UC Berkeley

UC Berkeley Previously Published Works

Title

Programmed DNA destruction by miniature CRISPR-Cas14 enzymes.

Permalink

https://escholarship.org/uc/item/81032581

Journal

Science (New York, N.Y.), 362(6416)

ISSN

0036-8075

Authors

Harrington, LB

Burstein, D

Chen, JS

et al.

Publication Date

2018-11-01

DOI

10.1126/science.aav4294

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Programmed DNA destruction by miniature CRISPR-Cas14 enzymes

Lucas B. Harrington[1],*,†,

David Burstein[2],*,‡,

Janice S. Chen[1],†,

David Paez-Espino[3],

Enbo Ma[1],

Isaac P. Witte[1],

Joshua C. Cofsky[1],

Nikos C. Kyrpides[3],

Jillian F. Banfield[2],[4],[5],

Jennifer A. Doudna[1],[5],[6],[7],[8],§

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.

[2]Department of Earth and Planetary Sciences, University of California, Berkeley, CA 94720, USA.

[3]Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA.

[4]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA.

[5]Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA.

[6]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[7]Department of Chemistry, University of California, Berkeley, CA 94720, USA.

[8]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.

§Corresponding author. Email: doudna@berkeley.edu

* These authors contributed equally to this work.

† Present address: Mammoth Biosciences, San Francisco, CA 94107, USA.

‡ Present address: School of Molecular Cell Biology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel.

Abstract

CRISPR-Cas systems provide microbes with adaptive immunity to infectious nucleic acids and are widely employed as genome editing tools. These tools

use RNA-guided Cas proteins whose large size (950 to 1400 amino acids) has been considered essential to their specific DNA- or RNA-targeting activities. Here we present a set of CRISPR-Cas systems from uncultivated archaea that contain Cas14, a family of exceptionally compact RNA-guided nucleases (400 to 700 amino acids). Despite their small size, Cas14 proteins are capable of targeted single-stranded DNA (ssDNA) cleavage without restrictive sequence requirements. Moreover, target recognition by Cas14 triggers nonspecific cutting of ssDNA molecules, an activity that enables high-fidelity single-nucleotide polymorphism genotyping (Cas14-DETECTR). Metagenomic data show that multiple CRISPR-Cas14 systems evolved independently and suggest a potential evolutionary origin of single-effector CRISPR-based adaptive immunity.

Introduction

Competition between microbes and viruses stimulated the evolution of CRISPR-based adaptive immunity to provide protection against infectious agents (*1*, *2*). In class 2 CRISPR-Cas systems, a single 100- to 200-kDa CRISPR-associated (Cas) protein with multiple functional domains carries out RNA-guided binding and cutting of DNA or RNA substrates (*3*, *4*). To determine whether simpler, smaller RNA-guided proteins occur in nature, we queried terabase-scale metagenomic datasets (*5*–*9*) for uncharacterized genes proximal to both a CRISPR array and *cas1*, the gene that encodes the universal CRISPR integrase (*10*, *11*). This analysis identified a diverse family of CRISPR-Cas systems that contain *cas1*, *cas2*, *cas4*, and a previously unrecognized gene, *cas14*, encoding a 40- to 70-kDa polypeptide (Fig. 1A). We initially identified 24 different *cas14* gene variants that cluster into three subgroups (Cas14a, Cas14b, and Cas 14c) on the basis of comparative sequence analysis (Fig. 1, A and B, and figs. S1 and S2). Cas14 proteins are ~400 to 700 amino acids (aa), about half the size of previously known class 2 CRISPR RNA-guided enzymes (950 to 1400 aa) (Fig. 1, C and D). Though the identified Cas14 proteins exhibit considerable sequence diversity, all are united by the presence of a predicted RuvC nuclease domain, whose organization is characteristic of type V CRISPR-Cas DNA-targeting enzymes (Fig. 1D) (*3*, *12*, *13*).

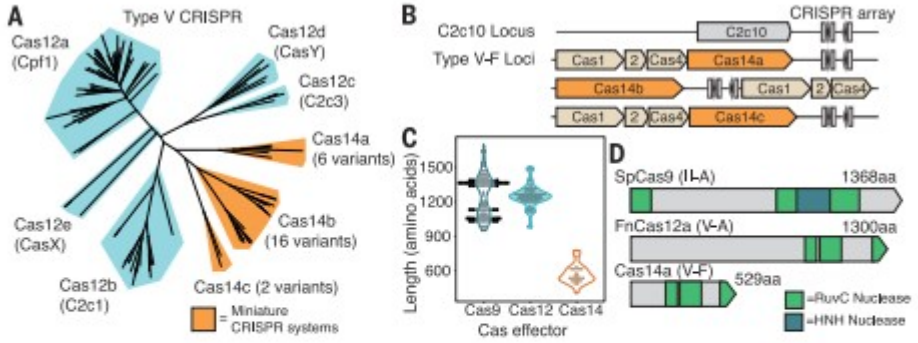


Fig. 1. Architecture and phylogeny of CRISPR-Cas14 genomic loci. (**A**) Phylogenetic tree of type V CRISPR systems. Newly identified miniature CRISPR systems are highlighted in orange. (**B**) Representative loci architectures for C2c10 and CRISPR-Cas14 systems. (**C**) Length distribution of Cas14a to Cas14c systems compared with Cas12a to Cas12e and Cas9. (**D**) Domain organization of Cas14a compared with Cas9 and Cas12a. Nuclease domains (RuvC and HNH) are indicated. Protein lengths are drawn to scale.

The Cas14 proteins we identified occur almost exclusively within DPANN, a superphylum of symbiotic archaea characterized by small cell and genome sizes (*14*, *15*). Phylogenetic comparisons showed that Cas14 proteins are widely diverse with similarities to C2c10 and C2c9, families of bacterial RuvC domain–containing proteins that are sometimes found near a CRISPR array but not together with other *cas* genes (Fig. 1B and fig. S1) (*3*). This observation and the small size of the *c2c10*, *c2c9*, and *cas14* genes made it improbable that these systems could function as stand-alone CRISPR effectors (*3*).

On the basis of their proximity to conserved genes responsible for creating genetic memory of infection (*cas1*, *cas2*, *cas4*) (fig. S3A), we explored whether CRISPR-Cas14 systems can actively acquire DNA sequences into their CRISPR arrays. Assembled metagenomic contiguous DNA sequences (contigs) for multiple CRISPR-Cas14 loci revealed that otherwise identical CRISPR systems showed diversity in their CRISPR arrays. These results are consistent with active adaptation to new infections, although without longitudinal sampling these data could also be explained by alternative biological mechanisms (Fig. 2A and fig. S3B) (*13*). The evidence suggesting acquisition of new DNA sequences led us to hypothesize that these CRISPR-Cas14 loci encode functional enzymes with nucleic acid targeting activity despite their small size. To test this possibility, we first investigated whether RNA components are produced from CRISPR-Cas14 loci. Environmental metatranscriptomic sequencing data were analyzed for the presence of RNA from the native archaeal host that contains CRISPR-Cas14a (Fig. 2B and fig. S4A). In addition to CRISPR RNAs (crRNAs), a highly abundant noncoding RNA was mapped to a ~130–base pair sequence located between *cas14a* and the adjacent CRISPR array. Notably, the 3′ end of this transcript was mostly complementary to the repeat segment of the crRNA (Fig. 2C and fig. S4B), as observed for trans-activating CRISPR RNAs (tracrRNAs) found in association with Cas9, Cas12b, and Cas12e CRISPR systems (*12*, *13*, *16*). In these

previously studied systems, the double-stranded RNA–cutting enzyme ribonuclease III (RNase III) generates mature tracrRNAs and crRNAs, but no genes encoding RNase III were present in *cas14*-containing reconstructed genomes (fig. S5A), nor did Cas14a cleave its own pre-crRNA when tested biochemically (fig. S5B). These observations imply that an alternative mechanism for CRISPR-associated RNA processing exists in these hosts.

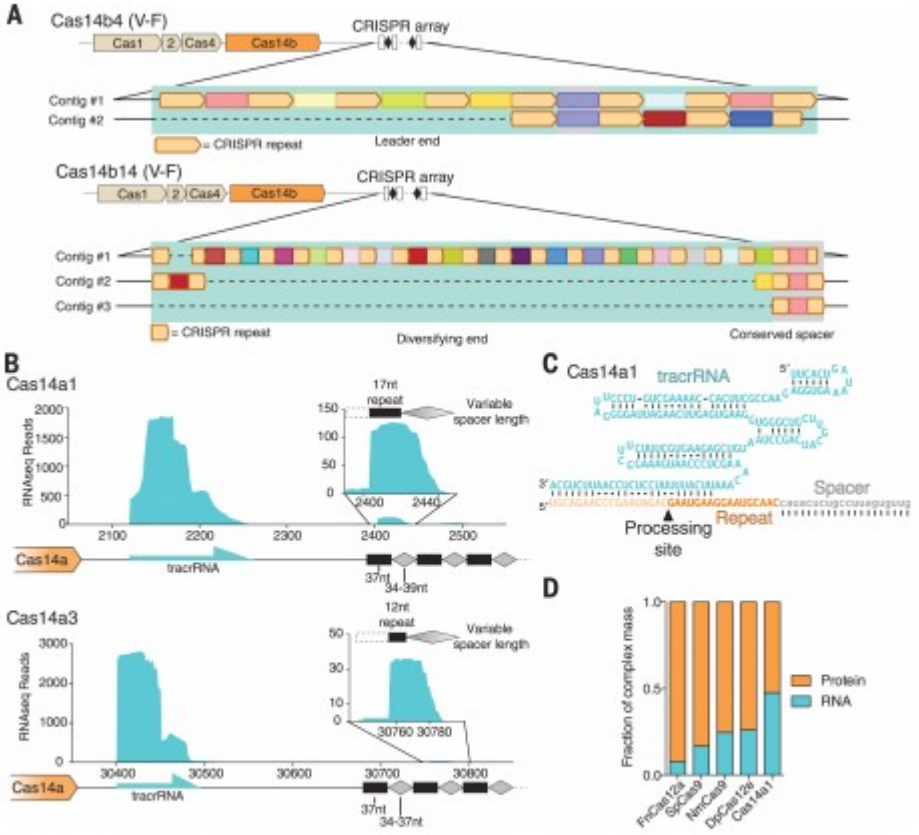


Fig. 2. CRISPR-Cas14a actively adapts and encodes a tracrRNA. (**A**) Spacer diversity for Cas14b4 and Cas14b14 with CRISPR repeats diagramed in tan and distinct spacers shown in different colors. (**B**) Metatranscriptomic reads mapped to Cas14a1 and Cas14a3 loci. The insets show an expanded view of the most abundant repeat and spacer sequence. nt, nucleotides. (**C**) In silico predicted structure of Cas14a1 crRNA and tracrRNA. Notably, RNase III orthologs were not identified in host genomes (fig. S5A). (**D**) Fraction of various CRISPR complex masses made up of RNA and protein.

To test whether the Cas14a proteins and associated RNA components can assemble together in a heterologous organism, we introduced a plasmid into *Escherichia coli* containing a minimal CRISPR-Cas14a locus that includes the *cas14* gene, the CRISPR array, and intergenic regions containing the putative tracrRNA. Affinity purification of the Cas14a protein from cell lysate and sequencing of copurifying RNA revealed a highly abundant mature crRNA as well as the putative tracrRNA, albeit in lower relative abundance than what was shown by environmental metatranscriptomics, suggesting that Cas14 associates with both crRNA and tracrRNA (fig. S5C). The calculated mass of the assembled Cas14a protein–tracrRNA–crRNA particle is 48% RNA by weight compared with just 17% for *Streptococcus*

pyogenes Cas9 (SpCas9) and 8% for *Francisella novicida* Cas12a (FnCas12a) (Fig. 2D), hinting at a central role of the RNA in the architecture of the Cas14a complex. Known class 2 CRISPR systems require a short sequence called a protospacer adjacent motif (PAM) to target double-stranded DNA (dsDNA) (*17*). To test whether Cas14a requires a PAM and can conduct dsDNA interference, we transformed *E. coli* expressing a minimal Cas14a locus with a dsDNA plasmid containing a randomized PAM region next to a sequence matching the target-encoding sequence (spacer) in the Cas14 array. Notably, no depletion of a PAM sequence was detected among *E. coli* transformants, suggesting that the CRISPR-Cas14a system is unable to target dsDNA, can do so without requiring a PAM, or is inactive in this heterologous host (fig. S6, A and B).

We next tested whether purified Cas14a-tracrRNA-crRNA complexes are capable of RNA-guided nucleic acid cleavage in vitro. All currently reconstituted DNA-targeting class 2 interference complexes are able to recognize both dsDNA and single-stranded DNA (ssDNA) substrates (*18–20*). We incubated purified Cas14a-tracrRNA-crRNA complexes with radiolabeled target oligonucleotides (ssDNA, dsDNA, and ssRNA) bearing a 20-nucleotide sequence complementary to the crRNA guide sequence or a noncomplementary ssDNA, and we analyzed these substrates for Cas14a-mediated cleavage. Only in the presence of a complementary ssDNA substrate was any cleavage product detected (Fig. 3A and fig. S7, A to C), and cleavage was dependent on the presence of both tracrRNA and crRNA, which could also be combined into a single-guide RNA (sgRNA) (Fig. 3B and fig. S8). The lack of detectable dsDNA cleavage suggests that Cas14a targets ssDNA selectively, although it is possible that some other host factor or sequence requirement could enable dsDNA recognition in the native host. Mutation of the conserved active site residues in the Cas14a RuvC domain eliminated cleavage activity (fig. S7, D and E), implicating RuvC as the domain responsible for DNA cutting. Moreover, Cas14a DNA cleavage was sensitive to truncation of the RNA components to lengths shorter than the naturally produced sequences (fig. S9, A to D). These results establish Cas14a as the smallest class 2 CRISPR effector demonstrated to conduct programmable RNA-guided DNA cleavage thus far.

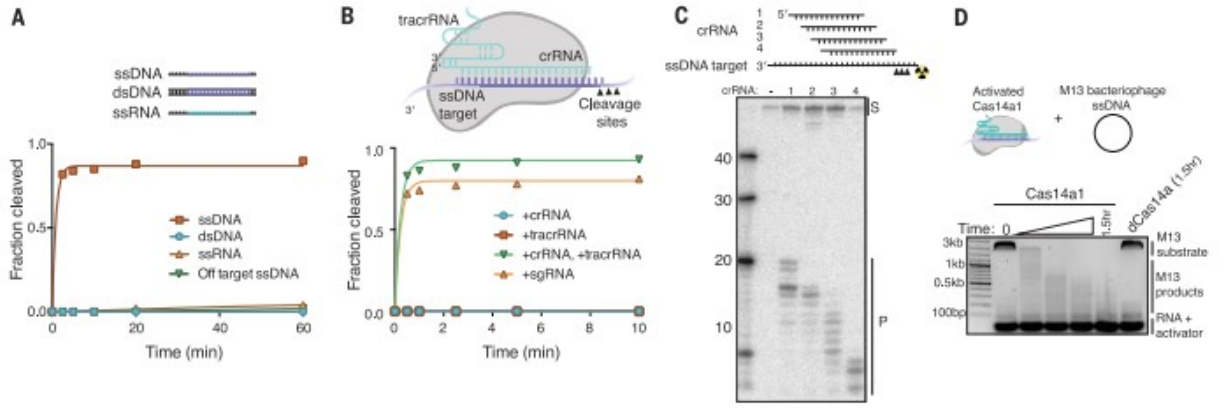


Fig. 3. CRISPR-Cas14a is an RNA-guided DNA endonuclease. (**A**) Cleavage kinetics of Cas14a1 targeting ssDNA, dsDNA, ssRNA, and off-target ssDNA. (**B**) Diagram of Cas14a RNP bound to target ssDNA and Cas14a1 cleavage kinetics of radiolabeled ssDNA in the presence of various RNA components. (**C**) Tiling of a ssDNA substrate by Cas14a1 guide sequences. (**D**) Cleavage of the ssDNA viral M13 genome with activated Cas14a1.

Although we were unable to identify a dsDNA PAM in vivo, we tested whether Cas14a requires a PAM for ssDNA cleavage in vitro by tiling Cas14a guides across a ssDNA substrate (Fig. 3C). Despite sequence variation adjacent to the targets of these different guides, we observed cleavage for all four sequences. Notably, the cleavage sites occur beyond the guide-complementary region of the ssDNA and shift in response to guide binding position (Fig. 3C). These data demonstrate Cas14a is a ssDNA-targeting CRISPR endonuclease that does not require a PAM for activation.

On the basis of the observation that Cas14a cuts outside of the crRNA/DNA targeting heteroduplex, we hypothesized that Cas14a might possess target-activated nonspecific ssDNA cleavage activity, similar to the RuvC-containing enzyme Cas12a (*20*, *21*). To test this possibility, we incubated Cas14a-tracrRNA-crRNA with a complementary activator DNA and an aliquot of M13 bacteriophage ssDNA bearing no sequence complementarity to the Cas14a crRNA or activator (Fig. 3D). The M13 ssDNA was rapidly degraded to small fragments, an activity that was eliminated by mutation of the conserved Cas14a RuvC active site, suggesting that activation of Cas14a results in nonspecific ssDNA degradation. However, we were unable to observe Cas14a-mediated interference against the ssDNA bacteriophage ΦX174 when we expressed Cas14a heterologously in *E. coli* (fig. S10, A to C), possibly due to the dissimilarity between *E. coli* and Cas14a's native archaeal host. To investigate the specificity of target-dependent nonspecific DNA cutting activity by Cas14a, we adapted a fluorophore-quencher (FQ) assay in which cleavage of dye-labeled ssDNA generates a fluorescent signal (Fig. 4A) (*22*). When Cas14a was incubated with various guide RNA–target ssDNA pairs, a fluorescent signal was observed only in the presence of the cognate target and showed strong preference for longer FQ-containing substrates (fig. S10D and Fig. 4A). We next tested Cas14a mismatch tolerance by tiling 2-nucleotide mismatches across the targeted region in various ssDNA substrates. Surprisingly, mismatches near the middle of the ssDNA target strongly inhibited Cas14a activity, revealing an internal seed

sequence that is distinct from the PAM-proximal seed region observed for dsDNA-targeting CRISPR-Cas systems (Fig. 4B and fig. S11, A to D). Moreover, DNA substrates containing strong secondary structure resulted in reduced activation of Cas14a (fig. S11E). Truncation of ssDNA substrates also resulted in reduced or undetectable trans cleavage (fig. S11F). Together, these results suggest a mechanism of fidelity distinct from dsDNA-targeting class 2 CRISPR systems, possibly using a mechanism similar to the ssRNA-targeting Cas13a enzymes (*23*–*25*).

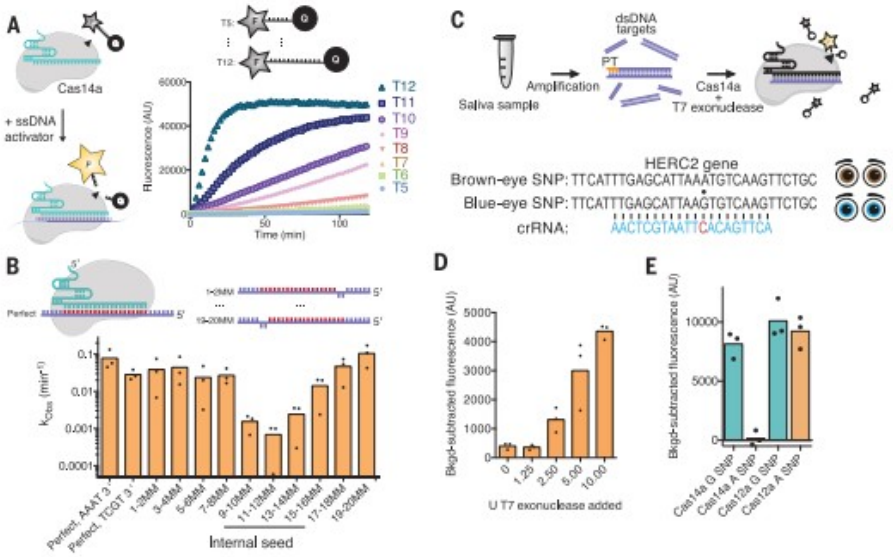


Fig. 4. High-fidelity ssDNA SNP detection by CRISPR-Cas14a. (**A**) FQ assay for detection of ssDNA by Cas14a1 and the cleavage kinetics for FQ substrates of various lengths. AU, arbitrary units. (**B**) Cleavage kinetics for Cas14a1 with mismatches tiled across the substrate (individual points represent replicate measurements). k_{Obs}, observed rate constant. (**C**) Diagram of Cas14-DETECTR strategy and *HERC2* eye color SNP. (**D**) Titration of T7 exonuclease and effect on Cas14a-DETECTR. Bkgd, background. (**E**) SNP detection using Cas14a-DETECTR with a blue-eye targeting guide for saliva samples from blue-eyed and brown-eyed individuals compared with ssDNA detection using Cas12a.

The target-dependent, nonspecific DNase activity of Cas12a serves as a DNA detection platform (DNA endonuclease-targeted CRISPR trans reporter, or DETECTR) for diagnostic uses (*20*, *26*). Although Cas12a exhibits low fidelity in discriminating against ssDNA substrates (*20*), Cas14a requires complementarity in the seed region for ssDNA substrate recognition. This improved specificity raised the possibility of using Cas14a for high-fidelity detection of DNA single-nucleotide polymorphisms (SNPs) without the constraint of a PAM sequence. To test this idea, DNA substrates were amplified using a phosphorothioate (PT)–containing primer to protect one strand from degradation by exonucleases. Upon addition of T7 exonuclease, the unmodified strand was degraded, leaving ssDNA substrates that can be detected by Cas14a (Fig. 4, C and D). As a proof of principle, we aimed to detect the human *HERC2* gene, which contains a SNP responsible for eye color (*27*). We amplified the *HERC2* gene from DNA in human saliva from both blue-eyed and brown-eyed individuals, using the PT amplification approach described above. When programmed with a guide RNA targeting the blue-eyed SNP, Cas12a failed to discriminate between the two ssDNA targets, exhibiting robust trans activity in both cases, whereas Cas14a exhibited strong activation in recognition of the blue-eyed SNP with near-background signal for the brown-eyed sample (Fig. 4E). The development of Cas14-DETECTR now allows for CRISPR-based detection of medically and

ecologically important ssDNA pathogens as well as high-fidelity detection of SNPs without the constraint of a PAM sequence.

Further investigation of compact type V systems in metagenomic data revealed a large diversity of systems that, like Cas14a to Cas14c, include a gene encoding a short RuvC-containing protein adjacent to acquisition-associated *cas* genes and a CRISPR array. We found 20 additional such systems in various uncultivated microbes that cluster into five main families (Cas14d to Cas14h). Excluding *cas14*g, which is related to *cas12*b, the *cas14*-like genes form separate clades on the type V effector phylogeny (fig. S12, A and B), suggesting that these families evolved from independent domestication events of TnpB, the transposase-associated protein implicated as the evolutionary ancestor of type V CRISPR effectors (*3*). Phylogenetic reconstruction of their associated *cas1* genes indicated that they too have different origins for the *cas14* subtypes (fig. S2). Altogether, we identified 38 CRISPR-Cas14 systems belonging to eight families (Cas14a to Cas14h) and eight additional systems that could not be clustered with our analysis (termed Cas14u) (data S3).

The small size of the Cas14 proteins described here and their resemblance to type V effector proteins suggest that RNA-guided ssDNA cleavage may have existed as an ancestral class 2 CRISPR system (*28*, *29*). In this scenario, a small, domesticated TnpB-like ssDNA interference complex may have gained additional domains over time, gradually improving dsDNA recognition and cleavage. Related to this hypothesis, smaller Cas9 orthologs exhibit weaker dsDNA-targeting activity than their larger counterparts but retain the ability to robustly cleave ssDNA (*19*). Aside from the evolutionary implications, the ability of Cas14 to specifically target ssDNA suggests a role in defense against ssDNA viruses or mobile genetic elements that propagate through ssDNA intermediates (*30*). A ssDNA-targeting CRISPR system would be particularly advantageous in certain ecosystems where ssDNA viruses constitute the vast majority of viral abundance (*31*). The unexpected finding that these miniature CRISPR proteins can conduct targeted DNA cleavage highlights the diversity of CRISPR systems hidden in uncultivated organisms. Ongoing exploration of these underrepresented microbial lineages will likely continue to reveal new, unexpected insights into this microscopic arms race and lead to continued development of valuable CRISPR-based technologies.

REFERENCES AND NOTES

1. R. Barrangou et al., Science 315, 1709–1712 (2007). 2. S. A. Jackson et al., Science 356, eaal5056 (2017). 3. S. Shmakov et al., Nat. Rev. Microbiol. 15, 169–182 (2017). 4. J. S. Chen, J. A. Doudna, Nat. Rev. Chem. 1, 0078 (2017). 5. C. T. Brown et al., Nature 523, 208–211 (2015). 6. K. Anantharaman et al., Nat. Commun. 7, 13219 (2016). 7. V. M. Markowitz et al., Nucleic Acids Res. 42, D568–D573 (2014). 8. A. J. Probst et al., Environ. Microbiol. 19, 459–474 (2017). 9. I. A. Chen et al., Nucleic Acids Res. 45, D507–D516 (2017). 10. I. Yosef, M. G. Goren, U. Qimron, Nucleic Acids Res. 40, 5569–5576 (2012). 11.

J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. A. Doudna, Nature 519, 193–198 (2015). 12. S. Shmakov et al., Mol. Cell 60, 385–397 (2015). 13. D. Burstein et al., Nature 542, 237–241 (2017). 14. C. Rinke et al., Nature 499, 431–437 (2013). 15. C. J. Castelle et al., Curr. Biol. 25, 690–701 (2015). 16. E. Deltcheva et al., Nature 471, 602–607 (2011). 17. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Microbiology 155, 733–740 (2009). 18. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. J. Sontheimer, Mol. Cell 60, 242–255 (2015). 19. E. Ma, L. B. Harrington, M. R. O'Connell, K. Zhou, J. A. Doudna, Mol. Cell 60, 398–407 (2015). 20. J. S. Chen et al., Science 360, 436–439 (2018). 21. B. Zetsche et al., Cell 163, 759–771 (2015). 22. A. East-Seletsky et al., Nature 538, 270–273 (2016). 23. L. Liu et al., Cell 170, 714–726.e10 (2017). 24. O. O. Abudayyeh et al., Science 353, aaf5573 (2016). 25. G. J. Knott et al., Nat. Struct. Mol. Biol. 24, 825–833 (2017). 26. S. Y. Li et al., Cell Res. 28, 491–493 (2018). 27. H. Eiberg et al., Hum. Genet. 123, 177–187 (2008). 28. E. V. Koonin, K. S. Makarova, F. Zhang, Curr. Opin. Microbiol. 37, 67–78 (2017). 29. K. S. Makarova et al., Nat. Rev. Microbiol. 13, 722–736 (2015). 30. O. Barabas et al., Cell 132, 208–220 (2008). 31. M. Yoshida et al., Front. Microbiol. 9, 75 (2018).

Acknowledgments: We thank N. Ma and K. Zhou for technical assistance and P. Harrington for graphic design assistance. Funding: D.B. was supported by a grant from the Paul Allen Frontiers Group. L.B.H., J.C.C., and J.S.C. were supported by U.S. National Science Foundation Graduate Research Fellowships. J.A.D. is an investigator of the Howard Hughes Medical Institute. This work was supported in part by a Frontiers Science award from the Paul Allen Institute to J.A.D. and J.F.B., a grant from the National Science Foundation (MCB-1244557 to J.A.D.), the Lawrence Berkeley National Laboratory's Sustainable Systems Scientific Focus Area funded by the U.S. Department of Energy (DE-AC02-05CH11231 to J.F.B.), and the Office of Science of the U.S. Department of Energy under contract no. DE-AC02–05CH11231. Author contributions: D.B. and D.P.-E. conducted the computational analysis. L.B.H., J.S.C., I.P.W., E.M., and J.C.C. designed and executed biochemical investigation of Cas14. L.B.H. designed and conducted experiments investigating Cas14 activity and assembly in *E. coli*. L.B.H. and D.B. conceived of the study. N.C.K., J.F.B., and J.A.D. supervised the research and experimental design. J.A.D., L.B.H., and D.B. wrote and revised the manuscript. All authors read, edited, and approved the manuscript. Competing interests: UC Regents have filed patents related to this work on which D.B., J.F.B., L.B.H., D.P.-E., J.S.C., and J.A.D. are inventors. L.B.H. and J.S.C. are cofounders of Mammoth Biosciences. I.P.W. is a consultant for Mammoth Biosciences. J.F.B. is a founder of Metagenomi. 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, Metagenomi, Mammoth Biosciences, and Inari. J.A.D. is a member of the

board of directors at Driver and Johnson & Johnson and has sponsored research projects by Roche Biopharma and Biogen. Data and materials availability: Data S1 specifies the accession numbers, coordinates, and samples of origin for all CRISPR-Cas14 systems described in this study.