Originally published 18 October 2018; corrected 14 November 2018



www.sciencemag.org/cgi/content/full/science.aav4294/DC1

Supplementary Materials for

Programmed DNA destruction by miniature CRISPR-Cas14 enzymes

Lucas B. Harrington*, David Burstein*, Janice S. Chen, David Paez-Espino, Enbo Ma, Isaac P. Witte, Joshua C. Cofsky, Nikos C. Kyrpides, Jillian F. Banfield Jennifer A. Doudna†

*These authors contributed equally to this work.
†Corresponding author. Email: doudna@berkeley.edu

Published 18 October 2018 on *Science* First Release
DOI: 10.1126/science.aav4294

This PDF file includes:

Materials and Methods
Figs. S1 to S12
Captions for Data S1 to S5
References (*32–42*)

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aav4294/DC1)

Data S1 to S5: Data S1 (.xlsx)
Data S2 (.fasta)
Data S3 and S4 (.zip)
Data S5 (.xlsx)

Correction: “Cas14a1” was mislabeled as “Cas14a3” in the original supplementary materials pdf, as well as in data files S2 to S5. These errors have been corrected.

Materials and Methods

Metagenomics and metatranscriptomics

The initial analysis was performed on previously assembled and binned metagenomes from two sites: the Rifle Integrated Field Research (IFRC) site, adjacent to the Colorado River near Rifle, Colorado (5, 6) and Crystal Geyser, a cold, CO_2-driven geyser on the Colorado Plateau in Utah (8). Metatranscriptomic data from IFRC site (5, 6) was used to detect transcription of non-coding elements in nature. Further mining of CRISPR-Cas14 systems was then performed on public metagenomes from IMG/M (7, 9, 32)

CRISPR-Cas computation analysis

The assembled contigs from the various samples were scanned with the HMMer suite (33) for known Cas proteins using Hidden Markov Model (HMMs) profiles (13). Additional HMMs were constructed for Cas14 proteins based on the MAFFT alignments of putative type V effectors that contained fewer than 800 aa, and were adjacent to acquisition *cas* genes and CRISPR arrays. These HMMs were iteratively refined by augmenting them with manually selected novel putative Cas14 sequences that were found using the existing Cas14 HMM models. The sequences of newly identified Cas14 orthologs are provided in Data S2. CRISPR arrays were identified using a local version of the CrisprFinder software (34) and CRISPRDetect (35). Phylogenetic trees of Cas1 and type V effector proteins were constructed using RAxML (36) with PROTGAMMALG as the substitution model and 100 bootstrap samplings. Trees were visualized using FigTree 1.4.1 (http://tree.bio.ed.ac.uk/software/figtree/), and iTOL v3 (37). Metatranscriptomic reads were mapped to assembled contigs using Bowtie2 (38).

RNase presence analysis was based on HMMs that were built from alignment of KEGG orthologous groups (KOs)(39).

Generation of expression plasmids, RNA, and DNA substrates

Minimal CRISPR loci for putative systems were designed by removing acquisition proteins and generating minimal arrays with a single spacer. These minimal loci were ordered as gBlocks (IDT) and assembled into a plasmid with a tetracycline-inducible promoter driving expression of the locus. Plasmid maps are available on Addgene. All RNA was *in vitro* transcribed using T7 RNA polymerase and PCR products as dsDNA templates. Resulting in vitro transcriptions were gel extracted and ethanol precipitated. DNA substrates were obtained from IDT and their sequences are available in Supplementary Data 5. For radiolabeled cleavage assays, DNA oligos were gel extracted from a PAGE gel before radiolabeling using T4 PNK (NEB). For FQ assays, DNA substrates were used without further purification.

***E. coli* RNAseq**

Small RNA sequencing was conducted as described previously with modification (40). *E. coli* NEB Stable3 was transformed with a plasmid expressing the Cas14a1 system with a tetracycline-inducible promoter upstream of the Cas14a1 ORF or the same plasmid with an N-terminal 10x-histidine tag fused to Cas14. Starters were grown up overnight in SOB, diluted 1:100 in 5mL fresh SOB containing 214 nM anhydrotetracycline and grown up overnight at 25°C. For sequencing of RNA pulled down with Cas14a, the plasmid containing an N-terminal His-tag fused to Cas14a1 was grown up at 18°C before lysis and purified as described in "Protein purification", stopping after the Ni-NTA elution. For total small RNA, cells were pelleted and RNA was extracted using hot phenol as previously described (40). Total nucleic acids were

treated with TURBO DNase and phenol extracted. The resulting RNA was treated with rSAP, which was heat inactivated before addition of T4 PNK. Adapters were ligated onto the small RNA using the NEBnext small RNA kit and gel-extracted on an 8% native PAGE gel. RNA was sequenced on a MiSeq with single-end 300 bp reads. For analysis, the resulting reads were trimmed using Cutadapt, discarding sequences <8 nt, and mapped to the reference plasmid using Bowtie2 (38).

PAM depletion assays

PAM depletion assays were conducted using methods previously described (13). Randomized plasmid libraries were generated using a primer containing a randomized PAM region adjacent to the target sequence. The randomized primers were hybridized with a primer that was complementary to the 3' end of the primer and the duplex was extended using Klenow Fragment (NEB). The dsDNA containing the target was digested with EcoRI and NcoI, ligated into pUC19 backbone, and transformed into *E. coli* DH5α. $>10^7$ cells were harvested. Next, *E. coli* NEBstable was transformed with either a CRISPR plasmid or an empty vector control and these transformed *E. coli* were made electrocompetent by repeated washing with 10% glycerol. These electrocompetent cells were transformed with 200 ng of the target library and plated on bioassay dishes containing selection for the target (carbenicillin, 100 mg l^{-1}) and CRISPR plasmid (chloramphenicol, 30 mg l^{-1}). Cells were harvested and prepared for amplicon sequencing on an Illumina MiSeq. The PAM region was extracted using Cutadapt and depletion values were calculated in python. PAMs were visualized using WebLogo(41).

Protein purification

Cas14a1 was purified as described previously with modification (42). *E. coli* BL21(DE3) RIL were transformed with 10xHis-MBP-Cas14a1 expression plasmid and grown up to OD_{600} =0.5 in

Terrific Broth (TB) and induced with 0.5 mM IPTG. Cells were grown overnight at 18°C, collected by centrifugation, resuspended in Lysis Buffer (50 mM Tris-HCl, pH 7.5, 20 mM imidazole, 0.5 mM TCEP, 500 mM NaCl) and broken by sonication. Lysate was batch loaded onto Ni-NTA resin, washed with the above buffer and eluted with Elution Buffer (50 mM Tris-HCl, pH 7.5, 300 mM imidazole, 0.5 mM TCEP, 500 mM NaCl). The MBP and His-tag were removed by overnight incubation with TEV protease at 4°C. The resulting protein was exchanged into Buffer A (20 mM HEPES, pH 7.5, 0.5 mM TCEP, 150 mM NaCl), loaded over tandem MBP, heparin columns (GE, Hi-Trap), and eluted with a linear gradient from Buffer A to Buffer B (20 mM HEPES, pH 7.5, 0.5 mM TCEP, 1250 mM NaCl). The resulting fractions containing Cas14a1 were loaded onto an Superdex 200 gel filtration column, concentrated, flash frozen, and stored at -80°C until use.

In vitro **cleavage assays**

Radiolabeled

Radiolabeled cleavage assays were conducted in 1× Cleavage Buffer (25 mM NaCl, 20 mM HEPES, pH 7.5, 1 mM DTT, 5% glycerol and 5mM $MgCl_2$). For cleavage assays with different divalent cations, 5mM $MgCl_2$ was replaced with 5mM of the indicated metal or EDTA. 100nM Cas14a1 was complexed with 125 nM crRNA and 125 nM tracrRNA for 10 min at RT. ~1nM radiolabeled DNA or RNA substrate was added and allowed to react for 30 min at 37°C unless otherwise noted for kinetic measurments. The reaction was stopped by adding 2x Quench Buffer (90% formamide, 25 mM EDTA and trace bromophenol blue), heated to 95°C for 2 min and run on a 10% polyacrylamide gel containing 7 M Urea and 0.5×TBE. Products were visualized by phosphorimaging.

RNA processing

RNA processing assays were conducted in 1× Cleavage Buffer (25 mM NaCl, 20 mM HEPES, pH 7.5, 1 mM DTT, 5% glycerol and 5 mM of the specified metal). 200 nM Cas14a1 was complexed with 100 nM crRNA and 100 nM tracrRNA for 30 min at 37°C. The reaction was stopped by adding 2x Quench Buffer (90% formamide, 25 mM EDTA and trace bromophenol blue), treated with proteinase K and run on a 10% polyacrylamide gel containing 7 M Urea and 0.5×TBE. Products were visualized by staining with SYBR Gold stain.

M13 DNA cleavage

M13 DNA cleavage assays were conducted in 100 mM NaCl, 20 mM HEPES, pH 7.5, 1 mM DTT, 5% glycerol and 5 mM $MgCl_2$. 250 nM Cas14a1 was complexed with 250 nM crRNA and 250 nM tracrRNA and 250 nM ssDNA activator. The reaction was initiated by addition of 5 nM M13 ssDNA plasmid and was quenched by addition of loading buffer supplemented with 10mM EDTA. Products were separated on a 1.5% agarose TAE gel prestained with SYBR Gold (Thermofisher).

FQ detection of trans-cleavage

FQ detection was conducted as previously described with modification (20). 100 nM Cas14a1 was complexed with 125 nM crRNA, 125 nM tracrRNA and 50 nM FQ probe in 1× Cleavage Buffer at 37°C for 10 min. The reaction was then initiated by addition of 2nM ssDNA activator for all reactions except for the RNA optimization experiments, where the variable RNA component was used to initiate. The reaction was monitored in a fluorescence plate reader for up to 120 minutes at 37°C with fluorescence measurements taken every 1 min (λ_{ex}: 485 nm; λ_{em}: 535 nm). The resulting data were background-subtracted using the readings taken in the absence of activator and fit to obtain K_{Obs}.

Cas14 and Cas12a-DETECTR assays

For Cas14a1 or Cas12a detection of the A/G SNP in the HERC2 gene, saliva samples were taken at three independent times by brown and blue-eyed individuals. For crude DNA extraction, saliva was pelleted and washed twice in phosphate buffered saline (1 × PBS), incubated for 5 min at 100°C, and centrifuged for 5 min at 10000×g.

DETECTR assays involved an initial PCR amplification followed by detection by Cas14a1 or LbCas12a. 50-*μ*L PCRs, consisting of 1 *μ*L template DNA, 10 *μ*L 5X Q5 Buffer, 0.48 *μ*M forward/reverse primers, 200 *μ*M (each) dNTPs, and 1 U Q5 DNA Polymerase underwent 25 cycles of amplification. The first four 5′ nucleotides of the forward primer were phosphorothioated to protect from degradation by T7 exonuclease in the subsequent detection step, while the reverse primer was unmodified. 2 *μ*L of the PCR product was transferred to a 384-well plate and combined directly in the plate with the DETECTR reaction mix. The Cas14a1 DETECTR reaction consisted of a final concentration of 100 nM Cas14a1, 125 nM sgRNA, 50 nM ssDNA-FQ reporter, and 5 U T7 exonuclease in a total reaction volume of 20 *μ*L. The LbCas12a DETECTR reaction consisted of a final concentration of 50 nM LbCas12a, 50 nM sgRNA, 50 nM ssDNA-FQ reporter, and 2.5 U T7 exonuclease in a total reaction volume of 20 *μ*L. Reactions were incubated in a fluorescence plate reader (Tecan Infinite Pro 200 M Plex) for 2 hours at 37°C with fluorescence measurements taken every 30 seconds (λex: 485 nm; λem: 535 nm).

Phage interference assays

For plaquing assays, *E. coli* strain C (ATCC 13706) was transformed with the specified CRISPR system and grown to saturation in SOB media containing 214 nM anhydrotetracycline (AtC)

overnight. The culture was diluted 1:1000 in soft agar overlay held at 42°C and containing 214 nM AtC. The mixture was poured on an LB agar plate containing chloramphenicol (30mg l^{-1}) and 214 nM AtC. After solidifying, the stock of Phi X174 (ATCC 13706B1) was serially diluted (1:10) and spotted onto the plate. For liquid growth assays, the overnight culture was diluted 1:100 in SOB containing chloramphenicol (30 mg l^{-1}) and 214 nM AtC, and a 1:1000 dilution of the phage stock was added to the media. The cultures were grown overnight in a Tecan plate reader at 37°C with shaking while OD600 was monitored.

Data Availability

Plasmids used in the study are available on Addgene (plasmid numbers 112500, 112501, 112502, 112503, 112504, 112505, 112506). All the sequences reported in this study for the first time have been deposited in NCBI. Nucleotide database accession and coordinates of each locus are specified in Supplemental Data 1. The Cas14 protein sequences used in this study are provided in Supplementary Data 2.

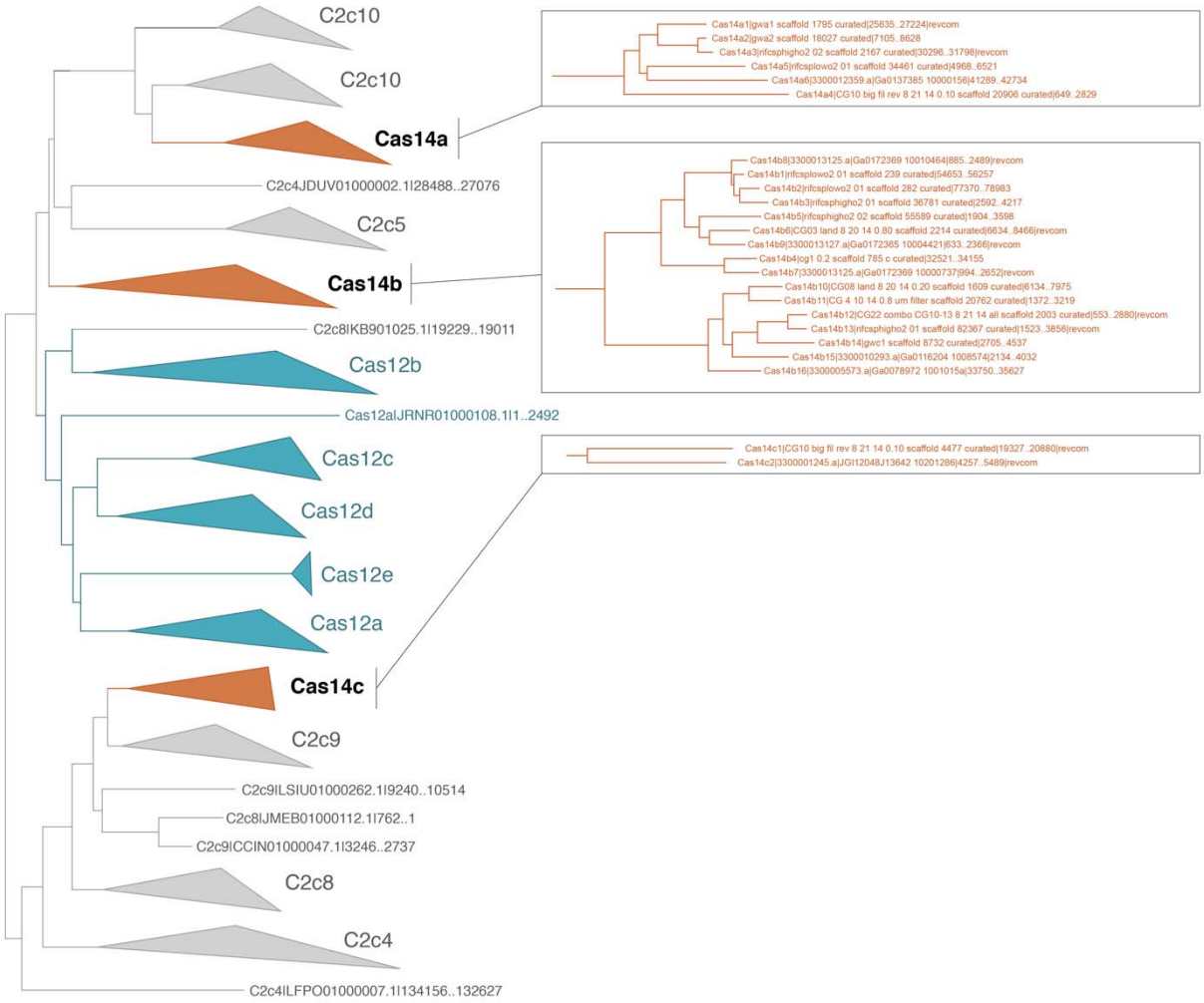


Fig. S1. Phylogenetic analysis of Cas14 orthologs

Maximum likelihood tree for known Type V CRISPR effectors and class 2 candidates containing a RuvC domain. Inset shows individual orthologs for each newly identified subtype.

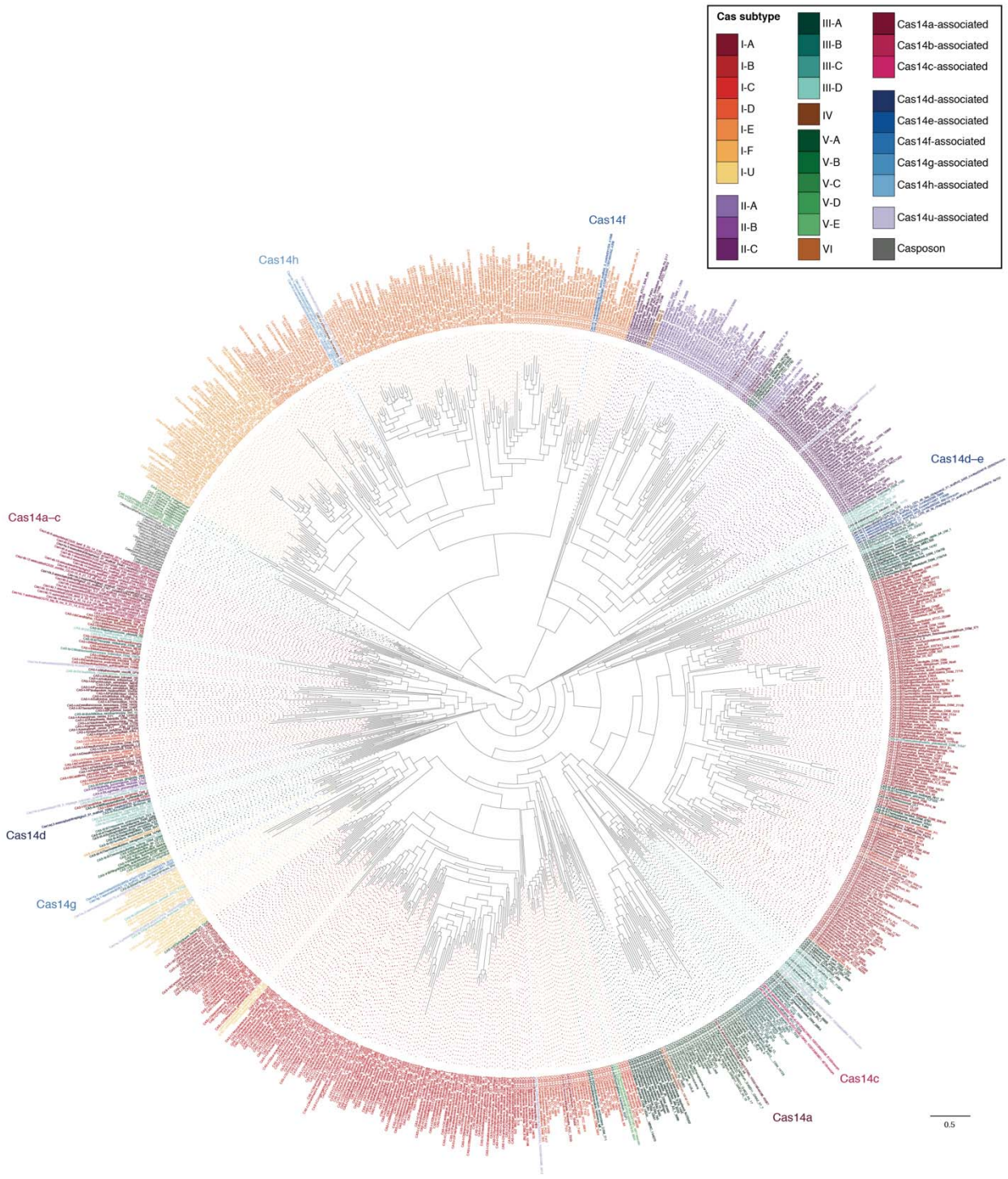


Fig. S2. Maximum likelihood tree for Cas1 from known CRISPR systems. Tree is rooted using Casposons as an outgroup. Newick format of this tree, including bootstrap values, is provided in Data S4.

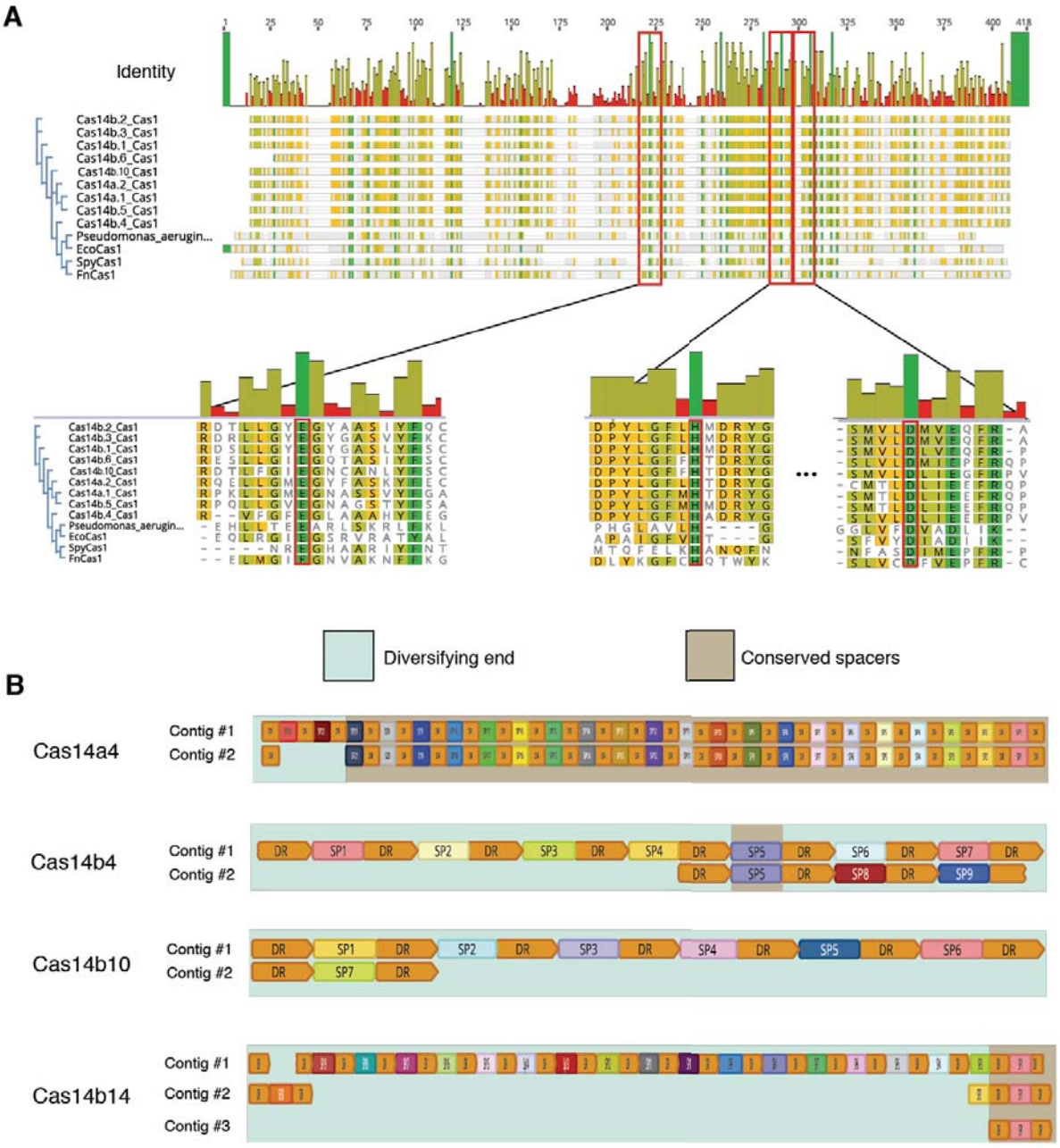


Fig. S3. Acquisition of new spacers by CRISPR-Cas14 systems

(A) Alignment of Cas14 Cas1 orthologs. Expansion shows conservation of previously implicated active site residues highlighted in red boxes. (B) Multiple CRISPR arrays assembled for various CRISPR-Cas14 systems revealing spacer diversity for these CRISPR systems. Orange arrows indicate repeats while variously colored boxes indicate unique spacers.

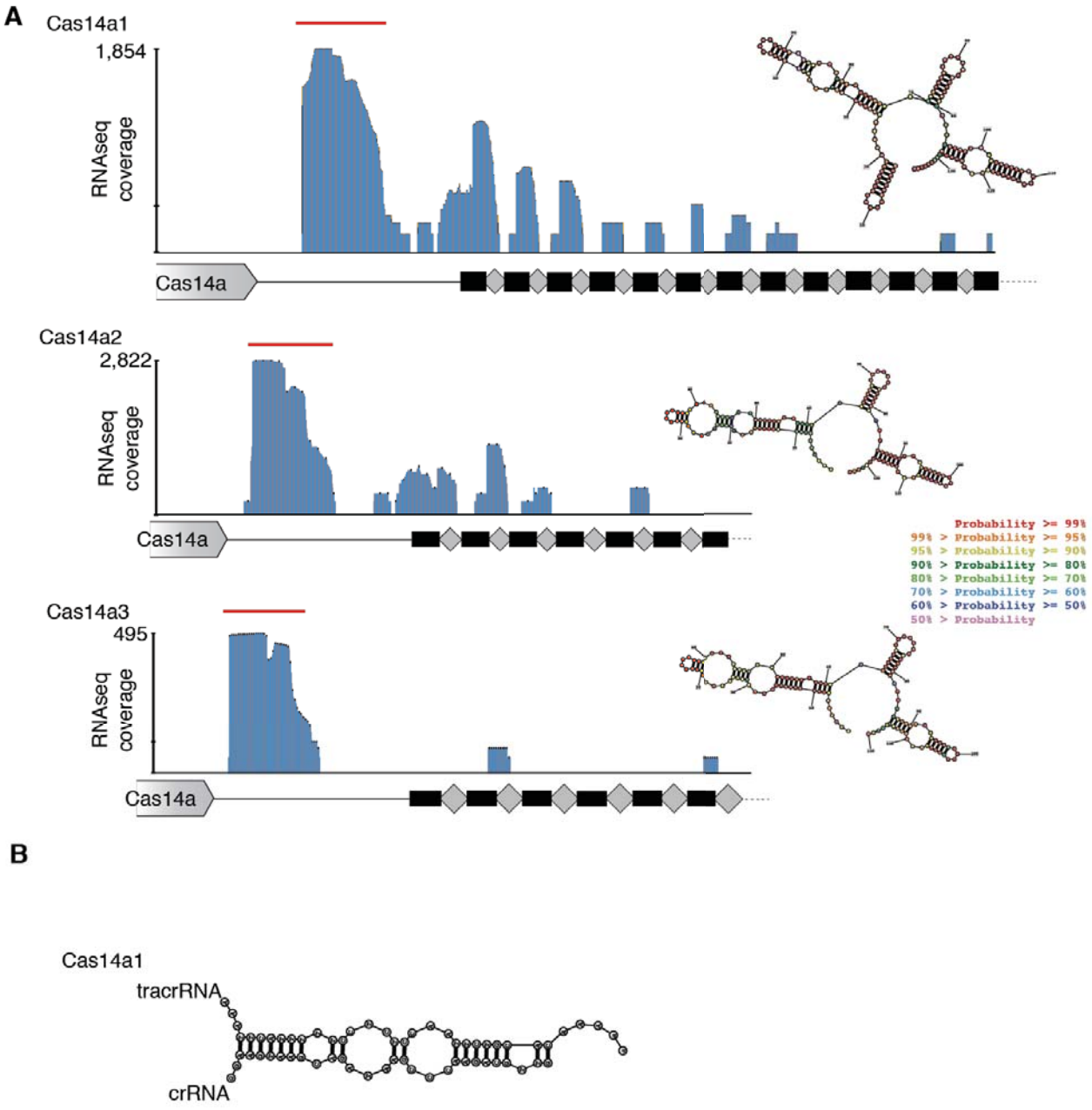


Fig. S4. Metatranscriptomics for CRISPR-Cas14 loci

(A) Environmental RNA sequencing reads for Cas14a orthologs. Location of Cas14 proteins and the CRISPR arrays are indicated below. RNA structures to the right show the *in silico* predicted structure of the tracrRNA identified from metatranscriptomics. (B) Predicted hybridization for Cas14a1 crRNA:tracrRNA duplex.

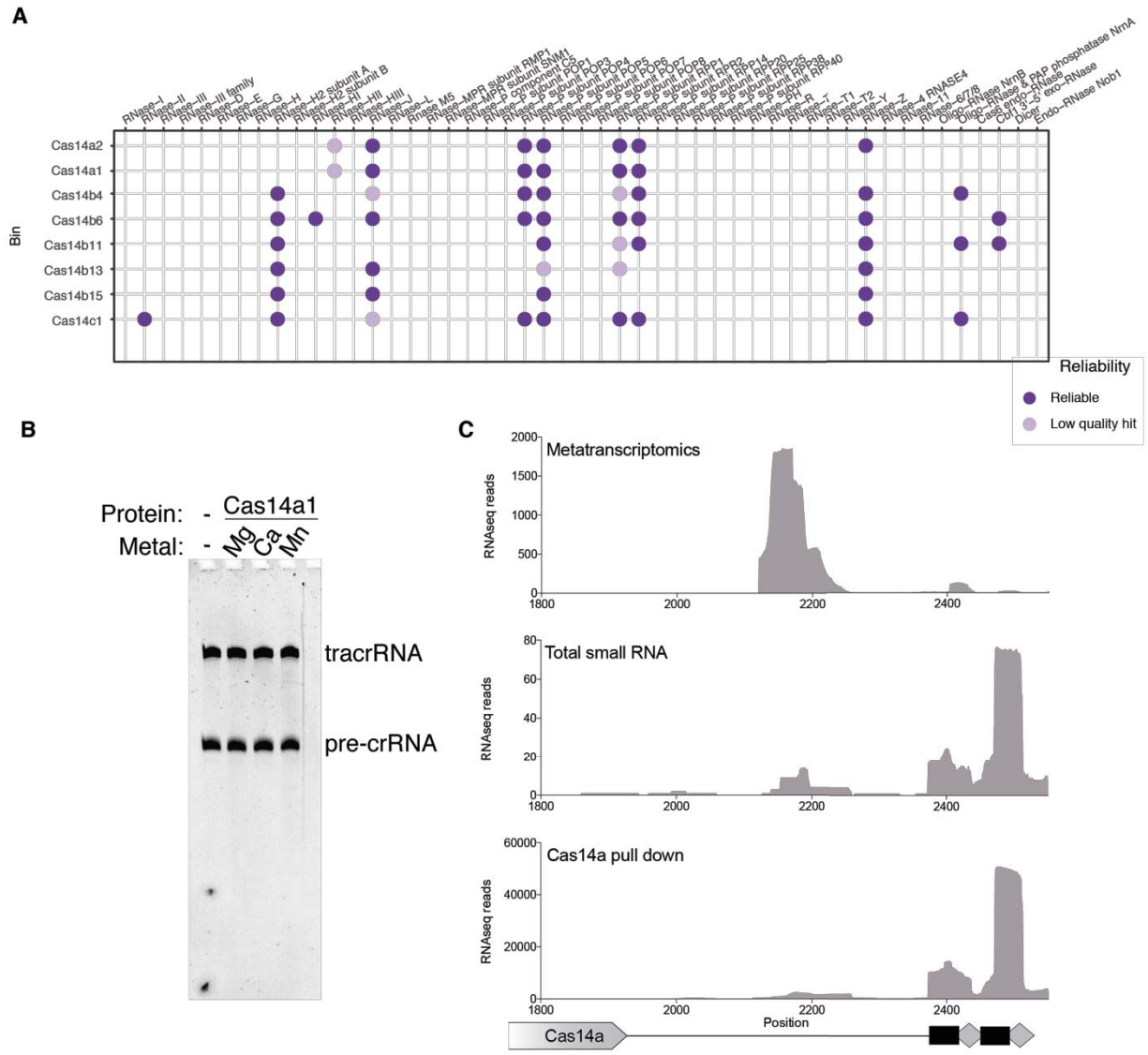


Fig. S5. RNA processing and heterologous expression by CRISPR-Cas14
(A) Presence of common RNase orthologs in Cas14 containing genomes. Light purple represents hits that were significantly shorter than the expected length for the given RNase. Note that RNase III is absent in all investigated genomes. (B) Urea-PAGE assessing the ability of Cas14a1 to cleave its pre-crRNA in the presence of various metals. (C) Small RNAseq reads from heterologous expression of Cas14a1 locus in *E. coli* (bottom two) compared to metatranscriptomic reads (top panel). Pull down refers to RNA that co-purified with Ni-NTA affinity purified Cas14a1.

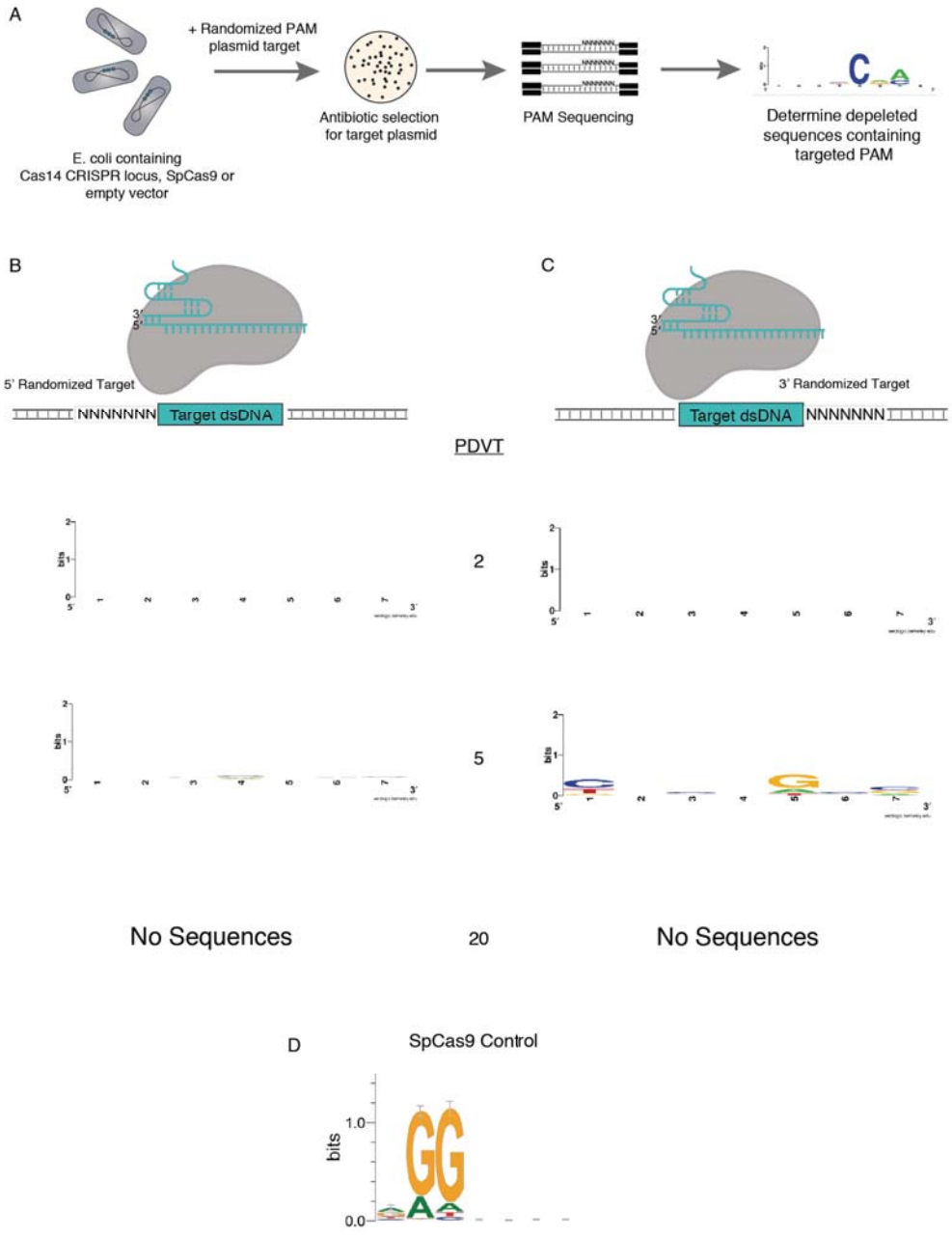


Fig. S6 Plasmid depletion by Cas14a1 and SpCas9

(A) Diagram outlining PAM discovery experiment. *E. coli* expressing the CRISPR system of interest is challenged with a plasmid containing a randomized PAM sequence flanking the target. The surviving (transformed) cells are harvested and sequenced along with a control harboring an empty vector. The depleted sequences are then sequenced and PAMs depleted more than the PAM Depletion Value Threshold (PDVT) are used to generate a WebLogo. (B–C) PAM sequences depleted by heterologously expressed Cas14a1 transformed with a target plasmid containing a randomized PAM sequence 5′ (B) or 3′ (C) of the target. “No sequences” indicates that no sequences were found to be depleted at or above the given PDVT.

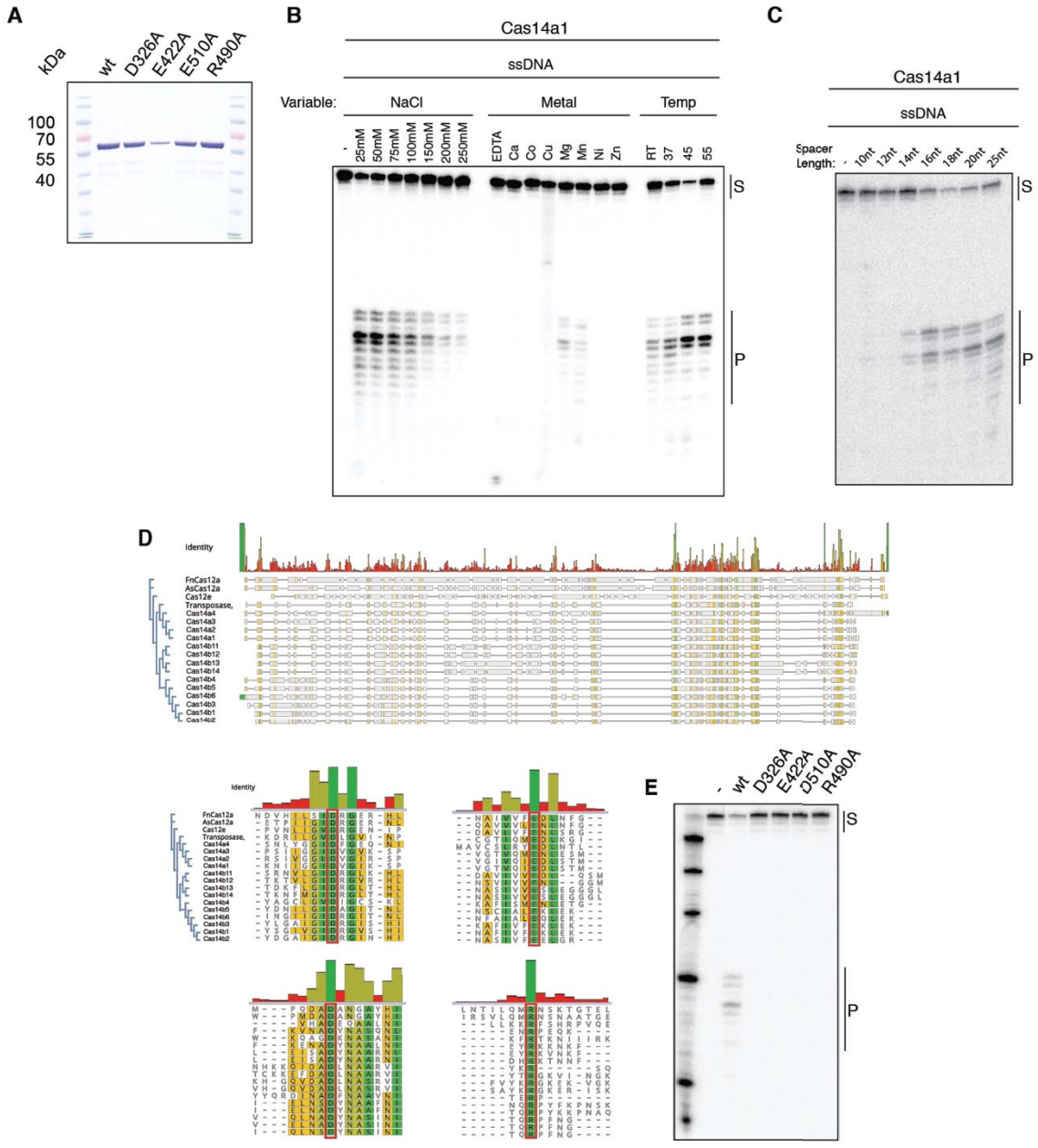


Fig. S7. Degradation of ssDNA by Cas14a1

(A) SDS-PAGE of purified Cas14a1 and Cas14a1 point mutants. (B) Optimization of salt, cation and temperature for Cas14a1 cleavage of ssDNA targets. (C) Radiolabled cleavage of ssDNA by Cas14a1 with spacer sequences of various lengths. (D) Alignment of Cas14 with previously studied Cas12 proteins to identify RuvC active site residues and (E) cleavage of ssDNA by purified Cas14a1 RuvC point mutants.

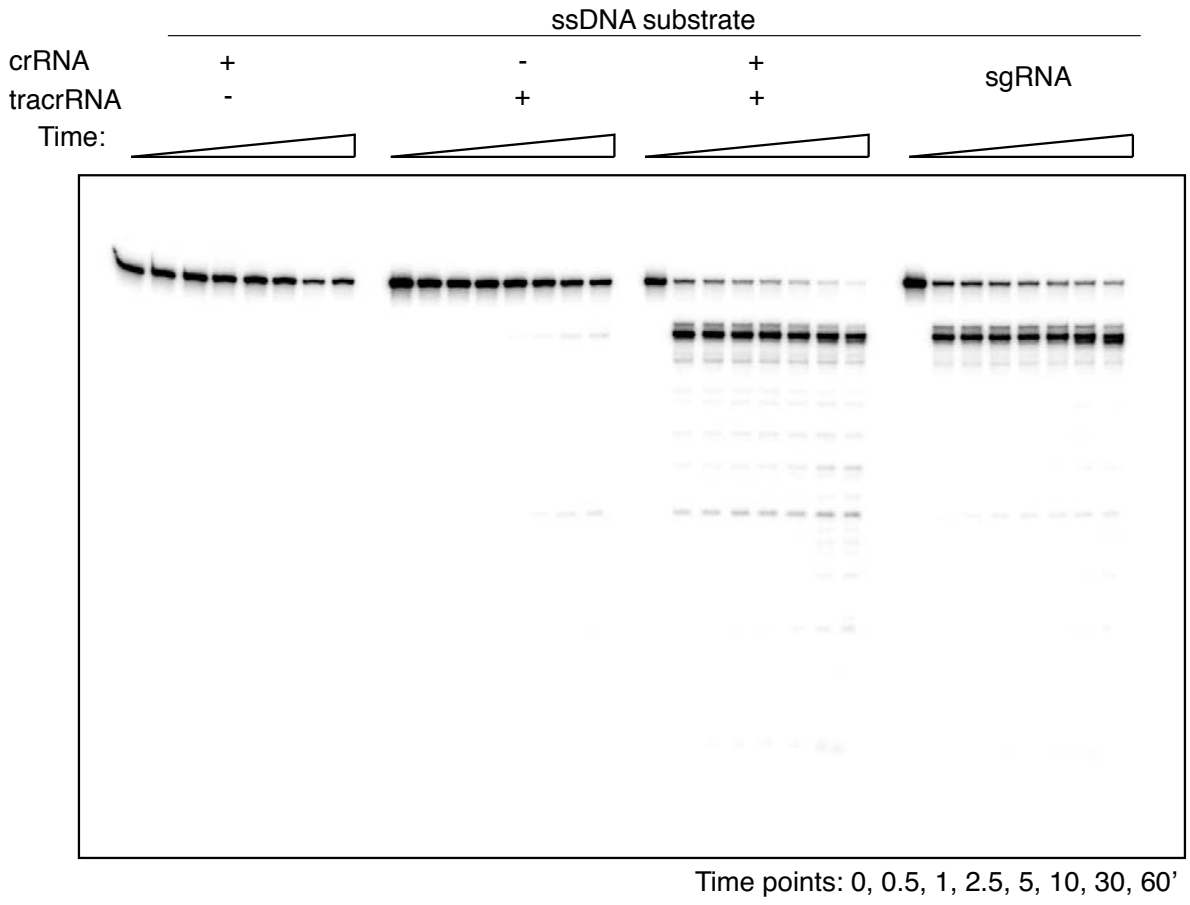


Fig. S8. Kinetics of Cas14a1 cleavage of ssDNA with various guide RNA components.

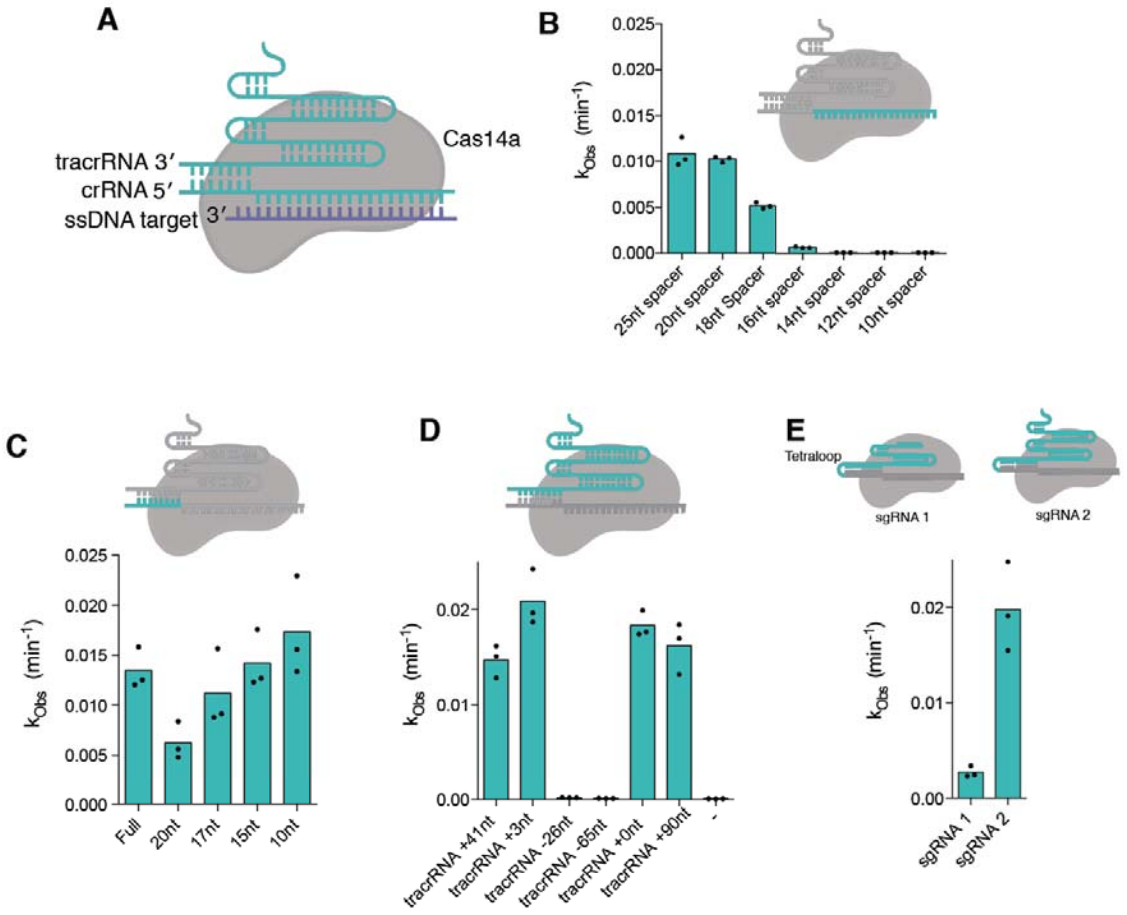


Fig. S9. Optimization of Cas14a1 guide RNA components

(A) Diagram of Cas14a1 targeting ssDNA. Impact on Cas14a1 cleavage of an FQ ssDNA substrate by varying the spacer length (B), repeat length (C), tracrRNA (D), and fusing the crRNA and tracrRNA together(E). For the tracrRNA variants, a “+nt” label refers to extensions at the 3′ end of the RNA, and for the single guide RNA (sgRNA) variants, sgRNA1 is truncated in the tracrRNA-derived region as compared to sgRNA 2.

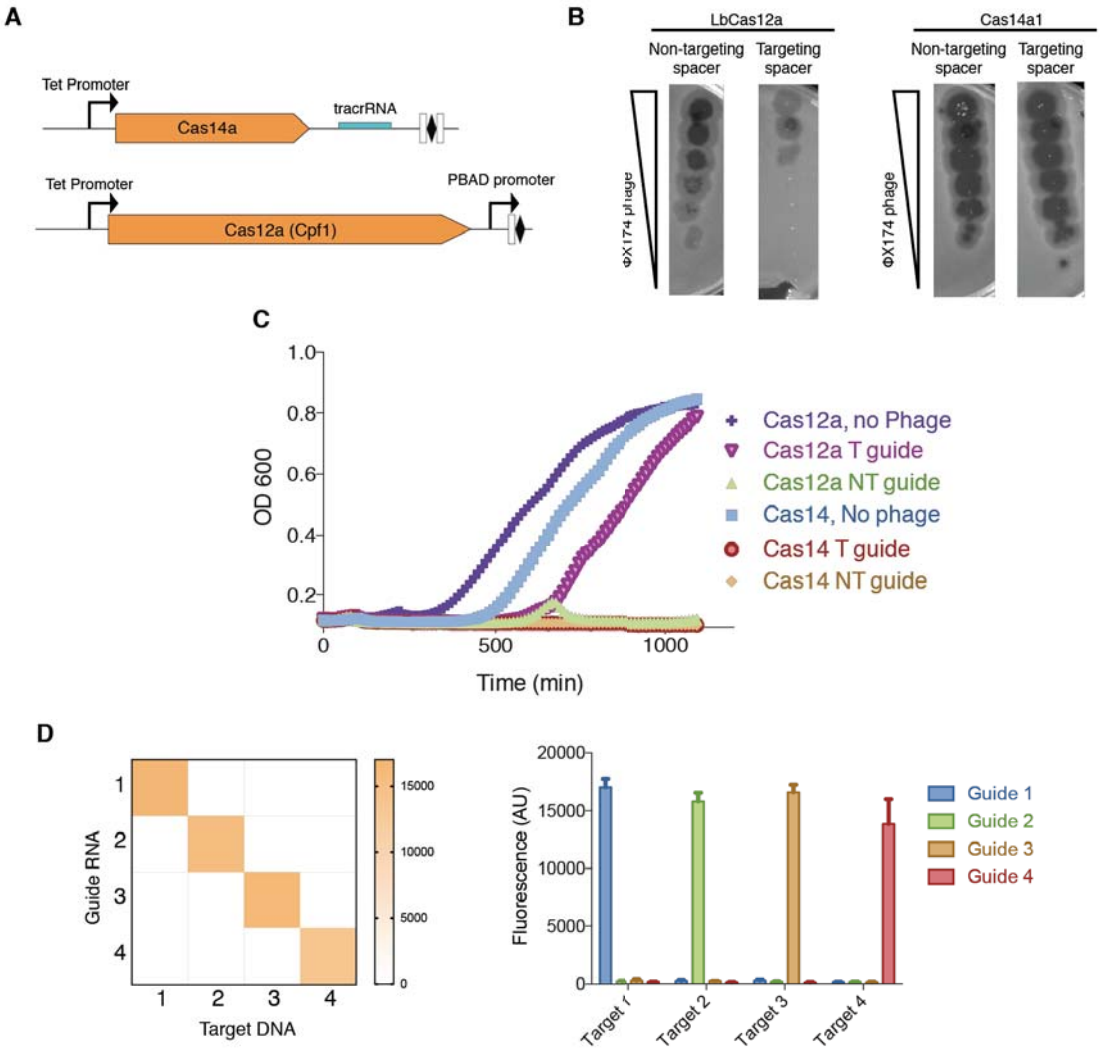
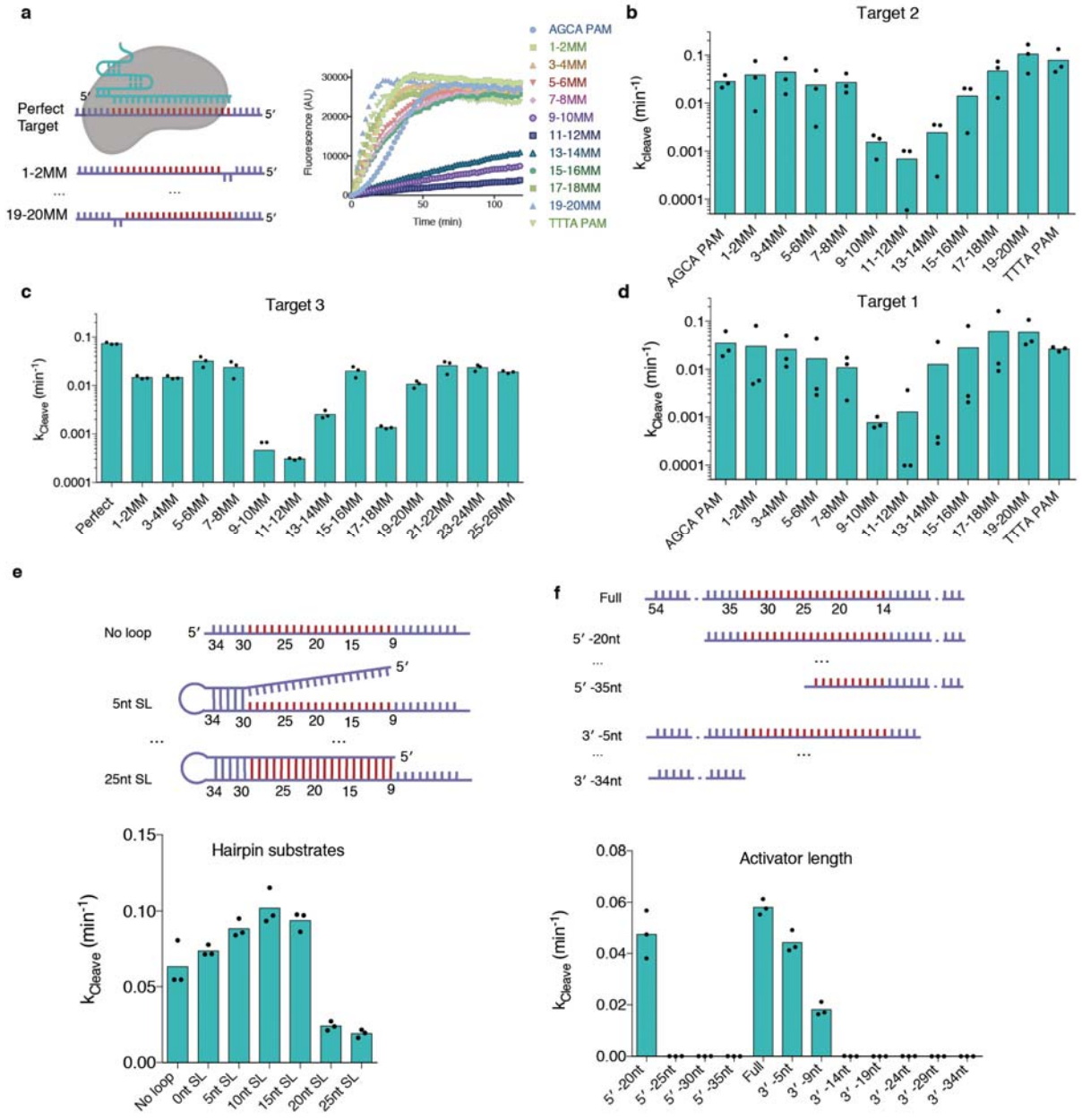


Fig. S10. Test of Cas14a1 mediated interference in a heterologous host

(A) Diagram of Cas14a1 and LbCas12a constructs to test interference in *E. coli*. (B) Plaques of ΦX174 spotted on *E. coli* revealing Cas12a- but not Cas14a1-mediated interference. Each spot represents a 10-fold dilution of the ΦX174 stock. (C) Growth curves of *E. coli* expressing Cas14a1 or LbCas12a infected with ΦX174 (T, targeting; NT, non-targeting). (D) Heat map showing the background-subtracted fluorescence resulting from cleavage of a ssDNA FQ reporter in the presence of various guide and target combinations after a 30min incubation.



Extended Data Figure 8 | Substrate requirements for Cas14 activation

Fig. S11. Impact of various activators on Cas14a1 cleavage rate

(A) Diagram of Cas14a1 targeting of ssDNA with position of mismatches used in panels A–D and raw rates for representative replicates of mismatch (MM) position for Target 1. Cleavage rates for Cas14a targeting substrates with mutations tiled across three different substrates (B–D). (E) Trans cleavage rates for substrates with increasing amounts of secondary structure. (F) Trans cleavage rates with truncated substrates. Points represent individual measurements.

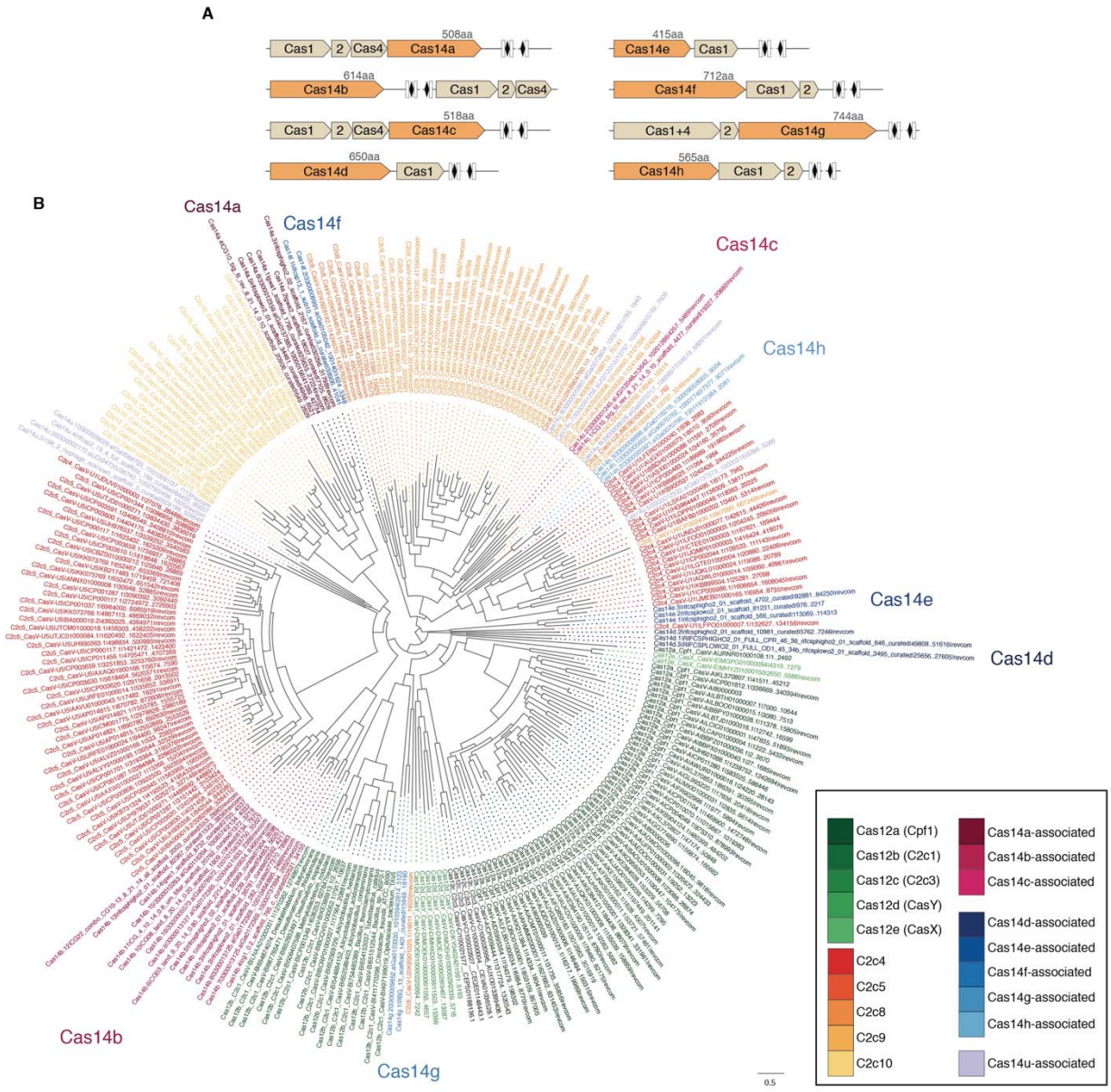


Fig. S12. Diversity of CRISPR-Cas14 systems

(A)Representative locus architecture for the eight Cas14 subtypes. Protein lengths are drawn to scale. (B) Maximum likelihood tree for Type V effectors including all identified subtypes of Cas14. Detailed tree in Newick format with bootstrap values is provided in Data S3.

Data S1. Accession numbers, coordinates and samples of origin for all CRISPR-Cas14 systems described in this study.

Data S2. Sequences of Cas14 proteins described in this study.

Data S3. Type V Cas effector tree, including the Cas14 proteins in Newick format

Data S4. Phylogenetic tree in Newick format of Cas1 from all known CRISPR-Cas types, including Cas1 associated with CRISPR-Cas14 systems.

Data S5. Oligonucleotides and plasmids used in this study

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). doi:10.1126/science.1138140 Medline

2. S. A. Jackson, R. E. McKenzie, R. D. Fagerlund, S. N. Kieper, P. C. Fineran, S. J. J. Brouns, CRISPR-Cas: Adapting to change. *Science* **356**, eaal5056 (2017). doi:10.1126/science.aal5056 Medline

3. 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). doi:10.1038/nrmicro.2016.184 Medline

4. J. S. Chen, J. A. Doudna, The chemistry of Cas9 and its CRISPR colleagues. *Nat. Rev. Chem.* **1**, 0078 (2017). doi:10.1038/s41570-017-0078

5. C. T. Brown, L. A. Hug, B. C. Thomas, I. Sharon, C. J. Castelle, A. Singh, M. J. Wilkins, K. C. Wrighton, K. H. Williams, J. F. Banfield, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015). doi:10.1038/nature14486 Medline

6. K. Anantharaman, C. T. Brown, L. A. Hug, I. Sharon, C. J. Castelle, A. J. Probst, B. C. Thomas, A. Singh, M. J. Wilkins, U. Karaoz, E. L. Brodie, K. H. Williams, S. S. Hubbard, J. F. Banfield, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016). doi:10.1038/ncomms13219 Medline

7. V. M. Markowitz, I.-M. A. Chen, K. Chu, E. Szeto, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, I. Pagani, S. Tringe, M. Huntemann, K. Billis, N. Varghese, K. Tennessen, K. Mavromatis, A. Pati, N. N. Ivanova, N. C. Kyrpides, IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**, D568–D573 (2014). doi:10.1093/nar/gkt919 Medline

8. A. J. Probst, C. J. Castelle, A. Singh, C. T. Brown, K. Anantharaman, I. Sharon, L. A. Hug, D. Burstein, J. B. Emerson, B. C. Thomas, J. F. Banfield, Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO_2 concentrations. *Environ. Microbiol.* **19**, 459–474 (2017). doi:10.1111/1462-2920.13362 Medline

9. I. A. Chen, V. M. Markowitz, K. Chu, K. Palaniappan, E. Szeto, M. Pillay, A. Ratner, J. Huang, E. Andersen, M. Huntemann, N. Varghese, M. Hadjithomas, K. Tennessen, T. Nielsen, N. N. Ivanova, N. C. Kyrpides, IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017). doi:10.1093/nar/gkw929 Medline

10. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40**, 5569–5576 (2012). doi:10.1093/nar/gks216 Medline

11. J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. A. Doudna, Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015). doi:10.1038/nature14237 Medline

12. 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). doi:10.1016/j.molcel.2015.10.008 Medline

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). doi:10.1038/nature21059 Medline

14. C. Rinke, P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J.-F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies, J. A. Dodsworth, B. P. Hedlund, G. Tsiamis, S. M. Sievert, W.-T. Liu, J. A. Eisen, S. J. Hallam, N. C. Kyrpides, R. Stepanauskas, E. M. Rubin, P. Hugenholtz, T. Woyke, Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013). doi:10.1038/nature12352 Medline

15. C. J. Castelle, K. C. Wrighton, B. C. Thomas, L. A. Hug, C. T. Brown, M. J. Wilkins, K. R. Frischkorn, S. G. Tringe, A. Singh, L. M. Markillie, R. C. Taylor, K. H. Williams, J. F. Banfield, Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015). doi:10.1016/j.cub.2015.01.014 Medline

16. E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, E. Charpentier, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011). doi:10.1038/nature09886 Medline

17. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009). doi:10.1099/mic.0.023960-0 Medline

18. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. J. Sontheimer, DNase H Activity of Neisseria meningitidis Cas9. *Mol. Cell* **60**, 242–255 (2015). doi:10.1016/j.molcel.2015.09.020 Medline

19. E. Ma, L. B. Harrington, M. R. O'Connell, K. Zhou, J. A. Doudna, Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell* **60**, 398–407 (2015). doi:10.1016/j.molcel.2015.10.030 Medline

20. 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). doi:10.1126/science.aar6245 Medline

21. 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). doi:10.1016/j.cell.2015.09.038 Medline

22. 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). doi:10.1038/nature19802 Medline

23. L. Liu, X. Li, J. Ma, Z. Li, L. You, J. Wang, M. Wang, X. Zhang, Y. Wang, The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170**, 714–726.e10 (2017). doi:10.1016/j.cell.2017.06.050 Medline

24. O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D. B. T. Cox, S. Shmakov, K. S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E. S. Lander, E. V. Koonin, F. Zhang, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016). doi:10.1126/science.aaf5573 Medline

25. G. J. Knott, A. East-Seletsky, J. C. Cofsky, J. M. Holton, E. Charles, M. R. O'Connell, J. A. Doudna, Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nat. Struct. Mol. Biol.* **24**, 825–833 (2017). doi:10.1038/nsmb.3466 Medline

26. S. Y. Li, Q.-X. Cheng, J.-K. Liu, X.-Q. Nie, G.-P. Zhao, J. Wang, CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491–493 (2018). doi:10.1038/s41422-018-0022-x Medline

27. H. Eiberg, J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K. W. Kjaer, L. Hansen, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum. Genet.* **123**, 177–187 (2008). doi:10.1007/s00439-007-0460-x Medline

28. E. V. Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017). doi:10.1016/j.mib.2017.05.008 Medline

29. K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. Mojica, R. M. Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen, E. V. Koonin, An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015). Medline

30. O. Barabas, D. R. Ronning, C. Guynet, A. B. Hickman, B. Ton-Hoang, M. Chandler, F. Dyda, Mechanism of IS200/IS605 family DNA transposases: Activation and transposon-directed target site selection. *Cell* **132**, 208–220 (2008). doi:10.1016/j.cell.2007.12.029 Medline

31. M. Yoshida, T. Mochizuki, S.-I. Urayama, Y. Yoshida-Takashima, S. Nishi, M. Hirai, H. Nomaki, Y. Takaki, T. Nunoura, K. Takai, Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku, Japan. *Front. Microbiol.* **9**, 75 (2018). doi:10.3389/fmicb.2018.00075 Medline

32. V. M. Markowitz, I.-M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. N. Ivanova, N. C. Kyrpides, IMG: The Integrated Microbial Genomes database and

comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012). doi:10.1093/nar/gkr1044 Medline

33. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39** (Suppl.), W29–W37 (2011). doi:10.1093/nar/gkr367 Medline

34. I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57 (2007). doi:10.1093/nar/gkm360 Medline

35. A. Biswas, R. H. J. Staals, S. E. Morales, P. C. Fineran, C. M. Brown, CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356 (2016). doi:10.1186/s12864-016-2627-0 Medline

36. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi:10.1093/bioinformatics/btu033 Medline

37. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016). doi:10.1093/nar/gkw290 Medline

38. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline

39. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999). doi:10.1093/nar/27.1.29 Medline

40. L. B. Harrington, D. Paez-Espino, B. T. Staahl, J. S. Chen, E. Ma, N. C. Kyrpides, J. A. Doudna, A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017). doi:10.1038/s41467-017-01408-4 Medline

41. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004). doi:10.1101/gr.849004 Medline

42. L. B. Harrington, K. W. Doxzen, E. Ma, J.-J. Liu, G. J. Knott, A. Edraki, B. Garcia, N. Amrani, J. S. Chen, J. C. Cofsky, P. J. Kranzusch, E. J. Sontheimer, A. R. Davidson, K. L. Maxwell, J. A. Doudna, A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* **170**, 1224–1233.e15 (2017). doi:10.1016/j.cell.2017.07.037 Medline