GENOME EDITING

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[1,2,7], Steven E. Jacobsen[6,8], Jillian F. Banfield[1,9], Jennifer A. Doudna[1,2,8,10,11,12]†

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 uses 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.

Competition between viruses and their host microbes fostered the evolution of CRISPR-Cas systems that employ nucleases and noncoding CRISPR RNAs (crRNAs) to target foreign nucleic acids by complementary base pairing (*1*). Processing of CRISPR array transcripts, which consist of repeats and spacer sequences acquired from viruses or other mobile genetic elements (MGEs) (*2*), generates mature crRNAs that guide Cas proteins to detect and destroy previously encountered viruses (*3*). 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 of five 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 to 80 kDa, about half the size of 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), here 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 that follow or precede a 2– to 5–base-pair (bp) protospacer-adjacent motif (PAM) for self versus nonself 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-stranded 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 (RNA-seq) analysis revealed transcription of the *cas*Φ gene and the reduced CRISPR array but no evidence of other noncoding 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 phages (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 to 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 to 12 nt (Fig. 2, B and C, and fig. S5, B and 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 and S7; supplementary text). Furthermore, CasΦ was found to cleave single-stranded DNA (ssDNA) but not ssRNA in cis or in trans (figs. S6B and 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 by 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- to 29-nt-long repeat and 20-nt spacer sequence of the crRNA were observed only in the presence of wild-type CasΦ; this was corroborated by RNA-seq analysis of native loci (Figs. 1D and 3, A and C; fig. S9). In control experiments, we found that pre-crRNA processing was strictly magnesium dependent (Fig. 3B; fig. S9), which is different from other CRISPR-Cas RNA processing reaction conditions and suggests 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

[1]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [3]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. [4]College of Natural Resources, University of California, Berkeley, Berkeley, CA, USA. [5]University of California, Berkeley—University of California, San Francisco Graduate Program in Bioengineering, Berkeley, CA, USA. [6]Department of Molecular, Cellular and Developmental Biology, University of California, Los Angeles, CA, USA. [7]Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Victoria, Australia. [8]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA. [9]Department of Earth and Planetary Sciences, University of California, Berkeley, Berkeley, CA, USA. [10]Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. [11]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [12]Gladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: doudna@berkeley.edu

Downloaded from https://www.science.org at University of California Los Angeles on September 14, 2021

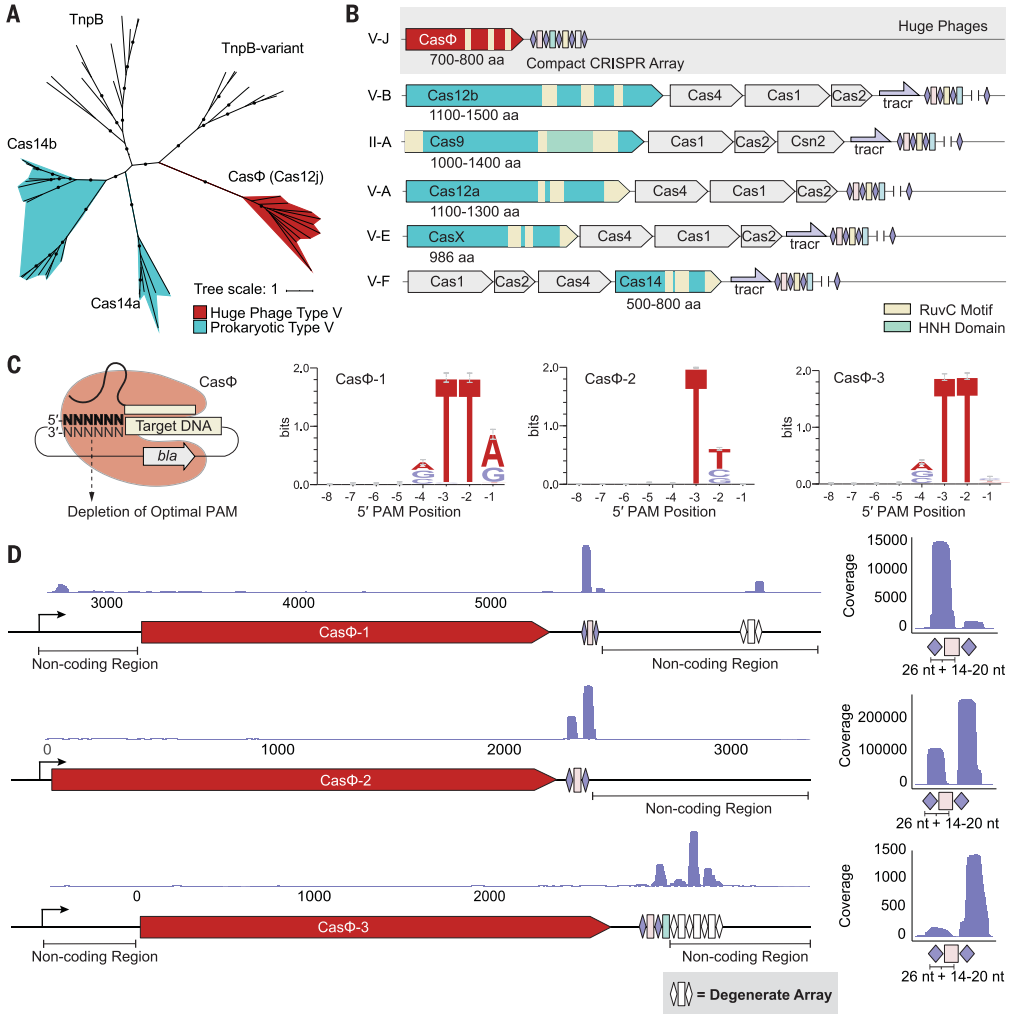

Fig. 1. CasΦ is a bona fide CRISPR-Cas system from huge phages. (**A**) Maximum likelihood phylogenetic tree of type V effector proteins and respective predicted ancestral TnpB nucleases. Bootstrap and approximate likelihood ratio test values of ≥90 are denoted on the branches with black circles. (**B**) 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**) (Left) RNA-seq results mapped onto the native genomic loci of CasΦ orthologs and their upstream and downstream noncoding regions cloned with reduced CRISPR arrays into expression plasmids. (Right) Enlarged view of RNA mapped onto the first repeat (diamond)-spacer (rectangle) pair. (**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 MGEs. (**F**) Predicted molecular weights of the RNP complexes of small CRISPR-Cas effectors and those functional in editing of mammalian cells.

RuvC-inactivating mutation and found it to be incapable of processing pre-crRNAs (Fig. 3B; fig. S9, A and B). Both wild-type and catalytically inactivated CasΦ proteins bound crRNA, and their reconstituted complexes with pre-crRNA had 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*). By 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 and then 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 is formed during the reaction catalyzed by CasΦ, in contrast to the acid-base– catalyzed processing reaction for Cas12a (Fig. 3, C and 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 coexpressed

Downloaded from https://www.science.org at University of California Los Angeles on September 14, 2021

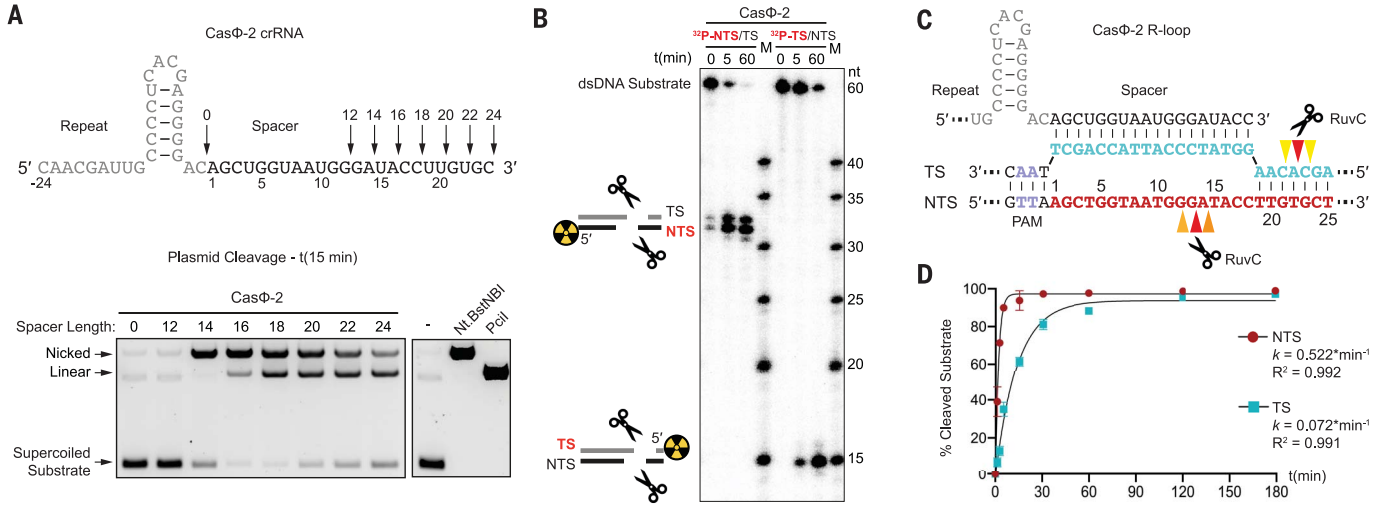

Fig. 2. CasΦ cleaves DNA. (**A**) Supercoiled plasmid cleavage assay testing CasΦ RNPs reconstituted with crRNAs of different spacer lengths. (**B**) Cleavage assay targeting dsDNA oligo-duplices for mapping of the cleavage structure. (**C**) Scheme illustrating the cleavage pattern. (**D**) NTS and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

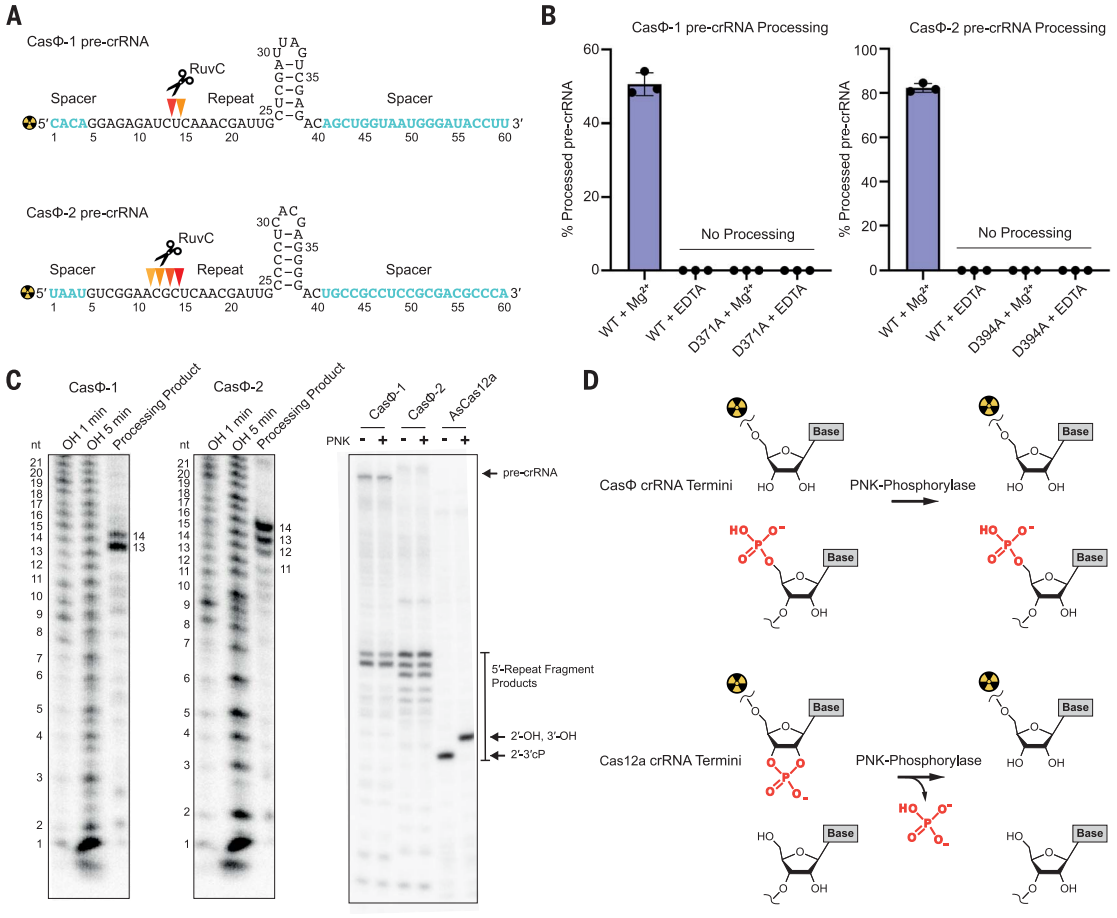

Fig. 3. CasΦ processes pre-crRNA within the RuvC active site. (**A**) Pre-crRNA substrates and processing sites (red triangles) derived from the OH ladder in panel C. (**B**) Pre-crRNA processing assay for CasΦ-1 and CasΦ-2 in dependence of Mg^{2+} and RuvC active site residue variation [Asp^{371}→Ala (D371A) and D394A] (n = 3 each; means ± SD; time = 60 min). Data are shown in fig. S9B. (**C**) (Left and middle) Alkaline hydrolysis ladder (OH) of the pre-crRNA substrate. (Right) T4 polynucleotide kinase (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.

with a crRNA in HEK293 cells (Fig. 4A). We found that CasΦ-2 and CasΦ-3 induced targeted disruption of a genomically integrated enhanced green fluorescent protein (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), which is comparable to levels initially reported for CRISPR-Cas9, CRISPR-Cas12a, and CRISPR-CasX (*7*, *8*, *20*). We next tested if CasΦ-2 could be delivered as ribonucleoproteins (RNPs) into plant protoplasts to edit the endogenous *Arabidopsis thaliana PDS3* gene (Fig. 4B; fig. S12). Next-generation sequencing revealed that CasΦ-2 introduced primarily 8- to 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

Downloaded from https://www.science.org at University of California Los Angeles on September 14, 2021

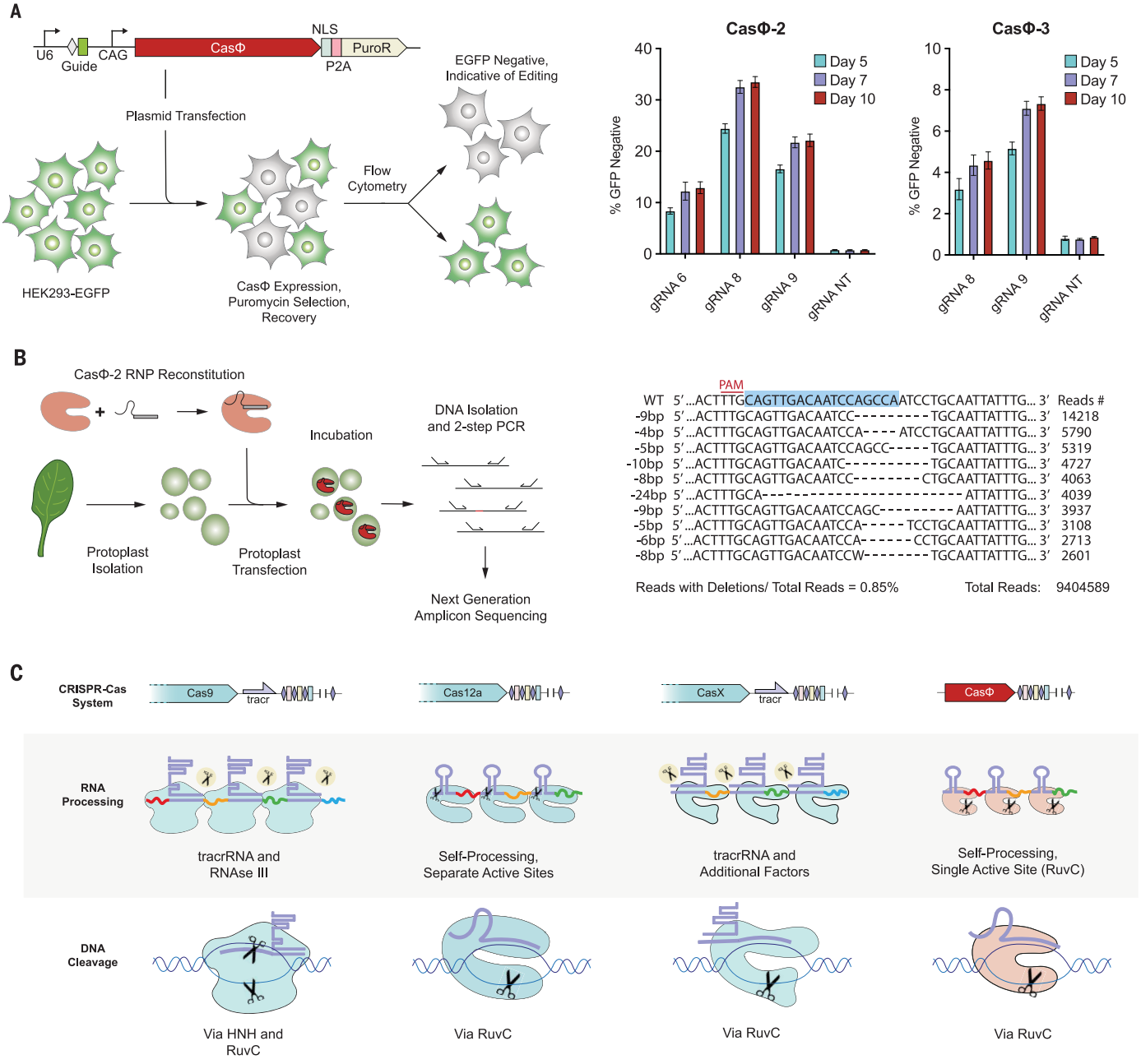

Fig. 4. CasΦ is functional for genome editing. (**A**) Experimental workflow of the GFP disruption assay (left) and GFP disruption using CasΦ-2 and CasΦ-3 and a nontargeting (NT) guide as a negative control (n = 3 each; means ± SD). (**B**) (Left) Experimental workflow of CasΦ2 RNP-mediated genome editing in *A. thaliana* mesophyll protoplasts. (Right) Amplicon sequencing data 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.

requirement will be particularly advantageous for both vector-based delivery into cells and a wider range of targetable genomic sequences, and as such it will provide 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 (Cas9) active sites 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 than in 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 they highlight huge phages as an exciting forefront for discovery and biotechnological applications for human health.

REFERENCES AND NOTES

1. R. Barrangou *et al.*, *Science* **315**, 1709–1712 (2007).
2. J. McGinn, L. A. Marraffini, *Nat. Rev. Microbiol.* **17**, 7–12 (2019).
3. F. Hille *et al.*, *Cell* **172**, 1239–1259 (2018).
4. B. Al-Shayeb *et al.*, *Nature* **578**, 425–431 (2020).
5. S. Shmakov *et al.*, *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
6. D. Gleditzsch *et al.*, *RNA Biol.* **16**, 504–517 (2019).
7. B. Zetsche *et al.*, *Cell* **163**, 759–771 (2015).
8. J.-J. Liu *et al.*, *Nature* **566**, 218–223 (2019).
9. J. S. Chen *et al.*, *Science* **360**, 436–439 (2018).
10. A. East-Seletsky *et al.*, *Nature* **538**, 270–273 (2016).
11. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
12. I. Fonfara, H. Richter, M. Bratovič, A. Le Rhun, E. Charpentier, *Nature* **532**, 517–521 (2016).

Downloaded from https://www.science.org at University of California Los Angeles on September 14, 2021

13. D. Burstein *et al*., *Nature* **542**, 237–241 (2017).
14. L. B. Harrington *et al*., *Science* **362**, 839–842 (2018).
15. W. X. Yan *et al*., *Science* **363**, 88–91 (2019).
16. S. Shmakov *et al*., *Mol. Cell* **60**, 385–397 (2015).
17. M. Nowotny, *EMBO Rep.* **10**, 144–151 (2009).
18. D. C. Swarts, J. van der Oost, M. Jinek, *Mol. Cell* **66**, 221–233.e4 (2017).
19. G. J. Knott, J. A. Doudna, *Science* **361**, 866–869 (2018).
20. P. Mali *et al*., *Science* **339**, 823–826 (2013).
21. S. Duffy, L. A. Shackelton, E. C. Holmes, *Nat. Rev. Genet.* **9**, 267–276 (2008).
22. M.-C. Lee, C. J. Marx, *PLOS Genet.* **8**, e1002651 (2012).
23. M. Lynch, *Annu. Rev. Microbiol.* **60**, 327–349 (2006).

ACKNOWLEDGMENTS

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:** P.P. was supported by the Paul G. Allen Frontiers Group. P.P. and B.A.-S. are supported by the Somatic Cell Genome Editing Program of the Common Fund of the NIH (NIH U01AI142817-02). B.A.-S. is supported by an NSF Graduate Research Fellowship (no. DGE 1752814). G.J.K. is supported by an NHMRC Investigator Grant (EL1, APP1175568) and previously an American Australian Association Fellowship. J.A.D. receives funding from the NSF (1817593). J.A.D. and S.E.J. are Investigators of the Howard Hughes Medical Institute. **Author contributions:** B.A.-S. and J.F.B. conceived the initial study, which was further developed with P.P. and J.A.D. P.P. and B.A.-S. designed experiments and analyzed data with input from J.A.D. and J.F.B. P.P., E.B-.R., and B.A.-S. cloned constructs with guidance from P.P. B.A.-S. conducted metagenomic assemblies, genome curation, protein and crRNA sequence analyses, active residue identification, and computational analyses for PAM depletion assays and RNA-seq. P.P. conducted PAM depletion assays, efficiency of transformation assays, performed RNA-seq experiments, purified proteins, and performed biochemical experiments. P.P., B.A.-S., C.A.T., and Z.L. designed genome editing experiments with input from S.E.J. C.A.T. performed tissue culture and flow cytometry, and Z.L. performed editing experiments in *A. thaliana*. B.F.C. and G.J.K. provided materials and experimental advice. B.A.-S. and P.P. wrote the manuscript with input from J.A.D. and J.F.B. P.P. and B.A.-S. prepared figures. The manuscript was reviewed and approved by all coauthors. **Competing interests:** The Regents of the University of California have patents pending for CRISPR technologies on which the authors are inventors. J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Intellia Therapeutics, Scribe Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Synthego, Algen Biotechnologies, Felix Biosciences, and Mammoth Biosciences. J.A.D. is a director at Johnson & Johnson and has sponsored research projects by Pfizer, Roche, AppleTree Partners, and Biogen. J.F.B. is a founder of Metagenomi. S.E.J. is a scientific cofounder of Inari Agriculture and S.E.J. and J.A.D. 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 J.A.D.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6501/333/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S4
References (*24–29*)
MDAR Reproducibility Checklist
Protein, repeat, and locus sequences for CRISPR-CasΦ

View/request a protocol for this paper from *Bio-protocol*.

5 February 2020; accepted 6 May 2020
10.1126/science.abb1400

Downloaded from https://www.science.org at University of California Los Angeles on September 14, 2021

Science

CRISPR-Cas# from huge phages is a hypercompact genome editor

Patrick PauschBasem Al-ShayebEzra Bisom-RappConnor A. TsuchidaZheng LiBrady F. CressGavin J. KnottSteven E. JacobsenJillian F. BanfieldJennifer A. Doudna

Science, 369 (6501),

Compact defense system in bacteriophages

The CRISPR-Cas system, naturally found in many prokaryotes, is widely used for genome editing. CRISPR arrays in the bacterial genome, derived from the genome of invading viruses, are used to generate a CRISPR RNA that guides the Cas enzyme to destroy repeat viral invaders. Recently, an unexpectedly compact CRISPR-Cas system was identified in huge bacteriophages. Pausch *et al*. show that even though this system lacks commonly found accessory proteins, it is functional. In addition to a CRISPR array, the only component of the system is an enzyme called CasF, which uses the same active site to process transcripts of the CRISPR arrays into CRISPR RNA and to destroy foreign nucleic acids. This system, which is active in human and plant cells, provides a hypercompact addition to the genome-editing toolbox.

Science this issue p. 333

View the article online
https://www.science.org/doi/10.1126/science.abb1400
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at University of California Los Angeles on September 14, 2021

Use of think article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.
Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works