CRISPR-CasΦ from huge phages is a hypercompact genome editor

Patrick Pausch[1,2 †], Basem Al-Shayeb[1,3 †], Ezra Bisom-Rapp[4], Connor A. Tsuchida[1,5], Zheng Li[6], Brady F. Cress[1,2], Gavin J. Knott[1,2], Steven E. Jacobsen[6,7], Jillian F. Banfield[1,8], Jennifer A. Doudna[1,2,7,9*]

[1] Innovative Genomics Institute, University of California, Berkeley, CA, USA

[2] Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

[3] Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[4] College of Natural Resources, University of California, Berkeley, CA, USA

[5] University of California, Berkeley - University of California, San Francisco Graduate Program in Bioengineering, University of California, Berkeley, CA, USA

[6] Department of Molecular, Cellular and Developmental Biology, University of California, Los Angeles, CA, USA.

[7] Howard Hughes Medical Institute, USA

[8] Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA

[9] Department of Chemistry, University of California, Berkeley, CA, USA

*Correspondence to: doudna@berkeley.edu

†Contributed equally to this manuscript

Abstract:

CRISPR-Cas systems are found widely in prokaryotes where they provide adaptive immunity against virus infection and plasmid transformation. We describe a minimal functional CRISPR-Cas system, comprising a single ~70 kilodalton protein, CasΦ, and a CRISPR array, encoded exclusively in the genomes of huge bacteriophages. CasΦ employs a single active site for both CRISPR RNA (crRNA) processing and crRNA-guided DNA cutting to target foreign nucleic acids. This hypercompact system is active *in vitro* and in human and plant cells with expanded target recognition capabilities relative to other CRISPR-Cas proteins. Useful for genome editing and DNA detection but with a molecular weight half that of Cas9 and Cas12a genome-editing enzymes, CasΦ offers advantages for cellular delivery that expand the genome editing toolbox.

One Sentence Summary:

Phage-derived CasΦ uses a single active site to process guide RNA and cut DNA for genome editing and nucleic acid detection.

Main Text:

Competition between viruses and their host microbes fostered the evolution of CRISPR-Cas systems that employ nucleases and non-coding CRISPR RNAs (crRNAs) to target foreign nucleic acids by complementary base pairing (*1*). Processing of CRISPR array transcripts, consisting of repeats and spacer sequences acquired from viruses or other mobile genetic elements (MGEs) (*2*), generates mature crRNAs that guide Cas proteins (*3*) to detect and destroy previously encountered viruses. Although found almost exclusively in microbial genomes, the recent discovery of ubiquitous huge bacteriophages (viruses of bacteria) revealed the surprising prevalence of CRISPR-Cas systems encoded in their genomes (*4*). These systems notably lack CRISPR spacer acquisition machinery (Cas1, Cas2 and Cas4 proteins) and generally harbor compact CRISPR arrays (median: 5 spacers per array), some of which target the genes of competing phages or phage hosts. CasΦ (Cas12j) is a family of Cas proteins encoded in the Biggiephage clade (*4*). CasΦ contains a C-terminal RuvC domain with remote homology to that of the TnpB nuclease superfamily from which type V CRISPR-Cas proteins are thought to have evolved (*4*, *5*) (fig. S1). However, CasΦ shares <7% amino acid identity with other type V CRISPR-Cas proteins and is most closely related to a TnpB group distinct from miniature type V (Cas14) proteins (Fig. 1A).

CasΦ's unusually small size of ~70-80 kDa, about half the size of the Cas9 and Cas12a (Fig. 1B), and its lack of co-occurring genes raised the question of whether CasΦ functions as a *bona fide* CRISPR-Cas system. We investigated three divergent CasΦ orthologs from metagenomic assemblies (fig. S2), hereafter referred to as CasΦ-1, CasΦ-2 and CasΦ-3. To

examine CasΦ's ability to recognize and target DNA in bacterial cells, we tested whether CasΦ could protect *Escherichia coli* from plasmid transformation. CRISPR–Cas systems target DNA sequences following or preceding a 2–5 base pair (bp) Protospacer Adjacent Motif (PAM) for self-versus-non-self discrimination (*6*). To determine whether CasΦ uses a PAM, we transformed a library of plasmids containing randomized regions adjacent to crRNA-complementary target sites, thereby depleting plasmids harboring functional PAMs. This revealed the crRNA-guided double-strand DNA (dsDNA) targeting capability of CasΦ and minimal T-rich PAM sequences, including 5′-TBN-3′ PAMs (where B is G, T, or C) depleted for CasΦ-2 (Fig. 1C).

We next used the *E. coli* expression system and plasmid interference assay to determine the components required for CRISPR-CasΦ system function. RNA-sequencing analysis revealed transcription of the *casΦ* gene and the reduced CRISPR array but no evidence of other non-coding RNA such as a trans-activating CRISPR RNA (tracrRNA) within the locus (Fig. 1D). In addition, CasΦ activity could be readily reprogrammed to target other plasmid sequences by altering the guide RNA (fig. S3). These findings suggest that in its native environment, CasΦ is a functional phage protein and *bona fide* CRISPR-Cas effector capable of cleaving crRNA-complementary DNA such as other phage (Fig. 1E). Furthermore, these results demonstrate that this single-RNA system is much more compact than other active CRISPR-Cas systems (Fig. 1F).

We next investigated the DNA recognition and cleavage requirements of CasΦ *in vitro*. RNA-seq revealed that the crRNA spacer, which is complementary to DNA targets, is 14-20 nucleotides (nt) long (Fig. 1D). Incubation of purified CasΦ (fig. S4) with crRNAs of different spacer sizes along with supercoiled plasmid or linear dsDNA revealed that DNA cleavage requires the presence of a cognate PAM and a spacer of ≥ 14 nt (Fig. 2A; fig. S5A). Analysis of

the cleavage products showed that CasΦ generated staggered 5′-overhangs of 8-12 nt (Fig. 2B, C; fig. S5B, C), similar to the staggered DNA cuts observed for other type V CRISPR-Cas enzymes including Cas12a and CasX (*7*, *8*). We also observed that CasΦ-2 and CasΦ-3 were more active *in vitro* than CasΦ-1, and the non-target strand (NTS) was cleaved faster than the target-strand (TS) within the RuvC active site (Fig. 2D; figs. S6A, S7; Supplementary Text). Furthermore, CasΦ was found to cleave ssDNA but not ssRNA *in cis* and *in trans* (fig. S6B, S8), suggesting that CasΦ may also target ssDNA MGEs or ssDNA intermediates. The trans-cleavage activity of CasΦ, observed only upon DNA recognition *in cis* (fig. S8), coupled with a minimal PAM requirement (Fig. 1C), may be useful for broader nucleic acid detection as previously demonstrated for type V and type VI Cas proteins (*9*–*11*).

CRISPR-CasΦ systems must produce mature crRNA to guide foreign DNA cleavage. Other type V CRISPR-Cas proteins process pre-crRNAs using an internal active site distinct from the RuvC domain (*12*) or by recruiting Ribonuclease III to cleave a pre-crRNA:tracrRNA duplex (*13*–*16*). The absence of a detectable tracrRNA for CasΦ hinted that CasΦ may catalyze crRNA maturation on its own. To test this possibility, we incubated purified CasΦ with substrates designed to mimic the pre-crRNA structure (Fig. 3A). Reaction products corresponding to a 26-29 nt-long repeat and 20 nt spacer sequence of the crRNA were observed only in the presence of wild type CasΦ, corroborated by RNA-seq analysis of native loci (Figs. 1D; 3A, C; fig. S9). In control experiments, we found that pre-crRNA processing is strictly magnesium-dependent (Fig. 3B; fig. S9), which is different from other CRISPR-Cas RNA processing reactions and suggested a distinct cleavage mechanism. Notably, the RuvC domain requires magnesium to cleave DNA (*17*), and some RuvC domains have been reported to have endoribonucleolytic activity (*15*). Based on these observations, we tested CasΦ containing a

RuvC-inactivating mutation and found it to be incapable of processing pre-crRNAs (Fig. 3B; fig. S9A, B). Both wild-type and catalytically inactivated CasΦ proteins bind crRNA, and their reconstituted complexes with pre-crRNA have similar elution profiles from a size exclusion column, suggesting no pre-crRNA binding or protein stability defect resulting from the RuvC mutation (fig. S10).

We hypothesized that if the RuvC domain is responsible for pre-crRNA processing, the products should contain 5′-phosphate and 2′- and 3′-hydroxyl moieties as observed in RNAs generated by the RuvC-related RNase HI enzymes (*17*). In contrast, other type V CRISPR-Cas enzymes process pre-crRNA by metal-independent acid-base catalysis in an active site distinct from the RuvC, generating 2′-3′-cyclic phosphate crRNA termini, as observed for Cas12a (*18*). Phosphatase treatment of CasΦ-generated crRNA followed by denaturing acrylamide gel analysis showed no change in the crRNA migration, distinct from the change in mobility detected for crRNA generated by Cas12a (Fig. 3C; fig. S9C). This result implies that no 2′-3′-cyclic phosphate was formed during the reaction catalyzed by CasΦ, in contrast to the acid-base catalyzed processing reaction by Cas12a (Fig. 3C, D). Together, these data demonstrate that CasΦ uses a single RuvC active site for both pre-crRNA processing and DNA cleavage.

The versatility and programmability of CRISPR-Cas systems for genome editing in virtually any organism have sparked a revolution in biotechnology and fundamental research (*19*). To investigate whether CasΦ can be harnessed for human genome editing, we performed a gene disruption assay (*8*) using CasΦ co-expressed with a crRNA in HEK293 cells (Fig. 4A). We found that CasΦ-2 and CasΦ-3, can induce targeted disruption of a genomically integrated EGFP gene (Fig. 4A; fig. S11). In one case, CasΦ-2 with an individual guide RNA was able to edit up to 33% of cells (Fig. 4A), comparable to levels initially reported for CRISPR–Cas9,

CRISPR–Cas12a, and CRISPR–CasX (*7*, *8*, *20*). We next tested if CasΦ-2 can be delivered as RNPs into plant protoplasts to edit the endogenous *Arabidopsis thaliana PDS3* gene (Fig. 4B; fig. S12). Next generation sequencing revealed that CasΦ-2 introduces primarily 8-10 bp deletions (Fig. 4B), consistent with the cleavage pattern observed *in vitro* (Fig. 2C). The small size of CasΦ in combination with its minimal PAM requirement will be particularly advantageous for both vector-based delivery into cells and a wider range of targetable genomic sequences, providing a powerful addition to the CRISPR-Cas toolbox.

Three other well-characterized Cas enzymes Cas9, Cas12a, and CasX, use one (Cas12a and CasX) or two active sites (Cas9) for DNA cutting and rely on a separate active site (Cas12a) or additional factors (CasX and Cas9) for crRNA processing (Fig. 4C). The finding that a single RuvC active site in CasΦ is capable of crRNA processing and DNA cutting suggests that size limitations of phage genomes, possibly in combination with large population sizes and higher mutation rates in phages compared to prokaryotes (*21–23*), led to a consolidation of chemistries within one catalytic center. Such compact proteins may be particularly amenable to engineering and laboratory evolution to create new functionalities for genome manipulation, and highlight huge phages as an exciting forefront for discovery and biotechnological applications for human health.

References and Notes:

1. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**, 1709–1712 (2007).

2. J. McGinn, L. A. Marraffini, Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol*. **17**, 7–12 (2019).

3. F. Hille, H. Richter, S. P. Wong, M. Bratovič, S. Ressel, E. Charpentier, The Biology of CRISPR-Cas: Backward and Forward. *Cell*. **172**, 1239–1259 (2018).

4. B. Al-Shayeb, R. Sachdeva, L.-X. Chen, F. Ward, P. Munk, A. Devoto, C. J. Castelle, M. R. Olm, K. Bouma-Gregson, Y. Amano, C. He, R. Méheust, B. Brooks, A. Thomas, A. Lavy, P. Matheus-Carnevali, C. Sun, D. S. A. Goltsman, M. A. Borton, A. Sharrar, A. L. Jaffe, T. C. Nelson, R. Kantor, R. Keren, K. R. Lane, I. F. Farag, S. Lei, K. Finstad, R. Amundson, K. Anantharaman, J. Zhou, A. J. Probst, M. E. Power, S. G. Tringe, W.-J. Li, K. Wrighton, S. Harrison, M. Morowitz, D. A. Relman, J. A. Doudna, A.-C. Lehours, L. Warren, J. H. D. Cate, J. M. Santini, J. F. Banfield, Clades of huge phages from across Earth's ecosystems. *Nature* (2020), doi:10.1038/s41586-020-2007-4.

5. S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O. O. Abudayyeh, J. S. Gootenberg, K. S. Makarova, Y. I. Wolf, K. Severinov, F. Zhang, E. V. Koonin, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).

6. D. Gleditzsch, P. Pausch, H. Müller-Esparza, A. Özcan, X. Guo, G. Bange, L. Randau, PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biology*. **16** (2019), pp. 504–517.

7. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. **163**, 759–771 (2015).

8. J.-J. Liu, N. Orlova, B. L. Oakes, E. Ma, H. B. Spinner, K. L. M. Baney, J. Chuck, D. Tan, G. J. Knott, L. B. Harrington, B. Al-Shayeb, A. Wagner, J. Brötzmann, B. T. Staahl, K. L. Taylor, J. Desmarais, E. Nogales, J. A. Doudna, CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*. **566**, 218–223 (2019).

9. J. S. Chen, E. Ma, L. B. Harrington, M. Da Costa, X. Tian, J. M. Palefsky, J. A. Doudna, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. **360**, 436–439 (2018).

10. A. East-Seletsky, M. R. O'Connell, S. C. Knight, D. Burstein, J. H. D. Cate, R. Tjian, J. A. Doudna, Two Distinct RNase Activities of CRISPR-C2c2 Enable Guide RNA Processing and RNA Detection. *Nature*. **538**, 270–273 (2016).

11. J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, F. Zhang, Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*. **356**, 438–442 (2017).

12. I. Fonfara, H. Richter, M. Bratovič, A. Le Rhun, E. Charpentier, The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*. **532**, 517–521 (2016).

13. D. Burstein, L. B. Harrington, S. C. Strutt, A. J. Probst, K. Anantharaman, B. C. Thomas, J. A. Doudna, J. F. Banfield, New CRISPR-Cas systems from uncultivated microbes. *Nature*. **542**, 237–241 (2017).

14. L. B. Harrington, D. Burstein, J. S. Chen, D. Paez-Espino, E. Ma, I. P. Witte, J. C. Cofsky, N. C. Kyrpides, J. F. Banfield, J. A. Doudna, Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*. **362**, 839–842 (2018).

15. W. X. Yan, P. Hunnewell, L. E. Alfonse, J. M. Carte, E. Keston-Smith, S. Sothiselvam, A. J. Garrity, S. Chong, K. S. Makarova, E. V. Koonin, D. R. Cheng, D. A. Scott, Functionally diverse type V CRISPR-Cas systems. *Science*. **363**, 88–91 (2019).

16. S. Shmakov, O. O. Abudayyeh, K. S. Makarova, Y. I. Wolf, J. S. Gootenberg, E. Semenova, L. Minakhin, J. Joung, S. Konermann, K. Severinov, F. Zhang, E. V. Koonin, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell*. **60**, 385–397 (2015).

17. M. Nowotny, Retroviral integrase superfamily: the structural perspective. *EMBO Rep*. **10**, 144–151 (2009).

18. D. C. Swarts, J. van der Oost, M. Jinek, Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell*. **66**, 221–233.e4 (2017).

19. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science*. **361**, 866–869 (2018).

20. P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, RNA-guided human genome engineering via Cas9. *Science*. **339**, 823–826 (2013).

21. S. Duffy, L. A. Shackelton, E. C. Holmes, Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet*. **9**, 267–276 (2008).

22. M.-C. Lee, C. J. Marx, Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet*. **8**, e1002651 (2012).

23. M. Lynch, Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol*. **60**, 327–349 (2006).

24. K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Č. Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, E. V. Koonin, Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol*., 1–17 (2019).

25. G. J. Knott, B. W. Thornton, M. J. Lobba, J.-J. Liu, B. Al-Shayeb, K. E. Watters, J. A. Doudna, Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nat. Struct. Mol. Biol*., 1 (2019).

26. C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, J. E. Corn, Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol*. **34**, 339–344 (2016).

27. 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).

28. D. C. Swarts, M. Jinek, Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Mol. Cell.* **73**, 589–600.e4 (2019).

29. J. C. Cofsky, D. Karandur, C. J. Huang, I. P. Witte, J. Kuriyan, J. A. Doudna, CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *bioRxiv* (2020), p. 2020.02.10.937540.

Acknowledgements: We thank Anabel Schweitzer and Hannah Spinner for providing plasmids (pBFC545, pBFC546, pBFC547 and the PAM library plasmid) and Juan Hurtado for providing the EGFP-HEK293 cell line. We thank Doudna Lab members as well as Alexander Crits-Christoph and Jacob West-Roberts for helpful discussions, and Rohan Sachdeva for bioinformatic support. **Funding:** PP is supported by the Paul G. Allen Frontiers Group and NIH Somatic Cell Genome Editing consortium (NIH U01AI142817-02). BAS is supported by an NSF Graduate Research Fellowship (No. DGE 1752814). GJK is supported by an NHMRC Investigator Grant (EL1, APP1175568) and previously an American Australian Association Fellowship. JAD receives funding from the William M. Keck Foundation, the National Multiple Sclerosis Society, the NIH (RM1HG009490, U01AI142817-02), NSF (1817593) JAD and SEJ are Investigators of the Howard Hughes Medical Institute. JAD is a Paul Allen Distinguished Investigator. **Author Contributions**: BAS and JFB conceived initial study, further developed with PP and JAD. PP and BAS designed experiments and analyzed data with input from JAD and JFB. PP, EBR, and BAS cloned constructs with guidance from PP. BAS conducted metagenomic assemblies, genome curation, protein and crRNA sequence analysis, active residue identification and computational analyses for PAM depletion assays and RNA-seq. PP conducted PAM depletion assays, EOT assays, performed RNA-seq experiments, purified proteins and performed biochemical experiments. PP, BAS, CT and ZL designed genome editing experiments

with input from SEJ. CT performed tissue culture and flow cytometry and ZL performed editing experiments in *A. thaliana*. BFC and GJK provided materials and experimental advice. BAS and PP wrote the manuscript with input from JAD and JFB. PP and BAS prepared figures. The manuscript was reviewed and approved by all co-authors. **Competing Interests**: The Regents of the University of California, Berkeley and Los Angeles have patents pending for CRISPR technologies on which the authors are inventors. JAD is a co-founder of Caribou Biosciences, Editas Medicine, Intellia Therapeutics, Scribe Therapeutics, and Mammoth Biosciences. JAD is a scientific advisory board member of Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Synthego, and Mammoth Biosciences. JAD is a director at Johnson & Johnson and has sponsored research projects by Pfizer, Roche Biopharma, and Biogen. JFB is a founder of Metagenomi. SEJ is a scientific co-founder of Inari Agriculture and SEJ and JAD are members of its scientific strategy board. **Data and materials availability:** All data are available in the manuscript or the supplementary material. Reagents are available through Addgene and upon request from JAD.

Supplementary Materials:

Materials and Methods
Supplementary Text
Figures S1-S13
Tables S1-S4

Other Supplementary Materials for this manuscript includes the following:

Protein and repeat sequences for CasΦ1-CasΦ10

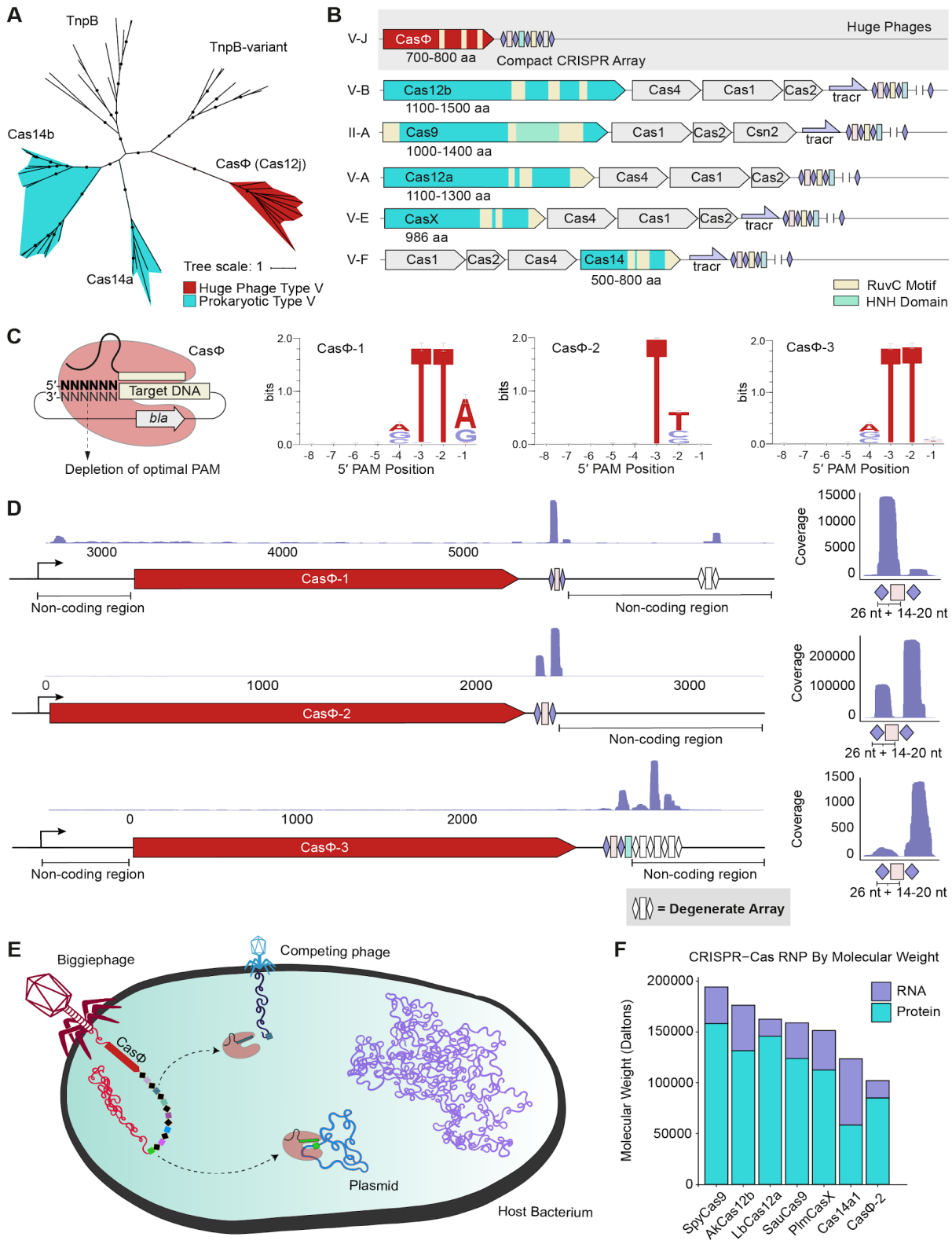

Fig. 1. CasΦ is a *bona fide* CRISPR-Cas system from huge phages. (**A**) Maximum Likelihood phylogenetic tree of type V effector proteins and respective predicted ancestral TnpB nucleases.

Bootstrap and approximate likelihood-ratio test values ≥ 90 are denoted on the branches with black circles. (**B**) Illustrations of genomic CRISPR-Cas loci of CasΦ, Cas14, and systems previously employed in genome editing applications. (**C**) Graphical representation of the PAM depletion assay and the resulting PAMs for three CasΦ orthologs. (**D**) RNA-sequencing results (left) mapped onto the native genomic loci of CasΦ orthologs and their upstream and downstream non-coding regions as cloned with reduced CRISPR-arrays into expression plasmids. Enlarged view of RNA mapped onto the first repeat-spacer pair (right). (**E**) Schematic of the hypothesized function of Biggiephage-encoded CasΦ in an instance of superinfection of its host. CasΦ may be used by the huge phage to eliminate competing mobile genetic elements. (**F**) Predicted molecular weights of the ribonucleoprotein (RNP) complexes of small CRISPR-Cas effectors and those functional in editing of mammalian cells.

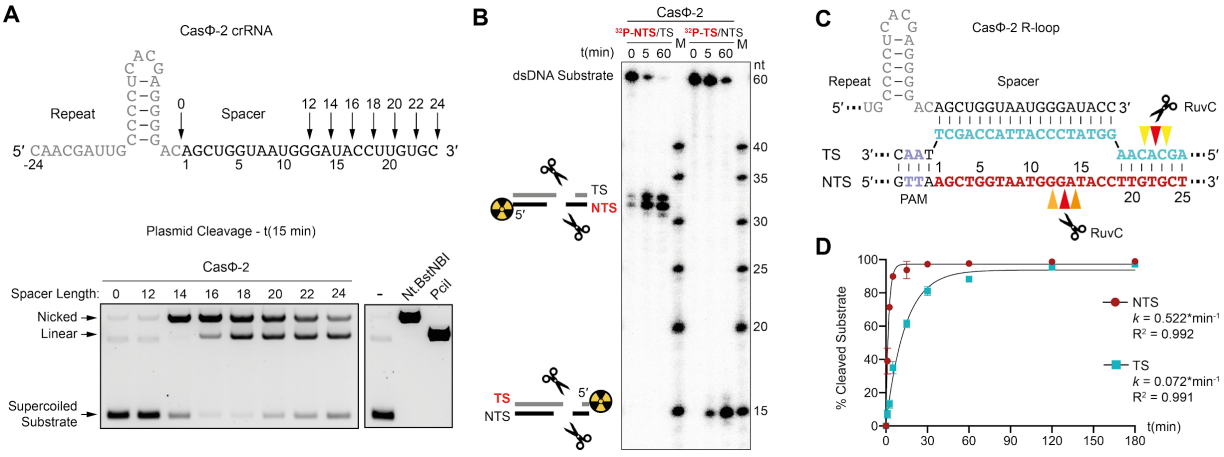

Fig. 2. CasΦ cleaves DNA. (**A**) Supercoiled plasmid cleavage assay testing CasΦ RNPs reconstituted with crRNAs of different spacer lengths. (**B**) Cleavage assay targeting dsDNA oligo-duplices for mapping of the cleavage structure. (**C**) Scheme illustrating the cleavage pattern. (**D**) NTS and TS DNA cleavage efficiency (n = 3 each, mean ± s.d.). Data is shown in fig. S7B.

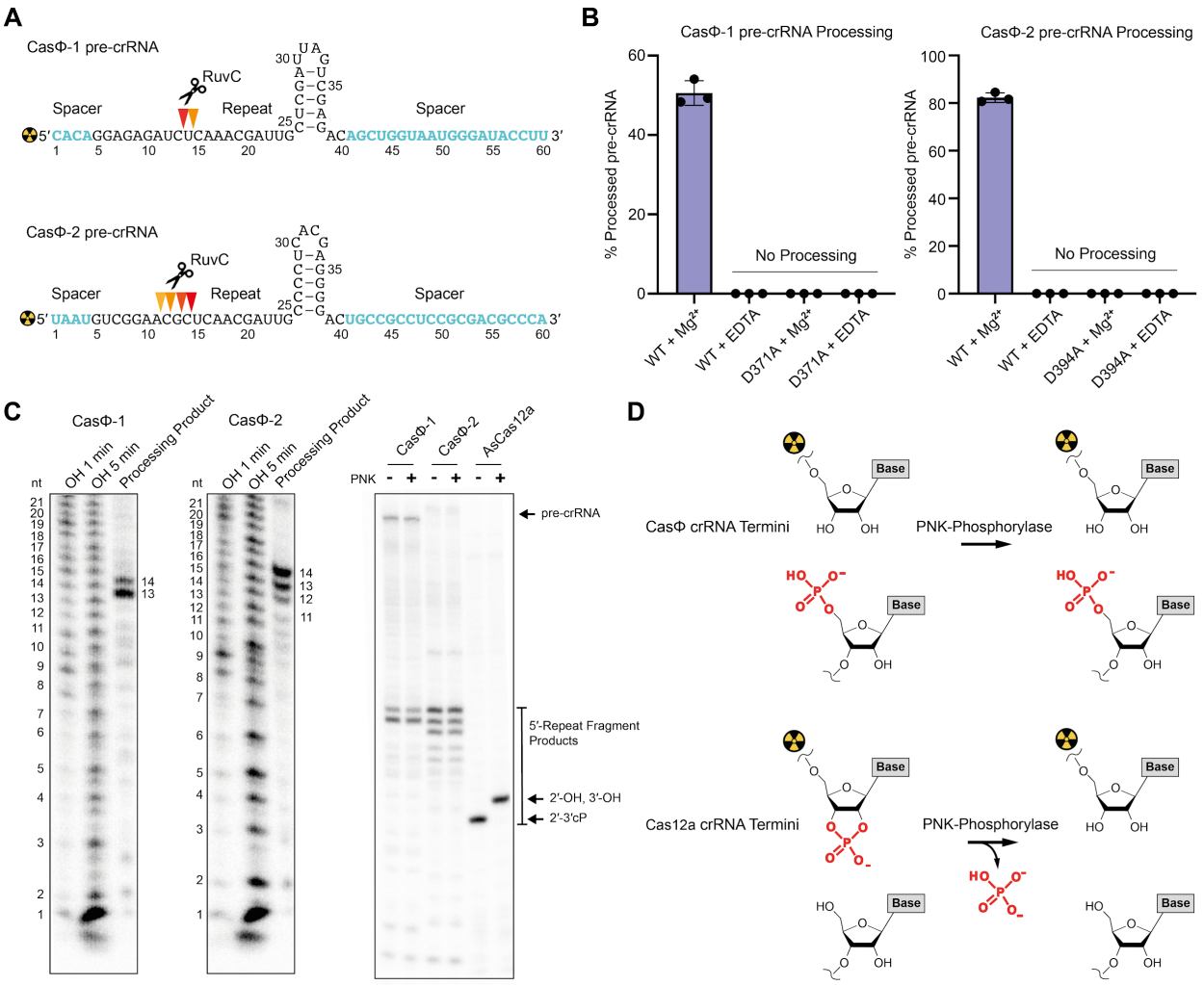

Fig. 3. CasΦ processes pre-crRNA within the RuvC active site. (**A**) pre-crRNA substrates and processing sites (red triangles) as derived from the OH-ladder in panel C. (**B**) Pre-crRNA processing assay for CasΦ-1 and CasΦ-2 in dependence of Mg^{2+} and RuvC active site residue variation (D371A and D394A) (n = 3 each, mean ± s.d.; t = 60 min). Data is shown in fig. S9B. (**C**) Left and middle: Alkaline hydrolysis ladder (OH) of the pre-crRNA substrate. Right: PNK-phosphatase treatment of the CasΦ and *Acidaminococcus sp*. Cas12a cleavage products. (**D**) Graphical representation of the mature crRNA termini chemistry of CasΦ and Cas12a and PNK-phosphorylase treatment outcomes.

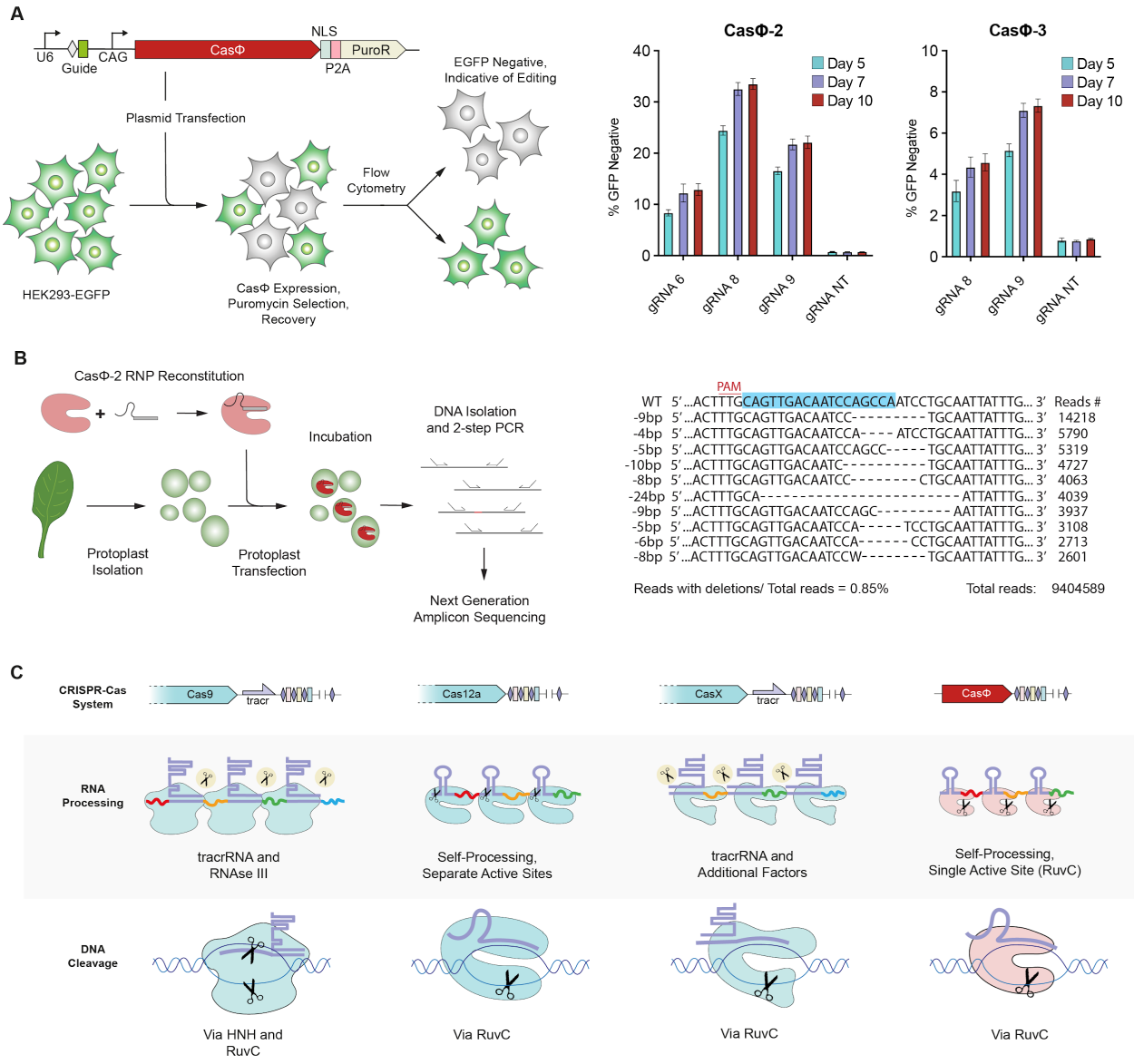

Fig. 4. CasΦ is functional for genome editing. (**A**) Experimental workflow of the GFP disruption assay (left) and GFP disruption using CasΦ-2 and CasΦ-3 and a non-targeting (NT) guide as a negative control (n = 3 each, mean ± s.d.). (**B**) Experimental workflow of CasΦ-2 RNP-mediated genome-editing in *A. thaliana* mesophyll protoplasts (left) and amplicon sequencing data (right) showing the most frequent deletions for gRNA33 in the targeted region (blue) within the *AtPDS3* gene. (**C**) Scheme illustrating the differences in RNA processing and DNA cutting for Cas9, Cas12a, CasX, and CasΦ.

Science
AAAS

Supplementary Materials for

CRISPR-CasΦ from huge phages is a hypercompact genome editor

Patrick Pausch[†], Basem Al-Shayeb[†], Ezra Bisom-Rapp, Connor A. Tsuchida, Zheng Li, Brady F. Cress, Gavin J. Knott, Steven E. Jacobsen, Jillian F. Banfield, Jennifer A. Doudna[*]

[†]These authors contributed equally to this manuscript
*correspondence to: doudna@berkeley.edu

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S4

Other Supplementary Materials for this manuscript includes the following:

Protein and repeat sequences for CasΦ1- CasΦ10

Materials and Methods

Metagenomic assemblies, genome curation, and CRISPR-CasΦ detection

Metagenomic sequencing data was assembled using previously described methods (*4*). Coding sequences (CDS) were predicted from sequence assemblies using prodigal with genetic code 11 (-m -g 11 -p single) and (-m -g 11 -p meta) and preliminary annotations and phage genome curations were performed as previously described (*4*). Bowtie2 v2.3.4.1 was used to map reads to the de novo assembled sequences, and we retained unplaced mate pairs of mapped reads with shrinksam (github.com/bcthomas/shrinksam). N-filled gaps and local misassemblies were identified and corrected, and unplaced or incorrectly placed paired reads allowed extension of contig ends. Local assembly changes and extensions were verified with further read mapping. A database of CasΦ sequences from (*4*) was generated using MAFFT v7.407 and hmmbuild. CDS from new assemblies were searched against the HMM database using hmmsearch with e-value < 1×10^{-5} and added to the database upon verification.

Phylogenetic analysis of type V systems

Cas protein sequences were collected from (*4*, *13–16*, *24*) and representatives from the TnpB superfamily were collected from (*24*) and top BLAST hits from RefSeq. The resulting set was clustered at 90% amino acid identity to reduce redundancy. A new alignment of CasΦ with the resulting sequence set was generated using MAFFT LINSI with 1000 iterations and filtered to remove columns composed of gaps in 95% of sequences. Poorly aligned sequences were removed and the resulting set was realigned. The phylogenetic tree was inferred using IQTREE v1.6.6 using automatic model selection and 1000 bootstraps.

crRNA sequence analysis

CRISPR-RNA (crRNA) repeats from Phage-encoded CRISPR loci were identified using MinCED (github.com/ctSkennerton/minced) and CRISPRDetect. The repeats were compared by

generating pairwise similarity scores using the Needleman-Wunsch algorithm followed by EMBOSS Needle. A heatmap was built using the similarity score matrix and hierarchical clustering produced dendrograms that were overlaid onto the heatmap to delineate different clusters of repeats.

Generation of plasmids

CasΦ loci, including an additional *E. coli* RBS upstream of *casΦ,* were ordered as G-blocks from Integrated DNA Technologies (IDT) and cloned using Golden Gate assembly (GG) under the control of a tetracycline-inducible promoter for RNA seq and PAM depletion plasmid interference experiments. Perfect repeat-spacer units of the by metagenomics identified CRISPR-arrays were reduced to a single repeat-spacer-repeat unit, amenable to stuffer-spacer exchange by GG-assembly (AarI-restriction sites). Subsequently, *CasΦ* gene sequences were subcloned by GG-assembly into pRSFDuet-1 (Novagen) within MCSI without tags for efficiency of transformation plasmid interference assays, or fused to a C-terminal hexa-histidine tag for protein purification. For plasmid interference assays, mini-CRISPR arrays (repeat-spacer-repeat, or repeat-spacer-HDV ribozyme) amenable to stuffer-spacer exchange by GG-assembly (AarI-restriction sites) were cloned into MCS II of pRSFDuet. For genome editing experiments in human cells, *casΦ* genes were ordered as G-blocks from IDT encoding codon optimized genes for expression in human cells. G-blocks were cloned via GG-assembly into the vector backbone of pBLO62.5, downstream fused to two SV40 NLSs via a GSG linker encoding sequence. The guide encoding sequence of pBLO62.5 was exchanged to encode for a single CRISPR-repeat of the respective homologue, followed by a 20 bp stuffer spacer sequence amenable to GG-assembly exchange using the restriction enzyme SapI. For production of NLS tagged CasΦ for *in planta* genome editing, *E. coli* codon optimized *casΦ* was cloned using GG assembly into MCSI of pRSFDuet-1 (Novagen) downstream fused to two SV40 NLS sequences and a hexa-histidine tag. A list of plasmids and a brief description is given in table S1. Plasmid sequences and maps will be made available on addgene. To reprogram the CasΦ vectors to target different loci, stuffer-spacer were exchanged via GG-assembly to encode the guide for

the selected target site (guide spacer sequences are listed in table S2). Mutations in the *casΦ* genes were introduced by GG-assembly to create *dcasΦ* genes.

PAM depletion DNA interference assay

PAM depletion assays were performed with both, CasΦ plasmids that either carried the whole CasΦ locus as derived from metagenomics (pPP049, pPP056 and pPP062), or with plasmids that contained only the *casΦ* gene and a mini CRISPR (pPP097, pPP102 and pPP107). Assays were performed as three individual biological replicates. Plasmids containing *casΦ* and mini CRISPRs were transformed into *E. coli* BL21(DE3) (NEB) and constructs containing CasΦ genomic loci were transformed into *E. coli* DH5α (QB3-Macrolab, UC Berkeley). Subsequently, electrocompetent cells were prepared by ice cold H_20 and 10 % glycerol washing. A plasmid library was constructed with 8 randomized nucleotides upstream (5') end of the target sequence (kind gift of Hannah Spinner). Competent cells were transformed in triplicate by electroporation with 200 ng library plasmids (0.1 mm electroporation cuvettes (Bio-Rad) on a Micropulser electroporator (Bio-Rad)). After a two-hour recovery period, cells were plated on selective media and colony forming units were determined to ensure appropriate coverage of all possible combinations of the randomized 5' PAM region. Strains were grown at 25 °C for 48 hours on media containing appropriate antibiotics (either 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol, or 100 μg/mL carbenicillin and 50 μg/mL kanamycin) and 0.05 mM isopropyl-β-D-thiogalactopyranoside (IPTG), or 200 nM anhydrotetracycline (aTc), depending on the vector to ensure propagation of plasmids and CasΦ effector production. Subsequently, propagated plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen).

PAM depletion sequencing analysis

Amplicon sequencing of the targeted plasmid was used to identify PAM motifs that are preferentially depleted. Sequencing reads were mapped to the respective plasmids and PAM randomized regions were extracted. The abundance of each possible 8 nucleotide combination was counted from the

aligned reads and normalized to the total reads for each sample. Enriched PAMs were computed by calculating the log ratio compared to the abundance in the control plasmids, and were used to produce sequence logos.

RNA preparation for RNAseq

Plasmids containing CasΦ loci were transformed into chemically competent *E. coli* DH5α (QB3-Macrolab, UC Berkeley). Preparations were performed as three individual biological replicates. Single colonies were picked to inoculate 5 mL starter cultures (LB, 34 μg/mL chloramphenicol) which were incubated at 37 °C shaking vigorously overnight. The next morning, main cultures were inoculated 1:100 (LB, 34 μg/mL chloramphenicol) and locus expression was induced with 200 nM aTc for 24 h at 16 °C. Cells were harvested by centrifugation, resuspended in lysis buffer (20 mM Hepes-Na pH 7.5 RT, 200 mM NaCl) and lysed using glass beads (0.1 mm glass beads, 4x 30 s vortex at 4 °C, interspaced by 30 s cool-down on ice). 200 μL cell lysis supernatant were transferred into Trizol for RNA extraction according to the manufacturer's protocol (Ambion). 10 μg RNA were treated with 20 units of T4-PNK (NEB) for 6 h at 37 °C for 2′-3′-dephosphorylation. Subsequently, 1 mM ATP was added and the sample was incubated for 1 h at 37 °C for 5′-phosphorylation before heat inactivation at 65 °C for 20 min and subsequent Trizol purification.

RNA analysis by RNAseq

cDNA libraries were prepared using the RealSeq-AC miRNA library kit illumina sequencing (somagenics). cDNA libraries were subjected to Illumina MiSeq sequencing, and raw sequencing data was processed to remove adapters and sequencing artifacts, and high-quality reads were maintained. The resulting reads were mapped to their respective plasmids to determine the CRISPR locus expression and crRNA processing, and coverage was calculated at each region.

Efficiency of transformation plasmid interference assay

CasΦ vectors were transformed into chemically competent *E. coli* BL21(DE3) (NEB). Individual colonies for biological replicates were picked to inoculate three 5 mL (LB, Kanamycin 50 μg/mL) starter cultures to prepare electrocompetent cells the following day. 50 mL (LB, Kanamycin 50 μg/mL) main cultures were inoculated 1:100 and grown vigorously shaking at 37 °C to an OD_{600} of 0.3. Subsequently, the cultures were cooled to room temperature and *casΦ* expression was induced with 0.2 mM IPTG. Cultures were grown to an OD_{600} of 0.6-0.7 at 25 °C, before preparation of electrocompetent cells by repeated ice-cold H_20 and 10% glycerol washes. Cells were resuspended in 250 μL 10% glycerol. 90 μL aliquots were flash frozen in liquid nitrogen and stored at -80 °C. The next day, 80 μL competent cells were combined with 3.2 μL plasmid (20 ng/μL pUC19 target plasmid, or 20 ng/μL pYTK001 control plasmid), incubated for 30 min on ice and split into three individual 25 μL transformation reactions. After electroporation in 0.1 mm electroporation cuvettes (Bio-Rad) on a Micropulser electroporator (Bio-Rad), cells were recovered in 1 mL recovery medium (Lucigen) supplemented with 0.2 mM IPTG, shaking at 37 °C for one hour. Subsequently, 10-fold dilution series were prepared and 5 μL of the respective dilution steps were spot-plated on LB-Agar containing the appropriate antibiotics. Plates were incubated overnight at 37 °C and colonies were counted the following day to determine the transformation efficiency. To assess the transformation efficiency, the mean and standard deviations were calculated from the cell forming units per ng transformed plasmids for the electroporation triplicates.

Protein production and purification

CasΦ overexpression vectors were transformed into chemically competent *E. coli* BL21(DE3)-Star (QB3-Macrolab, UC Berkeley) and incubated overnight at 37 °C on LB-Kan agar plates (50 μg/mL Kanamycin). Single colonies were picked to inoculate 80 mL (LB, Kanamycin 50 µg/mL) starter cultures which were incubated at 37 °C shaking vigorously overnight. The next day, 1.5 L TB-Kan medium (50 μg/mL Kanamycin) were inoculated with 40 mL starter culture and grown at 37 °C to an OD_{600} of 0.6, cooled down on ice for 15 min and gene expression was subsequently induced with 0.5 mM IPTG

followed by incubation overnight at 16 °C. Cells were harvested by centrifugation and resuspended in wash buffer (50 mM HEPES-Na pH 7.5 RT, 1 M NaCl, 20 mM imidazole, 5 % glycerol and 0.5 mM TCEP), subsequently lysed by sonication, followed by lysate clarification by centrifugation. The soluble fraction was loaded on a 5 mL Ni-NTA Superflow Cartridge (Qiagen) pre-equilibrated in wash buffer. Bound proteins were washed with 20 column volumes (CV) wash buffer and subsequently eluted in 5 CV elution buffer (50 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 500 mM imidazole, 5 % glycerol and 0.5 mM TCEP). The eluted proteins were concentrated to 1 mL before injection into a HiLoad 16/600 Superdex 200pg column (GE Healthcare) pre-equilibrated in size-exclusion chromatography buffer (20 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 5 % glycerol and 0.5 mM TCEP). Peak fractions were concentrated to 1 mL and concentrations were determined using a NanoDrop 8000 Spectrophotometer (Thermo Scientific). Proteins were purified at a constant temperature of 4 °C and concentrated proteins were kept on ice to prevent aggregation, snap frozen in liquid nitrogen and stored at -80 °C. AsCas12a was purified as previously described (*25*).

In vitro cleavage assays - spacer tiling

Plasmid targets were cloned by GG-assembly of spacer 2, found in the CRISPR-array of CasΦ-1, downstream to a cognate 5'-TTA PAM, or non-cognate 5'-CCA PAM into pYTK095 (Target sequences are given in table S3). Supercoiled plasmids were prepared by propagation of the plasmid overnight at 37 °C in *E. coli* Mach1 (QB3-Macrolab, UC Berkeley) in LB and Carbenicillin (100 μg/mL) and subsequent preparation using a Qiagen Miniprep kit (Qiagen). Linear DNA targets were prepared by PCR from the plasmid target. crRNA guides were ordered as synthetic RNA oligos from IDT (table S4), dissolved in DEPC H_20 and heated for 3 min at 95 °C before cool down at RT. Active RNP complexes were assembled at a concentration of 1.25 μM by mixing protein and crRNA (IDT) in a 1:1 molar ratio in cleavage buffer (10 mM Hepes-K pH 7.5 RT, 150 mM KCl, 5 mM $MgCl_2$, 0.5 mM TCEP) and incubation at RT for 30 min. Cleavage reactions were initiated by addition of DNA (10 nM) to preformed RNP (1 μM) in reaction buffer (10 mM Hepes-K pH 7.5 RT, 150 mM KCl, 5 mM $MgCl_2$, 0.5 mM TCEP). The

reactions were incubated at 37 °C, quenched with 50 mM EDTA and stored in liquid nitrogen. Samples were thawed and treated with 0.8 units proteinase K (NEB) for 20 min at 37 °C. Loading dye was added (Gel Loading Dye Purple 6X, NEB) and samples were analyzed by electrophoresis on a 1% agarose gel and stained with SYBR Safe (Thermo Fisher Scientific). For comparison to cleavage products, supercoiled plasmids were digested with PciI (NEB) for linearization and Nt.BstNBI (NEB) for plasmid nicking and open circle formation. Comparable cleavage assays under varied conditions ($n \geq 3$) showed consistent results.

In vitro cleavage assays - radiolabeled nucleic acids

Active CasΦ RNP complexes were assembled in a 1:1.2 molar ratio by diluting CasΦ protein to 4 μM and crRNA (IDT) to 5 μM in RNP assembly buffer (20 mM HEPES-Na pH 7.5 RT, 300 mM KCl, 10 mM $MgCl_2$, 20 % glycerol, 1 mM TCEP) and incubation for 30 min at RT. Substrates were 5′-end-labelled using T4-PNK (NEB) in the presence of ^{32}P-γ-ATP (Substrate sequences are given in table S3). Oligo-duplex targets were generated by combining ^{32}P-labelled and unlabelled complementary oligonucleotides in a 1:1.5 molar ratio. Oligos were hybridized to a DNA-duplex concentration of 50 nM in hybridization buffer (10 mM Tris-Cl pH 7.5 RT, 150 mM KCl), by heating for 5 min to 95 °C and a slow cool down to RT in a heating block. Cleavage reactions were initiated by combining 200 nM RNP with 2 nM substrate in reaction buffer (10 mM HEPES-Na pH 7.5 RT, 150 mM KCl, 5 mM $MgCl_2$, 10 % glycerol, 0.5 mM TCEP) and subsequently incubated at 37 °C. For trans-cleavage assays, guide complementary activator substrates were diluted in oligonucleotide hybridization buffer (10 mM Tris pH 7.8 RT, 150 mM KCl) to a concentration of 4 μM, heated to 95 °C for 5 min, and subsequently cooled down at RT to allow duplex formation for double stranded activator substrates. Cleavage reactions were set up by combining 200 nM RNP with 100 nM activator substrate and incubation for 10 min at RT before addition of 2 nM ssDNA, or ssRNA, trans cleavage substrates. Reactions were stopped by addition of two volumes formamide loading buffer (96 % formamide, 100 μg/mL bromophenol blue, 50 μg/mL

xylene cyanol, 10 mM EDTA, 50 μg/mL heparin), heated to 95 °C for 5 min, and cooled down on ice before separation on a 12.5 % denaturing urea-PAGE. Gels were dried for 4 h at 80 °C before phosphor-imaging visualization using an Amersham Typhoon scanner (GE Healthcare). Technical replicates ($n \geq 2$) and comparable cleavage assays under varied conditions ($n \geq 3$) of biological replicates ($n \geq 2$) showed consistent results. Bands were quantified using ImageQuant TL (GE) and cleaved substrate was calculated from the intensity relative to the intensity observed at t = 0 min. Curves were fit to a One-Phase-Decay model in Prism 8 (graphpad) to derive the rate of cleavage.

In vitro pre-crRNA processing assay

Pre-crRNA substrates were 5′-end-labelled using T4-PNK (NEB) in the presence of ^{32}P-γ-ATP (Substrate sequences are given in table S3). Processing reactions were initiated by combining 50 nM CasΦ with 1 nM substrate in pre-crRNA processing buffer (10 mM Tris pH 8 RT, 200 mM KCl, 5 mM $MgCl_2$ or 25 mM EDTA, 10 % glycerol, 1 mM DTT) and subsequently incubated at 37 °C. Substrate hydrolysis ladders were prepared using the alkaline hydrolysis buffer according to the manufacturer's protocol (Ambion). 10 μL of the processing reaction products were treated with 10 units T4-PNK (NEB) for 1 h at 37 °C in the absence of ATP for termini chemistry analysis. Reactions were stopped by addition of two volumes formamide loading buffer (96 % formamide, 100 μg/mL bromophenol blue, 50 μg/mL xylene cyanol, 10 mM EDTA, 50 μg/mL heparin), heated to 95 °C for 3 min, and cooled down on ice before separation on a 12.5 %, or 20 %, denaturing urea-PAGE. Gels were dried for 4 h at 80 °C before phosphor-imaging visualization using an Amersham Typhoon scanner (GE Healthcare). Technical replicates ($n \geq 3$) and comparable cleavage assays under varied conditions ($n \geq 3$) of biological replicates ($n \geq 2$) showed consistent results. Bands were quantified using ImageQuant TL (GE) and processed RNA was calculated from the intensity at t = 60 min relative to the intensity observed at t = 0 min.

Analytical size exclusion chromatography

500 μL sample (5-10 μM protein, RNA, or reconstituted RNPs) were injected onto a S200 XK10/300 size exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated in SEC buffer (20 mM HEPES-Cl pH 7.5 RT, 250 mM KCl, 5 mM $MgCl_2$, 5 % glycerol and 0.5 mM TCEP). Prior to SEC, CasΦ RNP complexes were assembled by incubating CasΦ protein and pre-crRNA for 1 h in 2X pre-crRNA processing buffer (20 mM Tris pH 8 RT, 400 mM KCl, 10 mM $MgCl_2$, 20 % glycerol, 2 mM DTT).

Genome editing in human cells

The GFP HEK293 reporter cells were generated via lentiviral integration as previously described (*26*). Cells were routinely tested for absence of mycoplasma using the MycoAlert Mycoplasma Detection Kit (Lonza), according to the manufacturer's protocol. GFP HEK293 reporter cells were seeded into 96-well plates and transfected at 60-70% confluency the next day according to the manufacturer's protocol with lipofectamine 3000 (Life Technologies) and 200 ng of plasmid DNA encoding the CasΦ gRNA and CasΦ–P2A–PAC fusion. As a comparison control, 200 ng of plasmid DNA encoding the SpyCas9 sgRNA and SpyCas9–P2A–PAC fusion was transfected identically, with target sequences adjusted for PAM differences. 24 hours post-transfection, successfully transfected cells were selected for by adding 1.5 μg/mL puromycin to the cell culture media for 72 hours. Cells were passaged regularly to maintain sub-confluent conditions and then analyzed on an Attune NxT Flow Cytometer with an autosampler. Cells were analyzed on the flow cytometer after 10 days to allow for clearance of GFP from cells.

Protoplast isolation and transfection

A. thaliana plants (Col-0 ecotype) were grown with 12 h light/12 h dark photoperiod under low light (75 μE m^{-2} s^{-1}) and mesophyll protoplasts were isolated from leaves of 4-week-old plants as described previously (*27*) . In brief, *A. thaliana* leaves were cut into 0.5-1 mm stripes with sharp razor

blades and submerged in enzyme solution (20 mM MES pH 5.7, 0.4 M mannitol, 20 mM KCl, 1.5% cellulase R10, 0.4% macerozyme R10, enzymes from Yakult Pharmaceutical Ind. Co., Ltd., Japan). The leaf stripes in enzyme solution were vacuum infiltrated for 30 min in dark and then incubated in dark for 3 h at room temperature. The protoplasts were released during this incubation. After the incubation, the enzyme/protoplast solution was diluted with equal volume of W5 solution (2 mM MES pH 5.7, 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl), and filtered through 70-μm nylon mesh (Carolina Biological Supplies, cat 65222N) into round bottom tubes. Protoplasts were collected by centrifuging the flow-through at 100 g for 2 min at 4 °C. Supernatant was removed and protoplasts (pellet) were resuspended in W5 solution at 2 x 10^5 cells/ml. Resuspended protoplasts were kept on ice for 30 min for resting. During the resting, the protoplasts were re-collected at the bottom of tubes by gravity. Then the supernatant was removed as much as possible and the protoplasts were resuspended with MMG solution (4 mM MES PH 5.7, 0.4 M mannitol, 15 mM $MgCl_2$) to the same volume (2 x 10^5 cells/ml). Sterile and RNase-free reagents were used for protoplast isolation. Active CasΦ-2 RNP complexes were reconstituted by diluting CasΦ-2-NLS protein, purified as described above, to 4 μM and gRNA to 5 μM in RNP assembly buffer as described above and incubated for 30 min at RT. 26 μL of 4 μM RNP were first added to a round-bottom 2 mL tube. Then 200 μL of protoplasts (at 2 x 10^5 cells/mL) were added to the tube. 2 μL of 5 μg/μL salmon sperm DNA was added and mixed gently by tapping the tube 3-4 times. Then, 228 μL of fresh, sterile and RNase free PEG-$CaCl_2$ solution (40% PEG4000, 0.2 M mannitol, 100 mM $CaCl_2$) was added to the protoplast-RNP mixture and mixed well by gently tapping tubes. The protoplasts with PEG solution were incubated at room temperature for 10 min, then 880 μL of W5 solution was added and mixed with the protoplasts by inverting the tube 2-3 times to stop the transfection. Protoplasts were harvested by centrifugation at 100 rcf for 2 min, resuspended in 1 mL WI solution (4 mM MES pH 5.7, 0.5 M mannitol, 20 mM KCl) and plated into 6-well plates pre-coated with 5% calf serum. The lids of the 6-well plates were closed to begin the incubation of the protoplasts. For control samples, 10 μg of HBT-sGFP plasmid (ABRC stock CD3-911) were added to 200 μl protoplasts and followed the same transfection and plating procedure as stated above. For the initial RNP screening experiment, the protoplasts were

incubated at RT for 12 h, then moved to 37 °C for 2.5 h. Then, the protoplasts were moved back to room temperature and incubated for a total duration of 36 h. For the independent experiment where gRNA28, gRNA31 and gRNA33 were tested, the protoplasts were incubated at RT for 12 hours, then moved to 37 °C for 2.5 h. Then, the protoplasts were moved back to room temperature and incubated for a total duration of 48 h. At the end of the incubations, the protoplasts were collected by a first centrifugation at 100 rcf for 2-3 min. Keeping the pellet, the supernatant was moved to another tube and went through another centrifugation at 3000 rcf for 3 min to collect any residue protoplasts. Pellets from these two centrifugations were combined and flash frozen for further analysis.

Amplicon sequencing

DNAs of protoplast samples were extracted using the Qiagen DNeasy plant mini kit. Amplicons were obtained by two rounds of PCR (2-step PCR). Amplification primers for the first round of PCR were designed to have the 3′ part of primer with sequences flanking a 200-300 bp fragment of the *AtPDS3* gene around the guide RNA of interest. The 5′ part of the primer contained sequences to be bound by common sequencing primers (for reading paired-end reads, read 1 and read 2). The primers were designed so that the gRNA sequence started from within 100 bp from the beginning of read 1. The first round of PCR was done with Phusion High-Fidelity Polymerase (ThermoFisher cat F530N). Half of all DNA from a protoplast transfection sample was used as the template, and 25 cycles of amplification were done for the first round. Then the reaction was cleaned by 1x Ampure XP beads (Beckman Coulter A63881). The elution from the cleanup was used as the template for the second round of PCR by Phusion High-Fidelity Polymerase with 12 cycles. The second round of PCR was designed so that indices were added to each sample. The samples were then purified by 0.8-1 X Ampure beads for 1-2 rounds until no primer dimers were seen, with fragments below 200 bp considered primer dimers. Then amplicons were sent for paired-end 150 bp next generation sequencing.

Amplicon sequencing result analysis

Reads were first quality- and adaptor-trimmed with trim-galore (version 0.4.4), then mapped to the *AtPDS3* genomic region. Sorted and indexed bam files were used as input files for further analysis by the CrispRvariants R package. Each mutation pattern with corresponding reads counts were exported by the CrispRvariants R package. After assessing all control samples, a criterion to classify reads containing deletion was established: only reads with ≥ 3 bp deletion of same pattern (deletion of same size starting with same location) with ≥ 100 reads counts from a sample were counted into the reads number with deletion. This criterion was established due to the fact that 1 bp indels and occasionally 2 bp deletions were observed with reads number >100 in control samples. Larger deletions were also observed at very low frequencies in control samples. These observations indicate that occasional PCR inaccuracy and low-quality sequencing in a small fraction of reads can result in the deletion patterns with corresponding read number ranges as stated above in control samples. These stringent criteria were employed so that the counted deletion signals were true signals indicating editing events, though it is possible that CasΦ-2 might be able to create 1-2 bp deletions at lower frequency.

Supplementary Text

To assess the role of the RuvC domain in DNA cleavage, the active site was mutated (D371A, D394A, or D413A) to produce a deactivated CasΦ variant (dCasΦ) that did not cleave dsDNA, ssDNA or ssRNA *in vitro* (fig. S6A, B). When expressed in *E. coli* along with crRNA, dCasΦ could not prevent transformation of a crRNA-complementary plasmid, consistent with a requirement for RuvC-catalyzed DNA cutting (fig. S3). This observation, together with the delayed cleavage of the TS after NTS cleavage (Fig. 2D; fig. S7), suggests that CasΦ cleaves each strand sequentially within the RuvC active site. Sequential strand cleavage is consistent with the dsDNA cutting mechanism of the type V CRISPR-Cas proteins (*28*, *29*) that share closest evolutionary origin with CasΦ.

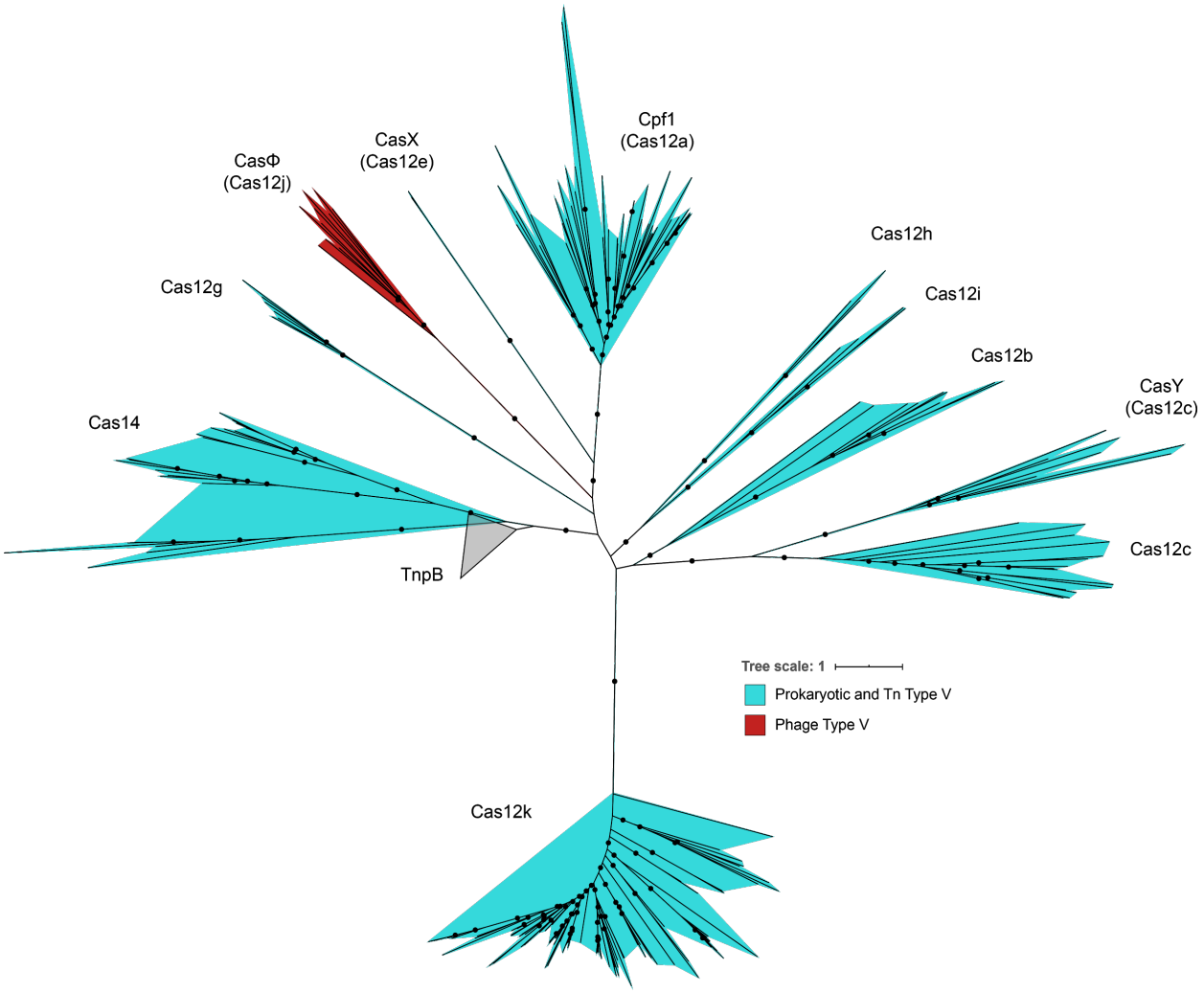

Fig. S1. Maximum likelihood phylogenetic tree of type V subtypes a-k. Phage-encoded CasΦ proteins are outlined in red, with prokaryote and transposon-encoded proteins in blue. Bootstrap and approximate likelihood ratio test values >90 are shown on the branches (circles).

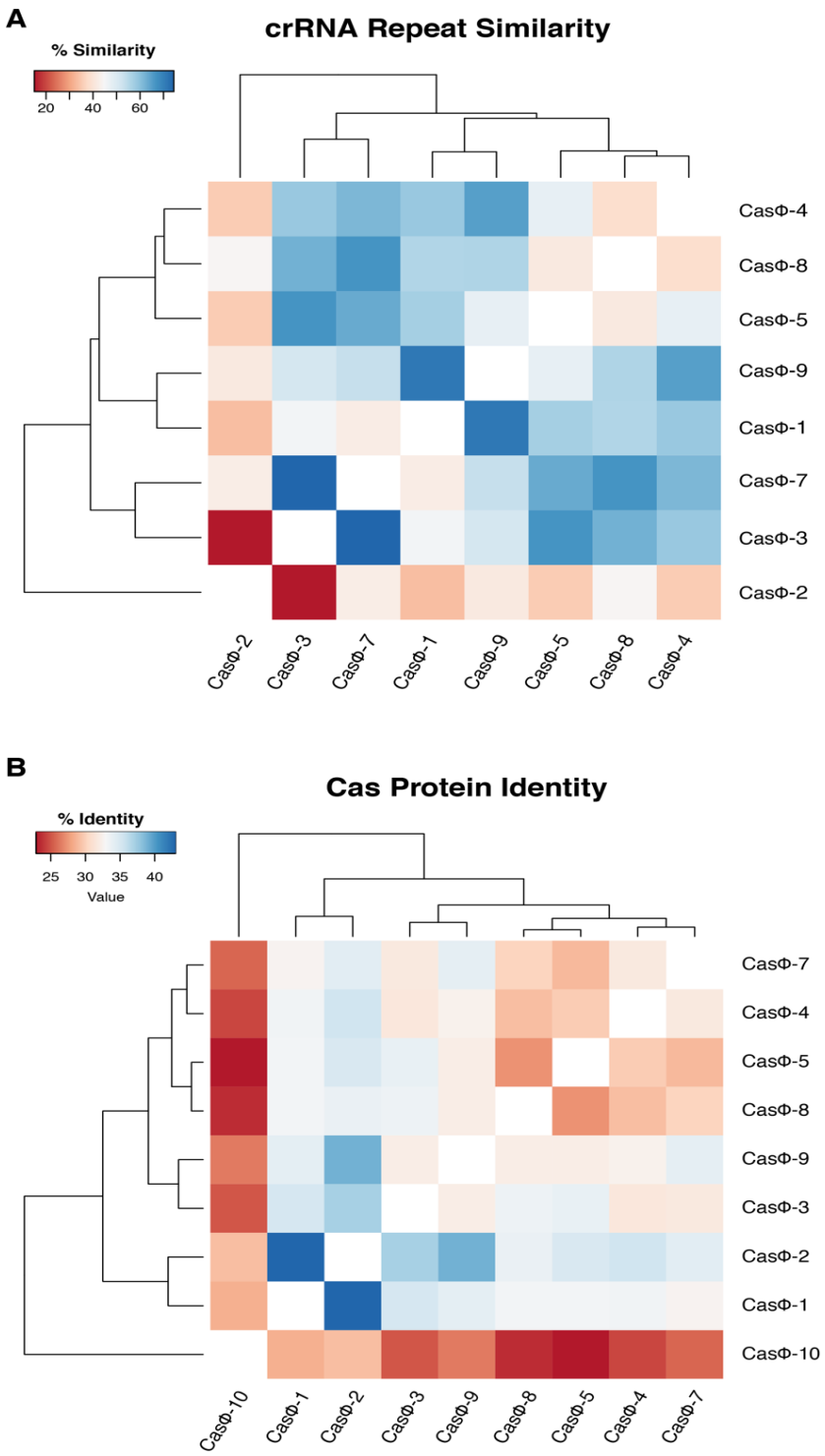

Fig. S2. CasΦ systems are highly diverse. A similarity matrix was built and visualized using a heatmap and hierarchical clustering dendrogram for CasΦ crRNA repeats (A) and CasΦ protein sequences (B). CasΦ-1, CasΦ-2, and CasΦ-3 were selected for analysis based on divergence of their protein and crRNA repeat sequences.

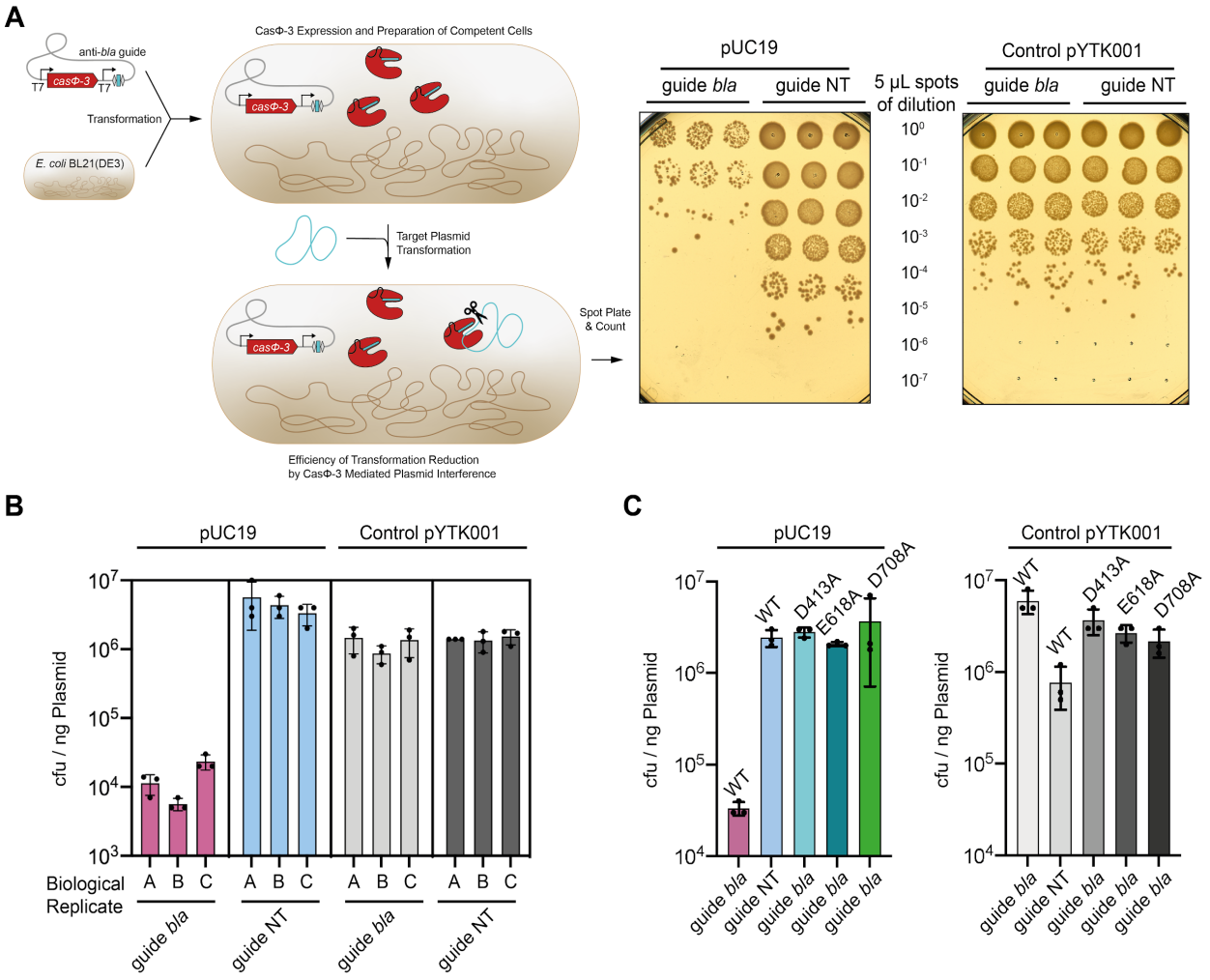

Fig. S3. CasΦ-3 protects against plasmid transformation. (**A**) Scheme illustrating the efficiency of transformation (EOT) assay. (**B**) EOT assay showing that CasΦ, programmed by a beta-lactamase (*bla*) gene targeting guide, reduces the efficiency of pUC19 transformation (red bars). Experiments were performed in three biological replicates and technical electroporation transformation triplicates (dots; n = 3 each, mean ± s.d.). Competent cells were tested for general transformation efficiency (grey bars) by transformation of pYTK095, which is not targeted by the tested *bla* and NT (non-targeting) guide. (**C**) EOT in dependence of CasΦ-3 RuvC active site residue variation (RuvCI: D413A; RuvCII: E618A; RuvCIII: D708A). N = 3 each, mean ± s.d.. Competent cells were tested for general transformation efficiency (grey bars).

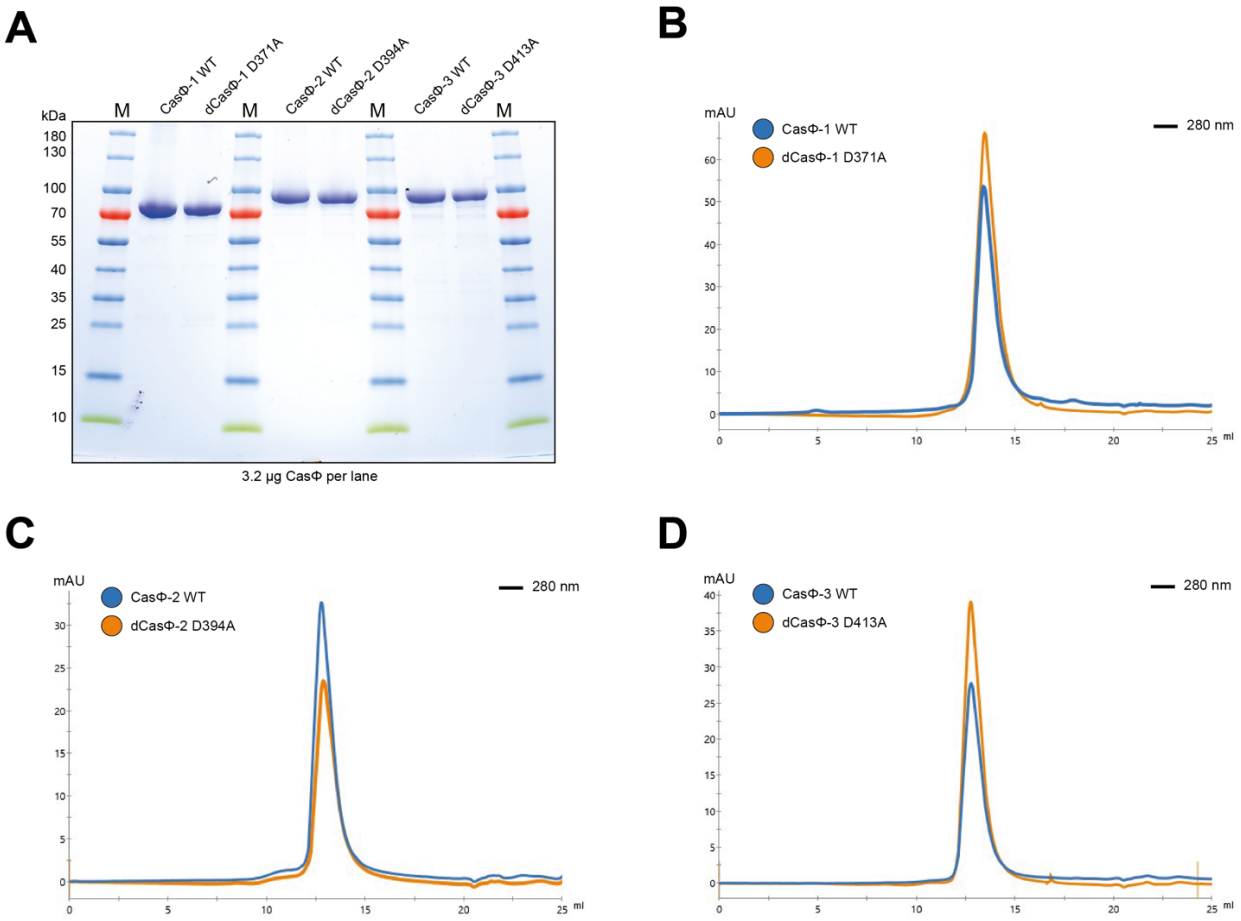

Fig. S4. Purification of apo CasΦ. (**A**) SDS-PAGE of the purified apo CasΦ orthologs and their dCasΦ variants. (**B**) Analytical size-exclusion chromatography (S200) of CasΦ-1 WT (blue trace) and dCasΦ-1 (orange trace). (**C**) Analytical size-exclusion chromatography (S200) of CasΦ-2 WT (blue trace) and dCasΦ-2 (orange trace). **D**) Analytical size-exclusion chromatography (S200) of CasΦ-3 WT (blue trace) and dCasΦ-3 (orange trace).

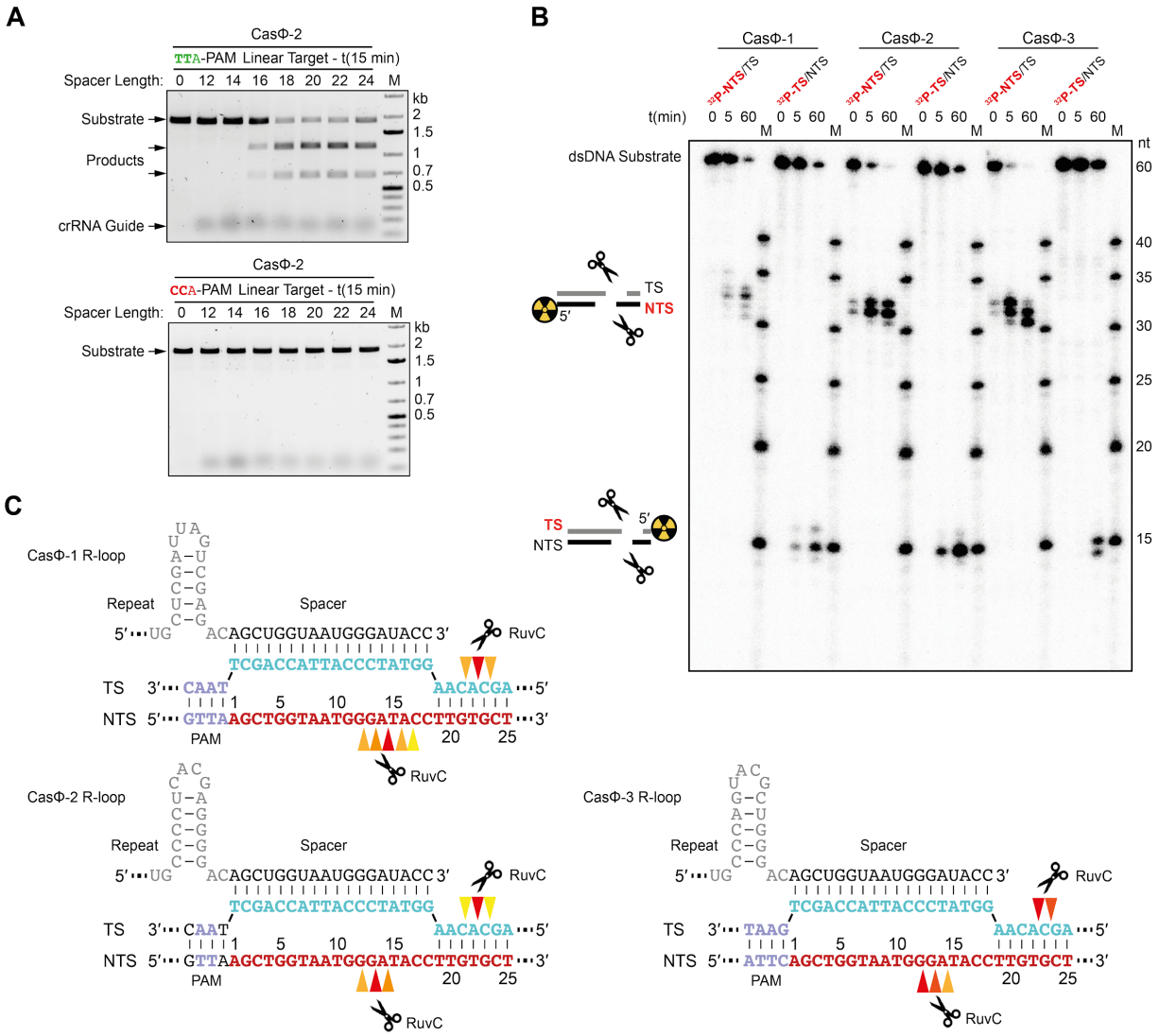

Fig. S5. CasΦ targets DNA *in vitro* to produce staggered cuts. (**A**) Linear PCR-fragment cleavage assay testing CasΦ RNPs reconstituted with crRNAs of different spacer lengths in presence of a cognate 5′-TTA-3′ PAM (above), or non-cognate 5′-CCA-3′ PAM (below). (**B**) Cleavage assay targeting dsDNA oligo-duplices for mapping of the cleavage structure. (**C**) Scheme illustrating the cleavage pattern of the staggered cuts. Shown are the proposed R-loop (replication loop) structures formed by CasΦ upon target DNA binding to the crRNA spacer.

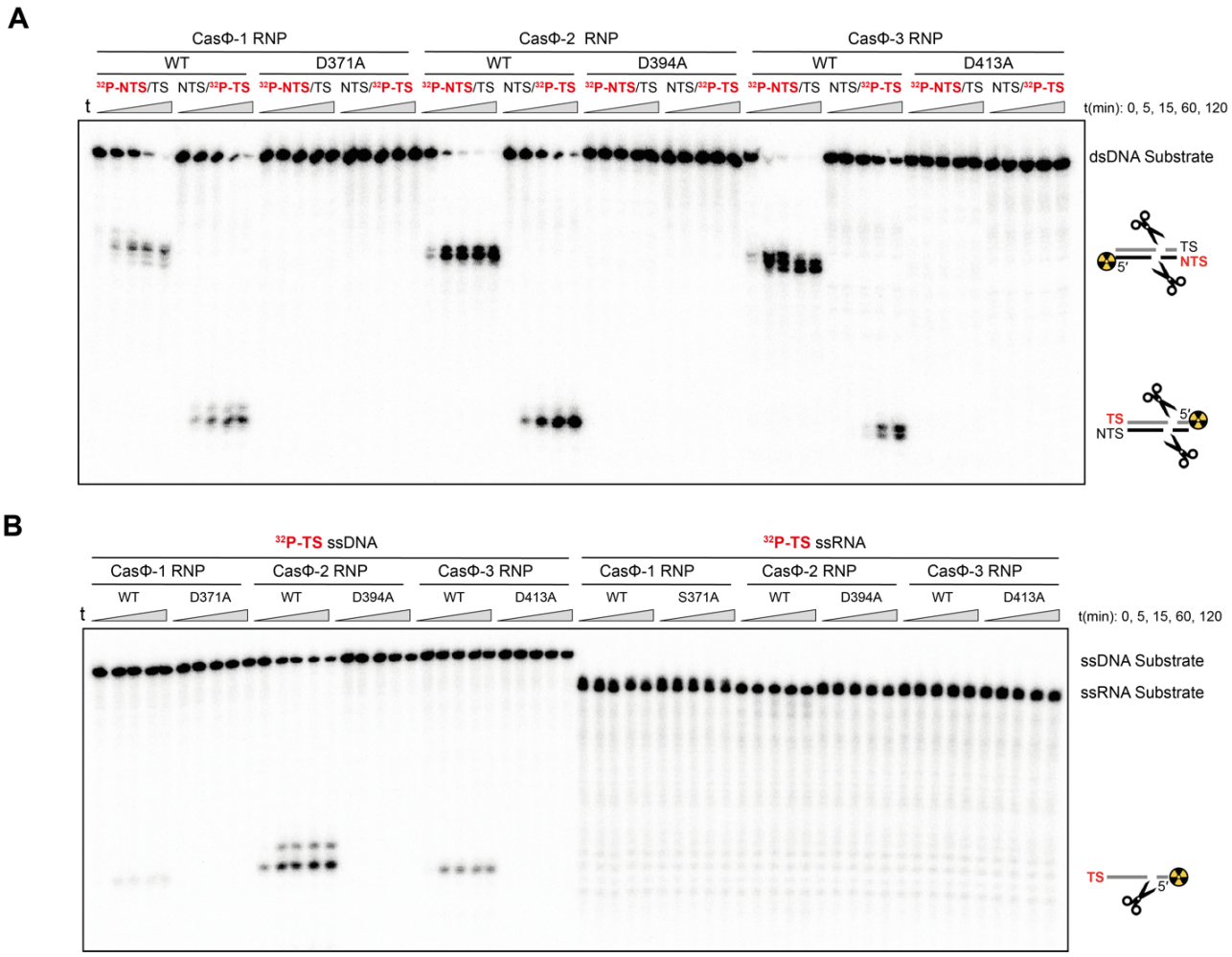

Fig. S6. CasΦ targets dsDNA and ssDNA, but not RNA *in vitro*. (**A**) Cleavage assay assessing the ability of CasΦ and dCasΦ variant (D371A, D394A and D413A) RNPs to cleave the target strand (TS), and non-target strand (NTS), of a dsDNA oligo duplex. (**B**) Cleavage assay testing the ability of CasΦ and dCasΦ variant (D371A, D394A and D413A) RNPs to target and cleave a single stranded DNA, or RNA, target strand.

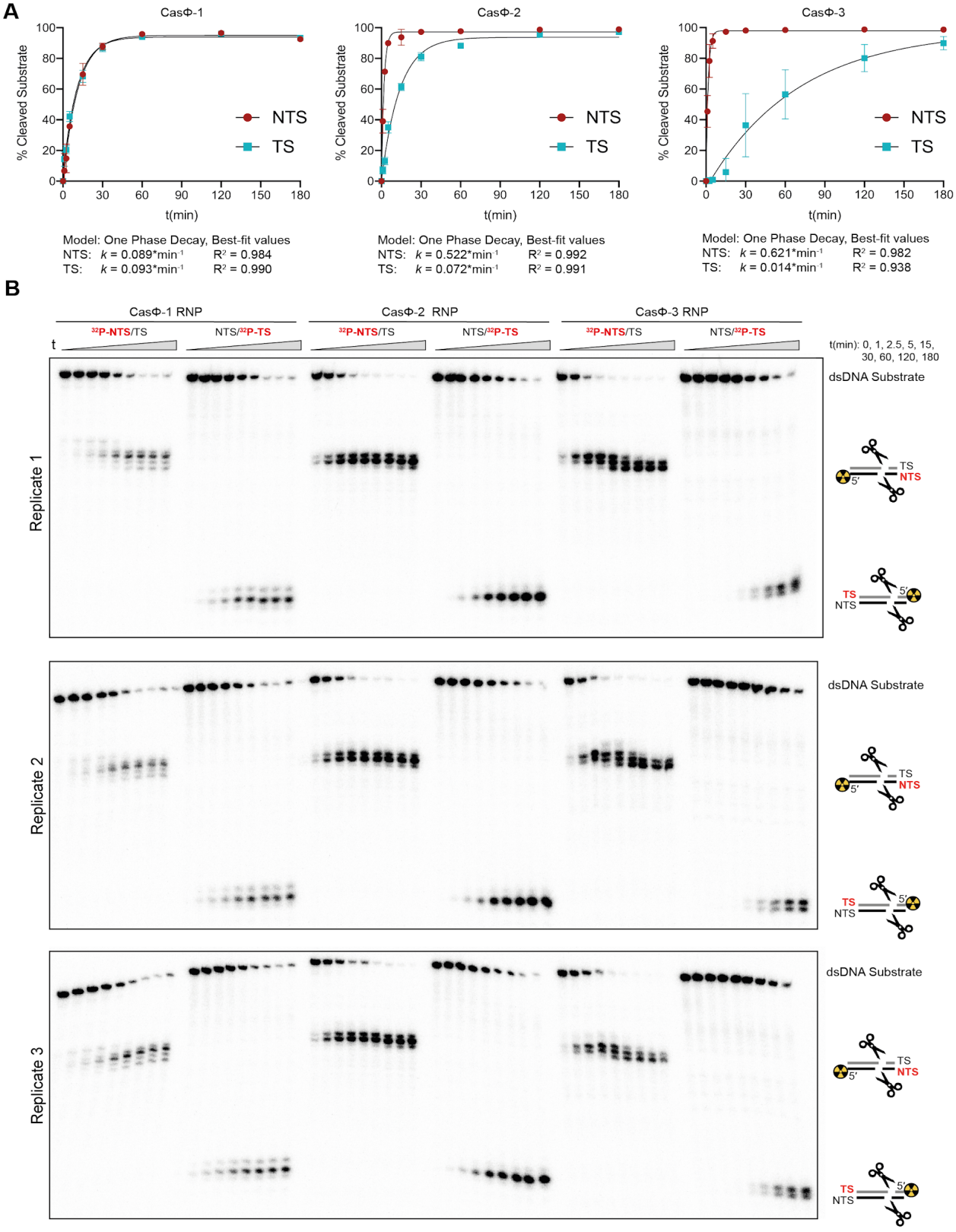

Fig. S7. Cleavage assay comparing TS and NTS cleavage efficiency by CasΦ. (**A**) Cleavage assay curves, fit to the One Phase Decay model using Prism 8 (GraphPad) (n = 3 each, mean ± s.d.). Cleaved

fractions are calculated based on the substrate band intensities at t = (0 min) (panel B) relative to the respective time point. (**B**) Urea-PAGE gels of the three independent reaction replicates (Replicates 1, 2 and 3). This panel also relates to main text Figure 2D for CasΦ-2.

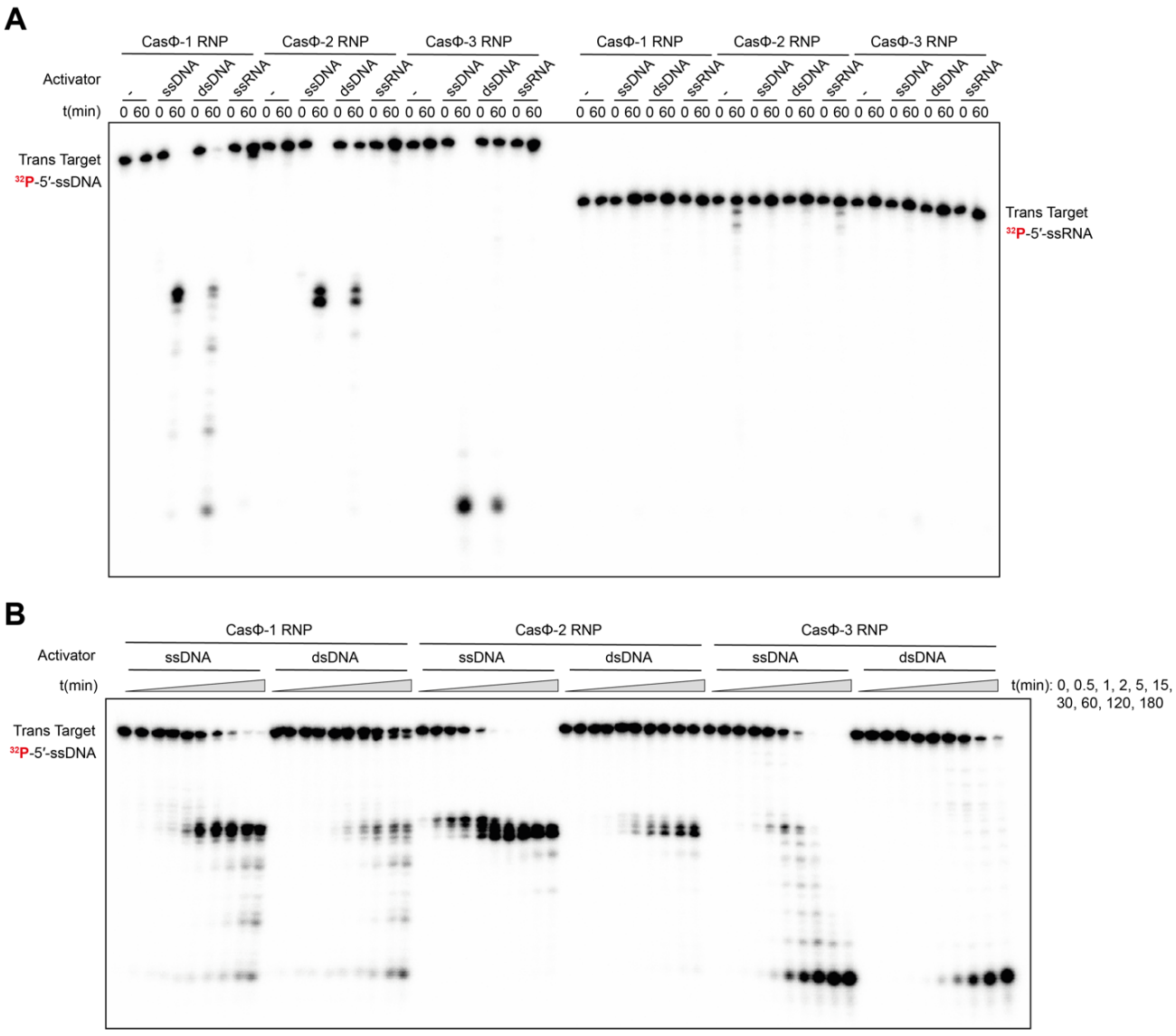

Fig. S8. CasΦ targets ssDNA, but not RNA, *in trans* upon activation *in cis*. (**A**) Cleavage assay comparing the trans-cleavage activities of CasΦ-1, CasΦ-2 and CasΦ-3 on ssDNA and ssRNA as targets *in trans* in dependence of either ssDNA, dsDNA, or ssRNA as activators *in cis*. (**B**) Cleavage assay comparing the trans-cleavage activity of CasΦ-1, CasΦ-2 and CasΦ-3.

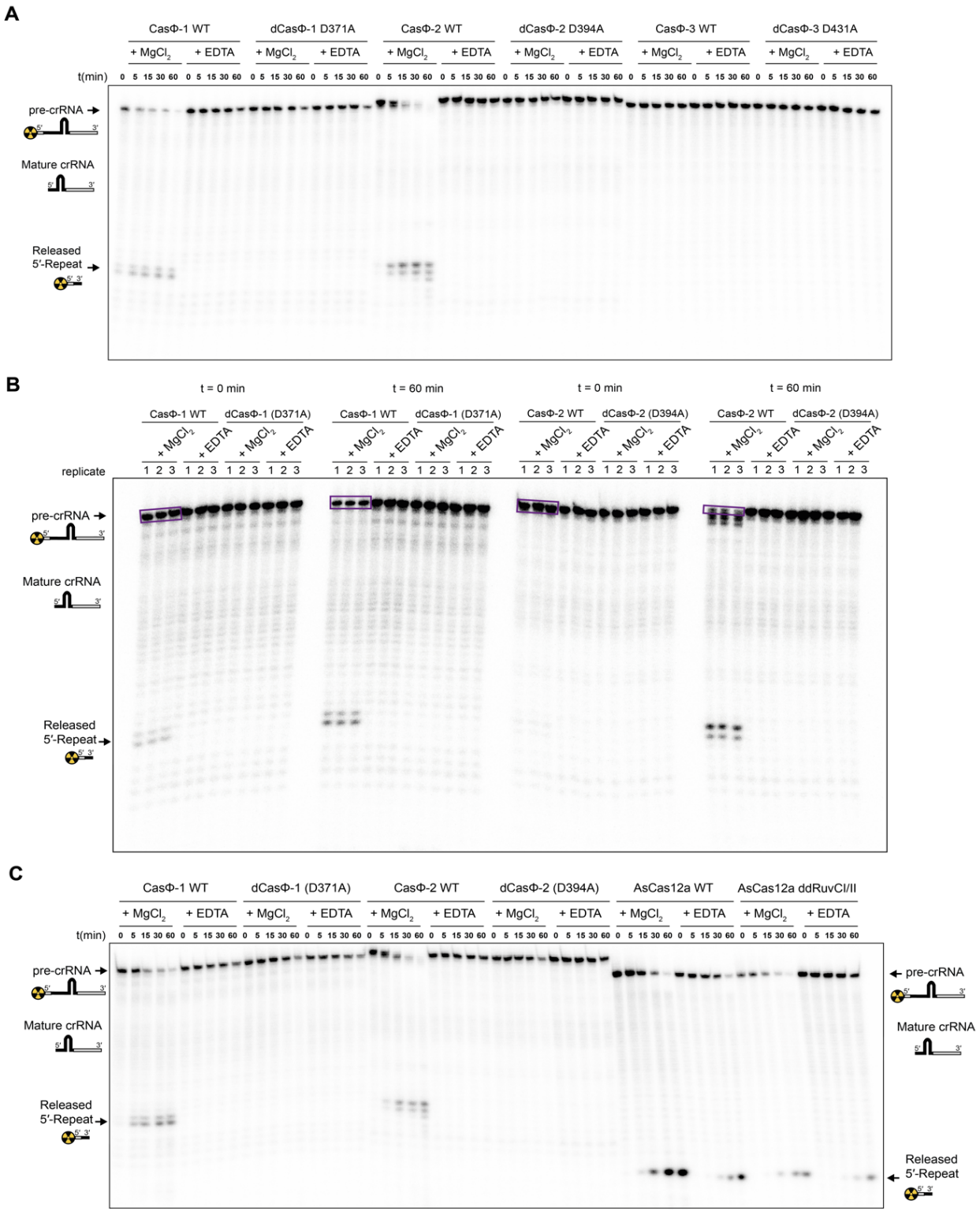

Fig. S9. CasΦ-1 and CasΦ-2, but not CasΦ-3, process pre-crRNA. (**A**) Assay testing pre-crRNA processing for CasΦ-1, CasΦ-2 and CasΦ-3 in dependence of Mg^{2+} and RuvC active site catalytic

residues (dCasΦ variants). (**B**) Processing reaction replicates for CasΦ-1 and CasΦ-2 at t = 0 min and t = 60 min. Purple squares indicate quantified bands. This panel relates to main text Figure 3B. (**C**) Assay testing pre-crRNA processing for CasΦ-1, CasΦ-2 and AsCas12a in dependence of Mg^{2+} and RuvC active site catalytic residues (dCasΦ variants).

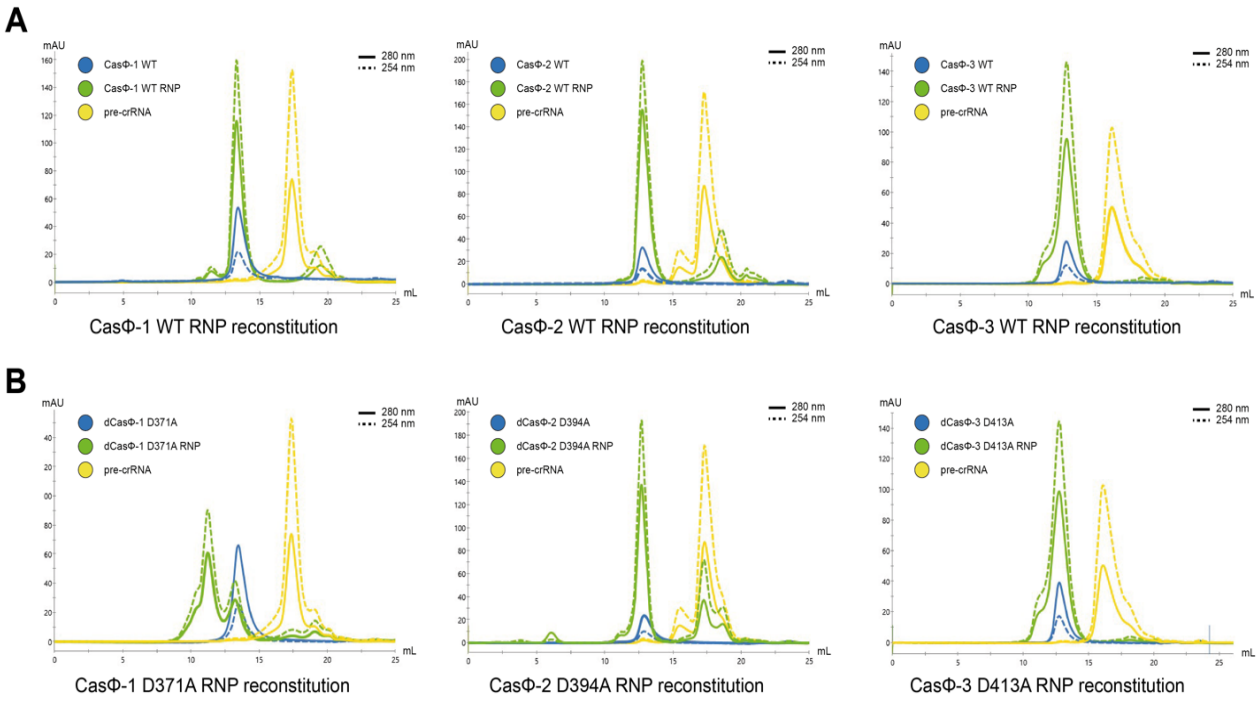

Fig. S10. CasΦ WT and dCasΦ proteins form RNPs with pre-crRNA. (**A**) Analytical size-exclusion chromatography (S200) of wild-type proteins (blue trace), pre-crRNA (yellow trace), and their respective reconstituted RNP (green trace). (**B**) Analytical size-exclusion chromatography (S200) of dCasΦ variant proteins (blue trace), pre-crRNA (yellow trace), and their respective reconstituted RNP (green trace).

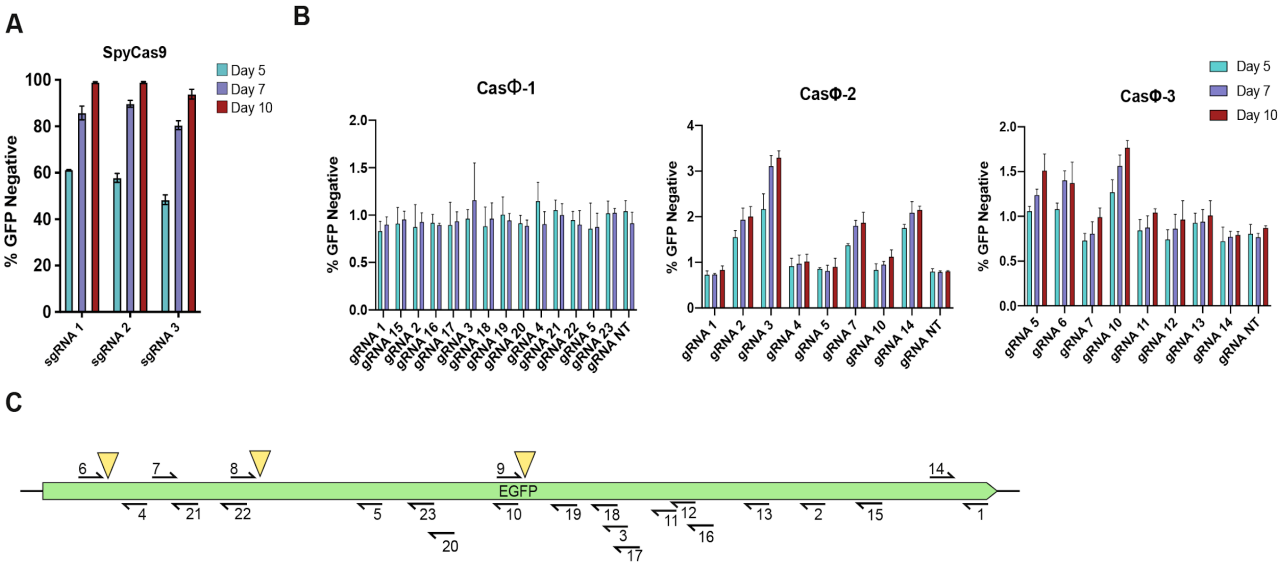

Fig. S11. CasΦ mediated EGFP gene disruption in HEK293 cells. (**A**) EGFP disruption by three *Streptococcus pyogenes* Cas9 guides (n = 3 each, mean ± s.d.) (**B**) CasΦ guides that had an EGFP disruption below 5 % (n = 3 each, mean ± s.d.). (**C**) EGFP map showing the target sites and orientation of guides (arrows and numbers). Yellow triangles indicate the best guides for gene disruption (relates to main text figure 4A). Guide sequences are listed in table S2.

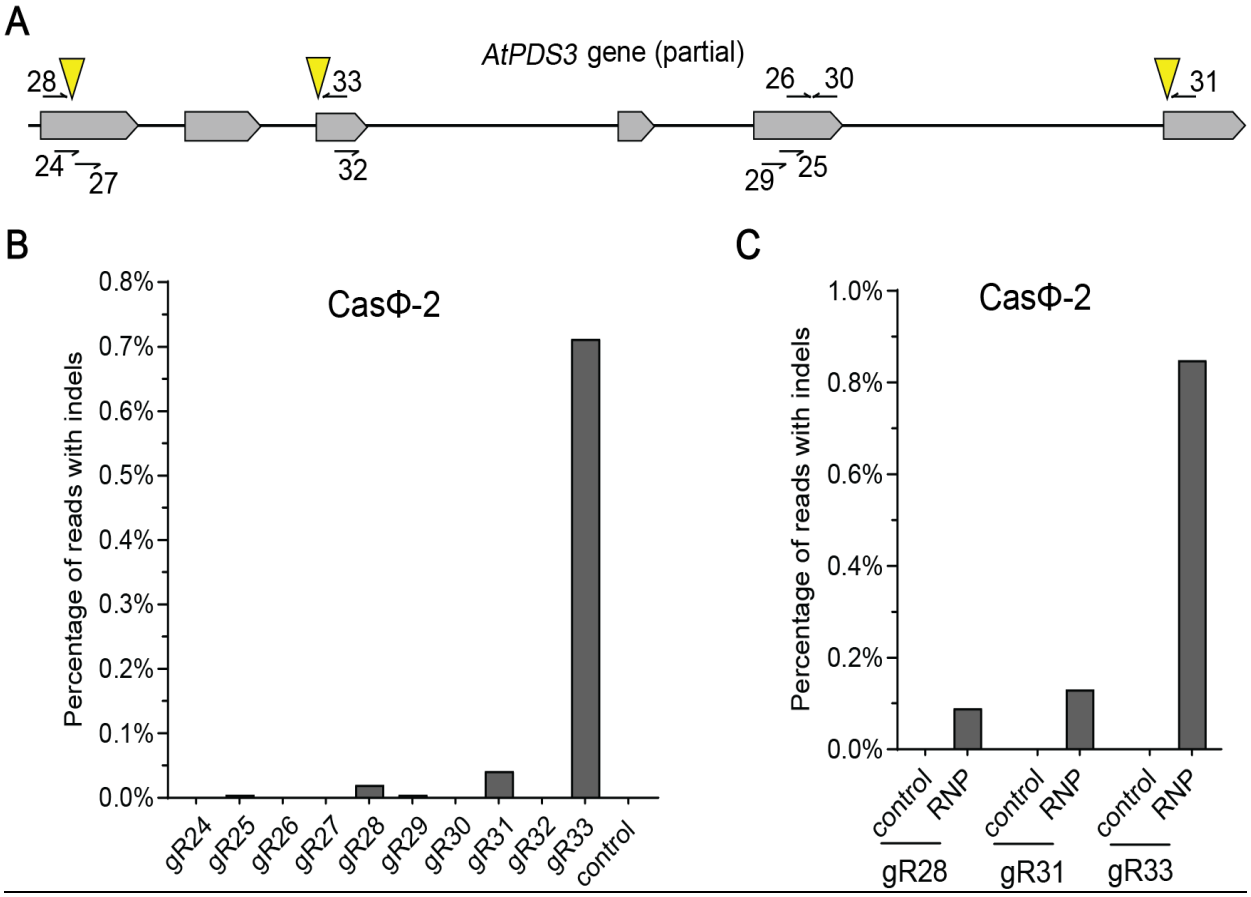

Fig. S12. CasΦ-2 mediated gene editing in plant cells. (A) (B) *AtPDS3* gene (partial) map showing the location of guide RNA target sites and the orientation of guide RNAs (numbers and arrows).Yellow triangles indicate guides that resulted in gene editing events. (B) Initial screening of CasΦ-2 RNP with gRNA24-33 for gene editing activity in *A. thaliana* protoplasts (n = 1, each). Control sample DNA was amplified and sequenced for target region of gRNA33. (C) An independent round of gene editing experiments were performed with gRNA28, gRNA31 and gRNA33 (n = 1, each). Three control samples were amplified and sequenced for the target regions of gRNA28, gRNA31 and gRNA33.

Supplementary Table S1: Plasmids

ID #	Assay	Features	Selection marker
pPP048	PAM-depletion, RNA-seq	pBLO1-backbone derived plasmid containing CasΦ-1 locus under control of tetracycline inducible promoter. Contains mini-CRISPR non-targeting AarI-GG stuffer spacer.	Chloramphenicol
pPP053	PAM-depletion, RNA-seq	pBLO1-backbone derived plasmid containing CasΦ-2 locus under control of tetracycline inducible promoter. Contains mini-CRISPR non-targeting AarI-GG stuffer spacer. *CasΦ* gene was codon optimized for expression in *E. coli*.	Chloramphenicol
pPP060	PAM-depletion, RNA-seq	pBLO1-backbone derived plasmid containing CasΦ-3 locus under control of tetracycline inducible promoter. Contains mini-CRISPR non-targeting AarI-GG stuffer spacer.	Chloramphenicol
pPP049	PAM-depletion	pBLO1-backbone derived plasmid containing CasΦ-1 locus under control of tetracycline inducible promoter. Contains mini-CRISPR with PAM-library targeting spacer.	Chloramphenicol
pPP056	PAM-depletion	pBLO1-backbone derived plasmid containing CasΦ-2 locus under control of tetracycline inducible promoter. Contains mini-CRISPR with PAM-library targeting spacer. *CasΦ* gene was codon optimized for expression in *E. coli*.	Chloramphenicol
pPP062	PAM-depletion	pBLO1-backbone derived plasmid containing CasΦ-3 locus under control of tetracycline inducible promoter. Contains mini-CRISPR with PAM-library targeting spacer.	Chloramphenicol
pPP094	PAM-depletion	pRSF-Duet1 derived plasmid containing *CasΦ-1* in MCS1 and a mini CRISPR (repeat-spacer-repeat) in MCS2. Contains non-targeting AarI-GG stuffer spacer.	Kanamycin
pPP100	PAM-depletion	pRSF-Duet1 derived plasmid containing *CasΦ-2* in MCS1 and a mini CRISPR (repeat-spacer-repeat) in MCS2. Contains non-targeting AarI-GG stuffer spacer.	Kanamycin
pPP106	PAM-depletion	pRSF-Duet1 derived plasmid containing *CasΦ-3* in MCS1 and a mini CRISPR (repeat-spacer-repeat) in MCS2. Contains non-targeting AarI-GG stuffer spacer.	Kanamycin
pPP097	PAM-depletion	pRSF-Duet1 derived plasmid containing *CasΦ-1* in MCS1 and a mini CRISPR (repeat-spacer-repeat) in MCS2. Contains mini-CRISPR with PAM-library targeting spacer.	Kanamycin
pPP102	PAM-depletion	pRSF-Duet1 derived plasmid containing *CasΦ-2* in MCS1 and a mini CRISPR (repeat-spacer-repeat) in MCS2. Contains mini-CRISPR with PAM-library targeting spacer.	Kanamycin
pPP107	PAM-depletion	pRSF-Duet1 derived plasmid containing *CasΦ-3* in MCS1 and a mini CRISPR (repeat-spacer-repeat) in MCS2.Contains mini-CRISPR with PAM-library targeting spacer.	Kanamycin
pPP190	Efficiency of transformation	pRSF-Duet1 derived plasmid containing *CasΦ-3* in MCS1 and a mini CRISPR (repeat-spacer-HDVrz) in MCS2. Contains non-targeting AarI-GG stuffer spacer.	Kanamycin
pPP192	Efficiency of transformation	pRSF-Duet1 derived plasmid containing *CasΦ-3* in MCS1 and a mini CRISPR (repeat-spacer-HDVrz) in MCS2. Contains *bla* gene targeting spacer.	Kanamycin
pPP240	Efficiency of transformation	pRSF-Duet1 derived plasmid containing *dCasΦ-3* (D413A) in MCS1 and a mini CRISPR (repeat-spacer-HDVrz) in MCS2. Contains *bla* gene targeting spacer.	Kanamycin
pPP242	Efficiency of transformation	pRSF-Duet1 derived plasmid containing *dCasΦ-3* (E618A) in MCS1 and a mini CRISPR (repeat-spacer-HDVrz) in MCS2. Contains *bla* gene targeting spacer.	Kanamycin
pPP244	Efficiency of transformation	pRSF-Duet1 derived plasmid containing *dCasΦ-3* (D708A) in MCS1 and a mini CRISPR (repeat-spacer-HDVrz) in MCS2. Contains *bla* gene targeting spacer.	Kanamycin

pPP076	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally hexa-histidine tagged *CasΦ-1* in MCS1.	Kanamycin
pPP085	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally hexa-histidine tagged *CasΦ-2* in MCS1.	Kanamycin
pPP089	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally hexa-histidine tagged *CasΦ-3* in MCS1.	Kanamycin
pPP378	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally hexa-histidine tagged d*CasΦ-1* (D371A) in MCS1.	Kanamycin
pPP381	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally hexa-histidine tagged d*CasΦ-2* (D394A) in MCS1.	Kanamycin
pPP384	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally hexa-histidine tagged d*CasΦ-3* (D431A) in MCS1.	Kanamycin
pPP338	Cleavage assay	pYTK095 derived plasmid containing a protospacer corresponding to position 1-24 of spacer 2 from the CasΦ-1 CRISPR-array, as derived from metagenomic analysis. PAM: 5-TTA (cognate)	Ampicillin
pPP341	Cleavage assay	pYTK095 derived plasmid containing a protospacer corresponding to position 1-24 of spacer 2 from the CasΦ-1 CRISPR-array, as derived from metagenomic analysis. PAM: 5-CCA (non-cognate)	Ampicillin
pPP394	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. SapI-GG stuffer spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP441	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. SapI-GG stuffer spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP444	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. SapI-GG stuffer spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP400	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #1 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP403	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #15 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP406	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #2 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)

pPP409	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #16 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP412	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #17 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP415	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #3 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP417	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #18 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP420	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #19 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP423	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #20 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP426	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #4 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP428	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #21 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP429	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #22 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP432	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #5 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)

pPP435	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #23 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP438	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-1 under the control of CMV-CAG promoter. CasΦ-1 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. Non-targeting scrambled guide spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP447	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #1 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP449	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #2 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP452	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #3 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP455	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #4 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP458	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #5 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP460	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #6 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP463	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #9 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP466	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #10 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)

pPP468	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #7 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP471	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #8 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP473	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #14 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP475	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-2 under the control of CMV-CAG promoter. CasΦ-2 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. Non-targeting scrambled guide spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP478	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #5 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP481	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #6 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP484	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #7 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP487	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #8 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP490	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #9 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP493	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #10 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)

pPP495	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #11 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP498	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #12 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP501	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #13 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP504	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. EGFP targeting guide #14 spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP506	Genome editing	Plasmid contains *Homo sapiens* codon optimized CasΦ-3 under the control of CMV-CAG promoter. CasΦ-3 is C-terminally fused to two SV40 NLS sequences, a FLAG tag and linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a repeat-spacer unit terminated by a poly-T sequence. Non-targeting scrambled guide spacer.	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pBFC545	Genome editing	Plasmid contains SpyCas9 under the control of CMV-CAG promoter. SpyCas9 is N-terminally and C-terminally fused to SV40 NLS sequences, a FLAG tag and C-terminally linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a spacer-repeat-tracr unit terminated by a poly-T sequence. EGFP targeting single-guide (sg) #1	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pBFC546	Genome editing	Plasmid contains SpyCas9 under the control of CMV-CAG promoter. SpyCas9 is N-terminally and C-terminally fused to SV40 NLS sequences, a FLAG tag and C-terminally linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a spacer-repeat-tracr unit terminated by a poly-T sequence. EGFP targeting single-guide (sg) #2	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pBFC547	Genome editing	Plasmid contains SpyCas9 under the control of CMV-CAG promoter. SpyCas9 is N-terminally and C-terminally fused to SV40 NLS sequences, a FLAG tag and C-terminally linked to PuroR via the P2A peptide sequence. Locus for guide expression is controlled by U6-promoter and constitutes a spacer-repeat-tracr unit terminated by a poly-T sequence. EGFP targeting single-guide (sg) #3	Ampicillin (*E. coli*) Puromycin (*H. Sapiens*)
pPP537	Protein purification	pRSF-Duet1 derived plasmid containing C-terminally 2xSV40 NLS and hexa-histidine tagged *CasΦ-2* in MCS1.	Kanamycin

Supplementary Table S2: Guide sequences for *in vivo* experiments

ID #	Assay	Description	Spacer sequence
g_Lib	PAM-depletion	CasΦ guide targeting the PAM depletion library plasmid next to the randomized PAM sequence	CTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGT
g_*bla*	Efficiency of transformation	CasΦ guide targeting the beta lactamase gene (*bla*) of pUC19	AACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCG
sg_1	Genome editing	SpyCas9 single-guide targeting the EGFP gene	GGCGAGGGCGATGCCACCTA
sg_2	Genome editing	SpyCas9 single-guide targeting the EGFP gene	TTCAAGTCCGCCATGCCCGA
sg_3	Genome editing	SpyCas9 single-guide targeting the EGFP gene	GGTGAACCGCATCGAGCTGA
g_1	Genome editing	CasΦ guide targeting the EGFP gene	CTTGTACAGCTCGTCCATGC
g_2	Genome editing	CasΦ guide targeting the EGFP gene	TCGGGCAGCAGCACGGGGCC
g_3	Genome editing	CasΦ guide targeting the EGFP gene	TAGTTGTACTCCAGCTTGTG
g_4	Genome editing	CasΦ guide targeting the EGFP gene	TGGCCGTTTACGTCGCCGTC
g_5	Genome editing	CasΦ guide targeting the EGFP gene	AAGAAGTCGTGCTGCTTCAT
g_6	Genome editing	CasΦ guide targeting the EGFP gene	ACCGGGGTGGTGCCCATCCT
g_7	Genome editing	CasΦ guide targeting the EGFP gene	AGCGTGTCCGGCGAGGGCGA
g_8	Genome editing	CasΦ guide targeting the EGFP gene	ATCTGCACCACCGGCAAGCT
g_9	Genome editing	CasΦ guide targeting the EGFP gene	GAGGGCGACACCCTGGTGAA
g_10	Genome editing	CasΦ guide targeting the EGFP gene	ACCAGGGTGTCGCCCTCGAA
g_11	Genome editing	CasΦ guide targeting the EGFP gene	TTCTGCTTGTCGGCCATGAT

g_12	Genome editing	CasΦ guide targeting the EGFP gene	ACCTTGATGCCGTTCTTCTG
g_13	Genome editing	CasΦ guide targeting the EGFP gene	TGCTGGTAGTGGTCGGCGAG
g_14	Genome editing	CasΦ guide targeting the EGFP gene	GTGACCGCCGCCGGGATCAC
g_15	Genome editing	CasΦ guide targeting the EGFP gene	GGGTCTTTGCTCAGCTTGGA
g_16	Genome editing	CasΦ guide targeting the EGFP gene	TGGCGGATCTTGAAGTTCAC
g_17	Genome editing	CasΦ guide targeting the EGFP gene	TGGCTGTTGTAGTTGTACTC
g_18	Genome editing	CasΦ guide targeting the EGFP gene	TACTCCAGCTTGTGCCCCAG
g_19	Genome editing	CasΦ guide targeting the EGFP gene	CCGTCCTCCTTGAAGTCGAT
g_20	Genome editing	CasΦ guide targeting the EGFP gene	CCGTCGTCCTTGAAGAAGAT
g_21	Genome editing	CasΦ guide targeting the EGFP gene	CCGTAGGTGGCATCGCCCTC
g_22	Genome editing	CasΦ guide targeting the EGFP gene	CCGGTGGTGCAGATGAACTT
g_23	Genome editing	CasΦ guide targeting the EGFP gene	AAGAAGATGGTGCGCTCCTG
g_24	Genome editing	CasΦ guide targeting the PDS3 gene	AAACGGGTTTTTGGAGGC
g_25	Genome editing	CasΦ guide targeting the PDS3 gene	CTATGCCAAGTAAACCTG
g_26	Genome editing	CasΦ guide targeting the PDS3 gene	TATGCCAAGTAAACCTGG

g_27	Genome editing	CasΦ guide targeting the PDS3 gene	AGGCACTTTCATCTGGAG
g_28	Genome editing	CasΦ guide targeting the PDS3 gene	GCCTTATCAAAACGGGTT
g_29	Genome editing	CasΦ guide targeting the PDS3 gene	TTGCTATGCCAAGTAAAC
g_30	Genome editing	CasΦ guide targeting the PDS3 gene	TAGGACATCTGGGAAGTC
g_31	Genome editing	CasΦ guide targeting the PDS3 gene	TTGTTCCGCAAAATAGCC
g_32	Genome editing	CasΦ guide targeting the PDS3 gene	AAAGTACCTGGCTGATGC
g_33	Genome editing	CasΦ guide targeting the PDS3 gene	CAGTTGACAATCCAGCCA
g_NT	Genome editing	Non-targeting scrambled CasΦ guide	CGTGATGGTCTCGATTGAGT

Supplementary Table S3: Substrate sequences for *in vitro* experiments

ID #	Assay	Description	Sequence (5′->3′)
rPP001	pre-crRNA processing	Contains from 5′->3′: 4 nt of spacer #1, 36 nt repeat #2 (grey), 20 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CACAGGAGAGAUCUCAAACGAUUGCUCGAUUAGUCGAGACAGCUGGUAAUGGGAUACCUU
rPP002	pre-crRNA processing	Contains from 5′->3′: 4 nt of spacer #2, 36 nt repeat #3 (grey), 20 nt of spacer #3 from the genomic CasΦ-2 CRISPR array	HO-UAAUGUCGGAACGCUCAACGAUUGCCCCUCACGAGGGGACUGCCGCCUCCGCGACGCCCA
rPP003	pre-crRNA processing	Contains from 5′->3′: 4 nt of leader, 36 nt repeat #1 (grey), 20 nt of spacer #1 from the genomic CasΦ-3 CRISPR array	HO-AUUAACCAAAACGACUAUUGAUUGCCCAGUACGCUGGGACUAUGAGCUUAUGUACAUCAA
rGJK008	pre-crRNA processing	AsCas12a pre-crRNA substrate	HO-GACCUUUUUAAUUUCUACUCUUGUAGAUAAAGUGCUCAUCAUUGGAAAACGU
pPP338	Spacer tiling cleavage assay	pYTK095 derived plasmid containing a protospacer corresponding to position 1-24 of spacer 2 from the CasΦ-1 CRISPR-array (blue sequence), as derived from metagenomic analysis. Cognate PAM: 5-TTA (red sequence)	circular dsDNA CCAATGCTTAATCAGTGAGGCACCTATCTCAGCG ATCTGTCTATTTCGTTCATCCATAGTTGCCTGACT CCCCGTCGTGTAGATAACTACGATACGGGAGGG CTTACCATCTGGCCCCAGTGCTGCAATGATACCG CGGGACCCACGCTCACCGGCTCCAGATTTATCAG CAATAAACCAGCCAGCCGGAAGGGCCGAGCGCA GAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA

			GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG CCATTGCTACAGGCATCGTGGTGTCACGCTCGTC GTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAA CGATCAAGGCGAGTTACATGATCCCCCATGTTGT GCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA CTCATGGTTATGGCAGCACTGCATAATTCTCTTA CTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC TGGTGAGTACTCAACCAAGTCATTCTGAGAATAG TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGT CAATACGGGATAATACCGCGCCACATAGCAGAA CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTC GGGGCGAAAACTCTCAAGGATCTTACCGCTGTTG AGATCCAGTTCGATGTAACCCACTCGTGCACCCA ACTGATCTTCAGCATCTTTTACTTTCACCAGCGTT TCTGGGTGAGCAAAAACAGGAAGGCAAAATGCC GCAAAAAAGGGAATAAGGGCGACACGGAAATG TTGAATACTCATACTCTTCCTTTTTCAATATTATT GAAGCATTTATCAGGGTTATTGTCTCATGAGCGG ATACATATTTGAATGTATTTAGAAAAATAAACAA ATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGC CACCTGTCATGACCAAAATCCCTTAACGTGAGTT TTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAG ATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG CGTAATCTGCTGCTTGCAAACAAAAAAACCACC GCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCA GCAGAGCGCAGATACCAAATACTGTTCTTCTAGT GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT GTAGCACCGCCTACATACCTCGCTCTGCTAATCC TGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC GTGTCTTACCGGGTTGGACTCAAGACGATAGTTA CCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG GGTTCGTGCACACAGCCCAGCTTGGAGCGAACG ACCTACACCGAACTGAGATACCTACAGCGTGAG CTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA AAGGCGGACAGGTATCCGGTAAGCGGCAGGGTC GGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG GGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT TTCGCCACCTCTGACTTGAGCGTCGATTTTTGTG ATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCC TTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGC GTTATCCCCTGATTCTGTGGATAACCGTGCGGCC GCCCCTTGTAG**TTA**<u>**AGCTGGTAATGGGATACCT** **TGTGC**</u>TACAGCGGCCGCGATTATCAAAAAGGAT CTTCACCTAGATCCTTTTAAATTAAAAATGAAGT TTTAAATCAATCTAAAGTATATATGAGTAAACTT GGTCTGACAGTTA
PCR-pPP338	Spacer tiling cleavage assay	PCR fragment of pYTK095 derived plasmid containing a protospacer corresponding to position 1-24 of spacer 2 from the CasΦ-1 CRISPR-array (blue sequence), as derived from metagenomic analysis. Cognate PAM: 5-TTA (red sequence)	<u>linear dsDNA</u> GCTCTTGCCCGGCGTCAATACGGGATAATACCGC GCCACATAGCAGAACTTTAAAAGTGCTCATCATT GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG ATCTTACCGCTGTTGAGATCCAGTTCGATGTAAC CCACTCGTGCACCCAACTGATCTTCAGCATCTTT TACTTTCACCAGCGTTTCTGGGTGAGCAAAAACA GGAAGGCAAAATGCCGCAAAAAAGGGAATAAG GGCGACACGGAAATGTTGAATACTCATACTCTTC CTTTTTCAATATTATTGAAGCATTTATCAGGGTT ATTGTCTCATGAGCGGATACATATTTGAATGTAT TTAGAAAAATAAACAAATAGGGGTTCCGCGCAC ATTTCCCCGAAAAGTGCCACCTGTCATGACCAAA ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT CAGACCCCGTAGAAAAGATCAAAGGATCTTCTT GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG CAAACAAAAAAACCACCGCTACCAGCGGTGGTT

			TGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC CGAAGGTAACTGGCTTCAGCAGAGCGCAGATAC CAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGG CCACCACTTCAAGAACTCTGTAGCACCGCCTACA TACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT GGACTCAAGACGATAGTTACCGGATAAGGCGCA GCGGTCGGGCTGAACGGGGGGTTCGTGCACACA GCCCAGCTTGGAGCGAACGACCTACACCGAACT GAGATACCTACAGCGTGAGCTATGAGAAAGCGC CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCG CACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGAC TTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGG GCGGAGCCTATGGAAAAACGCCAGCAACGCGGC CTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT GTGGATAACCGTGCGGCCGCCCCTTGTAG**TTAA** **GCTGGTAATGGGATACCTTGTGC**TACAGCGGC CGCGATTATCAAAAAGGATCTTCACCTAGATCCT TTTAAATTAAAAATGAAGTTTTAAATCAATCTAA AGTATATATGAGTAAACTTGGTCTGACAGTTACC AATGCTTAATCAGTGAGGCACCTATCTCAGCGAT CTGTCTATTTCGTTCATCCATAGTTGCCTGACTCC CCGTCGTGTAGATAACTACGATACGGGAGGGCT TACCATCTGGCCCCAGTGCTGCAATGATACCGCG GGACCCACGCTCACCGGCTCCAGATTTATCAGCA ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAG TTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCC ATTGCTACAGGCATCGTGGTGTCACGCTCGTCGT TTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG ATCAAGGCGAGTTACATGATCCCCCATGTTGTGC AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCG TTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT CATGGTTATGGCAGCACTGCATAATTCTCTTACT GTCATGCCATCCGTAAGATGCTTTTCTGTGACTG GTGAGTACTCAACCAAGTCATTCTGAGAATAGTG TATGCGGCG
PCR-pPP341	Spacer tiling cleavage assay	PCR fragment of pYTK095 derived plasmid containing a protospacer corresponding to position 1-24 of spacer 2 from the CasΦ-1 CRISPR-array (blue sequence), as derived from metagenomic analysis. Non-cognate PAM: 5-CCA (red sequence)	<u>linear dsDNA</u> GCTCTTGCCCGGCGTCAATACGGGATAATACCGC GCCACATAGCAGAACTTTAAAAGTGCTCATCATT GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG ATCTTACCGCTGTTGAGATCCAGTTCGATGTAAC CCACTCGTGCACCCAACTGATCTTCAGCATCTTT TACTTTCACCAGCGTTTCTGGGTGAGCAAAAACA GGAAGGCAAAATGCCGCAAAAAAGGGAATAAG GGCGACACGGAAATGTTGAATACTCATACTCTTC CTTTTTCAATATTATTGAAGCATTTATCAGGGTT ATTGTCTCATGAGCGGATACATATTTGAATGTAT TTAGAAAAATAAACAAATAGGGGTTCCGCGCAC ATTTCCCCGAAAAGTGCCACCTGTCATGACCAAA ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT CAGACCCCGTAGAAAAGATCAAAGGATCTTCTT GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG CAAACAAAAAAACCACCGCTACCAGCGGTGGTT TGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC CGAAGGTAACTGGCTTCAGCAGAGCGCAGATAC CAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGG CCACCACTTCAAGAACTCTGTAGCACCGCCTACA TACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT GGACTCAAGACGATAGTTACCGGATAAGGCGCA GCGGTCGGGCTGAACGGGGGGTTCGTGCACACA GCCCAGCTTGGAGCGAACGACCTACACCGAACT

			GAGATACCTACAGCGTGAGCTATGAGAAAGCGC CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCG CACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGAC TTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGG GCGGAGCCTATGGAAAAACGCCAGCAACGCGGC CTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT GTGGATAACCGTGCGGCCGCCCCTTGTAG**CCAA** **GCTGGTAATGGGATACCTTGTGC**TACAGCGGC CGCGATTATCAAAAAGGATCTTCACCTAGATCCT TTTAAATTAAAAATGAAGTTTTAAATCAATCTAA AGTATATATGAGTAAACTTGGTCTGACAGTTACC AATGCTTAATCAGTGAGGCACCTATCTCAGCGAT CTGTCTATTTCGTTCATCCATAGTTGCCTGACTCC CCGTCGTGTAGATAACTACGATACGGGAGGGCT TACCATCTGGCCCCAGTGCTGCAATGATACCGCG GGACCCACGCTCACCGGCTCCAGATTTATCAGCA ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAG TTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCC ATTGCTACAGGCATCGTGGTGTCACGCTCGTCGT TTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG ATCAAGGCGAGTTACATGATCCCCCATGTTGTGC AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCG TTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT CATGGTTATGGCAGCACTGCATAATTCTCTTACT GTCATGCCATCCGTAAGATGCTTTTCTGTGACTG GTGAGTACTCAACCAAGTCATTCTGAGAATAGTG TATGCGGCG
oPP510	radiolabeled cleavage assay	DNA non-target strand for CasΦ-1 and CasΦ-2 cleavage assays	HO- CGGCCGCCCCTTGTAGTTAAGCTGGTAATGGGAT ACCTTGTGCTACAGCGGCCGCG
oPP511	radiolabeled cleavage assay	DNA target strand for CasΦ-1 and CasΦ-2 cleavage assays	HO- CGCGGCCGCTGTAGCACAAGGTATCCCATTACCA GCTTAACTACAAGGGGCGGCCG
oPP596	radiolabeled cleavage assay	DNA non-target strand for CasΦ-3 cleavage assays	HO- CGGCCGCCCCTTGTAATTCAGCTGGTAATGGGAT ACCTTGTGCTACAGCGGCCGCG
oPP597	radiolabeled cleavage assay	DNA target strand for CasΦ-3 cleavage assays	HO- CGCGGCCGCTGTAGCACAAGGTATCCCATTACCA GCTGAATTACAAGGGGCGGCCG
rPP015	radiolabeled cleavage assay	RNA target strand for cleavage assays	HO- CGCUGUAGCACAAGGUAUCCCAUUACCAGCUU AACUACAAG
dGJK001	radiolabeled cleavage assay	Trans cleavage DNA substrate	HO- GTGGCCGTTTAAAAGTGCTCATCATTGGAAAACG TAGGATGGGCACCA
rGJK118	radiolabeled cleavage assay	Trans cleavage RNA substrate	HO-AGUAUUUAAUCGUUGCAAGAG GCGCUGCGUUU

Supplementary Table S4: crRNA sequences for *in vitro and in vivo* experiments

ID #	Assay	Description	Sequence (5′->3′)
rPP007	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array.	HO-CAACGAUUGCCCCUCACGAGGGGAC
rPP008	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 12 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGG
rPP009	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 14 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGGGA
rPP010	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 16 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGGGAUA
rPP011	Spacer tiling and radiolabeled cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGGGAUACC
rPP012	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 20 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGGGAUACCUU
rPP013	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 22 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGGGAUACCUUGU
rPP014	Spacer tiling cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 24 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-CAACGAUUGCCCCUCACGAGGGGACAGCUGG UAAUGGGAUACCUUGUGC
rPP016	radiolabeled cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-1 CRISPR array. 18 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-AAACGAUUGCUCGAUUAGUCGAGACAGCUGG UAAUGGGAUACC
rPP017	radiolabeled cleavage assay	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-3 CRISPR array. 18 nt of spacer #2 from the genomic CasΦ-1 CRISPR array	HO-UAUUGAUUGCCCAGUACGCUGGGACAGCUGG UAAUGGGAUACC
gRNA 24	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_24, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACAAACGG GUUUUUGGAGGC
gRNA 25	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_25, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACCUAUGC CAAGUAAACCUG
gRNA 26	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_26, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACUAUGCC AAGUAAACCUGG

gRNA 27	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_27, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACAGGCACUUUUCAUCUGGAG
gRNA 28	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_28, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACGCCUUAUCAAAACGGGUU
gRNA 29	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_29, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACUUGCUAUGCCAAGUAAAC
gRNA 30	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_30, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACUAGGACAUCUGGGAAGUC
gRNA 31	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_31, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACUUGUUCCGCAAAAUAGCC
gRNA 32	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_32, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACAAAGUACCUGGCUGAUGC
gRNA 33	Genome editing using RNP in A. thaliana	Contains from 5′->3′: 25 nt of 3′-repeat fraction (grey) from the genomic CasΦ-2 CRISPR array. 18 nt spacer targeting PDS3 gene. (corresponds to guides sequence g_33, see table S2)	HO-CAACGAUUGCCCCUCACGAGGGGACCAGUUGACAAUCCAGCCA