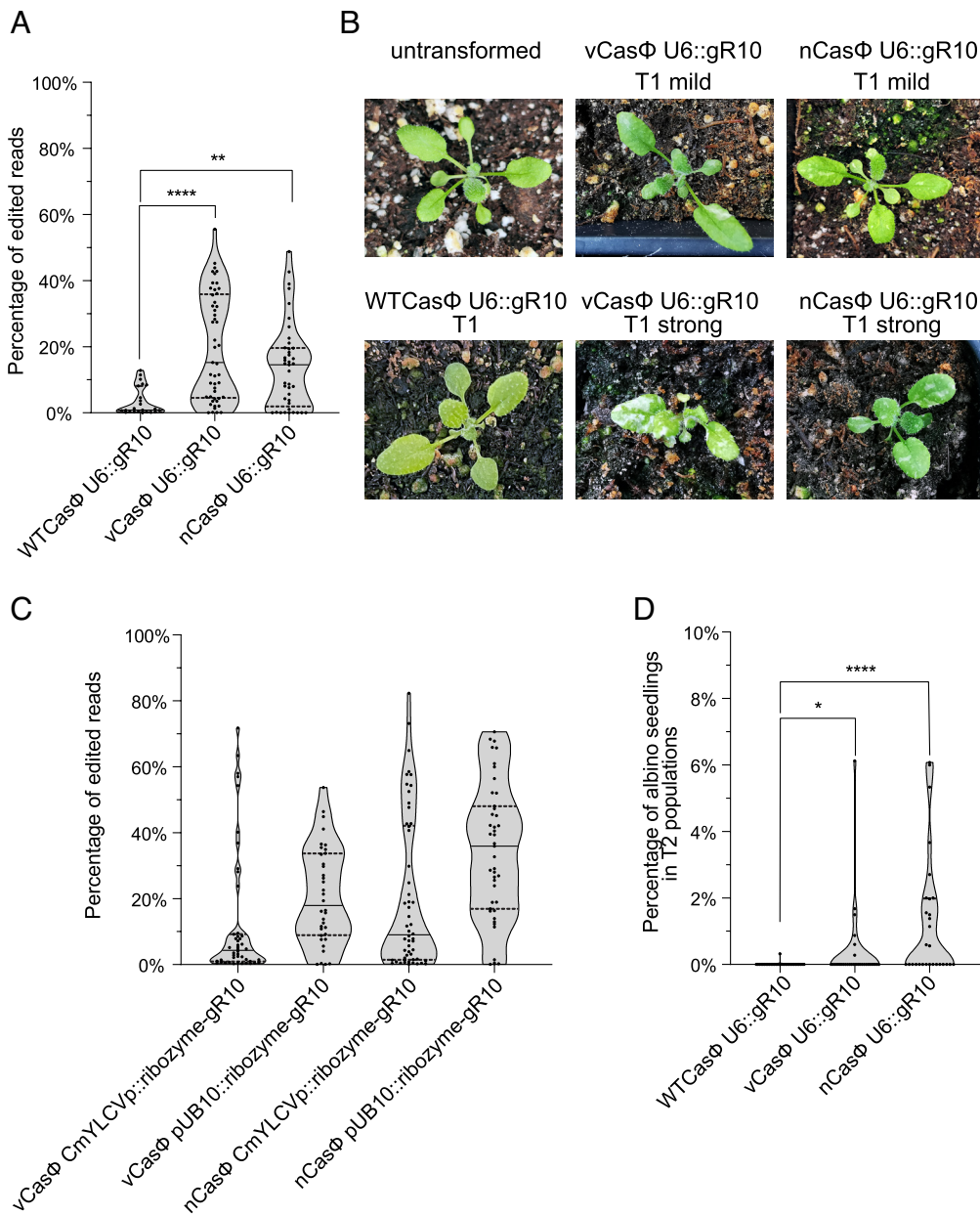


Fig. 5. Comparison of editing efficiency of the WTCasΦ with vCasΦ and nCasΦ variants in transgenic plants. (A) Leaf tissue of T1 transgenic plants of indicated constructs in the *rdp6-15* background was harvested for DNA extraction and amplicon sequencing analysis. gR10, *AtPDS3* gRNA10. The numbers of independent T1 plants (n) scored for each population are: n = 23 for WTCasΦ U6::gR10, n = 49 for vCasΦ U6::gR10, and n = 41 for nCasΦ U6::gR10. Mann-Whitney test was used to calculate the *P* value for each comparison indicated. **0.001 < *P* < 0.01, *****P* < 0.0001. (B) A *rdp6-15* plant (untransformed) is shown on the top left, with leaves appearing as uniformly green. Representative T1 transgenic plants in the *rdp6-15* background are shown for WTCasΦ U6::gR10 (Bottom Left), as well as vCasΦ U6::gR10 and nCasΦ U6::gR10 (top and bottom pictures of representative plants with mild and strong albino patches). (C) Leaf tissue of T1 transgenic plants expressing CasΦ variants with Pol-II promoters and *AtPDS3* gRNA10 flanked by ribozymes in the *rdp6-15* background was harvested for DNA extraction and amplicon sequencing analysis. gR10, *AtPDS3* gRNA10. The numbers of independent T1 plants (n) scored for each population are: n = 46 for vCasΦ CmYLCVp::ribozyme-gR10, n = 36 for vCasΦ pUB10::ribozyme-gR10, n = 54 for nCasΦ CmYLCVp::ribozyme-gR10, and n = 43 for nCasΦ pUB10::ribozyme-gR10. (D) Total seedling and albino seedling numbers were counted from random T2 populations of WTCasΦ U6::gR10, vCasΦ U6::gR10 and nCasΦ U6::gR10 in the *rdp6-15* mutant background. The percentage of albino seedlings from each T2 population are shown. gR10, *AtPDS3* gRNA10. The numbers of independent T2 populations (n) scored are: n = 31 for WTCasΦ U6::gR10, n = 30 for vCasΦ U6::gR10, and n = 29 for nCasΦ U6::gR10. Mann-Whitney test was used to calculate the *P* value of each comparison indicated. *0.01 < *P* < 0.05, *****P* < 0.0001. In (A), (C), and (D), truncated violin plots and all data points are shown, with median and quartiles indicated by solid and dashed line, respectively.

with white sectors were observed from the vCasΦ and nCasΦ expressing T1 populations (Fig. 5B), indicating strong editing activity in somatic cells. These white sectors were not observed in transgenic T1 plants expressing WT CasΦ (Fig. 5B). High target gene editing efficiencies were also observed in the T1 plants with vCasΦ and nCasΦ combined with the CmYLCV promoter or UBQ10 promoter-driven *AtPDS3* gRNA10 flanked by ribozymes (Fig. 5C). Consistent with the higher editing efficiency in the

T1 generation, vCasΦ and nCasΦ with *AtPDS3* gRNA10 also generated more albino seedlings in T2 populations compared to the WTCasΦ (Fig. 5D and SI Appendix, Fig. S7A and Table S2). Specifically, only one out of 31 T2 populations of WTCasΦ had albino seedlings at a percentage of 0.32% of total seedlings, while six out of 30 vCasΦ T2 populations and 14 out of 29 nCasΦ T2 populations showed albino seedlings, with highest percentage of albino seedlings at 6.12% and 6.07%, respectively (SI Appendix,

Table S2). Transgene-free albino seedlings were identified from these T2 populations, showing heritability of the mutations generated by the CasΦ variants (*SI Appendix, Fig. S7B*).

Off-Target Editing by the Engineered CasΦ-2 Variants vCasΦ and nCasΦ Is Rare. To evaluate in vivo off-target editing frequencies by CasΦ-2 in *Arabidopsis*, we performed whole genome sequencing on transgene-free T2 albino seedlings from T1 transgenic plants expressing vCasΦ or nCasΦ U6::*AtPDS3* gRNA10 in the *rdr6-15* mutant background. In T1 transgenic plants with ubiquitous CasΦ expression, unless off-target editing happens at a significant level in sequence contexts similar to the on-target loci, it is difficult to distinguish off-target editing effects from sequence variants generated from spontaneous mutation, PCR amplification errors, or sequencing errors. In these transgene-free T2 albino seedlings, inherited off-target edits will exist as heterozygous or homozygous mutations, thus making their detection robust by deep sequencing analysis at the whole genome level. Genomic DNA of albino transgene-free T2 seedlings from three independent T2 populations of vCasΦ and nCasΦ U6::*AtPDS3* gRNA10 in *rdr6-15*, together with four *rdr6-15* control seedlings, was sequenced to between 116-fold and 613-fold genome coverage, with >99% of genome covered by mapped reads (Fig. 6A and *SI Appendix, Table S3*). Variations relative to the reference Col-0 TAIR10 genome sequence were detected by

The Genome Analysis Toolkit (GATK) and Strelka2 (37, 38), and similar to previous observations (39–44), a large number of single nucleotide polymorphisms (SNPs) and additions or deletions of nucleotides (indels) were detected (*SI Appendix, Table S3*). However, the number of variants detected in the *rdr6-15* background plants relative to the Col reference genome was similar to that in the CasΦ-2 edited plants, suggesting that most of the variants detected are variations which already existed in the *rdr6-15* mutant background. We therefore filtered out all variants in the CasΦ-2 edited plants that were also present in the *rdr6-15* background. In addition, to select for heterozygous or homozygous mutations, variants with ratios of reference allele reads over variant allele reads larger than 3 (Ref/Alt>3) were discarded (Fig. 6A). Between 66 and 203 variants remained within the six albino seedlings tested after these filtering steps (*SI Appendix, Table S3*). On-target site mutations (*AtPDS3* gRNA10 target region) were reliably detected with this pipeline for all six albino seedlings, with two representative plants shown in Fig. 6B, *Left* panel. To test if other sequence variants might be due to CasΦ editing, we utilized Cas-OFFinder (45) to predict potential off-target sites with a TBN PAM sequence, allowing up to four base pair mismatches and two base pairs bulges relative to the *AtPDS3* gRNA10 spacer sequence. We observed no overlap of the predicted off-target sites and the sequence variants detected in the albino seedlings, suggesting that CasΦ-2 editing is highly specific

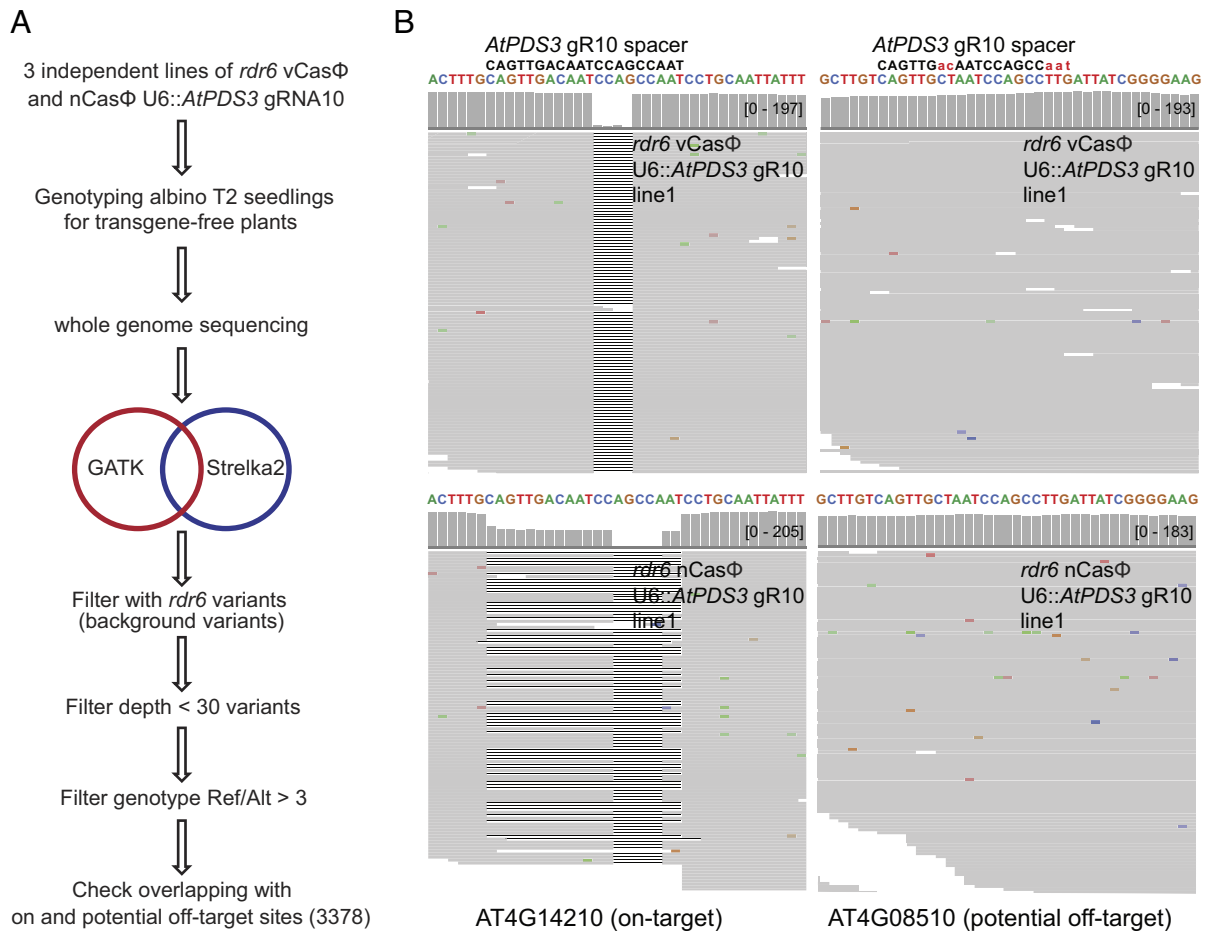


Fig. 6. Evaluation of in vivo off-target editing by the vCasΦ and nCasΦ variants through whole genome sequencing. GATK and Strelka2 were used to identify sequence variants compared to the TAIR10 reference genome sequence. Ref/Alt, the ratio of reads with reference sequence over reads with variant sequence. (B) Screenshots of aligned reads and coverage at the *AtPDS3* (*AT4G14210*) gRNA10 target region (*Left*) and a potential off-target site (*AT4G08510*) (*Right*). Top screenshots, from vCasΦ U6::*AtPDS3* gRNA10 line 1. Bottom screenshots, from nCasΦ U6::*AtPDS3* gRNA10 line 1. Capitalized and colored sequences are the reference genomic sequences at these two loci. *AtPDS3* gRNA10 spacer sequence is shown in black letters with uncanceled red letters showing the mismatched nucleotides between the *AtPDS3* gRNA10 spacer and the potential off-target site.

(SI Appendix, Table S3). An example of detailed reads at a predicted off-target site is shown in Fig. 6 B, Right panel and SI Appendix, Fig. S8 A, Right panel.

We also manually inspected each of the sequence variants detected genome wide, which indicated that most of them are located within long stretches of repeated single or di-nucleotides, and reads with similar variants were readily detected in the *rd6-15* plants, suggesting that most of the detected variants are likely due to imprecise PCR amplification, sequencing reaction or mapping outcomes. After removing these sequence variants at simple repeats, between 6 and 18 high-confidence variants remained in the six albino plants, with more SNP than indels (SI Appendix, Fig. S8B). However, there was no overlap of the variant loci except for a single SNP between two albino seedlings (SI Appendix, Table S4). These data suggest that the high-confidence variants we detected are rare and likely due to spontaneous mutation events, meaning that CasΦ-2 shows little or no off-target activity.

CasΦ-2 Is Able to Perform Gene Editing in Maize Protoplasts.

To test gene editing by CasΦ-2 in maize, an agronomically important monocot species with a more complex genome than *Arabidopsis*, 11 guide RNAs were designed targeting the maize *Lc* gene (SI Appendix, Table S5). RNPs of WTCasΦ, vCasΦ, and nCasΦ with *Lc* gRNAs were transfected into maize protoplasts. By amplicon sequencing, editing of target loci by *Lc* gRNA3 to gRNA7 was detected (SI Appendix, Fig. S9). Consistent with results in *Arabidopsis*, vCasΦ and nCasΦ yielded much higher target editing efficiency compared to the WTCasΦ, with up to about 7% editing for the best guide RNA. These data suggest that CasΦ systems might be useful in many different plant species.

Discussion

In this study, we explored the use of CasΦ-2 for plant genome engineering. We showed that heritable gene edits could be generated in *Arabidopsis* using WTCasΦ at a low frequency. We also showed that the vCasΦ and nCasΦ protein variants exhibited much higher editing efficiency than WTCasΦ in *Arabidopsis* and maize protoplasts and also a higher frequency of heritable edits in transgenic *Arabidopsis* plants. Given that CasΦ-2 was able to edit the dicot plant *Arabidopsis*, the monocot plant maize, as well as mammalian cells (19), it is likely that CasΦ-2 could be utilized in most eukaryotic systems. The wide temperature range of enzymatic activity of CasΦ-2 may also make it useful for plants and other organisms that grow at cooler temperatures. Since Pol-II promoters can be used for CasΦ-2 guide RNA transcription, the vast array of Pol-II promoters could be useful for expression of gRNAs in specific conditions, such as in different tissues, at higher levels with strong promoters, or in an inducible fashion. CasΦ-2-based editing can also be applied in both RNP and plasmid expression formats. We also found that CasΦ-2 editing was highly specific, with no convincing off-target editing detected by genome wide analysis.

We observed that transgene silencing limited the editing efficiency of CasΦ-2, because editing frequencies were higher when utilizing the *rd6* mutant. This suggests that the level of expression of either the CasΦ-2, its gRNA precursors, or both, is limiting for high efficiency editing. This also suggests that tuning down the transgene silencing machinery and improvements in transgene designs are likely to lead to more efficient editing. By comparing editing efficiencies at the *FWA* locus in protoplasts derived from the WT methylated strain or the *fwa* unmethylated epiallele strain, we also found that methylated compact chromatin negatively affected editing by CasΦ-2, presumably because the more highly compact chromatin limits access of CasΦ-2 to the DNA. Fusion

of CasΦ-2 protein to chromatin remodelers, as previously reported for the Cas9 protein (46), or other methods, could facilitate successfully editing at difficult target sites, or might increase editing efficiency at all sites.

We observed that more than half of the guide RNAs we designed yielded no detectable editing events by amplicon sequencing, and only a few of the guide RNAs lead to efficient editing. Similarly, strong differences in editing efficiencies between different guide RNAs were observed when CasΦ-2 was used to target an enhanced green fluorescent protein (EGFP) gene in HEK293 cells (19). In addition, variable activities were observed with in vitro substrate DNA digestion assays with CasΦ-2 RNPs reconstituted with different guide RNAs (SI Appendix, Fig. S4). However, with the best performing guide RNAs, reasonable editing efficiencies by CasΦ-2 and its variants were observed. In mammalian cells, more than 30% of cells lost EGFP signal due to the editing of the WTCasΦ-2 with gRNA8 (19). In *Arabidopsis*, with the CasΦ-2 protein variants and *AtPDS3* gRNA10, 10% to 30% editing efficiency was observed in protoplasts (Fig. 4D) and 40 to 80% of somatic editing efficiency was observed in multiple T1 transgenic plants (as indicated by the upper quartile in Fig. 5C). However, in the future it will be important to systematically define gRNA design principles to increase these frequencies.

Amplicon sequencing and in vivo editing results indicated that CasΦ-2 edits consist of small deletions of roughly nine base pairs, consistent with biochemical data showing that CasΦ-2 produces an 8 to 12 nucleotide staggered cut (19). CasΦ-2 could therefore be useful for making in-frame deletions to create partial loss of function alleles, or to remove small sections of specific domains of proteins. Indeed, one of the heritable edits found in this study was a two amino acid deletion resulting in an in-frame deletion of the coding region of the *AtPDS3* gene. The ability to make small deletions may also be useful for promoter analysis by removing individual transcription factor binding sites, or by promoter deletion scanning by tiling gRNAs across a promoter.

In summary, this study demonstrates that the compact CasΦ editing system can be used for gene editing in plants and can generate heritable mutations with high specificity. Compared to the Cas9 and Cas12a systems, which has been optimized with community efforts for years, CasΦ-2 is still less efficient in overall editing efficiency. However, with well-performing guide RNAs, a reasonable level of editing efficiency can be obtained with the CasΦ system. In addition, the vCasΦ and nCasΦ variants demonstrated the effectiveness of protein engineering to potentially improve the editing efficiency further in the future. With its hypercompact protein size, wide working temperature range, and T-rich minimal PAM, CasΦ should be a useful supplement to the plant gene editing toolbox and should be useful in both basic research and agricultural biotechnology.

Materials and Methods

Plant Materials and Growth Condition. To grow *Arabidopsis* plants for protoplast preparation, Col-0 and *fwa-4* epi-mutant plants were grown under a 12-h light/12-h dark photoperiod and with low light condition for 3 to 4 wk. To grow maize plants for protoplast isolation, B73 seeds were soaked in water overnight and planted in half peat moss and half vermiculite. After 3 d of light incubation, emerged seedlings were transferred to dark until the second leaf was 10 to 15 cm long.

Agrobacterium mediated *Arabidopsis* plant transformation was performed as previously described (47), and transgenic T1 plants were screened with half MS plates with 40 μg/mL hygromycin B. For 28 °C treatment of T1 plants, including the T1 plants in SI Appendix, Fig. S1B (version2 construct *AtPDS3* gRNA10), stratified seeds on half MS plates with hygromycin B were incubated at 28 °C and resistant T1 plants were transferred to soil after about a week and put back to 28 °C

for a total of 2 wk incubation at 28 °C. T1 plants were then moved to a greenhouse (23 °C) for the rest of the life cycle. To support the growth of albino seedlings in T2 generation, 3% sucrose was supplemented to half MS plates when needed.

Protein Purification. WTCasΦ, vCasΦ, and nCasΦ proteins with nuclear localization signal (NLS) were purified as previously described (19, 22).

RNP Reconstitution. Guide RNAs were synthesized as 25 nucleotide repeat + 20 nucleotide spacer as shown in *SI Appendix, Table S5*. Lyophilized RNA was dissolved by adding DEPC-treated H₂O to a concentration of 0.5 mM. The dissolved RNA was incubated at 65 °C for 3 min, then cooled down to room temperature. For RNP reconstitution, heated and cooled RNA was added to 2 × Cleavage Buffer (2 × CB buffer, 20 mM Hepes-Na, 300 mM KCl, 10 mM MgCl₂, 20% glycerol, 1 mM TCEP, pH 7.5) to a final concentration of 5 μM and vortexed to mix. Then, WTCasΦ, vCasΦ, or nCasΦ proteins were added to a final concentration of 4 μM and mixed by pipetting. This solution was then incubated at room temperature for 30 min. The resulting solution contains 4 μM of RNP in 2 × CB buffer.

Plasmids Used in this Study. Plasmids generated in this study are listed in *SI Appendix, Table S6*. The CasΦ-2 × SV40NLS-2 × FLAG coding sequence (without IV2 intron) was *Arabidopsis* codon optimized and synthesized by IDT (Integrated DNA Technologies). The HBT-pcoCas9 vector (addgene52254) backbone with the FLAG and SV40 NLS (for version 1) or without the FLAG and SV40 NLS (for version 2) was amplified and assembled by TAKARA in-fusion HD cloning kit (cat639650) with the synthesized CasΦ sequence amplified as two fragments as well as the IV2 intron sequence amplified from the HBT-pcoCas9 vector. pCAMBIA1300-pYAO-cas9-MCS vector (48) was digested by KpnI and EcoRI to remove the pYAO-cas9 cassette. Then, pcoCasΦ with IV2 intron, NLS, and FLAG tag coding sequence was amplified from HBT_pcoCasPhi_version 1 and version 2 plasmids and cloned into the digested vector together with PCR amplified pUB10 and rbcS-E9 terminator by TAKARA in-fusion HD cloning kit (cat639650). Subsequently, various guide RNA cassettes were cloned into the SpeI site of the obtained vectors by TAKARA in-fusion reaction or by restriction digestion and ligation of PCR amplified or IDT synthesized DNA fragments. Construction of the nCasΦ and vCasΦ binary vectors was performed by first digesting the pC1300_pUB10_pcoCasPhi_E9t_MCS_version2 vector with KpnI to remove pUB10 promoter and the first part of CasΦ coding sequence. Then, pUB10 promoter and CasΦ sequences containing the desired mutations were PCR amplified as two fragments with desired mutations added by overlapping primers and cloned back into the digested vector by an in-fusion reaction. Primers used for cloning and synthesized DNA fragments used for cloning are listed in *SI Appendix, Table S7*. The control plasmid HBT-sGFP for protoplast transfection was previously described (49) and obtained from ABRC (stock CD3-911).

Protoplast Isolation and Transfection. *Arabidopsis* mesophyll protoplast isolation was performed as previously described (49). For RNP transfection into *Arabidopsis* protoplasts, 26 μL of 4 μM RNP was first added to a round bottom 2 mL tube, followed by 200 μL of protoplasts (2 × 10⁵ cells/mL). Then, 2 μL of 5 μg/μL salmon sperm DNA was added and mixed gently by tapping the tube 3 to 4 times. Finally, 228 μL of fresh, sterile, and RNase free PEG-CaCl₂ solution (49) was added to the protoplast-plasmid mixture and mixed well by gently tapping the tube. The protoplasts with PEG solution were incubated at room temperature for 10 min, then 880 μL of W5 solution (49) was added and mixed with the protoplasts by inverting the tube two to three times to stop the transfection. Protoplasts were harvested by centrifuging the tubes at 100 rcf for 2 min and resuspended in 1 mL of WI solution. They were then plated in 6-well plates pre-coated with 5% calf serum. These 6-well plates were then incubated at room temperature for 48 h.

For plasmid transfections into *Arabidopsis* protoplasts, the concentrations of plasmids were determined by nanodrop. Then 20 to 50 μg of plasmids (plasmid amounts were the same within each experiment) were added to the bottom of each transfection tube, and the volume of the plasmids was supplemented with H₂O to reach 20 μL. 200 μL of protoplasts were added followed by 220 μL of fresh and sterile polyethylene glycol (PEG)-CaCl₂ solution. The mixture was mixed well by gently tapping tubes and incubated at room temperature for 10 min. 880 μL of W5 solution was added and mixed with the protoplasts by inverting the tube two to three times to stop the transfection. Protoplasts were harvested by centrifuging the tubes at 100 rcf for 2 min and resuspended in 1 mL of WI solution. They were then plated in 6-well plates pre-coated with 5% calf serum. These 6-well plates were then incubated at room temperature or at 28 °C for 48 h.

For maize protoplast isolation, middle 8 to 10 cm of the second leaf from etiolated maize seedlings were sliced into thin strips (about 0.5 mm wide) and immersed in enzyme solution (0.6 M mannitol, 10 mM MES pH 5.7, 1.5% cellulase R10, 0.3% macerozyme R10, 1 mM CaCl₂, 5 mM BME, 0.1% BSA). Then the enzyme solution with leaf stripes was put under vacuum for half an hour and then placed in dark for 2.5 h. The solution was then gently swirled to release the protoplasts and filtered through a 40 μm cell strainer. The cells were collected by centrifugation at 150 g for 5 min and resuspended in 10 mL washing solution (0.6 M mannitol, 4 mM MES pH 5.7, 20 mM KCl). Then the cells were rested on ice for at least half an hour. After resting, cell pellets were resuspended with washing solution to 2 × 10⁵ cells/mL in cold washing buffer. For RNP transfection into maize protoplasts, 13 μL 4 μM RNP and 1 μL 5 μg/μL salmon sperm DNA were added to 100 μL protoplasts. Then 114 μL PEG solution (40% PEG 4000, 0.2 M mannitol, 0.1 M CaCl₂) was added and mixed by tapping. After 10 min of incubation, transfections were stopped by adding 880 μL washing solution and gently inverting several times. Transfected protoplasts were collected by 3 min of centrifugation at 150 g and resuspended in 1 mL incubation buffer (50). Then, the protoplasts were plated in 6-well plates pre-coated with 5% calf serum. These 6-well plates were then incubated at room temperature for 48 h.

Amplicon Sequencing. DNA was extracted from protoplast samples and leaves of transgenic plants with Qiagen DNeasy plant mini kit (Qiagen 69106). The amplicon was obtained using two rounds of PCR. Amplification primers for the first round of PCR were designed to have the 3' sequence of the primers flanking a 200 to 300 bp fragment of the genomic area targeted by the guide RNA of interest. Primers for the first round of amplification are listed in *SI Appendix, Table S7*. The 5' part of the primer contained a sequence which will be bound by common sequencing primers. After 25 cycles of the first round of PCR amplification, the reaction was purified using 1 × Ampure XP beads (Beckman Coulter A63881). The eluate was used as template for the second round of PCR amplification of 12 cycles. The second round of PCR was designed so that indexes were added to each sample. The samples were then purified using 0.8 × Ampure XP beads. Part of the purified libraries were run on a 2% agarose gel to check for size and absence of primer dimer (fragments below 200 bp considered as primer dimer). Then amplicons were sent for next generation sequencing.

Amplicon Sequencing Result Analysis. Reads were first quality and adaptor trimmed using *Trim Galore* and then mapped to the target genomic region by the BWA aligner (v0.7.17, BWA-MEM algorithm). Sorted and indexed bam files were used as input files for further analysis by the CrispRvariants R package (v1.14.0). Each mutation pattern with corresponding read counts was exported by the CrispRvariants R package. After assessing all control samples, a criterion to classify reads as edited reads was established: Only reads with a ≥ 3 bp deletion or insertion (indel, mainly as deletions) of the same pattern (indels of same size starting at the same location) with ≥ 100 read counts from a sample were counted as edited reads. This criterion is established due to the observation of 1 bp indels and occasionally 2 bp indels with read numbers > 100 in control samples. Also, larger indels that occur at very low frequencies (much lower than 100 reads) were observed in control samples. These observations indicate that occasional PCR inaccuracy and low-quality sequencing in a small fraction of reads can result in the indel patterns with corresponding read number ranges as stated above in control samples with the typical sequencing depth in our experiments (1 to 5 million reads/sample). By employing such stringent criteria, it is believed that the editing signals counted are true signals indicating editing events. Additionally, for *FWA* gRNA5 and gRNA6 targeted regions, there are long stretches of adenines near these target regions. Due to the high error rate of polymerases amplifying long stretches of adenines, reads with indels only within these stretches of adenines were not counted as true deletions. For amplicon sequencing result analysis of maize protoplast, the criterion to classify reads as edited reads was established: Only reads with a ≥ 3bp deletion or insertion (indel, mainly as deletions) of the same pattern (indels of same size starting at the same location) with ≥ 10 read counts from a sample were counted as edited reads. The adjustment of the criterion was based on the fact that this amplicon sequencing was of less depth (performed by the iSeq 100) compared to the amplicon libraries of *Arabidopsis* protoplasts.

Off-Target Analysis. DNA from single *Arabidopsis* seedlings was extracted with the Qiagen DNeasy plant mini kit and sheared to 300 bp size with a Covaris. Library

preparation was performed with Tecan Ovation Ultralow V2 DNA-seq kit. For variant calling, whole genome sequencing reads were aligned to the TAIR10 reference genome using BWA mem (v0.7.17) (51) with default parameters. GATK (4.2.0.0) (37) MarkDuplicatesSpark was used to remove PCR duplicate reads. Then GATK HaplotypeCaller was used as call raw variants. Raw SNPs were filtered with QD (Variant Confidence/Quality by Depth) < 2.0, FS (Phred-scaled p-value using Fisher's exact test to detect strand bias) > 60.0, MQ (RMS Mapping Quality) < 40.0, and SOR (Symmetric Odds Ratio of 2x2 contingency table to detect strand bias) > 4.0. Raw InDels were filtered with QD < 2.0, FS > 200.0, and SOR > 10.0 and used for base quality score recalibration. The recalibrated bam file was further applied to GATK and Strelka (v2.9.2) (38) for SNPs/InDel calling. Only SNPs/InDels called by both GATK and Strelka were used for further analysis. The intersection of GATK and Strelka SNPs/InDels were filtered by removing identical SNPs/InDels in the *rd6-15* background by BedTools (v2.26.0) (52). Variants with coverage lower than 30 reads were removed. Variant loci at which the ratio of the reads with the WT allele over the reads with variant allele larger than three was removed. For heterozygous alleles, at a coverage level of 30 reads, the chance of observing WT reads /mutation allele reads > 3 is 0.26% (binomial distribution, one-tailed $P < 0.0026$).

Data, Materials, and Software Availability. All high-throughput sequencing data generated in this study are accessible at the National Center for Biotechnology information Gene Expression Omnibus via series accession GSE206798 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206798>).

- R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- J. McGinn, L. A. Marraffini, Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).
- A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-based technologies for the manipulation of Eukaryotic genomes. *Cell* **168**, 20–36 (2017).
- A. Pickar-Oliver, C. A. Gersbach, The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
- G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
- K. S. Makarova *et al.*, Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
- Z. Feng *et al.*, Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* **23**, 1229–1232 (2013).
- J.-F. Li *et al.*, Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **31**, 688–691 (2013).
- V. Nekrasov, B. Staskawicz, D. Weigel, J. D. G. Jones, S. Kamoun, Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 691–693 (2013).
- Q. Shan *et al.*, Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
- K. Xie, Y. Yang, RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* **6**, 1975–1983 (2013).
- B. Zetsche *et al.*, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- X. Tang *et al.*, A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* **3**, 17103 (2017).
- A. Endo, M. Masafumi, H. Kaya, S. Toki, Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Sci. Rep.* **6**, 38169 (2016).
- Z. Zhong *et al.*, Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. *Mol. Plant* **11**, 999–1002 (2018).
- H. Kim *et al.*, CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* **8**, 14406 (2017).
- B. Al-Shayeb *et al.*, Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).
- P. Pausch *et al.*, CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
- M. Jinek *et al.*, Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
- H. Nishimasu *et al.*, Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
- P. Pausch *et al.*, DNA interference states of the hypercompact CRISPR-CasΦ effector. *Nat. Struct. Mol. Biol.* **28**, 652–661 (2021).
- G. Qin *et al.*, Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in *Arabidopsis* by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res.* **17**, 471–482 (2007).
- T. Dalmay, A. Hamilton, S. Rudd, S. Angell, D. C. Baulcombe, An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543–553 (2000).
- P. Mourrain *et al.*, *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
- E. Allen *et al.*, Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36**, 1282–1290 (2004).

Previously published data were used for (Fig. 2A) of this work (GSE100010 GSE124546 GSE155503) (32–34).

ACKNOWLEDGMENTS. We thank Suhua Feng and Mahnaz Akhavan for support with high-throughput sequencing at the UCLA Broad Stem Cell Research Center BioSequencing Core Facility. S.E.J. and J.A.D. are investigators of the HHMI. P.P. receives funding from the European Regional Development Fund under grant agreement number 01.2.2-CPVA-V-716-01-0001 with the Central Project Management Agency (CPVA), Lithuania, and from the Research Council of Lithuania (LMTLT) under grant agreement number S-MIP-22-10.

Author affiliations: ^aDepartment of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095; ^bLife Sciences Center - European Molecular Biology Laboratory (LSC-EMBL) Partnership for Genome Editing Technologies, Vilnius University - Life Sciences Center, Vilnius 10257, Lithuania; ^cInnovative Genomics Institute, University of California, Berkeley, CA 94720; ^dDepartment of Plant and Microbial Biology, University of California, Berkeley, CA 94720; ^eDepartment of Human Genetics, University of California, Los Angeles, CA 90095; ^fHHMI, University of California, Berkeley, CA 94720; ^gDepartment of Molecular and Cell Biology, University of California, Berkeley, CA 94720; ^hCalifornia Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720; ⁱDepartment of Chemistry, University of California, Berkeley, CA 94720; ^jMolecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; ^kGladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA 94158; and ^lHHMI, University of California, Los Angeles, CA 90095

- L. Rivero *et al.*, Handling *Arabidopsis* plants: Growth, preservation of seeds, transformation, and genetic crosses. *Methods Mol. Biol.* **1062**, 3–25 (2014).
- A. Příbylová, L. Fischer, D. E. Pyott, A. Bassett, A. Molnar, DNA methylation can alter CRISPR/Cas9 editing frequency and DNA repair outcome in a target-specific manner. *New Phytol.* **235**, 2285–2299 (2022).
- T. Weiss *et al.*, Epigenetic features drastically impact CRISPR-Cas9 efficacy in plants. *Plant Physiol.* **190**, 1153–1164 (2022).
- T. Kinoshita *et al.*, One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**, 521–523 (2004).
- W. J. Soppe *et al.*, The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802 (2000).
- W. Liu *et al.*, RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in *Arabidopsis*. *Nat. Plants* **4**, 181–188 (2018).
- J. Gallego-Bartolomé *et al.*, Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in *Arabidopsis*. *Cell* **176**, 1068–1082.e19 (2019).
- Z. Zhong *et al.*, DNA methylation-linked chromatin accessibility affects genomic architecture in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2023347118 (2021).
- T. Čermák *et al.*, A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* **29**, 1196–1217 (2017).
- Y. Gao, Y. Zhao, Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **56**, 343–349 (2014).
- A. McKenna *et al.*, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- S. Kim *et al.*, Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
- Z. Feng *et al.*, Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4632–4637 (2014).
- H. Zhang *et al.*, The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* **12**, 797–807 (2014).
- X. Tang *et al.*, A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol.* **19**, 84 (2018).
- J. Li *et al.*, Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. *Plant Biotechnol. J.* **17**, 858–868 (2019).
- W. Xu *et al.*, Comprehensive analysis of CRISPR/Cas9-mediated mutagenesis in *Arabidopsis thaliana* by genome-wide sequencing. *Int. J. Mol. Sci.* **20**, 4125 (2019).
- X. Wang *et al.*, Whole-genome sequencing reveals rare off-target mutations in CRISPR/Cas9-edited grapevine. *Hortic. Res.* **8**, 114 (2021).
- S. Bae, J. Park, J.-S. Kim, Cas-OffFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
- X. Ding *et al.*, Improving CRISPR-Cas9 genome editing efficiency by fusion with chromatin-modulating peptides. *Cris. J.* **2**, 51–63 (2019).
- X. Zhang, R. Henriques, S.-S. Lin, Q.-W. Niu, N.-H. Chua, Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**, 641–646 (2006).
- L. Yan *et al.*, High-efficiency genome editing in *Arabidopsis* using YAO promoter-driven CRISPR/Cas9 system. *Mol. Plant* **8**, 1820–1823 (2015).
- S. D. Yoo, Y. H. Cho, J. Sheen, *Arabidopsis* mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
- Y. Niu, R. W. Shultz, M. M. D. Unson, M. A. Kock, J. P. J. Casey, "Novel plant cells, plants, and seeds". US Patent WO 2018/140899 (2018).
- H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).