RESEARCH ARTICLE

MICROBIOLOGY

Diverse enzymatic activities mediate antiviral immunity in prokaryotes

Linyi Gao[1,2,3], Han Altae-Tran[1,2,3], Francisca Böhning[1,2], Kira S. Makarova[4], Michael Segel[1,2,3,5,6], Jonathan L. Schmid-Burgk[1,2,3,5,6], Jeremy Koob[1,2], Yuri I. Wolf[4], Eugene V. Koonin[4], Feng Zhang[1,2,3,5,6]*

Bacteria and archaea are frequently attacked by viruses and other mobile genetic elements and rely on dedicated antiviral defense systems, such as restriction endonucleases and CRISPR, to survive. The enormous diversity of viruses suggests that more types of defense systems exist than are currently known. By systematic defense gene prediction and heterologous reconstitution, here we discover 29 widespread antiviral gene cassettes, collectively present in 32% of all sequenced bacterial and archaeal genomes, that mediate protection against specific bacteriophages. These systems incorporate enzymatic activities not previously implicated in antiviral defense, including RNA editing and retron satellite DNA synthesis. In addition, we computationally predict a diverse set of other putative defense genes that remain to be characterized. These results highlight an immense array of molecular functions that microbes use against viruses.

Bacterial and archaeal viruses are the most abundant, and possibly the most diverse, biological entities on Earth (*1*, *2*). To resist frequent and varied attacks by viruses, prokaryotes possess multiple antiviral defense systems. These include the adaptive immune system CRISPR-Cas, which provides immunity by memorizing past infection events (*3*), and a variety of innate immune systems, such as restriction-modification (RM) systems that target specific, predefined sequences in the viral DNA; abortive infection (Abi) systems that induce cell dormancy or death upon viral infection; and additional systems with mechanisms that have not yet been elucidated (*4*). Antiviral defense systems range in complexity from a single small protein (e.g., certain types of Abi systems) to 10 or more proteins acting in concert (e.g., type I and type III CRISPR-Cas systems). Conversely, viruses have evolved strategies to counteract many of these defense systems, including anti-CRISPR and antirestriction proteins (*5*, *6*). Given the vast diversity of viruses and their complex patterns of coevolution with defense systems (*7*–*9*), more types of defense systems with diverse mechanisms can be expected to exist than are currently known.

[1]Howard Hughes Medical Institute, Cambridge, MA 02139, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. [5]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: zhang@broadinstitute.org

Domain-independent prediction of uncharacterized defense systems

Many antiviral defense genes in bacterial and archaeal genomes show a distinctive tendency to cluster together in defense "islands" (*7*, *10*). As a consequence, an uncharacterized gene whose homologs consistently occur next to, for instance, RM genes has an increased likelihood of being involved in defense (*11*, *12*). Using this principle, a recent analysis (*4*) identified and validated 10 previously unknown defense systems, based on the requirement that each (putative) system contain at least one annotated protein domain that is enriched in defense islands.

We hypothesized that additional, unknown defense systems exist that either lack annotated domains or only contain domains that are not typically associated with defense but have been co-opted in specific instances to perform defense functions. To test this hypothesis, we developed an expanded computational approach in which putative defense systems are predicted independent of domain annotations (Fig. 1A). We analyzed all bacterial and archaeal genomes available in GenBank as of November 2018, collectively encoding 620 million proteins. To identify candidate defense genes, we first compiled a list of all genes within 10 kb or 10 open reading frames away from known defense systems (materials and methods). This initial list (n = 8.7 × 10^6), which evidently contained both defense genes and nondefense ones, was clustered to yield 6 × 10^5 representative sequences ("seeds"). To distinguish between defense and nondefense seeds, we identified all homologs of each seed present in GenBank and analyzed their gene neighborhoods. The seed was predicted to be a defense gene if these neighborhoods resembled those of known defense genes—in particular, if a high percentage of homologs were located in proximity to known defense genes and displayed context diversity (Fig. 1B, fig. S1, and materials and methods). All clustering and homolog detection steps were performed on the basis of amino acid sequences, without invoking existing domain annotations and thus allowing the prediction of previously unknown types of defense genes.

After all filtering and curation steps, we identified a total of 7472 seeds (table S1) that represented putative defense genes, along with 4555 seeds for known defense genes under the same analysis parameters (Fig. 1C and table S2). These seeds were analyzed with additional, more sensitive analysis of their domain content (table S3). Of the uncharacterized genes, 1687 (23%) had either no annotated domains or contained only domains of unknown function (DUFs), and an additional 2756 (37%) contained only domains that are different from the characteristic domains of known defense genes. These results suggest the existence of a diverse set of defense genes with mechanisms that remain to be investigated.

Candidate defense systems exhibit antiviral activity in a heterologous system

To characterize the functional diversity among the predicted defense genes, we selected 48 candidate systems to test experimentally for defense activity. Candidate systems were prioritized on the basis of the presence of predicted molecular functions not previously implicated in defense; broad phylogenetic distribution; the presence of at least one protein larger than 300 amino acids (to increase the likelihood of the presence of enzymes); and, for multigene systems, conservation of the component genes. Because wild-type bacterial strains are likely to harbor multiple active defense systems, thereby maintaining phage resistance even if one of the systems were knocked out (*13*), we elected to assay activity by heterologous reconstitution. For each system, one to four homologs were selected, cloned from the source organism into the low-copy vector pACYC, and transformed into *Escherichia coli* (Fig. 2A), comprising a total of 395 kb of exogenous DNA (see tables S4 to S11 for sequence, accession, and source organism information). Three previously identified defense systems, BREX type I (*13*, *14*), Druantia type I (*4*), and the Abi reverse transcriptase RT-Abi-P2 (*15*) were included as positive controls. Each system was then challenged with a diverse panel of coliphages with double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), or single-stranded RNA (ssRNA) genomes, and phage sensitivity of the bacteria was compared to that observed with the empty vector control (Fig. 2, B and C).

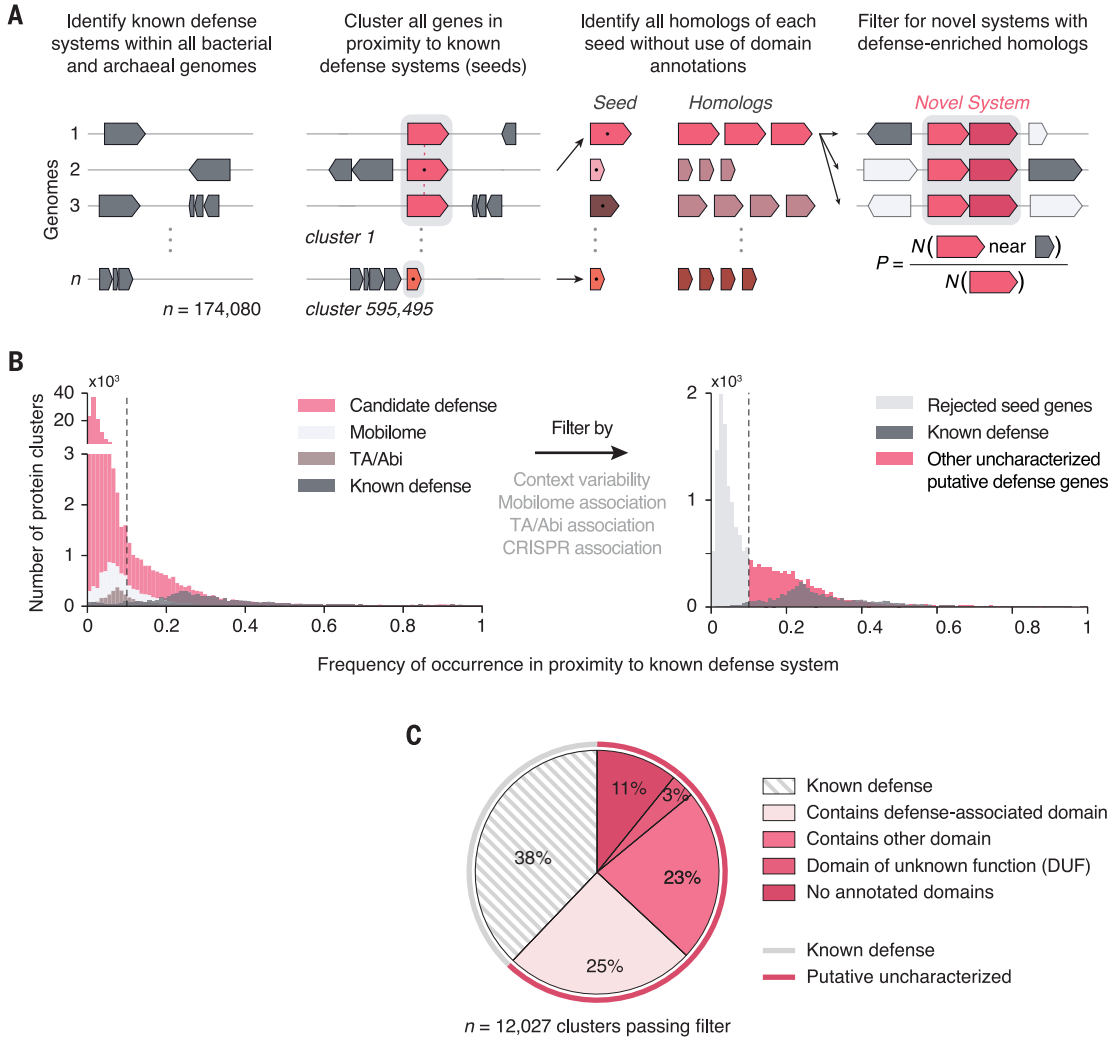

Fig. 1. Domain-independent prediction of putative antiviral defense systems. (**A**) Computational pipeline to identify uncharacterized putative defense systems across all sequenced bacterial and archaeal genomes. Defense systems were predicted on the basis of analysis of amino acid sequences, independent of domain annotations. (**B**) Histograms of defense association frequencies before filtering and after neighborhood context–based filtering (minimum 50 homologs). Seeds to the right of the dashed line (0.1) were selected for further analysis. TA, toxin-antitoxin. (**C**) Pie chart of the domain diversity among predicted defense genes, based on additional analysis using HHpred against pfam domains.

We observed antiphage activity for 29 of the 48 tested candidates (60%) (fig. S2). Systems from source organisms outside the Enterobacteriaceae family, which consists of *Escherichia* and closely related genera including *Salmonella* and *Klebsiella*, had little to no activity, suggesting the importance of host compatibility. The most active representative in each of these 29 systems (representing 4% of the uncharacterized defense seeds) was further tested with an expanded panel of phages in two *E. coli* strains (Fig. 2D and fig. S3). All 29 systems were active against at least one dsDNA phage, and four were active against ssDNA phages (M13 or φX174). Phage specificity was typically narrow and varied widely across systems. The abundance of these defense systems among the sequenced bacterial and archaeal genomes spans two orders of magnitude, ranging from ~0.1 to ~10% of the genomes (Fig. 2D). Overall, 32% of all sequenced bacterial and archaeal genomes contain at least one of these defense systems, which are broadly distributed across bacterial and archaeal phyla (fig. S4).

RADAR contains a divergent adenosine deaminase that edits RNA in response to phage infection

We identified a two-gene cassette consisting of an adenosine triphosphatase (ATPase) (~900 residues) and a divergent adenosine deaminase (~900 residues) that was active against dsDNA phages T2, T3, T4, and T5. Because deaminase activity had not been previously implicated in antiviral defense, we focused on this system for further investigation. The system appears in diverse defense contexts and forms three subtypes (Fig. 3A and fig. S5A). In most cases, it consists of the ATPase and deaminase only, but some variants also include a small membrane protein, either a SLATT domain (*16*) or the type VI-B CRISPR ancillary protein Csx27 (*17*). Mutations in the ATPase Walker B motif or in the putative divalent metal cation-binding HxH motif of the deaminase abolished defense activity, whereas the SLATT domain membrane protein was required for resistance against phage T5 but not against phage T2 (Fig. 3B).

Given the large size of the deaminase compared with typical metabolic adenosine deaminases and its sequence divergence due to large insertions in the deaminase domain (fig. S5B), we hypothesized that it acts on nucleic acids rather than on free nucleosides or nucleotides. To test this hypothesis, we performed whole-transcriptome sequencing and found an enrichment of A-to-G substitutions in sequencing reads at specific sites in the presence of phage, whereas C, G, or U bases were not affected (Fig. 3C and fig. S6A), consistent with RNA editing of adenosine to inosine. Furthermore, the overall expression of phage genes, including early genes, was reduced by ~100-fold even at a multiplicity of infection (MOI) of 2 (Fig. 3D). Because most of the cells in the culture were expected to be infected, this suggested that defense activity occurs early in the infection cycle, which was not evident from efficiency of plating alone.

RNA editing occurred only when both the defense system and the phage were present; expression of the defense system without the phage resulted in a near-baseline level of

editing, and no editing was detected in the absence of the system. Mutations in the ATPase or deaminase active sites abolished editing, and no DNA editing was detected (fig. S6B). Editing sites were broadly distributed throughout the *E. coli* transcriptome (Fig. 3E, figs. S6A and S7, and table S12), and editing could also be induced by coexpressing specific phage proteins with the system (fig. S8 and table S13). RNA secondary-structure predictions indicated a characteristic stem-loop structure at strong editing sites; specific adenosines in loops were edited with up to ~90% frequency, whereas adenosines in the stem were not edited within the limit of detection (Fig. 3E and fig. S7). Finally, some of the editing sites are likely to be deleterious to the host cell, resulting in nonsynonymous mutations such as at the UAA stop codon of the transfer messenger RNA (tmRNA) (fig. S8B), which rescues ribosomes stalled during translation (*18*).

On the basis of these results, we named this system phage restriction by an adenosine deaminase acting on RNA (RADAR). Growth kinetics at varying phage MOI revealed a threshold MOI above which RADAR-expressing cells had a lower optical density at 600 nm

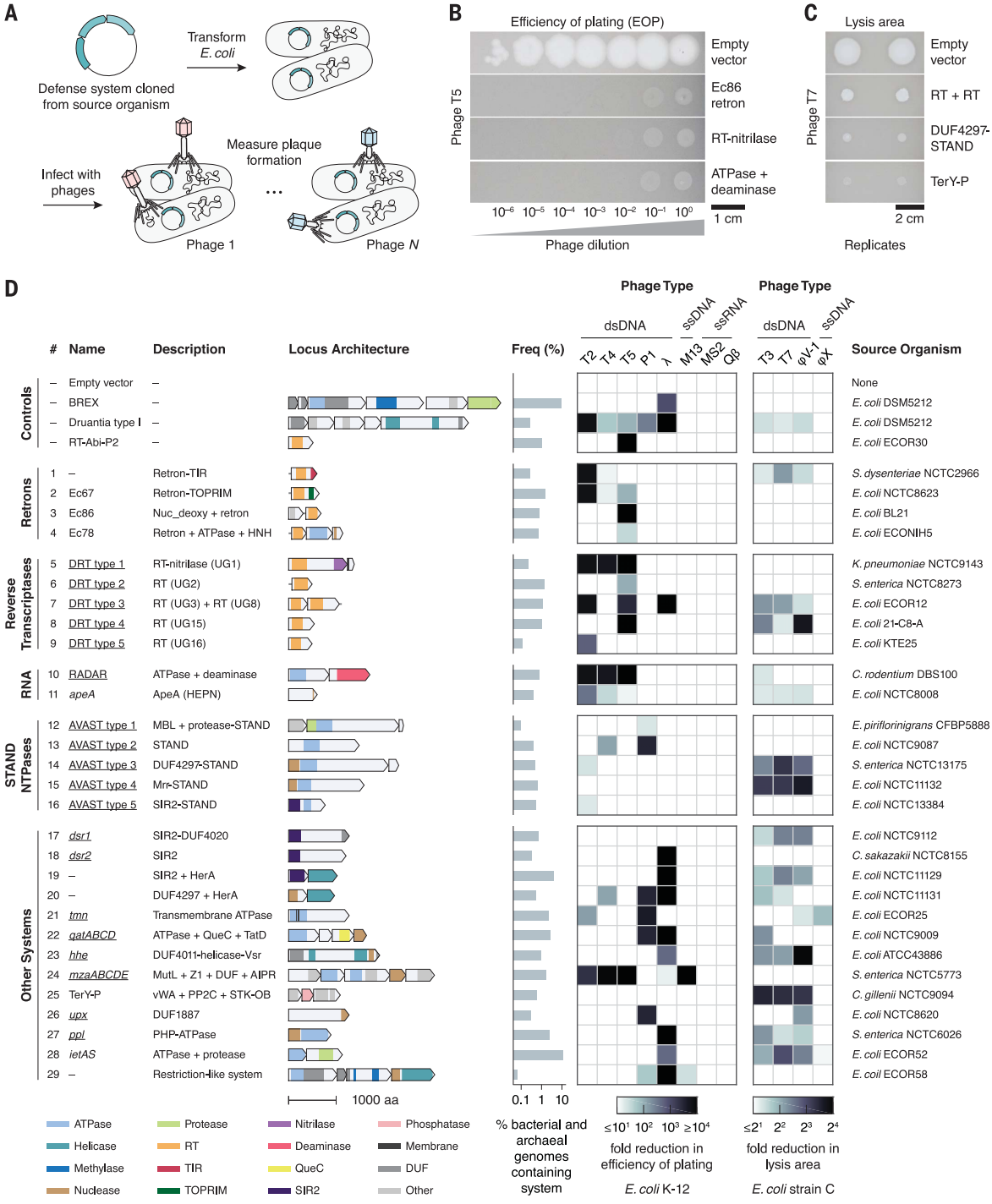

Fig. 2. Candidate defense systems exhibit antiviral activity in a heterologous system. (**A**) Experimental validation pipeline using phage plaque assays on *E. coli* heterologously expressing a cloned candidate defense system. (**B**) Example plaques and (**C**) zones of lysis for six candidate defense systems. (**D**) Antiphage activity across a panel of 12 coliphages with dsDNA, ssDNA, or ssRNA genomes (mean of two replicates). The bar graph shows the abundance of each system in sequenced bacterial and archaeal genomes. Domains: RT, reverse transcriptase; TIR, Toll/interleukin-1 receptor homology domain; TOPRIM, topoisomerase-primase domain; QueC, 7-cyano-7-deazaguanine synthase-like domain; SIR2, sirtuin; membrane, transmembrane helix; DUF, domain of unknown function. Proposed gene names: DRT, defense-associated reverse transcriptase; RADAR, phage restriction by an adenosine deaminase acting on RNA; AVAST, antiviral ATPase/NTPase of the STAND superfamily; *dsr*, defense-associated sirtuin; *tmn*, transmembrane NTPase; *qat*, QueC-like associated with ATPase and TatD DNase; *hhe*, HEPN, helicase, and Vsr endonuclease; *mza*, MutL, Z1, and AIPR; *upx*, uncharacterized (P)D-(D/E)-XK defense protein; *ppl*, polymerase/histidinol phosphatase-like. aa, amino acids; HerA, helicase; MBL, metallo β-lactamase.

(OD600) compared with the empty vector control, suggestive of RADAR-mediated growth arrest (Fig. 3F). Together with the abundance and broad distribution of editing sites in the host transcriptome (figs. S6 and S7), these results are consistent with an editing-dependent Abi mechanism that is activated by phage.

A widespread family of defense systems containing reverse transcriptases

We discovered that a family of uncharacterized reverse transcriptases (RTs) are active defense systems. Although most RTs in prokaryotes are components of mobile retroelements, distinct clades of RTs that lack the hallmarks of mobility also exist, including 16 unknown groups (UGs) (*19*–*22*). We independently identified many of these uncharacterized RTs through our pipeline, suggesting that they might be defense genes (Fig. 4A). Indeed, six of these candidates (UG1, UG2, UG3, UG8, UG15, and UG16) provided robust protection

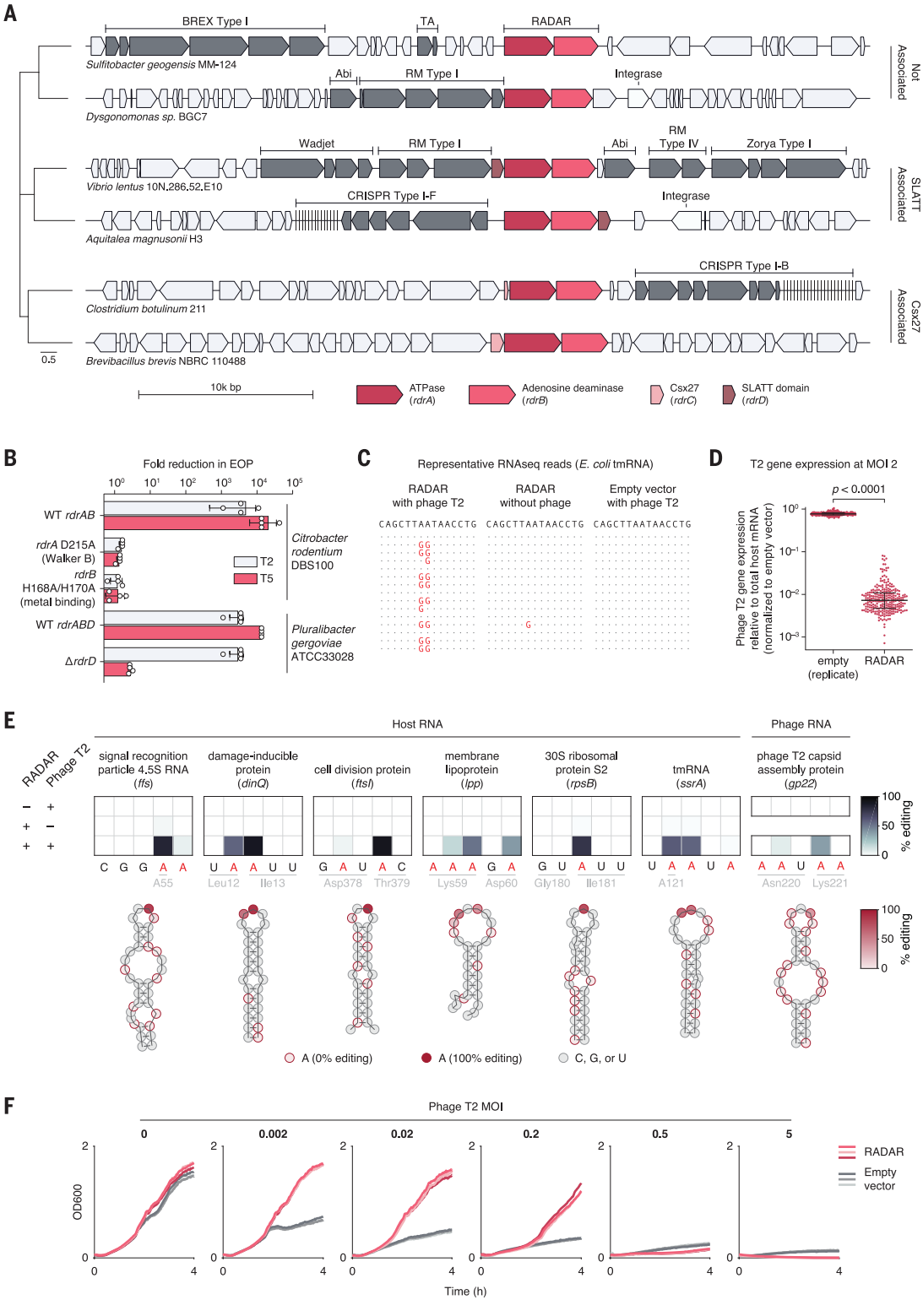

Fig. 3. RADAR mediates RNA editing in response to phage infection. (**A**) Examples of genomic loci containing three subtypes of RADAR (standalone, Csx27-associated, and SLATT-associated). (**B**) Essentiality of the core RADAR genes *rdrAB* and the accessory gene *rdrD* against phages T2 and T5. D215A, Asp^{215}→Ala; H168A, His^{168}→Ala; H170A, His^{170}→Ala; WT, wild type. (**C**) Representative RNA sequencing (RNAseq) reads from *E. coli* expressing either RADAR or an empty vector control. (**D**) Expression of phage T2 RNA relative to total host RNA in *E. coli* containing RADAR. Each dot represents a phage gene. Cells were infected at a MOI of 2. The *p* value was determined by a Wilcoxon signed-rank test. (**E**) Representative editing sites in the host and phage transcriptomes, with corresponding predicted RNA secondary structures. (**F**) Growth kinetics of RADAR-containing *E. coli* in comparison with an empty vector control under varying MOI by phage T2.

against dsDNA phages. In all cases, mutations in the RT active site [(Y/F)xDD to (Y/F)xAA, where x is any amino acid] abolished activity (Fig. 4B and fig. S9, A and B). We named these genes defense-associated RTs (DRTs).

Each of these RT systems displayed a distinct pattern of phage resistance (Fig. 2D). Moreover, whereas UG2 (*drt2*), UG15 (*drt4*), and UG16 (*drt5*) act as individual genes, the UG3 (*drt3a*) and UG8 (*drt3b*) RTs are components of the same defense system (DRT type 3), with both RTs required for defense activity. Like RADAR, some subtypes of the UG1 (DRT type 1) and DRT type 3 systems are also associated with small membrane proteins (Fig. 4A). Moreover, DRT type 1 encompasses a much larger protein (~1200 residues) than the other five RTs and also contains a C-terminal nitrilase domain. Mutation of the catalytic cysteine of the nitrilase to alanine (C1119A) abolished activity (Fig. 4B). Nitrilases typically function in processes unrelated to defense, such as nucleotide metabolism and small-molecule biosynthesis (*23*). Thus, DRT type 1, which is divergent from typical nitrilases and forms a distinct clade in the phylogenetic tree of the nitrilase family (fig. S10), exemplifies a non-defense domain that was apparently co-opted for a defense function.

To further characterize these RTs, we performed whole-transcriptome sequencing of RT-expressing *E. coli* during phage infection. These experiments revealed substantial differences in phage gene expression across the different RTs (Fig. 4C). For instance, DRT type 1 strongly suppressed the expression of phage late genes, such as capsid proteins, whereas early and middle genes were not substantially affected, suggesting that it is active before the late stage of infection

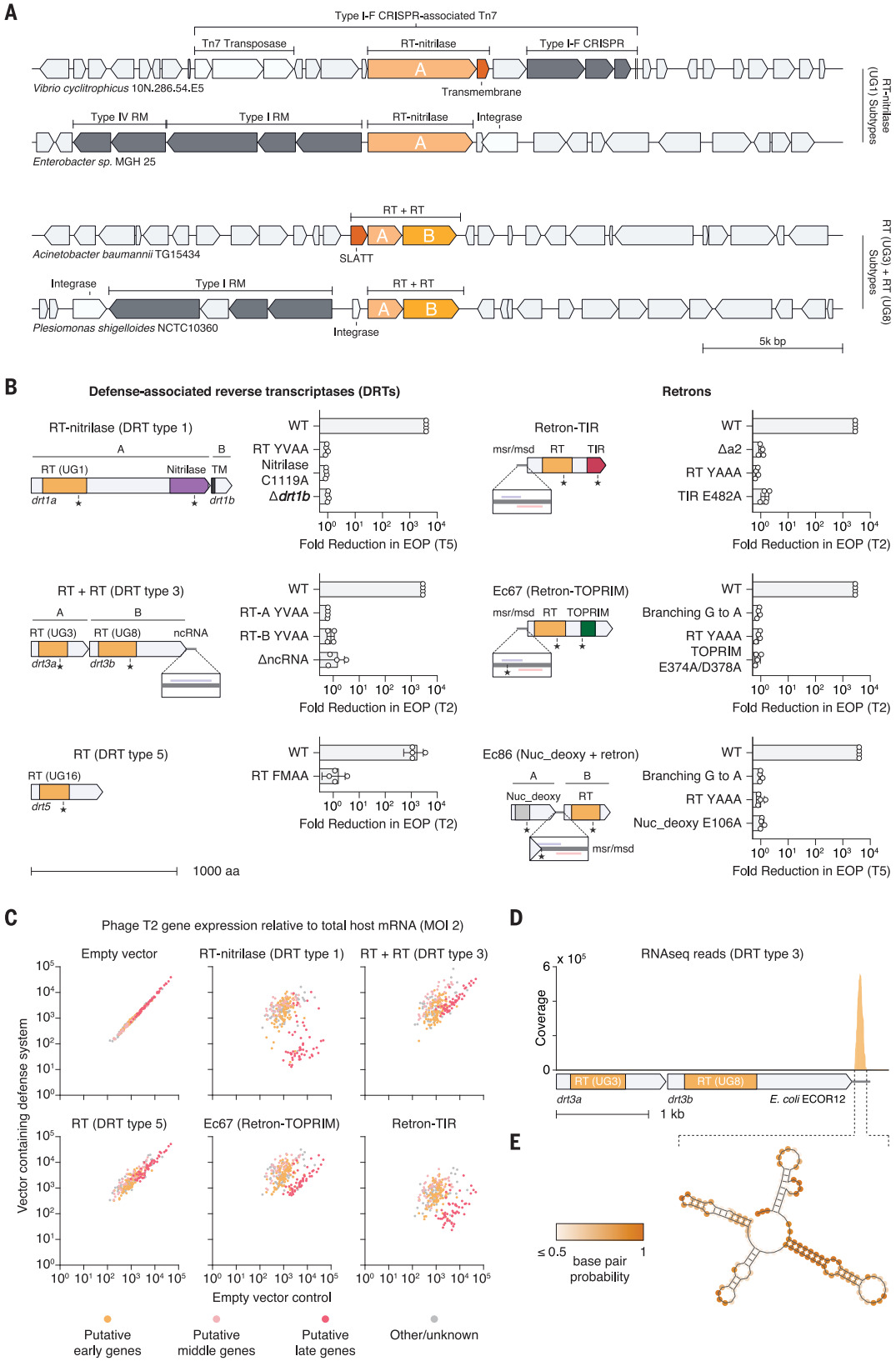

Fig. 4. Diverse families of RTs mediate antiviral defense. (**A**) Examples of genomic loci containing two RT-based defense systems (DRT type 1 and type 3), with two representative subtypes shown for each system. (**B**) Essential components of non-retron RTs (left panel) and retrons (right panel). TM, transmembrane; ncRNA, noncoding RNA; msr/msd: genes encoding msRNA and msDNA, respectively; a2, retron 5′ inverted repeat. (**C**) Effect of defense RTs on the expression of phage T2 genes in *E. coli* infected at an MOI of 2. (**D**) RNAseq reads mapping to the DRT type 3 system. (**E**) Predicted secondary structure of the highly expressed noncoding RNA identified in (D).

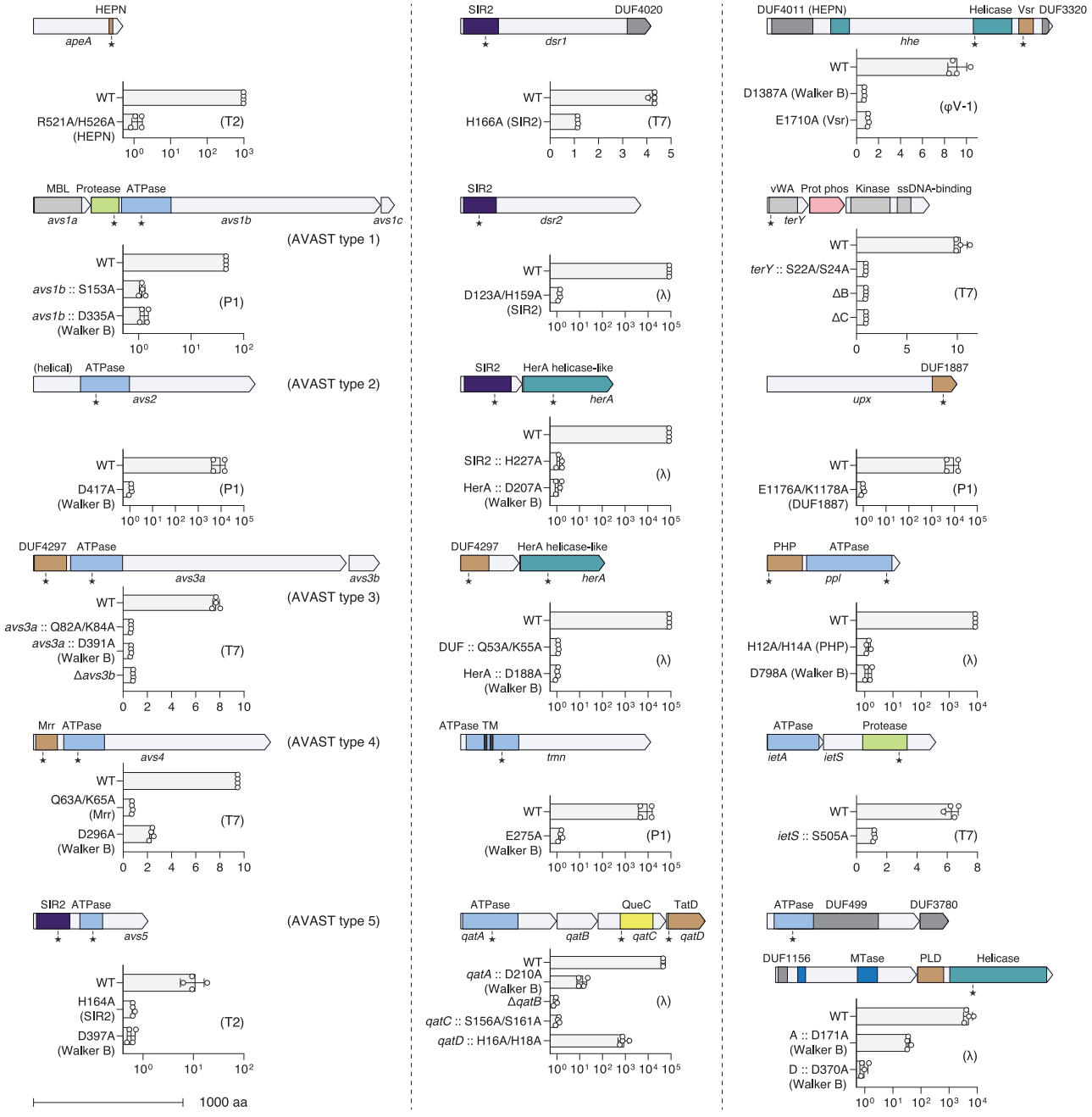

Fig. 5. Domain architectures and mutational analysis of additional defense systems. Graphics show domains identified by using HHpred, and stars indicate locations of active site mutations. Bar graphs (n = 4 replicates per bar) show either $\log_{10}$ fold change of efficiency of plating (for phages T2, P1, and λ) or $\log_2$ fold change in the area of the zone of lysis (for phages T7 and φV-1) relative to the empty vector control. MBL, metallo β-lactamase; SIR2, sirtuin; HerA, helicase; QueC, 7-cyano-7-deazaguanine synthase-like domain; Vsr, very short patch repair endonuclease; TatD, DNase; vWA, von Willebrand factor type A; Prot phos, serine/threonine protein phosphatase; PHP, polymerase/histidinol phosphatase; MTase, methyltransferase; PLD, phospholipase D; DUF, domain of unknown function.

but does not prevent the injection of phage DNA into the host cell. By contrast, DRT type 3 did not strongly suppress expression of any of the phage genes, despite growing at a rate similar to DRT type 1 during phage infection (fig. S11A). Transcriptome sequencing also identified a highly expressed, structured noncoding RNA at the 3′ end of the DRT type 3 system that is required for activity (Fig. 4, B, D, and E).

Retrons mediate antiphage defense

We also found that retrons, a distinct class of RTs that produce extrachromosomal satellite DNA [multicopy ssDNA (msDNA)], are active antiphage defense systems. The retron msDNA is produced from the 5′ untranslated region of its own mRNA and is covalently linked to an internal guanosine of the RNA through a 2′-5′ phosphodiester bond (*24*). First identified more than 30 years ago, retrons have been harnessed for bacterial genome engineering (*25*), but their native biological function has remained unknown. We found that the original *E. coli* retrons Ec67 (*26*) and Ec86 (*27*), as well as a homolog of the Ec78 retron (*28*) and a previously uncharacterized TIR (Toll/interleukin-1 receptor) domain–associated retron, mediate defense against dsDNA phages. The Ec86

retron is natively present in the widely used laboratory *E. coli* strain BL21. Mutations in the (Y/F)xDD active site motif of the RT, as well as at the branching guanosine, abolished activity, indicating that the defense function depends on msDNA synthesis (Fig. 4B and fig. S9C). Furthermore, perturbations to the msDNA also abolished activity (fig. S11), suggesting that its structure, and not simply formation, is essential for the defense activity. Indeed, a single nucleotide mismatch in the msDNA hairpin reduced activity by 100- to 1000-fold, but introducing a second mutation on the complementary strand to restore the structure of the msDNA also restored wild-type activity (fig. S11). Notably, these retrons are associated with other domains, including TOPRIM (topoisomerase-primase) (*29*); TIR (*30*), a nucleoside deoxyribosyltransferase–like enzyme; and the Septu defense system (*4*), all of which are required for activity (Fig. 4B).

Additional molecular functions of defense systems

We investigated several additional systems with diverse components (Fig. 5 and fig. S12). These include a three-gene system containing a von Willebrand factor A (vWA) metal ion binding protein, a serine/threonine protein phosphatase, and a serine/threonine protein kinase that provided strong protection against T7-like phages (T3, T7, and φV-1). This system, called the TerY-phosphorylation triad (TerY-P), has been previously analyzed computationally in the context of tellurite resistance–associated stress response and might operate as a phosphorylation switch that couples the activities of the kinase and the phosphatase (*31*).

Additional systems include proteins containing a SIR2 (sirtuin) deacetylase domain that is also present in the recently discovered Thoeris system (*4*) and has also been detected in the same neighborhoods with prokaryotic Argonaute proteins (*32*); ApeA, a predicted HEPN-family Abi protein (*33*) and a putative ancestor of the type VI CRISPR effector Cas13; a ~1300-residue P-loop ATPase containing an unusual insertion of two transmembrane helices into the ATPase domain, similar to the KAP ATPases (*34*); and a four-gene cassette containing a 7-cyano-7-deazaguanine synthase-like protein (QueC), suggestive of small-molecule biosynthesis. All of these components are essential for defense activity (Fig. 5). Further investigation is required to understand the mechanisms by which these systems sense and respond to phage infection.

Finally, we also demonstrate defense functions for several predicted nucleoside triphosphatases (NTPases) of the STAND (signal transduction ATPases with numerous associated domains) superfamily (Fig. 5). This expansive superfamily consists of multidomain proteins that include eukaryotic ATPases and GTPases involved in programmed cell death and various forms of signal transduction (*35*, *36*). Typically, STAND NTPases contain a C-terminal helical sensor domain that, on target recognition, induces oligomerization by means of ATP or GTP hydrolysis, leading to activation of the N-terminal effector domain. The role of the STAND NTPases in prokaryotes has long remained enigmatic (*35*, *37*); the few for which experimental data are available contain a helix-turn-helix domain and have been shown to regulate transcription (*36*). Several STAND NTPases were active against dsDNA phages (Fig. 2D); these proteins contained different putative effector domains, including DUF4297 [a putative PD-(D/E)xK-family nuclease], an Mrr-like nuclease, SIR2, a trypsin-like serine protease, and an uncharacterized helical domain. We named these systems antiviral ATPases/NTPases of the STAND superfamily (AVAST). As homologs of essential eukaryotic programmed cell death effectors, AVAST systems are likely to function through an Abi mechanism, i.e., by causing growth arrest or programmed cell death in infected hosts.

Discussion

These findings substantially expand the space of protein domains, molecular functions, and interactions that are used by bacteria and archaea in antiviral defense. Some of these functions, including RNA editing, have not been previously implicated in defense mechanisms. The high success rate of defense system prediction based on the evolutionary conservation of their proximity to previously identified defense genes supports the defense island concept (*4*, *7*, *10*) and demonstrates its growing utility at the time of rapid expansion of sequence databases. Furthermore, the computational approach implemented in this work provided for a substantial expansion of the range of the identified putative defense systems. Many of these previously unknown systems contain enzymatic activities as well as predicted sensor components that potentially could be engineered for new biotechnology applications.

Despite similarities in domain architectures among some of the identified defense systems, their phage specificities differ significantly, emphasizing the importance of multiple defense mechanisms for the survival of prokaryotes in the arms race against viruses. These observations are compatible with the concept of distributed microbial immunity, according to which defense systems encoded in different genomes collectively protect microbial communities from the diverse viromes they confront (*38*). Additionally, several of the identified defense systems incorporate molecular functions from typically nondefense sources, highlighting the versatility of activities that are recruited for antiviral defense. These include the RADAR deaminase, nitrilases, RTs, and retrons. The demonstration of defense functions for multiple RTs, which are generally associated with mobile genetic elements (MGEs), is consistent with the "guns for hire" paradigm whereby enzymes are shuttled between MGEs and defense systems during microbial evolution (*8*). Finally, most of these defense systems do not appear to be substantially enriched within prophages, suggesting that they are dedicated host defense genes, rather than virus superinfection exclusion modules (fig. S13 and materials and methods).

The overall patchy pattern of phage specificity observed for the different defense systems was unexpected. In some cases, the same system exhibited widely varying levels of protection against similar phages; for instance, retron Ec67 and DRT type 3 offered full protection against phage T2 but poor protection against phage T4, which is ~82% identical to T2. We hypothesize that phage-encoded inhibitors or antidefense genes that have yet to be discovered may play an important role in determining the specificity of defense systems. Indeed, many phage-encoded proteins have no known function and remain to be investigated for antidefense activity.

With the exception of RADAR, we do not yet know the mechanisms of most of the identified defense systems. The range of domains contained in these systems indicates that they use diverse biochemical activities. Additional experimental characterization is required to elucidate the effector functions for these systems and the molecular basis of antiphage action and specificity. The identification of these defense systems, as well as others we have predicted computationally, provides a foundation for further mechanistic investigation.

The results described here have broad implications for understanding antiviral resistance and host-virus interactions in natural populations of microbes, as well as for technological applications such as the development of antibacterial therapeutics, nucleic acid editing, molecular detection, and targeted cell destruction.

REFERENCES AND NOTES

1. C. A. Suttle, *Genome* **56**, 542–544 (2013).
2. A. G. Cobián Güemes *et al.*, *Annu. Rev. Virol.* **3**, 197–214 (2016).
3. F. Hille *et al.*, *Cell* **172**, 1239–1259 (2018).
4. S. Doron *et al.*, *Science* **359**, eaar4120 (2018).
5. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, *Nat. Rev. Microbiol.* **11**, 675–687 (2013).
6. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, *Nature* **493**, 429–432 (2013).
7. K. S. Makarova, Y. I. Wolf, E. V. Koonin, *Nucleic Acids Res.* **41**, 4360–4377 (2013).
8. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, *Nat. Rev. Genet.* **21**, 119–131 (2020).
9. G. Faure *et al.*, *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
10. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, *J. Bacteriol.* **193**, 6039–6056 (2011).

11. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5307–E5316 (2018).
12. S. A. Shmakov *et al.*, *Nat. Protoc.* **14**, 3013–3031 (2019).
13. J. Gordeeva *et al.*, *Nucleic Acids Res.* **47**, 253–265 (2019).
14. T. Goldfarb *et al.*, *EMBO J.* **34**, 169–183 (2015).
15. R. Odegrip, A. S. Nilsson, E. Haggård-Ljungquist, *J. Bacteriol.* **188**, 1643–1647 (2006).
16. A. M. Burroughs, D. Zhang, D. E. Schäffer, L. M. Iyer, L. Aravind, *Nucleic Acids Res.* **43**, 10633–10654 (2015).
17. K. S. Makarova, L. Gao, F. Zhang, E. V. Koonin, *FEMS Microbiol. Lett.* **366**, fnz088 (2019).
18. C. D. Rae, Y. Gordiyenko, V. Ramakrishnan, *Science* **363**, 740–744 (2019).
19. S. Zimmerly, L. Wu, *Microbiol. Spectr.* **3**, A3–A0058, 2014 (2015).
20. N. Toro, R. Nisa-Martínez, *PLOS ONE* **9**, e114083 (2014).
21. K. K. Kojima, M. Kanehisa, *Mol. Biol. Evol.* **25**, 1395–1404 (2008).
22. D. M. Simon, S. Zimmerly, *Nucleic Acids Res.* **36**, 7219–7229 (2008).
23. H. C. Pace, C. Brenner, *Genome Biol.* **2**, S0001 (2001).
24. A. J. Simon, A. D. Ellington, I. J. Finkelstein, *Nucleic Acids Res.* **47**, 11007–11019 (2019).
25. F. Farzadfard, T. K. Lu, *Science* **346**, 1256272 (2014).
26. B. C. Lampson *et al.*, *Science* **243**, 1033–1038 (1989).
27. D. Lim, W. K. Maas, *Cell* **56**, 891–904 (1989).
28. T. M. Lima, D. Lim, *Plasmid* **38**, 25–33 (1997).
29. L. Aravind, D. D. Leipe, E. V. Koonin, *Nucleic Acids Res.* **26**, 4205–4213 (1998).
30. S. Horsefield *et al.*, *Science* **365**, 793–799 (2019).
31. V. Anantharaman, L. M. Iyer, L. Aravind, *Mol. Biosyst.* **8**, 3142–3165 (2012).
32. K. S. Makarova, Y. I. Wolf, J. van der Oost, E. V. Koonin, *Biol. Direct* **4**, 29 (2009).
33. V. Anantharaman, K. S. Makarova, A. M. Burroughs, E. V. Koonin, L. Aravind, *Biol. Direct* **8**, 15 (2013).
34. L. Aravind, L. M. Iyer, D. D. Leipe, E. V. Koonin, *Genome Biol.* **5**, R30 (2004).
35. D. D. Leipe, E. V. Koonin, L. Aravind, *J. Mol. Biol.* **343**, 1–28 (2004).
36. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, *Structure* **17**, 172–182 (2009).
37. E. V. Koonin, L. Aravind, *Cell Death Differ.* **9**, 394–404 (2002).
38. A. Bernheim, R. Sorek, *Nat. Rev. Microbiol.* **18**, 113–119 (2020).

ACKNOWLEDGMENTS

We thank A. Ladha, J. Strecker, S. Kannan, J. Kreitz, and G. Faure for valuable discussions and experimental assistance; R. Macrae for a critical reading of the manuscript; S. Richards and R. Belliveau for assistance with ordering bacterial strains; and the entire Zhang lab for support and advice. **Funding:** K.S.M. and E.V.K. are supported by intramural funds of the U.S. Department of Health and Human Services (to National Library of Medicine). F.Z. is supported by the National Institutes of Health (grants 1R01-HG009761, 1R01-MH110049, and 1DP1-HL141201); the Howard Hughes Medical Institute; the Open Philanthropy Project, the Harold G. and Leila Mathers and Edward Mallinckrodt, Jr. Foundations; the Poitras Center for Psychiatric Disorders Research at MIT; the Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT; and by the Phillips family and J. and P. Poitras.
Author contributions: L.G. and F.Z. conceived of the project. L.G. and H.A.-T. performed defense island analysis. L.G. cloned defense systems and performed transcriptome sequencing. L.G. performed plaque assays with assistance from F.B., M.S., and J.K. L.G. and F.B. performed bacterial density and phage fragmentation assays. H.A.-T. performed phylogenetic distribution and prophage analysis. L.G., H.A.-T., F.B., K.S.M., J.L.S.-B., Y.I.W., E.V.K., and F.Z. analyzed data. F.Z. supervised the research and experimental design. L.G., H.A.-T., E.V.K., and F.Z. wrote and revised the manuscript with input from all authors. **Competing interests:** F.Z. is a scientific advisor and cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences. L.G, J.L.S.-B., and F.Z. are co-inventors on US provisional application no. 62/928,269, which includes bacterial defense systems described in this manuscript. **Data and materials availability:** Expression plasmids are available from Addgene (nos. 157879 to 157912) under the Uniform Biological Material Transfer Agreement. The genome assembly of *E. coli* phage ϕV-1 has been deposited in GenBank (accession number MT542512). All other data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6507/1077/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S13
References (*39*–*60*)
MDAR Reproducibility Checklist

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

29 October 2019; accepted 6 July 2020
10.1126/science.aba0372

Diverse enzymatic activities mediate antiviral immunity in prokaryotes

Linyi Gao, Han Altae-Tran, Francisca Böhning, Kira S. Makarova, Michael Segel, Jonathan L. Schmid-Burgk, Jeremy Koob, Yuri I. Wolf, Eugene V. Koonin and Feng Zhang

Science **369** (6507), 1077-1084.
DOI: 10.1126/science.aba0372

Prokaryotic enzymes for viral defense

The arms race between prokaryotes and viruses provides a strong evolutionary force to create diverse enzymatic activities that mediate antiviral immune responses. These immune components often cluster together in the host genomes, leading to expanded defense systems. Taking advantage of the evolutionary modularity of defense systems, Gao *et al.* bioinformatically predicted defense genes in most available bacterial and archaeal genomes. In addition, they reconstituted the newly identified systems in Escherichia coli and verified their defense functions against specific bacteriophages. In particular, they characterized defense functions for several predicted nucleoside triphosphatases.

Science, this issue p. 1077