PROKARYOTIC IMMUNITY

Functionally diverse type V CRISPR-Cas systems

Winston X. Yan[1]*, Pratyusha Hunnewell[1]*, Lauren E. Alfonse[1], Jason M. Carte[1], Elise Keston-Smith[1], Shanmugapriya Sothiselvam[1], Anthony J. Garrity[1], Shaorong Chong[1], Kira S. Makarova[2], Eugene V. Koonin[2], David R. Cheng[1], David A. Scott[1]†

Type V CRISPR-Cas systems are distinguished by a single RNA-guided RuvC domain-containing effector, Cas12. Although effectors of subtypes V-A (Cas12a) and V-B (Cas12b) have been studied in detail, the distinct domain architectures and diverged RuvC sequences of uncharacterized Cas12 proteins suggest unexplored functional diversity. Here, we identify and characterize Cas12c, -g, -h, and -i. Cas12c, -h, and -i demonstrate RNA-guided double-stranded DNA (dsDNA) interference activity. Cas12i exhibits markedly different efficiencies of CRISPR RNA spacer complementary and noncomplementary strand cleavage resulting in predominant dsDNA nicking. Cas12g is an RNA-guided ribonuclease (RNase) with collateral RNase and single-strand DNase activities. Our study reveals the functional diversity emerging along different routes of type V CRISPR-Cas evolution and expands the CRISPR toolbox.

Competition between prokaryotes and viruses has led to the evolution of diverse defense strategies, with more being identified through the mining of growing genomic and metagenomic sequence databases (*1*–*3*). Class 2 CRISPR-Cas systems are of particular interest, because their programmable single-effector nucleases have enabled genome engineering and nucleic acid detection tools (*4*–*8*). Class 2 systems include types II, V, and VI, which are based on Cas9, Cas 12, and Cas13 effectors, respectively (*9*–*11*). Cas9 contains an HNH nuclease domain inserted into a RuvC nuclease domain (*12*–*14*), and the two domains together cleave double-stranded DNA (dsDNA). Cas12 contains a single RuvC nuclease domain that cleaves dsDNA adjacent to protospacer adjacent motif (PAM) sequences (*15*) and single-stranded DNA (ssDNA) nonspecifically (*16*). Cas13 contains two HEPN domains that cleave RNA exclusively (*10*, *17*, *18*).

We aggregated more than 10 terabytes of sequence data and generated a database of 293,985 putative CRISPR-Cas systems (*19*). From this database we identified type V systems with predicted effectors ranging in size from 720 to 1093 amino acids, each of which contained a C-terminal RuvC domain (fig. S1). The classification tree of type V effectors splits into three major branches: (i) Cas12a, -c, -d, and -e; (ii) Cas12b and its distant homologs; and (iii) subtype V-U variants closely related to transposon-encoded TnpB. Predicted type V effectors in this study showed weak sequence similarity ($E > 10^{-3}$) with previously characterized ones. Combined with differences in locus organization and subsequently uncovered functional differences, our work supports the assignment of separate subtypes, V-G, V-H, and V-I (Fig. 1A, fig. S1, and table S1). Whereas Cas12h and -i cluster with Cas12b, albeit at a large evolutionary distance, Cas12g clusters with the predicted subtype V-U effectors and TnpBs (Fig. 1A). The subtype V-U effectors, including the recently identified Cas14a, -b, and -c (subtype V-F), are much smaller than the typical CRISPR effectors and show greater similarity to TnpB (*11*, *20*). Cas14a and Cas12g appear to have evolved from distinct TnpB ancestors (Fig. 1A). Thus, experimental characterization of subtype V-G is of particular interest to elucidate the routes of evolution of TnpB proteins into functional CRISPR effectors.

To functionally characterize the type V-G, -H, and -I systems, we used an *Escherichia coli* negative selection screen, in which RNA-guided interference activity of reconstituted CRISPR-Cas systems reduces bacterial viability at 37°C (*19*). Each screen included: (i) an effector plasmid carrying predicted Cas genes; (ii) a CRISPR array library targeting pACYC184 and *E. coli* essential genes; and (iii) a noncoding plasmid containing concatenated *cas* gene-flanking noncoding sequences for the unbiased detection of trans-activating crRNA (tracrRNA) elements (Fig. 1, B and C).

In vivo screening of the compact subtype V-G effector, Cas12g1 (767 amino acids), revealed interference activity that specifically targeted the sense DNA strand of actively transcribed substrate regions (Fig. 2A, fig. S2A, and table S2). Analysis of target-flanking sequences revealed no PAM requirements for interference (fig. S2, B to D). Mutation of the RuvC-I motif of Cas12g1 [Asp^{513}→Ala (D513A)] or omission of the noncoding plasmid substantially decreased interference activity (Fig. 2B and fig. S2, E to G). RNA sequencing of screen samples revealed a tracrRNA expressed from the noncoding plasmid and a mature crRNA from the CRISPR array library (Fig. 2, C and D, and fig. S3). However, purified Cas12g1 was incapable of processing its pre-crRNA in vitro with or without tracrRNA, suggesting that additional endogenous factors are required for in vivo crRNA biogenesis (figs. S4 and S5 and table S3).

By investigating the mechanism of in vivo interference by subtype V-G systems, we found that ternary complexes consisting of Cas12g1, tracrRNA, and in vivo screen–validated crRNAs showed no cleavage of cognate ssDNA or dsDNA substrates at 37°C (fig. S6, A and B, and tables S4 to S8). The Cas12g1 locus originates from a hot spring metagenome but, although the ternary complex is thermostable [complex melting temperature (T_m) = 74°C] (fig. S7), we observed no ssDNA or dsDNA cleavage at 42°C, 50°C, or 60°C (fig. S6, C to H). Given the transcriptional association of Cas12g1 in vivo interference indicated by the preferential targeting of sense-strand DNA (Fig. 2A), we assessed cleavage of sense ssDNA (containing crRNA spacer-complementary target) or antisense ssDNA substrates in the presence of sense or antisense RNA transcripts. The Cas12g1 ternary complex efficiently cleaved the sense ssDNA in the presence of sense RNA (hereafter, target RNA), and this activity increased in efficiency from 37° to 50°C (Fig. 2E and fig. S8A). No ssDNA cleavage was observed for any other DNA-RNA substrate combination (fig. S8, B to D). In the presence of target RNA, Cas12g1 ternary complex also cleaved unrelated collateral ssDNA (Fig. 2E and fig. S9), demonstrating that target RNA activates nonspecific collateral ssDNA cleavage in trans by Cas12g1.

The weak ssDNA cleavage observed at 37°C is likely not responsible for the robust Cas12g1 interference activity observed in vivo. Thus, we investigated the intrinsic ribonuclease (RNase) activity of Cas12g1 and observed strong target RNA cleavage with the ternary complex at 37°C (Fig. 2F), and this was further enhanced at 50°C (fig. S10 and tables S4 to S8). At 50°C, detectable target RNA cleavage was observed at ternary complex concentrations as low as 125 pM (fig. S11A), with no background cleavage of nontarget RNA at the highest complex concentration tested (250 nM) (figs. S10B and S11B). Cas12g1 ternary complex also cleaved dye-labeled collateral RNA accompanying unlabeled target RNA at target concentrations as low as 100 pM, demonstrating that the stand-alone RNA detection sensitivity of Cas12g1 is comparable to that of the highest performing Cas13 variants (Fig. 2F and fig. S11, C and D) (*21*). Both RNA and ssDNA cleavage by Cas12g1 are metal ion dependent and require an intact RuvC domain that was previously known to cleave only DNA (Fig. 2, G and H, and figs. S4 and S12). The thermostability and nucleic acid detection sensitivity of Cas12g1 has the potential to enhance the performance and durability of nucleic acid diagnostic methods, such as SHERLOCK and DETECTR (*16*, *18*, *22*). Additionally, the small size of Cas12g1 is likely to facilitate delivery for diverse in vivo transcriptome engineering applications (*23*, *24*).

We next investigated subtype V-H and V-I systems containing effectors Cas12h (870 to 933

[1]Arbor Biotechnologies, Cambridge, MA 02139, USA.
[2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: dscott@arbor.bio

Downloaded from http://science.sciencemag.org/ on February 23, 2019

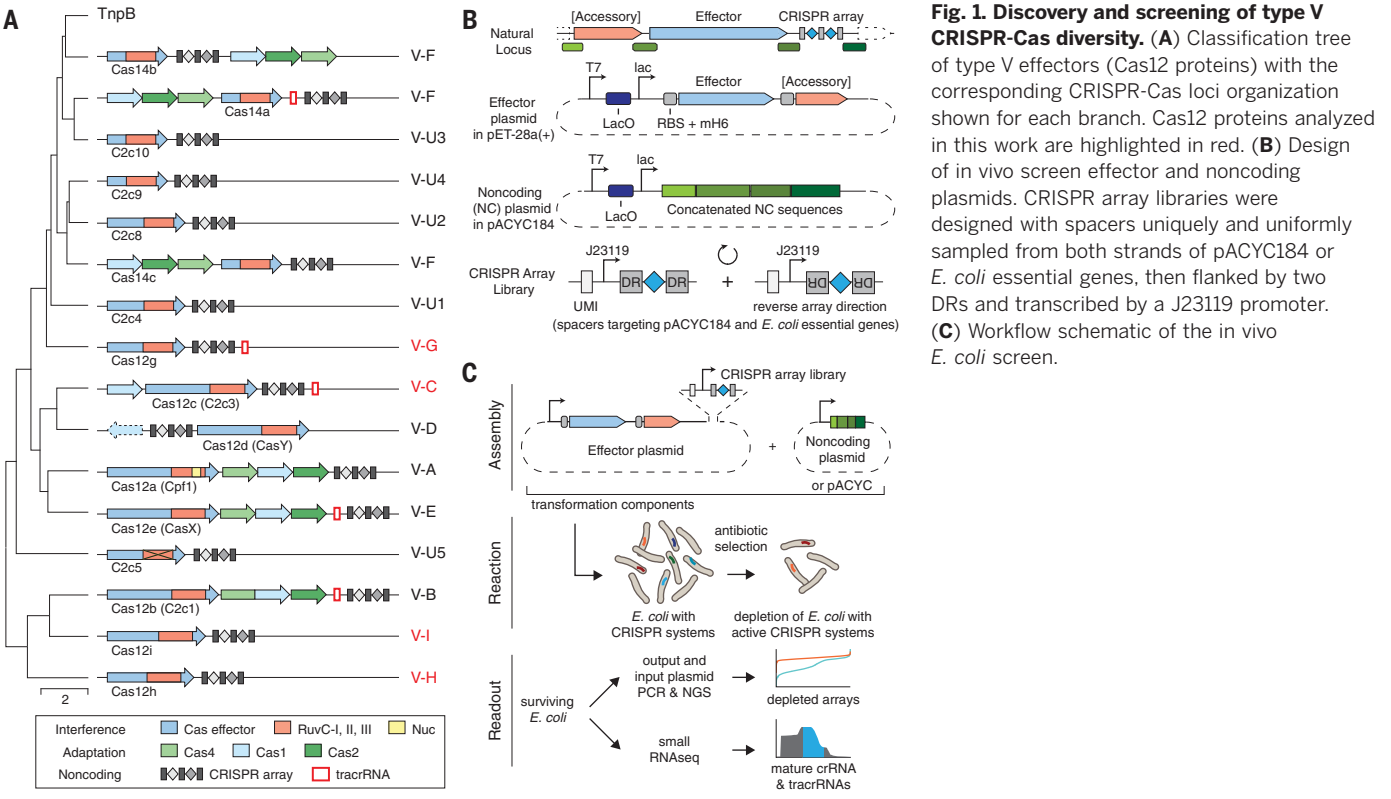


Fig. 1. Discovery and screening of type V CRISPR-Cas diversity. (**A**) Classification tree of type V effectors (Cas12 proteins) with the corresponding CRISPR-Cas loci organization shown for each branch. Cas12 proteins analyzed in this work are highlighted in red. (**B**) Design of in vivo screen effector and noncoding plasmids. CRISPR array libraries were designed with spacers uniquely and uniformly sampled from both strands of pACYC184 or *E. coli* essential genes, then flanked by two DRs and transcribed by a J23119 promoter. (**C**) Workflow schematic of the in vivo *E. coli* screen.

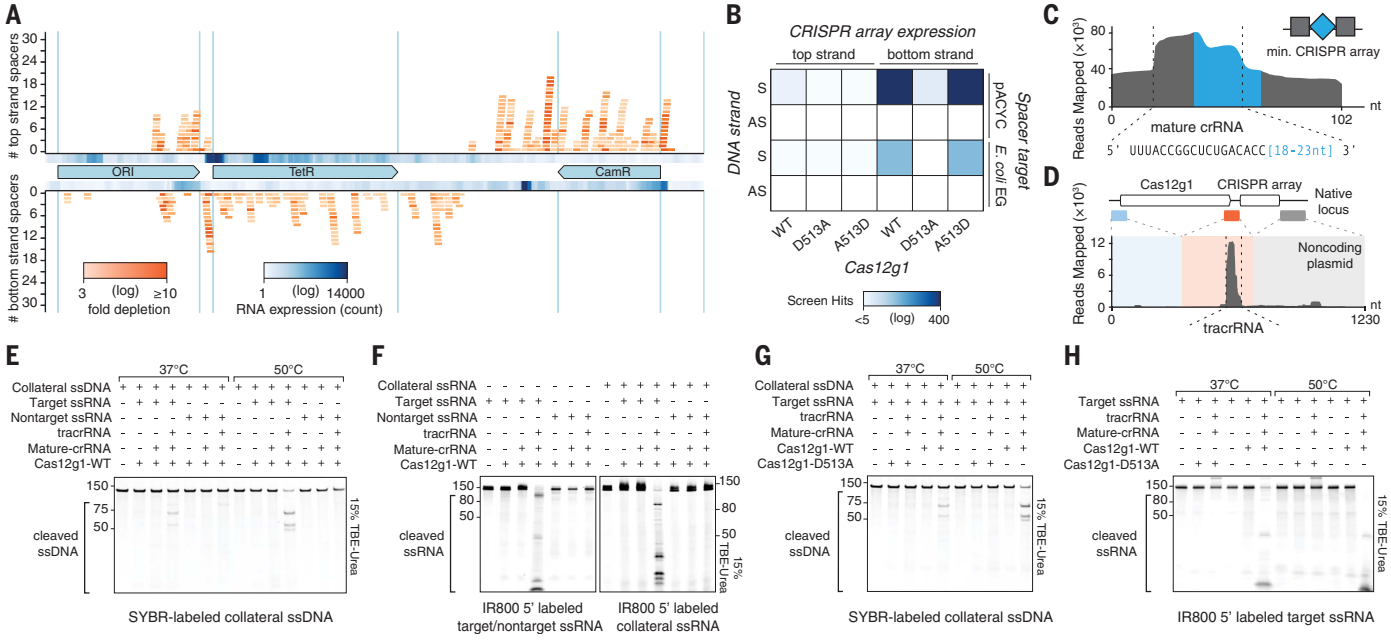


Fig. 2. Cas12g displays RNA-activated target cleavage of RNA and collateral trans-cleavage of RNA and ssDNA. (**A**) Strongly depleted CRISPR arrays from in vivo screening of Cas12g1 and its noncoding plasmid mapped to pACYC184. (**B**) Heatmap showing strongly depleted CRISPR arrays (screen hits) to evaluate RuvC and substrate strand dependencies of Cas12g1 (S, sense; AS, antisense; EG, essential genes). A513D was cloned from the D513A construct to rescue its activity. Strongly depleted CRISPR arrays in negative control screens without the effector were subtracted from this and similar analyses. (**C** and **D**) Mature crRNA (C) and tracrRNA (D) identified from small RNA sequencing of in vivo screen samples containing Cas12g1 and noncoding plasmid. The schematic above tracrRNA shows construction of noncoding plasmid from native locus sequences. (**E** and **F**) Target ssRNA activated collateral ssDNA cleavage at 37°C and 50°C (E) and target and collateral ssRNA cleavage at 37°C (F). (**G** and **H**) Cleavage assays targeting collateral ssDNA (G) and ssRNA (H) with purified RuvC mutant dCas12g1 D513A.

Downloaded from http://science.sciencemag.org/ on February 23, 2019

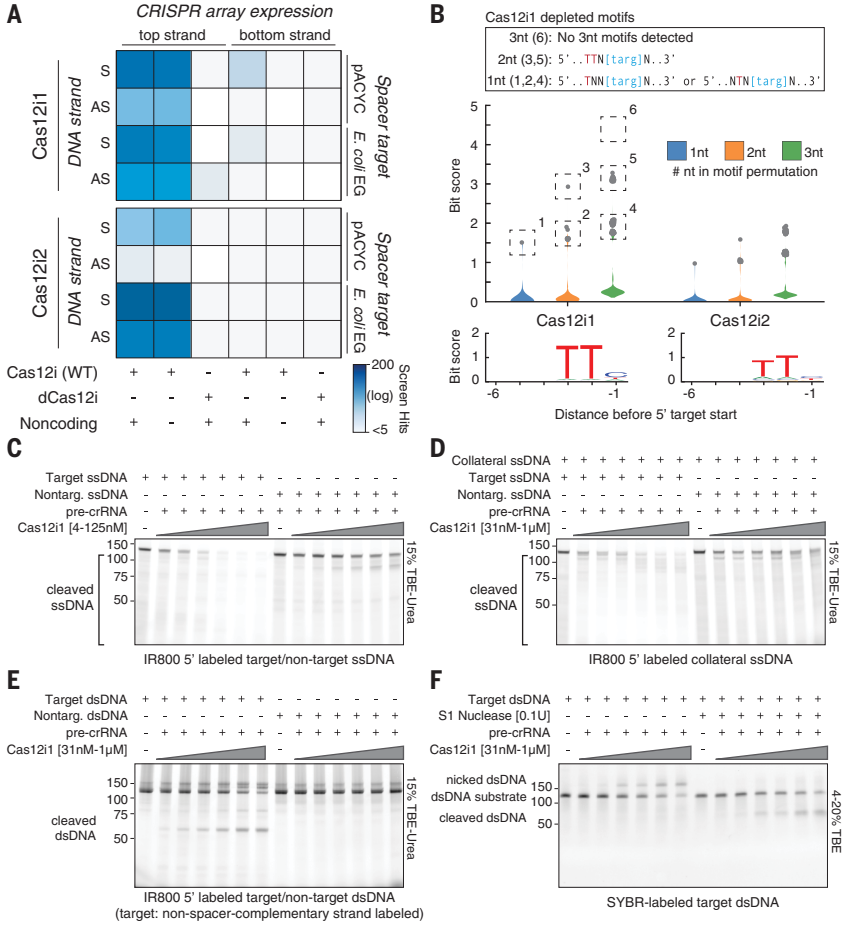


Fig. 3. In vivo and in vitro activity of Cas12i. (**A**) Evaluation of a minimal active system for Cas12i, with heatmaps showing strongly depleted CRISPR arrays from in vivo screening in different Cas12i system compositions (S, sense; AS, antisense; EG, essential genes). (**B**) (Top) Distribution of bit scores for all permutations of 1- to 3-nucleotide (nt) motifs within the target and 15-nt flanking sequences corresponding to strongly depleted in vivo arrays, calculated as described in (*19*). The box above describes motif analysis for Cas12i1 as an example. (Bottom) Web logos from target-flanking sequences. (**C** to **E**) Titration of a Cas12i1 binary complex on target and nontarget ssDNA (C), collateral ssDNA with target and nontarget ssDNA (D), and target and nontarget dsDNA (E). (**F**) S1 nuclease treatment to resolve dsDNA nicks (induced by Cas12i1) into dsDNA breaks.

amino acids) and Cas12i (1033 to 1093 amino acids), respectively. These effectors show distant similarity to Cas12b, with substantial truncation of N-terminal regions responsible for PAM recognition and DNA unwinding (*25*, *26*). In vivo screening of Cas12h1 (870 amino acids), Cas12i1 (1093 amino acids), and Cas12i2 (1054 amino acids) demonstrated robust and broadly distributed targeting of both strands of dsDNA substrates that was dependent on an intact RuvCI domain (Fig. 3A and figs. S13A and S14, A to F). The noncoding plasmid was not required, indicating that, unlike subtype V-B systems, the minimal V-H and V-I interference modules include only the effector and crRNA (Fig. 3A and fig. S14, G and H). Analysis of target-flanking sequences corresponding to strongly depleted arrays from in vivo screens showed that dsDNA interference by Cas12h1 depends on a 5′ RTR PAM (fig. S13B), whereas Cas12i1 and Cas12i2 prefer a 5′ TTN PAM (Fig. 3B).

Small RNA sequencing of Cas12i1 in vivo screen samples demonstrated biogenesis of a mature crRNA (fig. S15), which was confirmed in vitro using purified Cas12i1 and a minimal pre-crRNA (DR-spacer-DR-spacer-DR) (fig. S16 and table S3). Binary complexes containing Cas12i1 and pre-crRNAs efficiently cleaved target containing ssDNA substrates (Fig. 3C) as well as labeled collateral ssDNA in the presence of unlabeled target ssDNA, consistent with collateral ssDNA cleavage activity (Fig. 3D).

We observed Cas12i1-mediated cleavage of dsDNA under denaturing conditions, which was suggestive of dsDNA nicking. While reactions containing dsDNA with a labeled non-spacer-complementary strand showed robust DNA cleavage over a wide range of binary complex concentrations, those containing dsDNA with the spacer-complementary strand labeled showed only small amounts of cleavage at the highest concentrations tested (Fig. 3E and fig. S17, A and B). Under nondenaturing conditions, Cas12i1 cleavage reactions yielded products with lower electrophoretic mobility than the input dsDNA, and these products were then converted to double-strand breaks by S1 nuclease treatment, consistent with nicking of dsDNA substrates (Fig. 3F). These results suggested that Cas12i1 preferentially nicks the non-spacer-complementary strand, and it cleaves the spacer-complementary strand with a lower efficiency to yield a dsDNA break. Together, the small size, autonomous processing of multiplexed crRNAs, and nicking activity of Cas12i could enhance double-nicking applications for high-fidelity genome editing (*27*).

Subtype V-C loci have been previously observed but never characterized due to incomplete genomic data (*10*). With our expanded database, we detected and synthesized in vivo screen plasmids for complete subtype V-C systems containing the effectors OspCas12c (from *Oleiphilus* sp. HI0009), Cas12c1, and Cas12c2. All these systems showed broad and symmetrical targeting of both DNA strands, consistent with autonomous dsDNA interference (Fig. 4A and fig. S18). RNA sequencing of screening samples for the minimal subtype V-C systems demonstrated pre-crRNA processing and highly expressed tracrRNAs (fig. S19). A 5′ TG PAM was required for Cas12c1 and OspCas12c, and a minimal 5′ TN PAM was required for Cas12c2 (Fig. 4B). The single-nucleotide TN PAM for Cas12c2 dsDNA targeting complements recently engineered Cas9 effectors with minimal PAMs (*28*), potentially expanding the target space for genome editing.

We have presented here a framework for systematic discovery, screening, and characterization of class 2 CRISPR-Cas systems, and we demonstrated a range of activities for four type V CRISPR-Cas subtypes, including target and collateral cleavage of ssRNA and ssDNA as well as dsDNA nicking and cleavage (Fig. 4C). These findings reveal the transition in the properties of Cas12 proteins along the proposed evolutionary path from TnpB to large type V effectors. Additionally, future applications could include

Downloaded from http://science.sciencemag.org/ on February 23, 2019

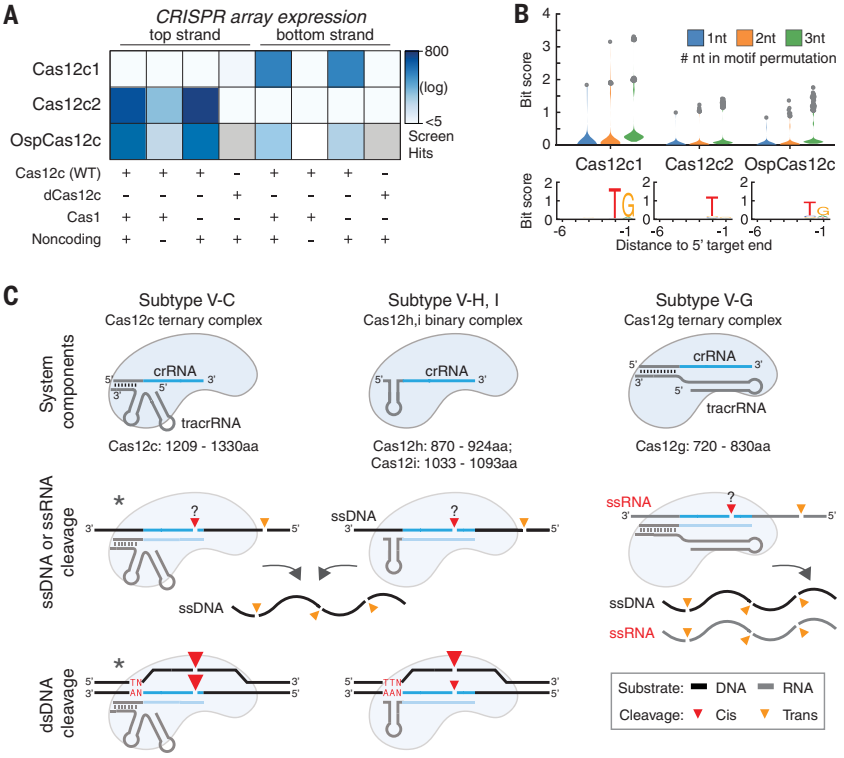


Fig. 4. In vivo dsDNA interference by Cas12c. (**A**) Evaluation of a minimal active system for Cas12c, with heatmaps showing strongly depleted CRISPR arrays from in vivo screening in different Cas12c system compositions. Gray boxes indicate data not available. (**B**) (Top) Distribution of bit scores for all permutations of 1- to 3-nt motifs within the target and 15-nt flanking sequences corresponding to strongly depleted arrays. (Bottom) Web logos from target-flanking sequences. (**C**) Overview of minimal components and interference mechanisms of Cas12g, -h, -i, and -c. Asterisks denote putative mechanisms subject to additional validation.

expanded genomic targeting via the minimal Cas12c2 PAM, high-fidelity genome editing using Cas12i nicking (*27*), or sensitive and durable nucleic acid detection via collateral cleavage by the thermostable Cas12g1 (*18*, *22*). We anticipate that our discovery framework will yield new CRISPR-Cas variants as genomic and metagenomic sequence databases grow, expanding the understanding of CRISPR biology and the nucleic acid manipulation toolbox.

REFERENCES AND NOTES

1. D. Burstein *et al.*, *Nature* **542**, 237–241 (2017).
2. S. Doron *et al.*, *Science* **359**, eaar4120 (2018).
3. E. V. Koonin, K. S. Makarova, Y. I. Wolf, *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
4. R. Barrangou, P. Horvath, *Nat. Microbiol.* **2**, 17092 (2017).
5. R. Barrangou *et al.*, *Science* **315**, 1709–1712 (2007).
6. D. B. T. Cox, R. J. Platt, F. Zhang, *Nat. Med.* **21**, 121–131 (2015).
7. S. E. Klompe, S. H. Sternberg, Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems. *CRISPR J.* **1**, 141–158 (2018).
8. G. J. Knott, J. A. Doudna, *Science* **361**, 866–869 (2018).
9. K. S. Makarova *et al.*, *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
10. S. Shmakov *et al.*, *Mol. Cell* **60**, 385–397 (2015).
11. S. Shmakov *et al.*, *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
12. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586 (2012).
13. M. Jinek *et al.*, *Science* **337**, 816–821 (2012).
14. E. V. Koonin, K. S. Makarova, F. Zhang, *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
15. B. Zetsche *et al.*, *Cell* **163**, 759–771 (2015).
16. J. S. Chen *et al.*, *Science* **360**, 436–439 (2018).
17. O. O. Abudayyeh *et al.*, *Science* **353**, aaf5573 (2016).
18. A. East-Seletsky *et al.*, *Nature* **538**, 270–273 (2016).
19. W. X. Yan *et al.*, *Mol. Cell* **70**, 327–339.e5 (2018).
20. L. B. Harrington *et al.*, *Science* **362**, 839–842 (2018).
21. J. S. Gootenberg *et al.*, *Science* **360**, 439–444 (2018).
22. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
23. D. B. T. Cox *et al.*, *Science* **358**, 1019–1027 (2017).
24. S. Konermann *et al.*, *Cell* **173**, 665–676.e14 (2018).
25. D. Dong *et al.*, *Nature* **532**, 522–526 (2016).
26. T. Yamano *et al.*, *Cell* **165**, 949–962 (2016).
27. F. A. Ran *et al.*, *Cell* **154**, 1380–1389 (2013).
28. J. H. Hu *et al.*, *Nature* **556**, 57–63 (2018).

ACKNOWLEDGMENTS

We thank the entire Arbor Biotechnologies team for support and comments on this work. **Funding:** Arbor Biotechnologies is a privately funded company. K.S.M. and E.V.K. are supported by the intramural program of the U.S. Department of Health and Human Services (to the National Library of Medicine). **Author contributions:** W.X.Y. and D.A.S., with input from P.H., D.R.C., L.E.A., J.M.C., E.K.S., S.S., S.C., and A.J.G., conceived and designed the study. D.R.C. and D.A.S. designed and implemented the computational searches, with additional input from K.S.M. and E.V.K., including phylogenetic analysis and classification. W.X.Y., D.A.S., P.H., L.E.A., J.C., E.K.S., S.S., S.C., and A.J.G. performed all of the experimental work and analyzed the data. W.X.Y. and D.A.S. wrote the manuscript with input from E.V.K. and help from all authors. **Competing interests:** W.X.Y., P.H., L.E.A., J.M.C., E.K.S., S.S., S.C., A.J.G., D.R.C., and D.A.S. are employees and shareholders of Arbor Biotechnologies, Inc. W.X.Y., D.R.C., and D.A.S. are current or former officers and D.R.C. is a director of Arbor Biotechnologies. Arbor Biotechnologies has filed patents related to this work. **Data and materials availability:** All data are available in the manuscript or the supplementary material. All reagents are available to the academic community through Addgene. Sequencing data are available on the NCBI Sequence Read Archive under Bioproject ID PRJNA496291.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/88/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S10
References (*29*–*33*)

14 October 2018; accepted 20 November 2018
Published online 6 December 2018
10.1126/science.aav7271

Downloaded from http://science.sciencemag.org/ on February 23, 2019

Science

Functionally diverse type V CRISPR-Cas systems

Winston X. Yan, Pratyusha Hunnewell, Lauren E. Alfonse, Jason M. Carte, Elise Keston-Smith, Shanmugapriya Sothiselvam, Anthony J. Garrity, Shaorong Chong, Kira S. Makarova, Eugene V. Koonin, David R. Cheng and David A. Scott

Science **363** (6422), 88-91.
DOI: 10.1126/science.aav7271originally published online December 6, 2018

Additional, diverse CRISPR systems

CRISPR systems have been revolutionizing molecular biology. Mining the metagenomic database, Yan *et al.* systematically discovered additional subtypes of type V CRISPR-Cas systems. The additional Cas12 effectors displayed a range of activities, including target and collateral cleavage of single-stranded RNA and DNA, as well as double-stranded DNA nicking and cleavage. These diverse nuclease activities suggest how an ancient transposase may have evolved into various type V effectors and expand the nucleic acid detection and genome-editing toolbox.

Science, this issue p. 88

ARTICLE TOOLS http://science.sciencemag.org/content/363/6422/88

SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2018/12/05/science.aav7271.DC1

REFERENCES This article cites 32 articles, 11 of which you can access for free
http://science.sciencemag.org/content/363/6422/88#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.

Downloaded from http://science.sciencemag.org/ on February 23, 2019