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).

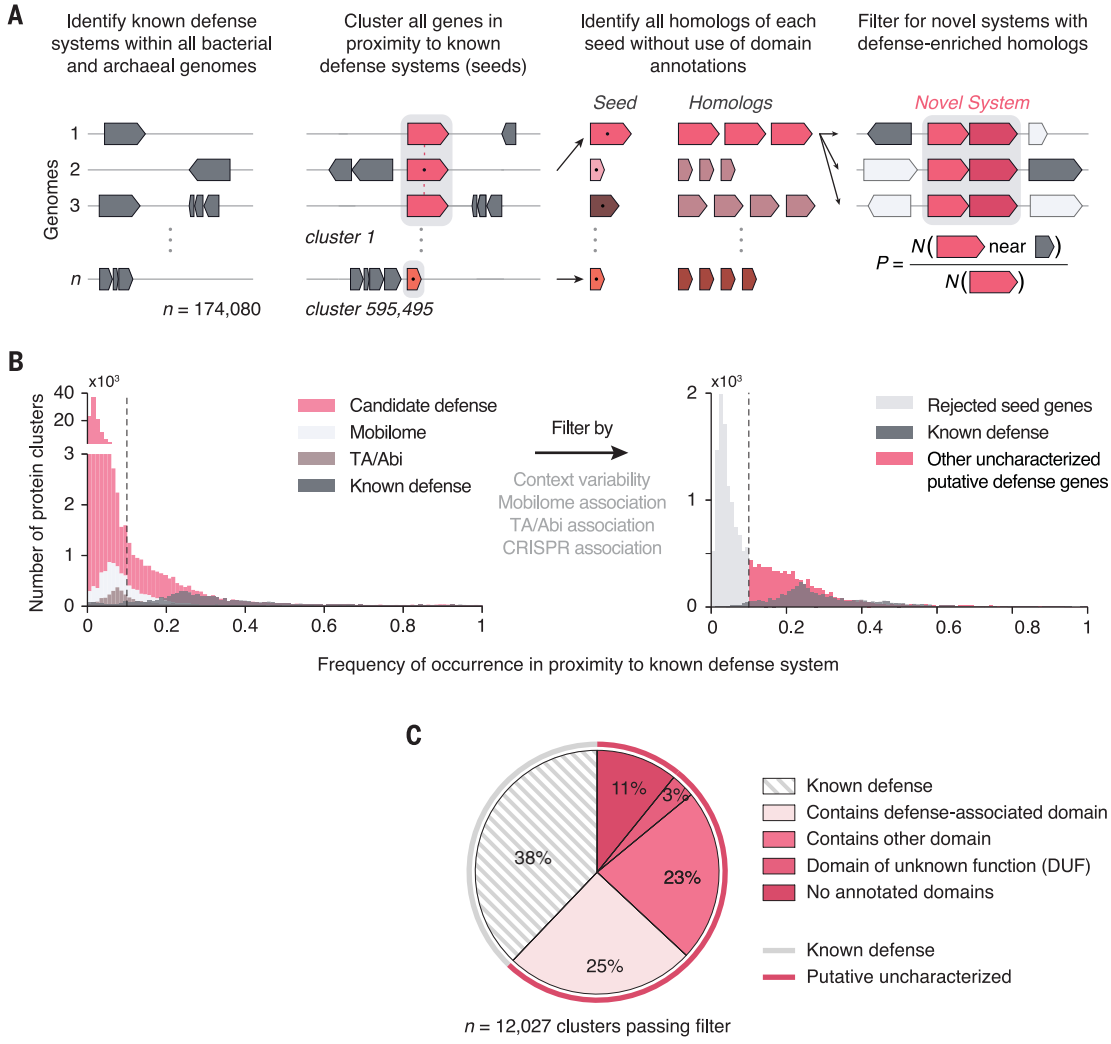

Fig. 1. Domain-independent prediction of putative antiviral defense systems. (**A**) Computational pipeline to identify uncharacterized putative defense systems across all sequenced bacterial and archaeal genomes. Defense systems were predicted on the basis of analysis of amino acid sequences, independent 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

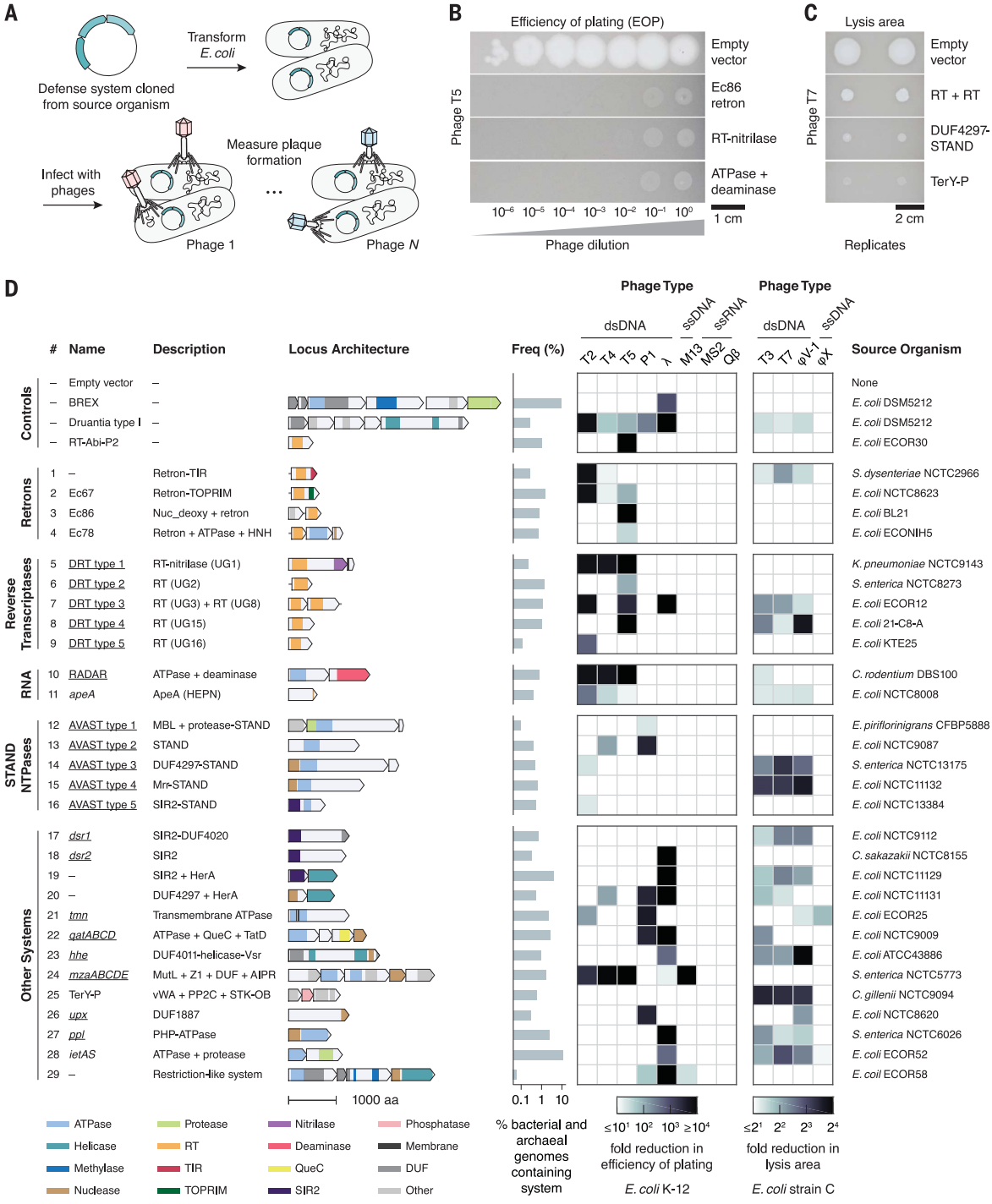

Fig. 2. Candidate defense systems exhibit antiviral activity in a heterologous system. (**A**) Experimental validation pipeline using phage plaque assays on *E. coli* heterologously expressing a cloned candidate defense system. (**B**) Example plaques and (**C**) zones of lysis for six candidate defense 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; mem, 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

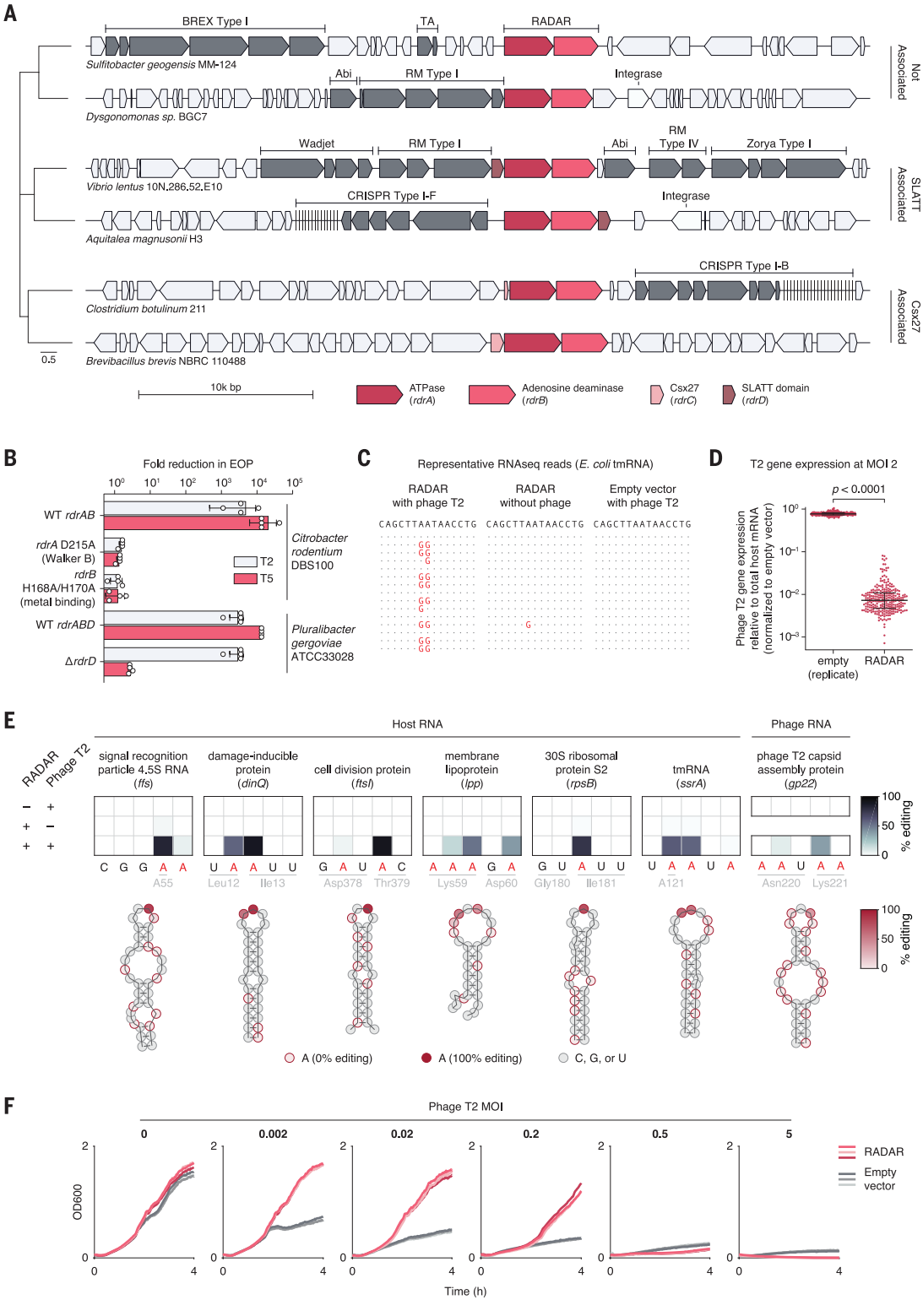

Fig. 3. RADAR mediates RNA editing in response to phage infection. (**A**) Examples of genomic loci containing three subtypes of RADAR (standalone, Csx27-associated, and SLATT-associated). (**B**) Essentiality of the core RADAR genes *rdrAB* and the accessory gene *rdrD* against phages T2 and T5. D215A, 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

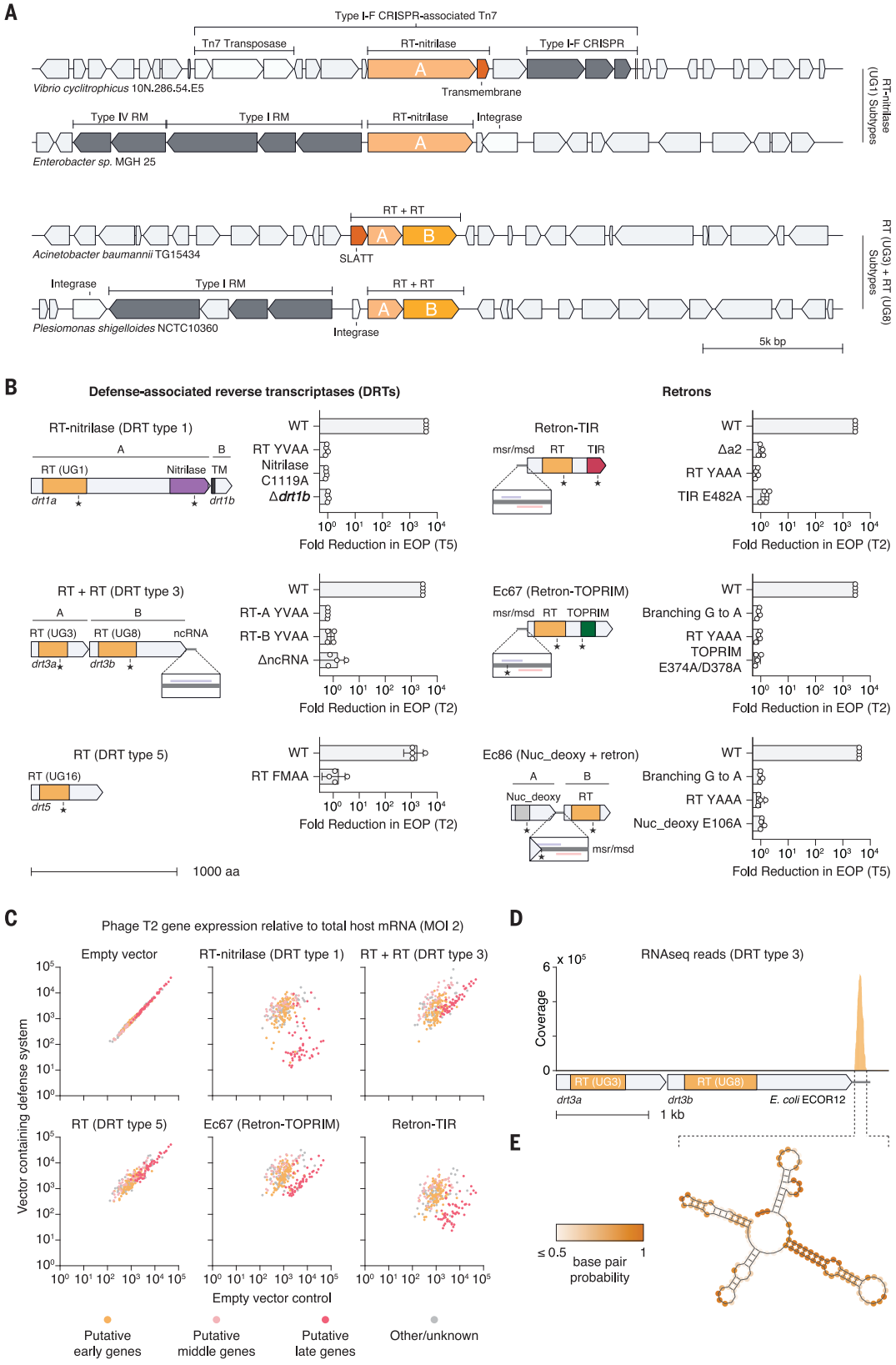

Fig. 4. Diverse families of RTs mediate antiviral defense. (**A**) Examples of genomic loci containing two RT-based defense systems (DRT type 1 and type 3), with two representative subtypes shown for each system. (**B**) Essential components of non-retron RTs (left panel) and retrons (right panel). TM, 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).

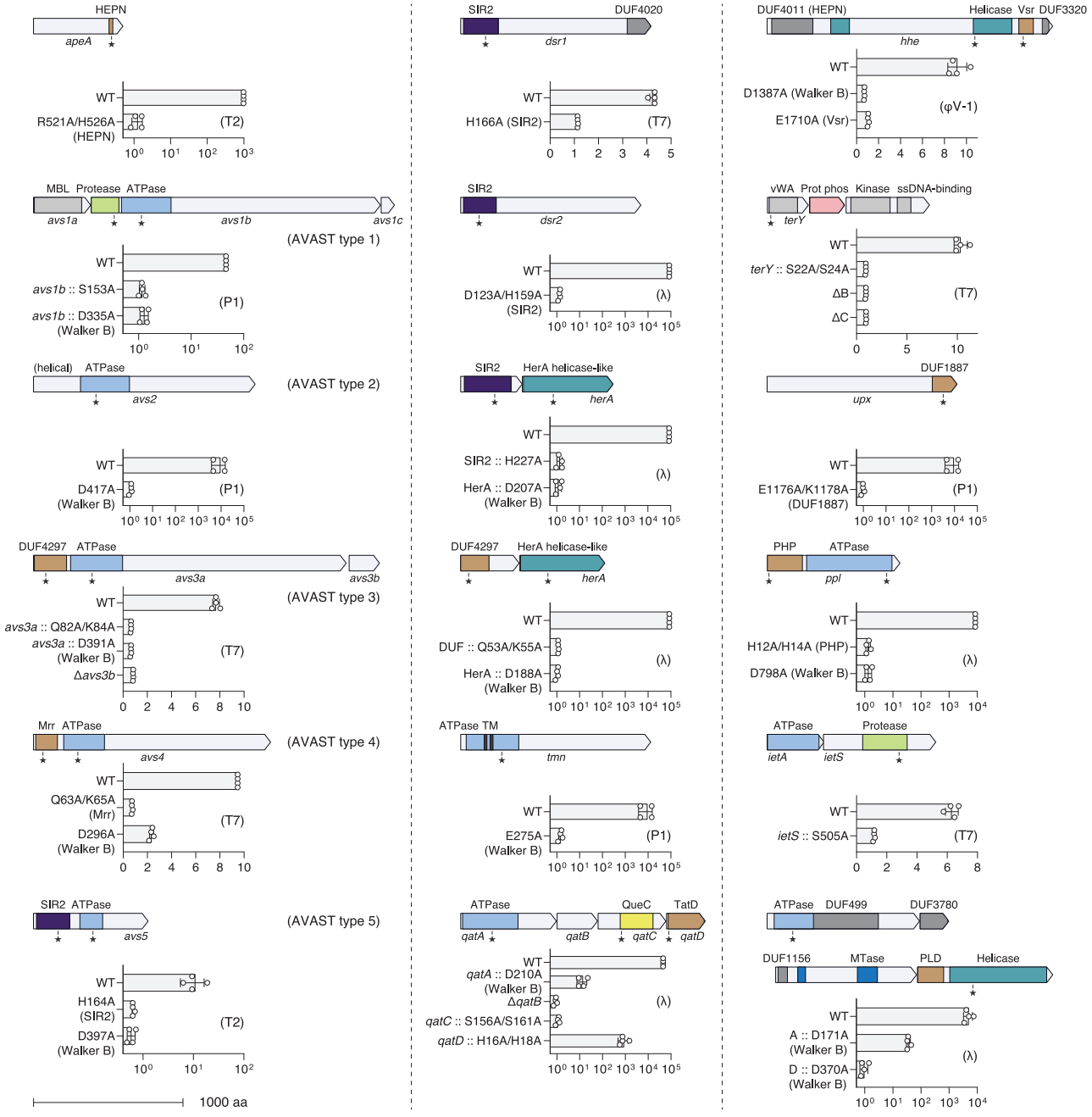

Fig. 5. Domain architectures and mutational analysis of additional defense systems. Graphics show domains identified by using HHpred, and stars indicate locations of active site mutations. Bar graphs (n = 4 replicates per bar) show either $\log_{10}$ fold change of efficiency of plating (for phages T2, 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

Supplementary Materials for

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, Feng Zhang*

*Corresponding author. Email: zhang@broadinstitute.org

Published 28 August 2020, *Science* **369**, 1077 (2020)
DOI: 10.1126/science.aba0372

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/content/369/6507/1077/suppl/DC1)

Materials and Methods

Detection of known defense systems. All bacterial and archaeal genomes (n = 174,080) were downloaded from Genbank (ftp://ftp.ncbi.nih.gov/genomes/genbank/) in November 2018. For genomes where gene annotations were incomplete or missing, genes were predicted using Prodigal (*39*). Known defense-related protein domains were annotated using RPSBLAST version 2.8.1 and the set of position-specific scoring matrices curated from the NCBI Conserved Domain Database (CDD) (*4, 10, 40, 41*). To reduce the false positive rate, a multi-gene system containing a ubiquitous protein domain was required to include two or more of its component genes in close proximity. For example, the type I restriction-modification endonuclease *hsdR* was called as a defense gene only if the corresponding methylase (*hsdM*) or specificity protein (*hsdS*) was also encoded in the vicinity. Genes were predicted for known defense systems including HsdRMS, McrBC, BREX, Druantia, Zorya, Wadjet, Thoeris, Hachiman, Lamassu, Gabjia, Septu, Shedu, Kiwa, pAgo, and other RM systems. Toxin-antitoxin systems were excluded from the set of known systems due to their overall low enrichment within defense islands (fig. S1).

Candidate novel defense genes. All translated protein-coding sequences within either 10 kb or 10 ORFs of known defense systems (whichever was greater), including the components of the known defense systems themselves, were compiled into a preliminary list (8.7 x 10^6 genes), which was expected to consist of both defense and non-defense genes. Highly similar sequences (at least 98% sequence identity and coverage) were discarded using the linclust option in MMseqs2 (*42, 43*) with parameters --min-seq-id 0.98 -c 0.98, resulting in a reduced list of 2.5 x 10^6 sequences. These sequences were then further clustered using the cascaded clustering option in MMSeqs2, yielding a final list of 6.0 x 10^5 representatives ('seeds').

Scoring candidate genes for defense enrichment. For each of the 6.0 x 10^5 seeds, a 'defense enrichment score' was computed as (number of homologs in proximity to one or more known defense systems) / (total number of homologs). A gene was considered to be located in proximity to a known defense system if it occurred no more than 5 kb or 5 ORFs away from the locus encoding that system. CRISPR-Cas systems were omitted from the defense score calculation due to their low defense island association (*10*). Candidate sequences with a defense enrichment score of 0.1 or higher were retained for subsequent analysis, with the exception of predicted mobilome components (such as transposons), which were discarded. This cut-off was chosen because more than 90% of the known defense genes scored higher than this value, whereas most mobilome, toxin-antitoxin, and other non-defense genes scored lower (Fig. 1B and fig. S1). To identify homologs of the candidate proteins, all 6.2 x 10^8 proteins in Genbank were tabulated, and highly similar proteins (at least 98% sequence identity and coverage) were removed, resulting in a reduced list of 1.3 x 10^8 proteins. Each seed sequence was then searched against this non-redundant protein sequence database using MMseqs2. To qualify as evidence of homology, the resulting alignments were required to have a minimum coverage of 70% and a maximum E value of 10^{-5} (parameters --cov-mode 0 -c 0.7 -e 0.00001). The set of identified homologs was further clustered at 90% sequence identity to perform stringent redundancy reduction. In order to accurately compute defense association frequencies, seeds with fewer than 50 homologs after redundancy reduction were discarded.

Filtering defense-enriched genes based on context diversity. To select for genes that are likely to encode components of independent defense modules, defense-enriched seeds were further required to have sufficient context diversity. For each seed, the number of homologs within 5 kb or 5 ORFs of different defense system categories was counted, and the seed was retained if the entropy of this list, defined as $\sum -p_i \ln p_i$, where p_i is the normalized frequency of category i, was at least 0.9. This value corresponds to

halfway between 2 and 3 non-zero entries in the case of a uniformly distributed frequency vector. Seeds were further filtered based on the proportion of homologs next to predicted toxin-antitoxin/Abi, mobilome, and CRISPR-Cas genes (fig. S1).

Refining the classification of putative defense genes. A total of 12,027 seeds passing filter was identified, consisting of both known and putative defense genes. To determine whether each gene was putative or known, the original classification was refined as follows. A list was compiled of the amino acid sequences of reported homologs of known systems, including 288,776 restriction-modification proteins from REBASE (*44*); 517 proteins for BREX (*14*); and 27,775 proteins for other recently-identified systems (*4, 45, 46*). This list was supplemented with additional curated homologs and, following redundancy reduction, searched against the putative defense seeds using MMseqs2. Seeds that matched one or more of these known defense genes (at least 70-80% coverage with a maximum E value of 10^{-5}) were labeled as known. A subset of labels were adjusted by an additional round of manual curation, resulting in a classification of 4,555 known and 7,472 putative defense genes.

Domain analysis of predicted defense genes. The 7,472 putative defense seeds were further analyzed with additional, more sensitive methods to assess their domain content. For each seed gene, a multiple sequence alignment (MSA) of its homologs was created using MAFFT (*47*). If the number of homologs was 1,000 or fewer, all homologs were included in the alignment; otherwise, 1,000 homologs were randomly selected for inclusion. MSAs were searched against the Pfam 32.0 database using HHpred (*48*), and domain predictions with at least 80% probability were retained. Of these 7,472 genes, 3,029 (41%) contained at least one pfam domain that has been reported to be defense-associated (*4, 10, 45*). Although some of these 3,029 proteins could be distant homologs of known defense proteins, many were included in this category because they contained ubiquitous pfam domains that are also employed by some known defense systems (in particular, AAA-family ATPases, helix-turn-helix (HTH) motifs, and (P)D-(D/E)xK-family nucleases); these are predicted to be uncharacterized defense genes. The remaining 59% either had no domain hits or contained only domains that were not in the set of defense-associated pfams.

From genes to defense systems. For each selected candidate defense protein, the gene neighborhoods of 30 homologs in proximity to known defense genes were randomly chosen and examined to identify conserved (predicted) operons that contained the seed and could be expected to constitute a minimal, intact defense system. Protein domains were predicted using HHpred, and the resulting prediction was used to infer the potential involvement of the respective proteins in the activity of the respective predicted defense system.

Estimation of defense system abundance. To estimate the abundance of each validated defense system in microbial genomes, we downloaded n = 205,214 genomes available in Genbank as of August 2019. For each defense system, initial protein sequence seeds encoded by the corresponding signature genes were taken from experimentally validated loci. Initial seeds were aligned and converted into HMM profiles. We then used a constrained 2 iteration HMM profile search to generate highly specific HMM profiles and retrieve related systems as follows. Each ORF of size 150aa or greater, with one or more hits, was searched against all HMM profiles using HMMER3.1 and assigned to the profile that had the highest scoring match. For each system, ORFs with profile hits with less than 500 bp of intergenic distance on the same strand were grouped into candidate loci. For multi-protein systems, a putative locus was considered a hit if every signature gene profile for the system had a match in the locus with a bit score of at least 25. For single gene systems, a locus was considered a hit if the protein had a match to the system's single signature gene profile with a bit score of at least 50 and an alignment coverage of at least 70%. Signature proteins from the

identified systems were separately clustered at 50% identity using MMseqs2 and subsequently aligned using MAFFT. The alignments were used to create a new set of signature gene profiles as input to the next iteration. For BREX and Type I RM, we used preexisting pfam profiles for the signature genes in place of iterative HMM profile searching. The final abundance was calculated as the number of hits for the given system divided by the number of genomes (*n*).

Bacteria and phage strains. Phages T2, T3, T4, T5, T7, P1, λ, φV-1, M13, φX174, MS2, and Qβ, as well as host *E. coli* strains K-12 (ATCC25404) and C (ATCC13706), were obtained from the American Type Culture Collection (ATCC). The genome of phage φV-1, originally isolated from a measles vaccine (*49, 50*), was sequenced and found to be 92% similar to enterobacteria phage 285P, a T7-like phage (*51*).

Cloning. To facilitate experimental validation using coliphages, the source organism of each candidate defense system was chosen to be as phylogenetically similar as possible to *E. coli*—in particular, from other strains of *E. coli* whenever possible. Candidate defense systems were cloned into the low-copy plasmid pACYC184. When possible, genomic DNA from source organisms was obtained from ATCC, NCTC, or DSMZ, and the genes of interest were amplified with Q5 (New England Biolabs) or Phusion Flash (Thermo Scientific) polymerase, using primers with 5' ends homologous to the ends of the plasmid backbone. Plasmids were assembled using the NEBuilder HiFi DNA Assembly mix (New England Biolabs). When the source organism was not readily available from public culture collections, genes were chemically synthesized (GenScript or GENEWIZ). When possible, the native promoter was retained. For source organisms outside of Enterobacteriaceae, the system was placed under a *bla* or *lac* promoter.

Sequence verification of plasmids. The full sequences of all plasmids were verified by high-throughput sequencing. To prepare sequencing libraries, 25-50 ng of each plasmid was mixed with purified Tn5 transposome loaded with Illumina adapters and incubated at 55 °C for 10 min in the presence of 5 mM $MgCl_2$ and 10 mM TAPS buffer (*52*). The quantity of Tn5 was titrated to generate an average fragment size of ~100-400 bp. Tagmentation reactions were subsequently treated with 0.5 volumes of 0.1% sodium dodecyl sulfate for 5 min at room temperature and amplified with KAPA HiFi HotStart polymerase using primers containing 8 nt i7 and i5 index barcodes. Barcoded amplicons were sequenced on a MiSeq (Illumina) with at least 150 cycles for the forward read. Reads were aligned to the reference plasmid sequence by the Geneious read mapper, and error-free plasmids were retained for subsequent experiments.

Competent cell production. *E. coli* strains K-12 and C were cultured in ZymoBroth, with 25 μg/mL chloramphenicol when appropriate, and made competent using Mix & Go buffers (Zymo) according to the manufacturer's recommended protocol.

Phage plaque assays. *E. coli* host strains were grown to saturation at 37 °C in Luria Broth (LB). To 10 mL top agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 7 g/L agar) was added chloramphenicol (final concentration 25 μg/mL) and 526 μL *E. coli* culture, and the mixture was poured on 10 cm LB-agar plates containing 25 μg/mL chloramphenicol. For phages T2, T4, T5, P1, λ, M13, MS2, and Qβ, dilutions of phage in phosphate buffered saline were spotted on the plates, and plaque counts were recorded after overnight incubation at 37 °C. If individual plaques were too small to be counted, the most concentrated dilution at which no plaque formation was visible was recorded as having a single plaque. For phages T3, T7, φV-1, and φX174, a total of 3 μL of phage containing 5 x 10^6 virions was spotted, and the area of the zone of lysis was measured after incubation at 37 °C for 68 hr. A total of 2-4 technical replicates was collected for each infection condition. Initial screening of defense system candidates was performed in *E. coli* K-12

(ATCC25404), excluding phage φX174 due to its inability to infect *E. coli* K-12; systems with observed defense activity were further tested as described above.

Phage cultivation. Phages T2, T3, T4, T7, φV-1, M13, φX174, MS2, and Qβ were propagated in liquid culture. The host *E. coli* strain for each phage was grown to an OD600 of 0.2 – 0.4 at 37 °C in LB and infected with a slab of top agar containing phage plaque from a previous lysis. Cultures were grown overnight at 37 °C with 250 rpm agitation. Phages T5, P1, and λ were propagated by the double agar overlay method; after overnight incubation at 37 °C, plaques were scraped in LB. For both liquid culture and double agar overlay, phage samples were centrifuged to pellet cellular debris, and the supernatant was filtered through with a 0.22 µm sterile filter.

Phage genome sequencing. DNA from phage φV-1 was isolated using QuickExtract DNA extraction solution (Epicentre) following the manufacturer's recommended protocol. After tagmentation and PCR amplification steps described earlier for plasmid sequence verification, the library was sequenced on a MiSeq with 200 cycles for the forward read and 110 cycles for the reverse read. Trimmed reads were assembled into contigs with SPAdes 3.13.0 using the --careful option, and contigs were subsequently scaffolded into a full genome using the genome sequence of enterobacteria phage 285P (*51*) as a reference.

Whole transcriptome sequencing. *E. coli* ATCC25404, containing either an empty vector or a candidate defense system, was grown to log phase in LB and diluted to an OD600 of 0.2. The culture was then split into two tubes, one of which was infected with phage T2 at an estimated MOI of 2. Both tubes were incubated at 37 °C for 1 hr with 250 rpm agitation. RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific) and treated with DNAse I, followed by a RiboMinus ribosomal RNA depletion kit (Thermo). Sequencing libraries were prepared using NEB Ultra II directional RNAseq library prep kit (New England Biolabs) and paired-end sequenced (2 x 75 cycles) with a NextSeq (Illumina). Adapter sequences were trimmed from sequencing reads using CutAdapt (with parameters --trim-n -q 20 -m 20 -a AGATCGGAAGAGC -A AGATCGGAAGAGC), and trimmed reads were aligned to the *E. coli* MG1655 reference genome using the Geneious read mapper.

Phage fragmentation. Phage fragments were amplified from the genome of phage T2 by PCR, cloned into an ampicillin-resistant plasmid after an IPTG-inducible T7 promoter, and sequenced verified as previously described. Each fragment was then transformed into NovaBlue(DE3) *E. coli* expressing the RADAR system from *Citrobacter rodentium* DBS100. Independent colonies for each fragment were grown to saturation at 37 °C in LB with 25 µg/mL chloramphenicol and 100 µg/mL ampicillin. Cultures were then diluted 1 to 5 in the same media, and IPTG was added to a final concentration of 0.5 mM. After 4h growth at 37 °C, cells were pelleted by centrifugation, and total RNA was extracted by a Direct-zol RNA purification kit (Zymo). The *E. coli* tmRNA was subsequently amplified by RT-PCR (QuantBio) and sequenced with a MiSeq (Illumina).

E. coli growth kinetics. Cells were grown to log phase in LB and diluted to an OD600 of 0.2. Cultures were infected with phage T2 at varying MOI at grown at 37 °C, and the OD600 was measured every 2 min for a total duration of 3-4 hr on a Synergy Neo2 plate reader (BioTek).

Classification of phage genes. Phage T2 genes were classified as putative early, middle, or late genes based on the closest promoter on the same strand, as annotated based on the genome of phage T4 (*53*). Genes that could not be unambiguously classified were labeled as unknown.

RNA secondary structure prediction. Minimum free energy RNA secondary structures were predicted using the Turner (2004) energy parameters at 37 °C (*54*).

Prophage analysis. Prophage and phage DNA sequences were downloaded from PHASTER (*55, 56*). All clusters (seed gene plus identified homologs) with hits matching the experimentally validated systems, as well as one cluster matching the *rexA* gene of phage lambda as a positive control, were searched against the PHASTER database with tblastn for near identical matches (≥ 95% identity). For each cluster, phage association frequency was calculated as the number of proteins in the cluster with unique matches to the PHASTER database divided by the total number of unique proteins in the cluster (number of proteins after clustering at 90% sequence identity). The cutoff for frequent phage association of a system was defined as half of the frequency for *rexA*. We note that PHASTER does not predict all instances of prophages and prophage remnants, and we have also considered an alternative approach of identifying prophage association based on proximity to integrases, which may allow a greater number of prophages to be identified. However, a challenge with the latter approach is that defense islands often appear to derive from mobile genetic elements other than prophages and contain many integrases that originate from non-phage sources (*e.g.*, CRISPR-associated transposases (*57, 58*)), leading to a high rate of false positives. The use of PHASTER provided the advantage of substantially reducing the false positives that would otherwise be expected for an approach based on integrase association.

Computational analysis of the RT (UG1) nitrilase domain. Homologs of the RT (UG1) defense gene were identified with a PSIBLAST search seeded on the experimentally validated sequence (WP_115196278.1), and highly similar homologs (≥ 90% identity) were removed. An MSA of the nitrilase domain was then created using MAFFT, and a custom position-specific scoring matrix (PSSM) was derived from this alignment. Bacterial and archaeal proteins in Genbank (redundancy-reduced at 98% sequence identity and coverage) were then searched against this profile with RPSBLAST, and the E-values of proteins with a match covering a minimum of 20% of the length of the profile were recorded. Known nitrilase enzymes were identified using a separate RPSBLAST search against the same set of Genbank proteins using 36 PSSMs from the CDD database (E-value ≤ 10^{-6}; minimum 40% profile coverage): cd07197, cd07564, cd07565, cd07566, cd07567, cd07568, cd07569, cd07570, cd07571, cd07572, cd07573, cd07574, cd07575, cd07576, cd07577, cd07578, cd07579, cd07580, cd07581, cd07582, cd07583, cd07584, cd07585, cd07586, cd07587, COG0388, pfam00795, PLN02504, PLN02747, PLN02798, PRK10438, PRK13286, PRK13825, TIGR00546, TIGR03381, and TIGR04048.

Establishing an abi response. Abortive infection (abi) systems, which are based on altruistic cell suicide or dormancy (*59*), typically induce non-specific or deleterious biochemical activity targeting the host cell that also interferes with the phage reproduction cycle. Abi responses can be characterized through traditional assays such as efficiency of the center of infection (ECOI), adsorption, host survival, and one-step growth curve measurements. However, because the events of phage DNA injection and expression of toxic early genes are likely to be deleterious to an infected cell even if the production of progeny phages is ultimately suppressed, these assays may not be informative in terms of distinguishing between abi vs. non-abi mechanisms. An alternative approach to establishing the existence of an abi response is to identify the biochemical activity of the defense system, which we have focused on for the RADAR system.

Gene knockouts vs. heterologous reconstitution. To further assess the feasibility of performing knockout experiments in the source bacterial strains for each defense system, we performed analyses which suggested that different defense systems with overlapping phage specificities often co-occur. For instance, *E. coli*

strain DSM5212 contains both BREX type I and Druantia type I (Fig. 2D), both of which were included as positive controls; if BREX were to be knocked out in this strain, the presence of Druantia would likely ensure that its phage resistance profile across the 12 phages in our assay would remain unchanged. Similarly, the SIR2 + HerA system from *E. coli* strain NCTC11129 primarily confers resistance to phage lambda (Fig. 2D); the source strain NCTC11129 additionally contains BREX type I, which also confers resistance against phage lambda. Collectively, these observations suggested that the knockout of a single defense system may not be sufficient to make its corresponding source strain phage-sensitive, motivating the use of heterologous reconstitution as the primary assay for defense activity.

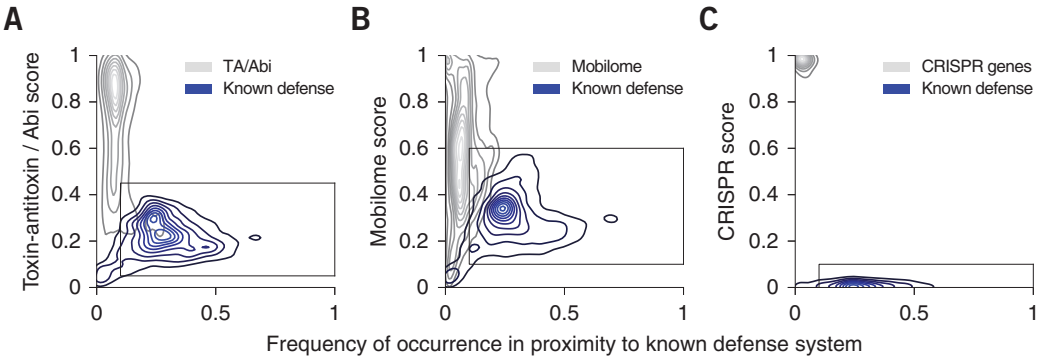

Figure S1: Selection of filtering thresholds for prediction of putative defense genes. Contour density plots for predicted (**A**) toxin-antitoxin / abi genes, (**B**) mobilome genes, and (**C**) CRISPR-Cas genes. Boxes indicated the parameter thresholds selected for filtering putative defense genes.

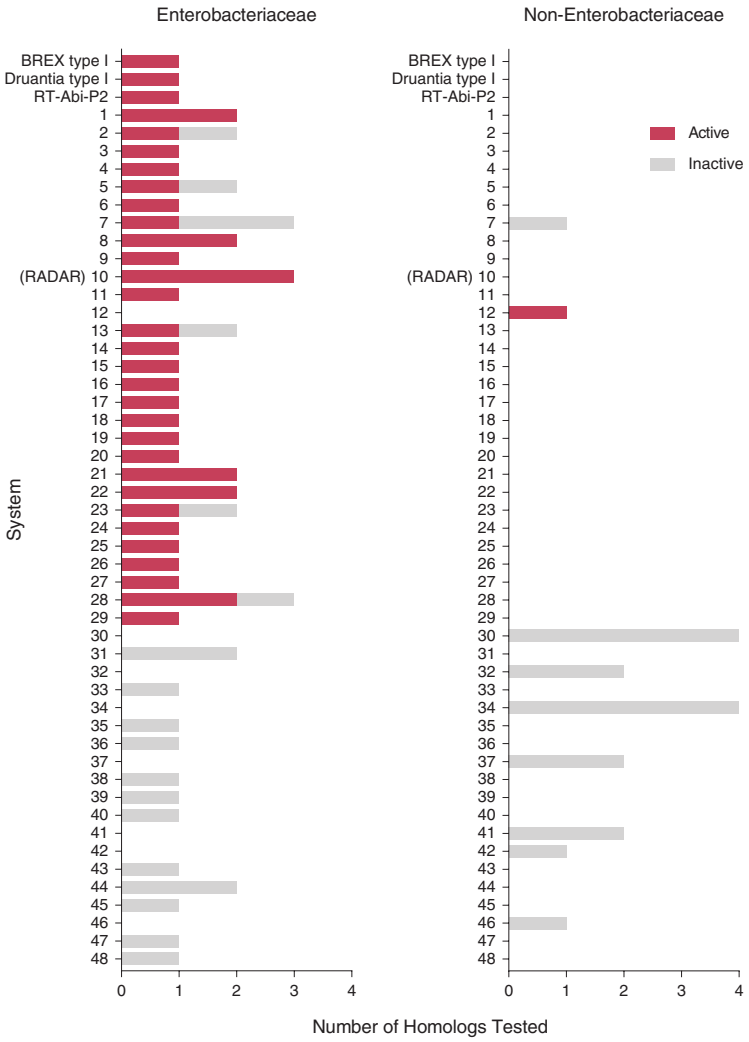

Figure S2: Summary of tested homologs of candidate defense systems, stratified by source organism (Enterobacteriaceae vs. non-Enterobacteriaceae). Systems 1-29 correspond to the numbering in Fig. 2D.

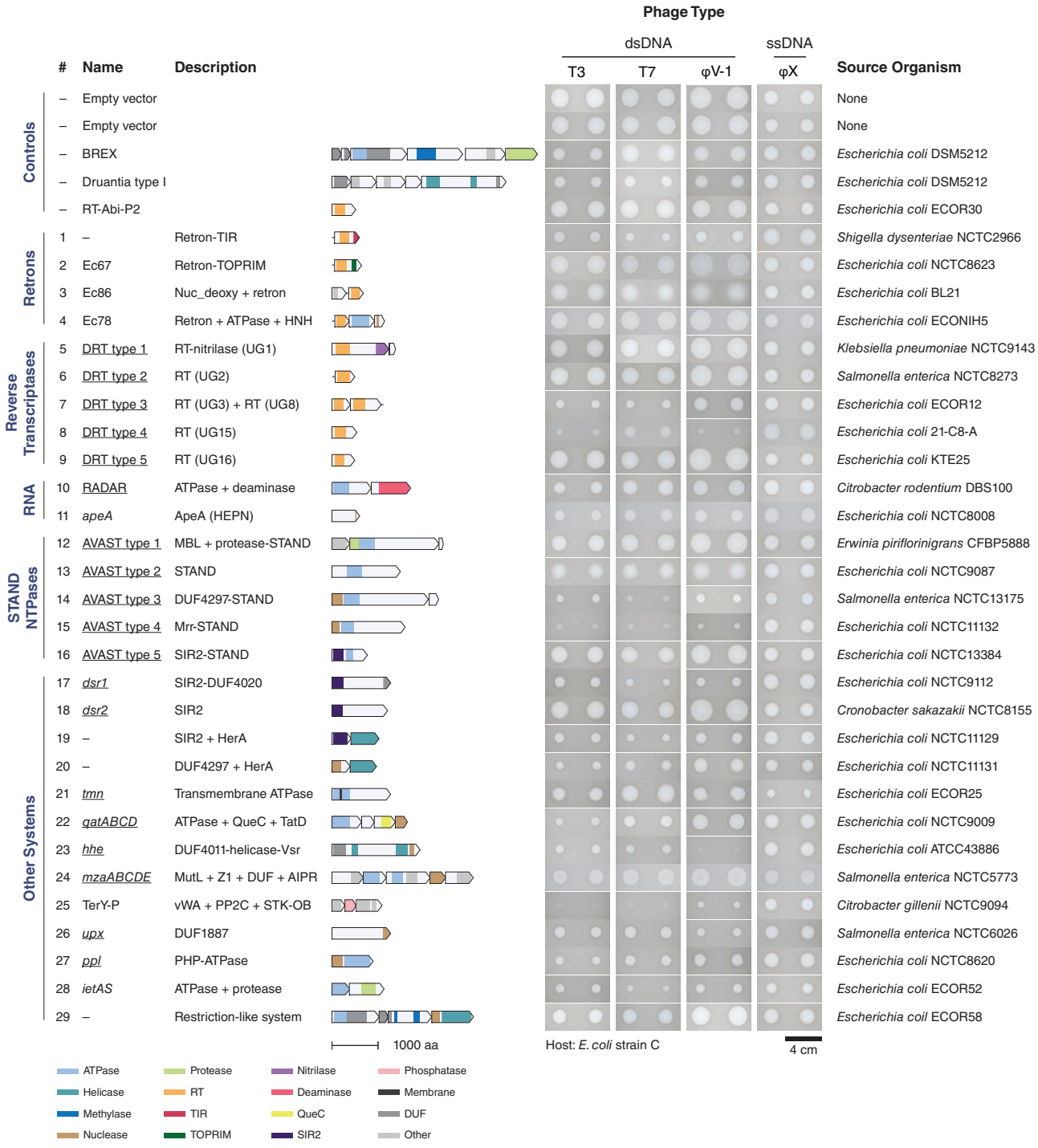

Figure S3: Representative zones of lysis for phages T3, T7, φV-1, and φX174 on *E. coli* strain C (n = 2 replicates each), corresponding to the right panel of **Figure 2D**. A total of 5 x 10^6 virions were deposited per spot.

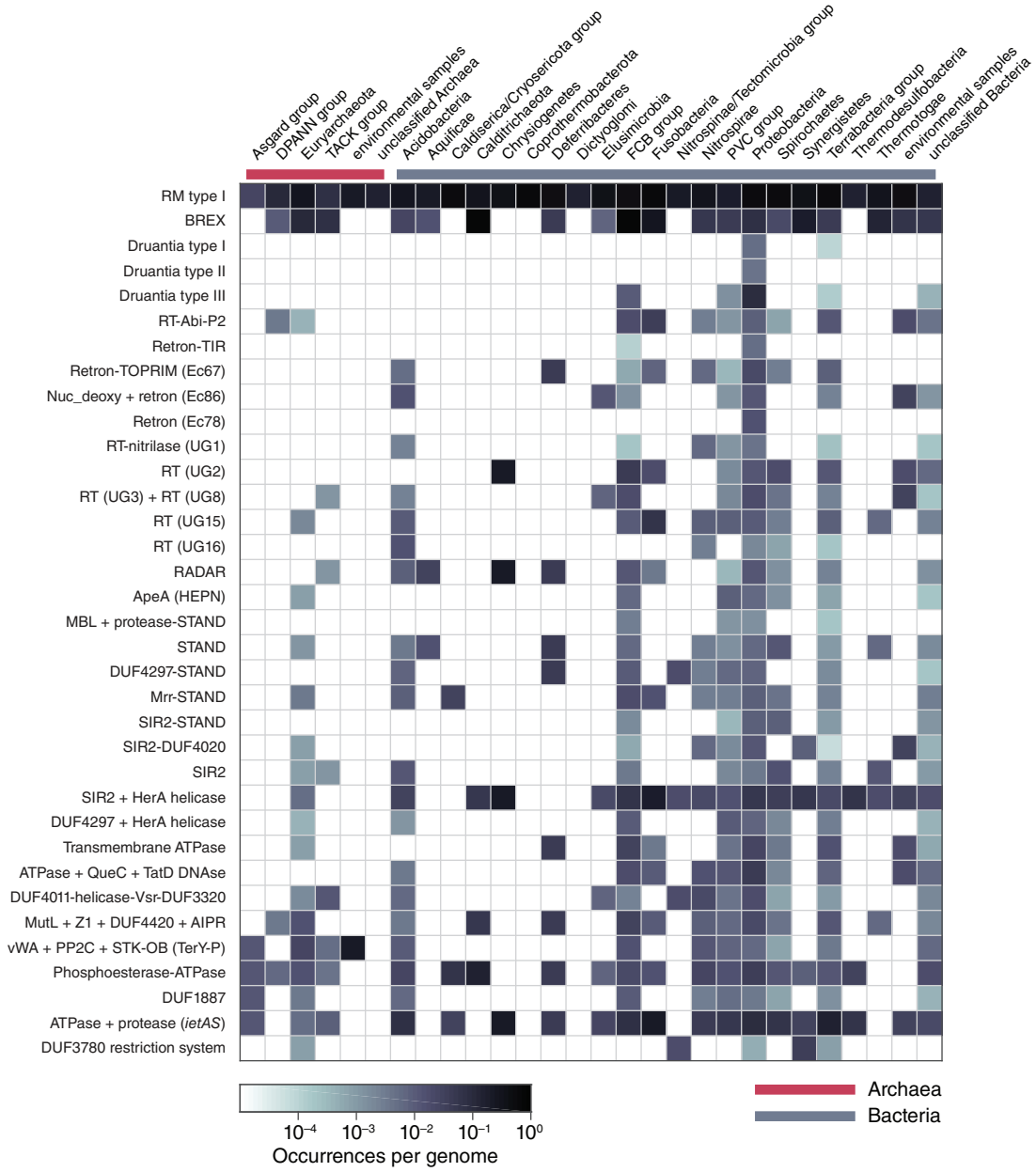

Figure S4: Abundance of validated defense systems within sequenced genomes, stratified by phylum. Defense system homologs were predicted using a two-step HMM-based search across all bacterial and archaeal genomes in Genbank (see **Methods**).

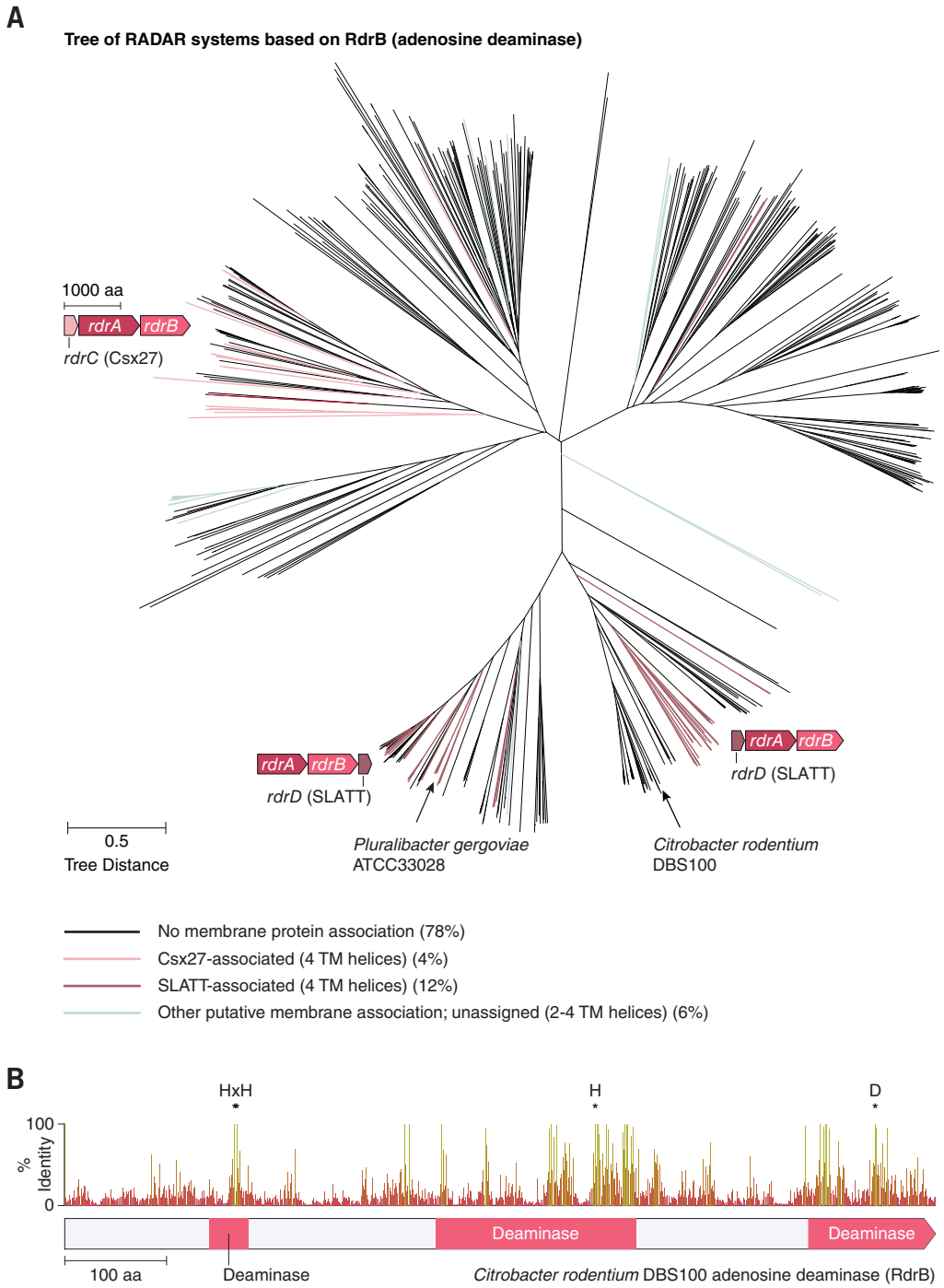

Figure S5: Domain and locus architecture of the RADAR deaminase. (**A**) Unrooted neighbor-joining tree of RdrB homologs with the Jukes-Cantor genetic distance model. Distinct clades of RADAR incorporate accessory membrane proteins RdrC (Csx27) or RdrD (SLATT). (**B**) RdrB contains a split deaminase domain (red) with uncharacterized insertions. Domain boundaries were predicted using HHpred. Percent identity was calculated from a multiple sequence alignment of 535 representative homologs with at most 98% pairwise similarity.

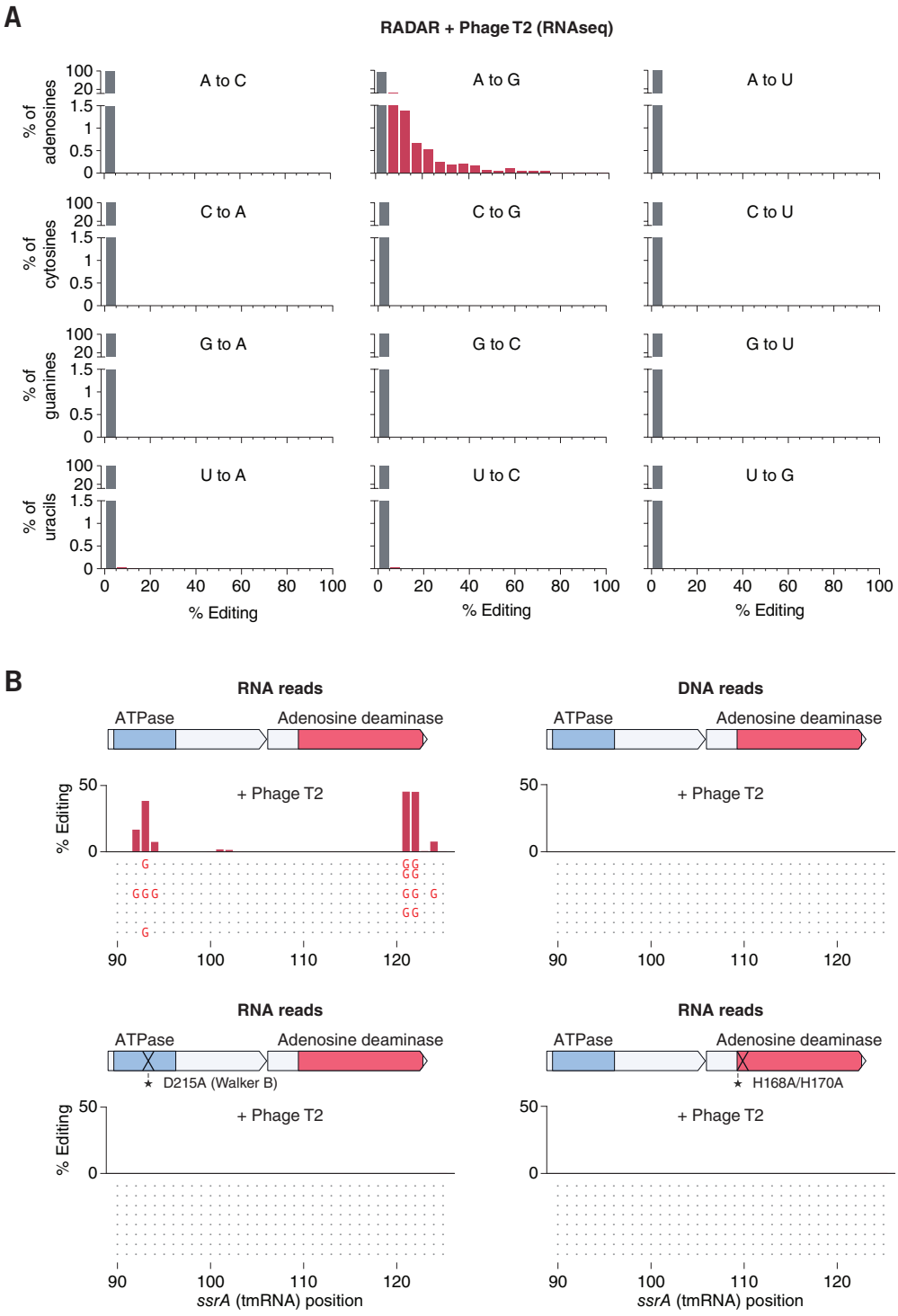

Figure S6: Deamination by the RADAR system occurs only on adenosines within RNA and requires both RADAR genes. (**A**) Empirical probability mass functions of editing frequency for each of the 12 possible RNA base changes, calculated using the highest-expressed mRNAs in the transcriptome of *E. coli* K-12 (ATCC25404) expressing the RADAR system from *Citrobacter rodentium* DBS100. Cells were harvested 1 hr after infection by phage T2 at an MOI of 2. (**B**) Editing frequency at a selected site within the transfer messenger RNA (tmRNA) locus (RNA or DNA). Sequences below the graphs show representative reads.

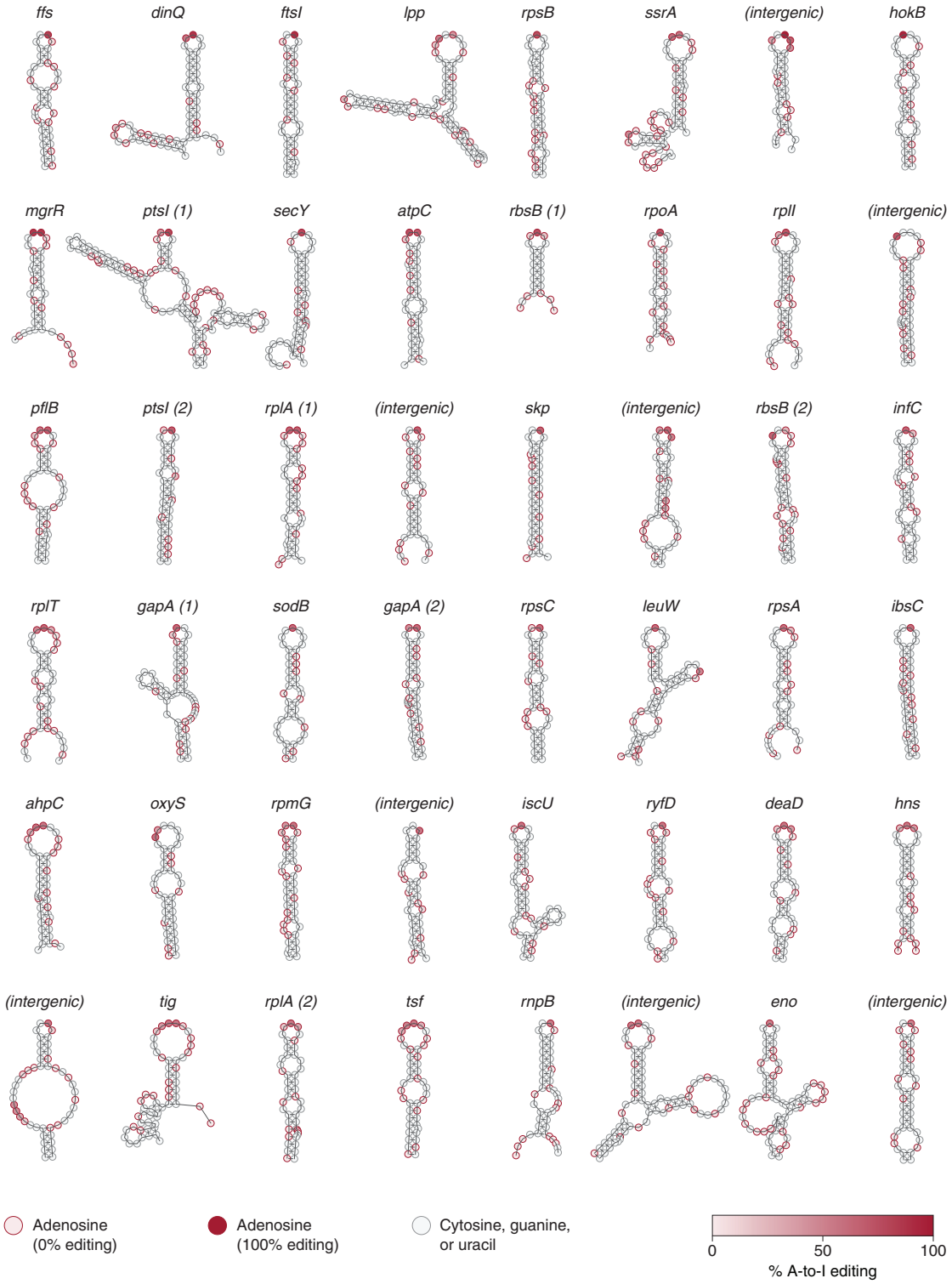

Figure S7: RADAR preferentially deaminates adenosines within loop regions of RNA stem-loops. Predicted RNA secondary structures of the 48 highest-expressed strong RADAR editing sites (≥50% editing).

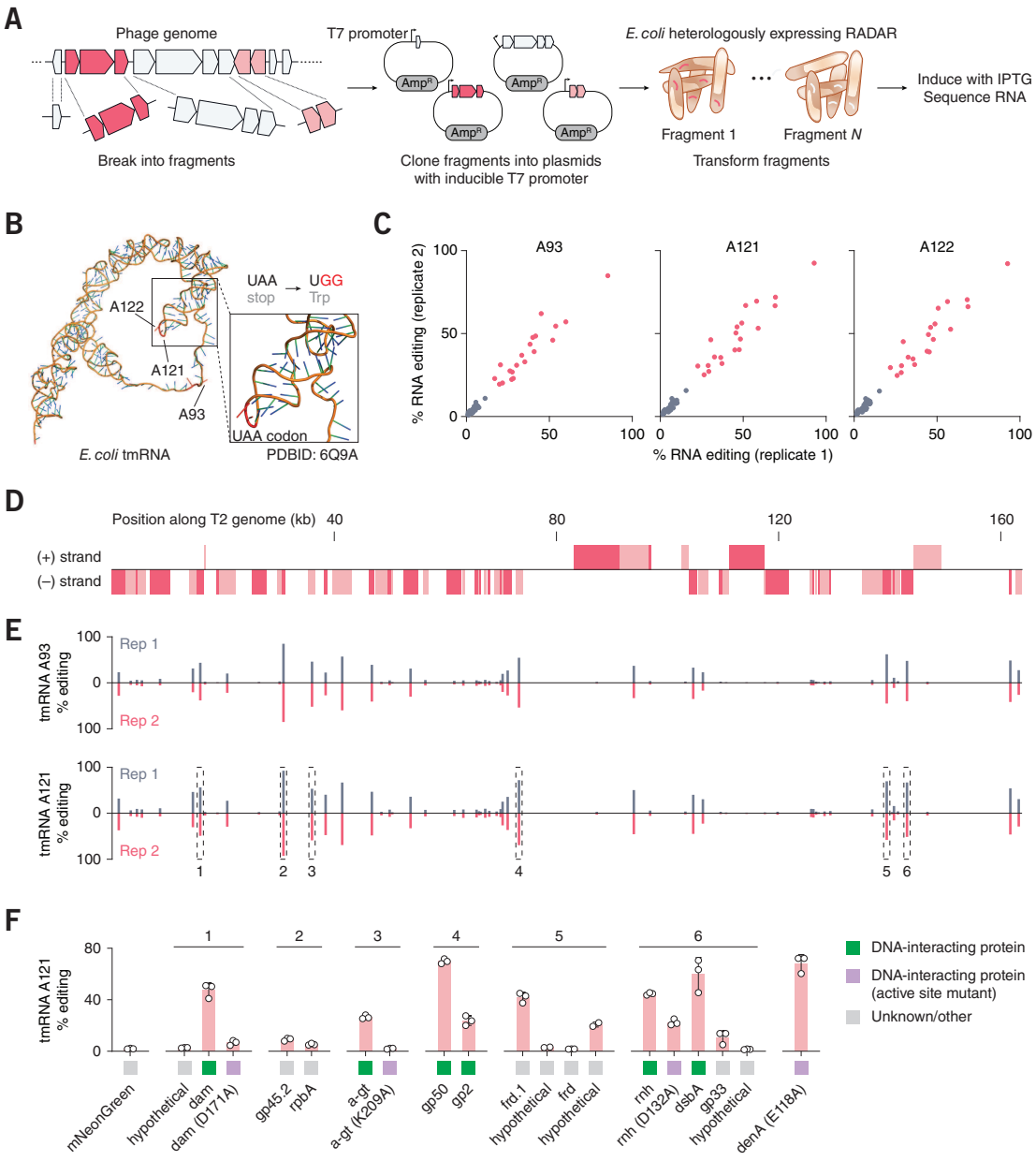

Figure S8: Effect of expression of specific phage genes on RNA editing by RADAR. (**A**) Phage genes were cloned after IPTG-inducible T7 promoter and transformed into *E. coli* heterologously expressing the RADAR system from *Citrobacter rodentium* DBS100. (**B**) Structure of *E. coli* transfer messenger RNA (tmRNA) (PDBID: 6Q9A), highlighting adenosines strongly edited by RADAR. (**C**) Scatter plots of RNA editing frequencies for two replicates. Each dot represents a different phage fragment. (**D**) Locations of fragments on the phage T2 genome. Each colored box represents a distinct fragment. (**E**) RNA editing frequencies of the fragments shown in (**D**) at A93 and A121 of the *E. coli* tmRNA. (**F**) RNA editing frequencies induced by expression of RADAR with individual genes within six of the highest-activity fragments identified in (**D**). Purple squares indicate active site mutants created by site-directed mutagenesis. *dam*: DNA adenine methyltransferase; *a-gt*: DNA alpha glucosyltransferase; *gp50*: head completion protein; *gp2*: DNA end protector protein; *frd*: dihydrofolate reductase; *rnh*: RNase H; *dsbA*: dsDNA binding protein; *denA*: endonuclease II.

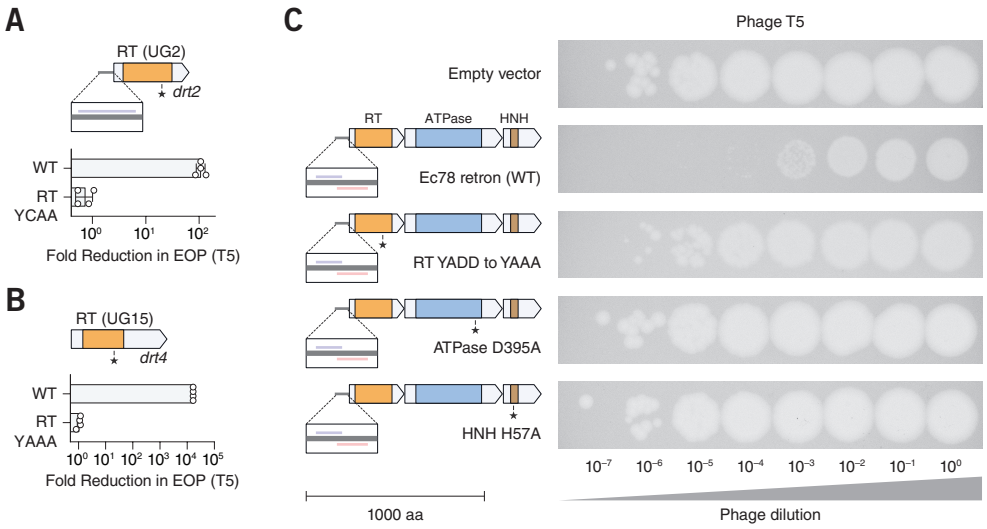

Figure S9: Mutational analysis of three RT-containing defense systems. Active site mutations abolish defense activity against phage T5 for the (**A**) DRT type 2, (**B**) DRT type 4, and (**C**) Ec78 retron systems. The ATPase and HNH proteins in Ec78 comprise the Septu defense system.

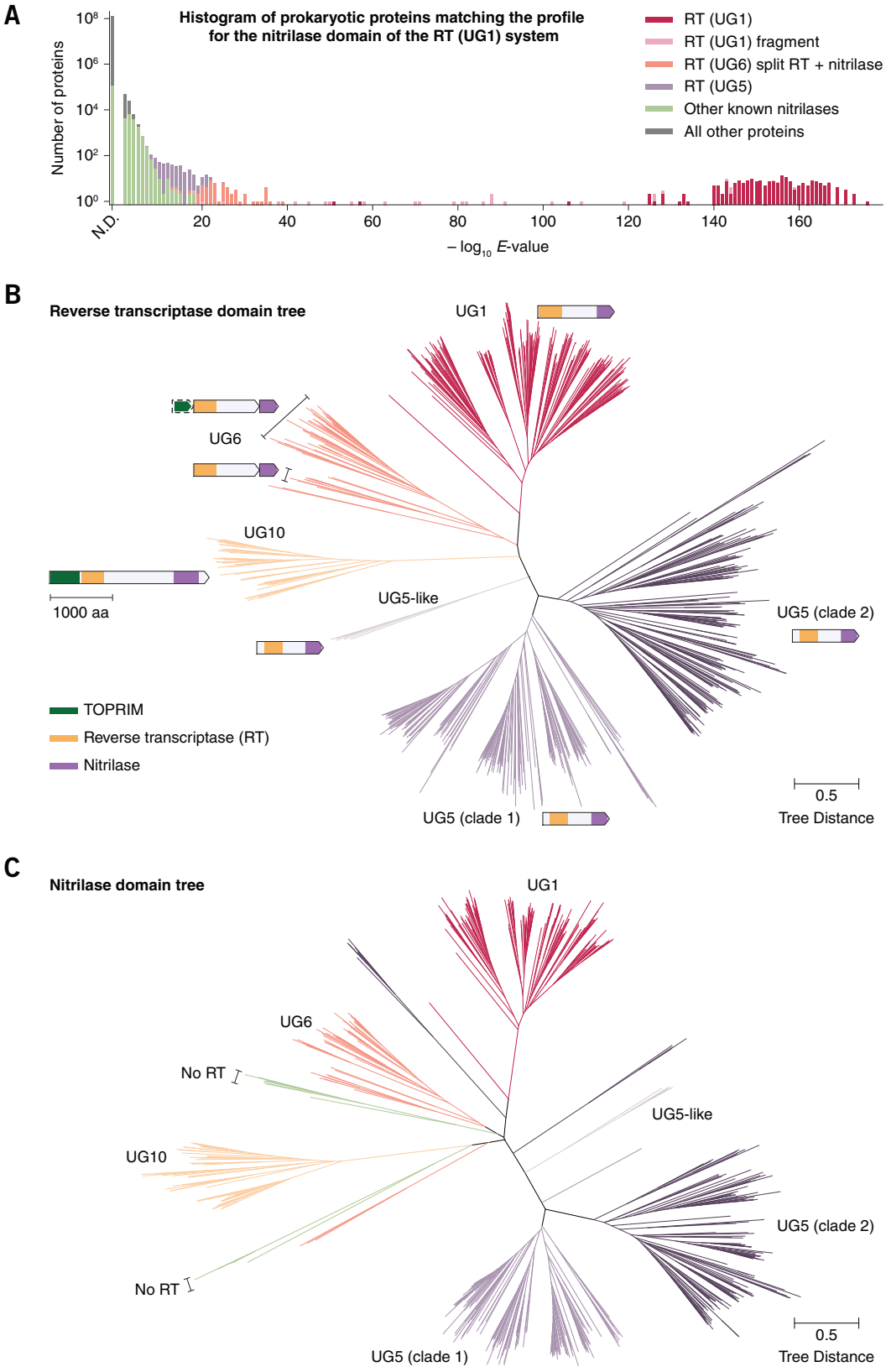

Figure S10: The nitrilase domain of the DRT type 1 (RT (UG1)) defense system forms a distinct clade among nitrilase enzymes. (**A**) Stacked histogram of *E*-values of sequence-profile matches (RPSBLAST) between prokaryotic proteins in Genbank and a custom position-specific scoring matrix (PSSM) for the RT (UG1) nitrilase domain (minimum 20% coverage). Proteins matching a known nitrilase PSSM from the CDD database (*E*-value $\leq 10^{-6}$; minimum 40% coverage) are shown in green. (**B**) Unrooted neighbor-joining tree of the RT domain in nitrilase-associated RTs ($n = 588$). Colors indicate distinct clades (cutoff tree distance 0.15). (**C**) Unrooted neighbor-joining tree of the nitrilase domain in proteins in (**B**) with the same color scheme (based on RT domain clade). Also included in the tree are the non-RT-associated nitrilases (green) that are most similar to the nitirlase domain in RT (UG1) among prokaryotic proteins.

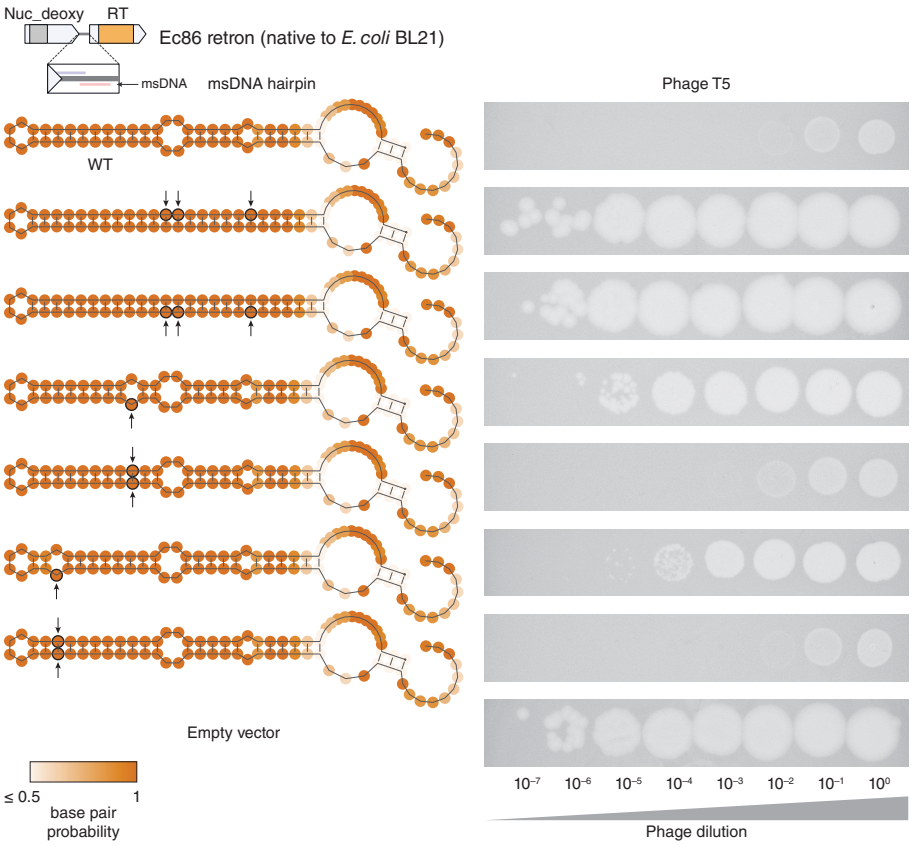

Figure S11: Effect of mutations in the multi-copy single-stranded DNA (msDNA) hairpin on defense activity for the Ec86 retron from *E. coli* BL21.

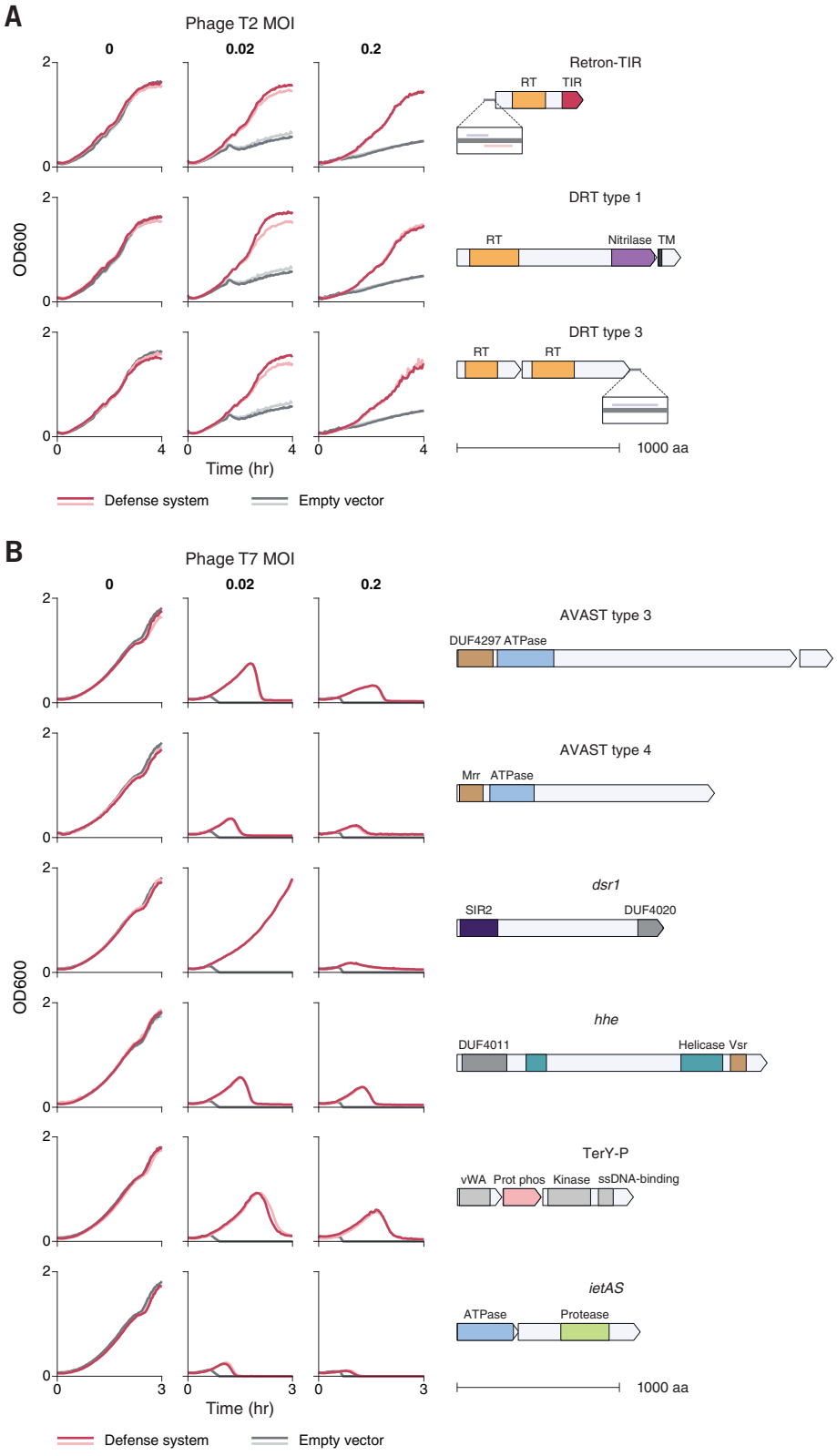

Figure S12: Bacterial densities over time for (**A**) retron-TIR, DRT type 1, and DRT type 3 defense systems infected with phage T2 and (**B**) additional defense systems infected with phage T7.

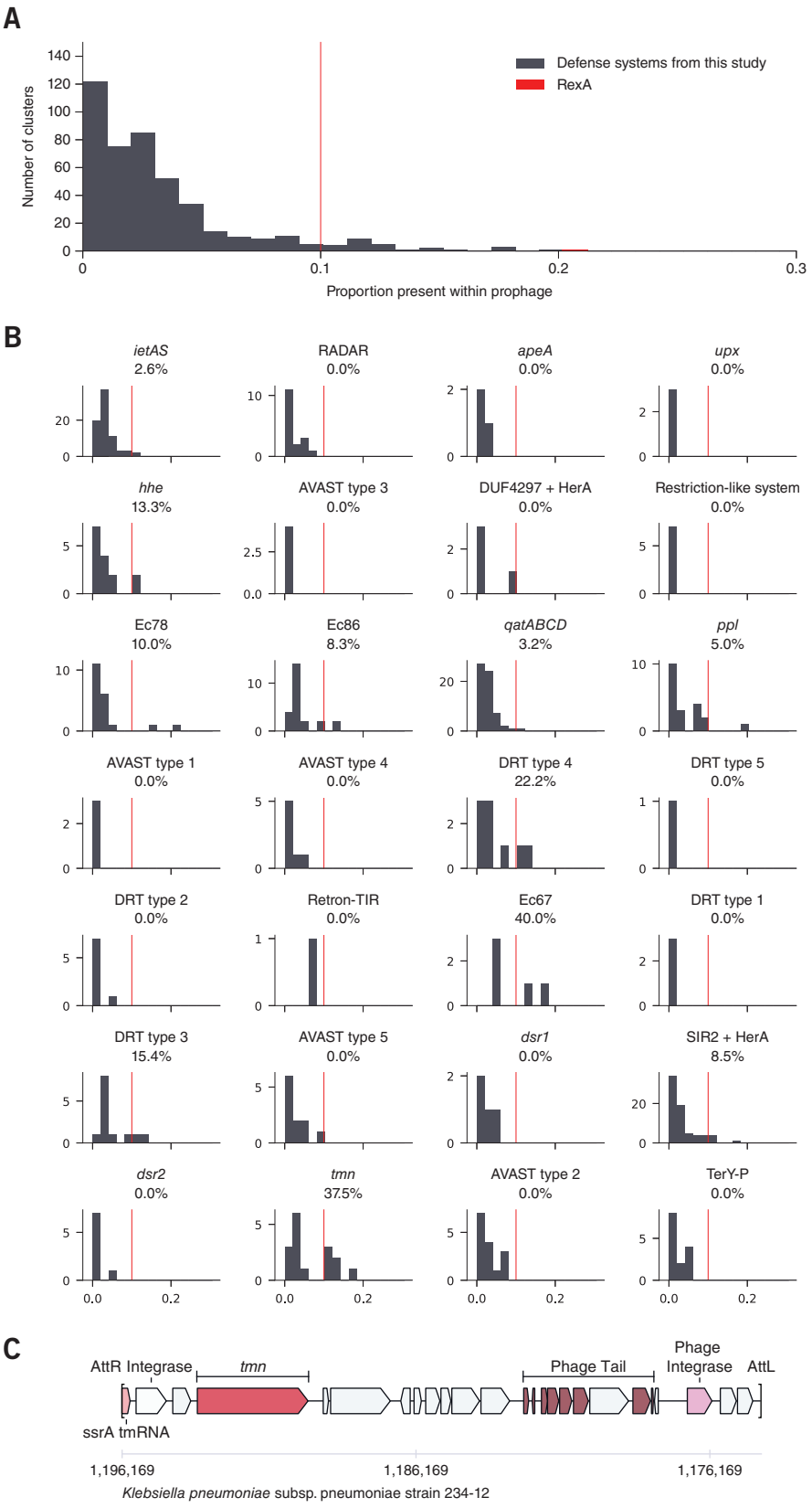

Figure S13: Phage and prophage association frequencies for validated defense system clusters. (**A**) Overall association frequency for 28 defense systems in this study. The *rexA* immunity gene from phage lambda is shown in red. (**B**) Per-system analysis of the distribution of phage association frequencies for each associated cluster in (**A**). (**C**) Example of the transmembrane ATPase (*tmn*) located within an incomplete prophage.

Table S1 (available online). Summary of predicted defense genes (n = 7,472).

Table S2 (available online). List of known defense-associated pfam domains used to tabulate Figure 1C.

Table S3 (available online). HHpred domain predictions (Pfam 32.0) of predicted defense genes.

Table S4. List of validated defense systems and their domain architectures.

#	WT	Mutants	Type	Name	Domain Architecture*
1	Fig. 2D	Fig. 4B	Retron	Retron-TIR	RT_retron-TIR
2	Fig. 2D	Fig. 4B	Retron	Ec67	RT_retron-TOPRIM
3	Fig. 2D	Fig. 4B	Retron	Ec86	Nuc_deoxy + RT_retron
4	Fig. 2D	fig. S9C	Retron	Ec78	RT_retron + ATPase_AAA + HNH
5	Fig. 2D	Fig. 4B	RT	DRT type 1	RT_UG1-nitrilase
6	Fig. 2D	fig. S9A	RT	DRT type 2	RT_UG2
7	Fig. 2D	Fig. 4B	RT	DRT type 3	RT_UG3 + RT_UG8
8	Fig. 2D	fig. S9B	RT	DRT type 4	RT_UG15
9	Fig. 2D	Fig. 4B	RT	DRT type 5	RT_UG16
10.A	Fig. 2D	Fig. 3B	RNA	RADAR	ATPase_AAA + ADA
10.B	Fig. 3B	Fig. 3B	RNA	RADAR	ATPase_AAA + ADA
11	Fig. 2D	Fig. 5	RNA	*apeA*	RNase_ApeA
12	Fig. 2D	Fig. 5	STAND	AVAST type 1	MBL + Protease_S1-ATPase_STAND
13	Fig. 2D	Fig. 5	STAND	AVAST type 2	ATPase_STAND
14	Fig. 2D	Fig. 5	STAND	AVAST type 3	Nuclease_DUF4297-ATPase_STAND
15	Fig. 2D	Fig. 5	STAND	AVAST type 4	Nuclease_Mrr-ATPase_STAND
16	Fig. 2D	Fig. 5	STAND	AVAST type 5	SIR2-ATPase_STAND
17	Fig. 2D	Fig. 5	Other	*dsr1*	SIR2-DUF4020
18	Fig. 2D	Fig. 5	Other	*dsr2*	SIR2
19	Fig. 2D	Fig. 5	Other	SIR2 + HerA	SIR2 + Helicase_HerA
20	Fig. 2D	Fig. 5	Other	DUF4297 + HerA	Nuclease_DUF4297 + Helicase_HerA
21	Fig. 2D	Fig. 5	Other	*tmn*	ATPase_AAA_TM
22	Fig. 2D	Fig. 5	Other	*qatABCD*	ATPase_AAA + QueC + DNase_TatD
23	Fig. 2D	Fig. 5	Other	*hhe*	HEPN_DUF4011-Helicase_SF1_Dna2-Nuclease_Vsr-DUF3320
24	Fig. 2D	--	Other	*mzaABCDE*	Ankyrin-sigma + ATPase_MutL + ATPase_AAA-Z1 + Nuclease_DUF4420 + AIPR
25	Fig. 2D	Fig. 5	Other	TerY-P	vWA + phosphatase_PP2C + STK-OB
26	Fig. 2D	Fig. 5	Other	*upx*	Nuclease_DUF1887
27	Fig. 2D	Fig. 5	Other	*ppl*	Phosphoesterase_PHP-ATPase_SMC
28	Fig. 2D	Fig. 5	Other	*ietAS***	ATPase_AAA + Protease_S8
29	Fig. 2D	Fig. 5	Other	Restriction-like system	ATPase_DUF499 + DUF3780 + Methylase_DUF1156 + Nuclease_PLD-Helicase_HepA

* Dashes (-) indicated domain fusions and (+) represents separate proteins.

** *ietAS* is also a previously-described plasmid stabilization toxin-antitoxin system (*60*).

Table S5. Source organism strains of validated defense systems and controls.

#	Source Organism	Strain	Promoter	Codon	Genes	bp
BREX type I	*Escherichia coli*	DSM5212	Native	Native	6	13703
Druantia type I	*Escherichia coli*	DSM5212	Native	Native	5	11823
RT-Abi-P2	*Escherichia coli*	ECOR30	Native	Native	1	1921
1	*Shigella dysenteriae*	NCTC2966	Native	Native	1	2064
2	*Escherichia coli*	NCTC8623	Native	Native	1	2038
3	*Escherichia coli*	BL21	Native	Native	2	2188
4	*Escherichia coli*	ECONIH5	Native	Native	3	3551
5	*Klebsiella pneumoniae*	NCTC9143	Native	Native	2	4451
6	*Salmonella enterica*	NCTC8273	Native	Native	1	1780
7	*Escherichia coli*	ECOR12	Native	Native	2	4995
8	*Escherichia coli*	21-C8-A	Native	Human	1	1838
9	*Escherichia coli*	KTE25	Native	Native	1	1608
10.A	*Citrobacter rodentium*	DBS100	Native	Native	2	5526
10.B	*Pluralibacter gergoviae*	ATCC33028	Native	Native	3	6689
11	*Escherichia coli*	NCTC8008	Native	Native	1	1981
12	*Erwinia piriflorinigrans*	CFBP5888	*bla*	Native	3	7246
13	*Escherichia coli*	NCTC9087	Native	Native	1	5109
14	*Salmonella enterica*	NCTC13175	Native	Native	2	7175
15	*Escherichia coli*	NCTC11132	Native	Native	1	4964
16	*Escherichia coli*	NCTC13384	Native	Native	1	3411
17	*Escherichia coli*	NCTC9112	Native	Native	1	4212
18	*Cronobacter sakazakii*	NCTC8155	Native	Native	1	4329
19	*Escherichia coli*	NCTC11129	Native	Native	2	3308
20	*Escherichia coli*	NCTC11131	Native	Native	2	3419
21	*Escherichia coli*	ECOR25	Native	Native	1	4415
22	*Escherichia coli*	NCTC9009	Native	Native	4	5408
23	*Escherichia coli*	ATCC43886	Native	Native	1	5958
24	*Salmonella enterica*	NCTC5773	Native	Native	5	9416
25	*Citrobacter gillenii*	NCTC9094	Native	Native	3	3605
26	*Salmonella enterica*	NCTC6026	Native	Native	1	4100
27	*Escherichia coli*	NCTC8620	Native	Native	1	3066
28	*Escherichia coli*	ECOR52	Native	Native	2	3676
29	*Escherichia coli*	ECOR58	Native	Native	4	9809

Table S6. PCR primers used to amplify validated defense systems and controls.

#	Sequence	
BREX type I	Fwd	gctaacttacattaattgcgttgcgcaACAGCACCACGTTCATCTTCC
	Rev	ccaaggggttatgctagttattgcgGTTCATTAAAATAGTTACTACGTTAATTCACACCC
Druantia type I	Fwd	gctaacttacattaattgcgttgcgcaGGTGAACGTTTGGTTGATAGGG
	Rev	ccaaggggttatgctagttattgcgCTCAATGGGCATAATTTTACATTGTGC
RT-Abi-P2	Fwd	gctaacttacattaattgcgttgcgcaACATCCCGTCATCATGCCATC
	Rev	ccaaggggttatgctagttattgcgCTCCTCGGAATAGAATGTTATGTTCG
1	Locus synthesized	
2	Fwd	gctaacttacattaattgcgttgcgcaCGCGCTATCACGTAAAATAGGC
	Rev	ccaaggggttatgctagttattgcgCGAAAAATCAGCCTTAGCGTTCATAAC
3	Fwd	gctaacttacattaattgcgttgcgcaGCTCATGTTATGCATGTGCATG
	Rev	ccaaggggttatgctagttattgcgATTAGGTCTTCGCTTTATTTAAAGGGTTC
4	Locus synthesized	
5	Fwd	gagctaacttacattaattgcgttgcgcaGTCCTTAAACACGACAAAACCTGTG
	Rev	cccaaggggttatgctagttattgcgCGCAATGTAACACCCACCC
6	Locus synthesized	
7	Fwd	gctaacttacattaattgcgttgcgcaTCTCAACTTCCCCAAATGTCCG
	Rev	cccaaggggttatgctagttattgcgTTAGCAAAATACGCCCACGAAGTC
8	Locus synthesized	
9	Locus synthesized	
10.A	Fwd	gctaacttacattaattgcgttgcgcaGAGGATTTATGCACAAAATCCTGATGC
	Rev	ccaaggggttatgctagttattgcgGATTTAATCTGTTGTTCCGAACGG
10.B	Fwd	gctaacttacattaattgcgttgcgcaTGTGGTTAGTTATCACAGCACTAACC
	Rev	ccaaggggttatgctagttattgcgGTGTATAAGAATCCGAGACCGAAC
11	Locus synthesized	
12	Fwd	ataaatgcttcaataatattgaaaaaggaagagtATGGTAGCGATAAAAATGTATCCGGC
	Rev	cccaaggggttatgctagttattgcgTCAATCCGTAGCCTCTTCATTCTCG
13	Fwd	gctaacttacattaattgcgttgcgcaGGGATTTCCACCACCTCCC
	Rev	ccaaggggttatgctagttattgcgTGCATAGCCAATGAAGATAAACGTG
14	Fwd	gctaacttacattaattgcgttgcgcaACAATTTTTTGCCATAAGACGCTTTC
	Rev	ccaaggggttatgctagttattgcgCATTAGGACTAGTAGAAAAGTCTTGGG
15	Fwd	gctaacttacattaattgcgttgcgcaGCGCAGCTGACAAAGATTGAC
	Rev	ccaaggggttatgctagttattgcgCGATAATAAAAAGGCTCCAATCCCTG
16	Fwd	gctaacttacattaattgcgttgcgcaACTAGCTAAGCAATAAGGGCG
	Rev	ccaaggggttatgctagttattgcgCAATCTCCGAGGTGGCCC
17	Fwd	gctaacttacattaattgcgttgcgcaTATTTTGCGTAGCTAGAACGCAATC
	Rev	ccaaggggttatgctagttattgcgTGGGTATTAGCTCATATCAGAACTAATACCC
18	Fwd	gctaacttacattaattgcgttgcgcaGTAAGACAAGGGTTGAGCAGGC
	Rev	ccaaggggttatgctagttattgcgCAATGGTGGGCTGATTAATTAGATGAG
19	Fwd	gctaacttacattaattgcgttgcgcaTAGCTATTGTGACTATGCTAACCATATG
	Rev	ccaaggggttatgctagttattgcgTTCAGTCTAAATACATACCTGTCGGG

20	Fwd	gctaacttacattaattgcgttgcgcaGTGCGCCTTATGTGATTACAACG
	Rev	ccaaggggttatgctagttattgcgCTCTCAGCCTAATGATTCCAGAATAG
21	Fwd	gctaacttacattaattgcgttgcgcaACCGTGCTGGCATGTTTTTAC
	Rev	ccaaggggttatgctagttattgcgAGGAAGATCCGTGACCAGGAG
22	Fwd	gctaacttacattaattgcgttgcgcaGAAATTATTTGGAATGGATGATGGCG
	Rev	ccaaggggttatgctagttattgcgACTTCTACCTCCCTTTAGAAAAGTTAATG
23	Fwd	gctaacttacattaattgcgttgcgcaCGGATTGAATCTGTTTATGAAATTTGGCTG
	Rev	ccaaggggttatgctagttattgcgCCGACAGTTGTCACTGTTCTTATTACC
24	Fwd	tgagctaacttacattaattgcgttgcgcaATGATGAAGATCACCTAAAATGATAGGTTG
	Rev	cccaaggggttatgctagttattgcgCAGCTGTTAATTGTATATTGATGCGATGC
25	Fwd	gctaacttacattaattgcgttgcgcaCGTGATGAATGAAGCGGCTAAATAC
	Rev	ccaaggggttatgctagttattgcgGTAAATCCTCGGGAAAACACAGG
26	Fwd	gctaacttacattaattgcgttgcgcaGGGCTGTTTGGTTGAATTAAAAATACG
	Rev	ccaaggggttatgctagttattgcgCCTTGATTTAAAACTATCAGTAGTAGGAACG
27	Fwd	gctaacttacattaattgcgttgcgcaGATGGACTGGTACTGTAGATTCACC
	Rev	ccaaggggttatgctagttattgcgCAAAGACGCAGAGGCCATCAG
28	Fwd	gctaacttacattaattgcgttgcgcaATAGAACGATGAAGGATGGAAGCTAC
	Rev	ccaaggggttatgctagttattgcgTTGTATTTTGTTGTGTATGGGCGG
29	Fwd	gctaacttacattaattgcgttgcgcaCGTGATTCAGTTCGCCAGAC
	Rev	ccaaggggttatgctagttattgcgCACTCGAAATGGATACCCTGAG

Table S7. Protein accession numbers of defense system components (proposed gene names underlined).

#	Gene	Name	Protein Accession	#	Gene	Name	Protein Accession
BREX type I	A	*brxA*	WP_085962535.1*	12	C	*avs1c*	WP_023654316.1
	B	*brxB*	WP_000566901.1	13	A	*avs2*	WP_063118745.1
	C	*brxC*	WP_001019648.1	14	A	*avs3a*	WP_126523998.1
	D	*pglX*	WP_021524842.1		B	*avs3b*	WP_126523997.1*
	E	*pglZ*	WP_001180895.1	15	A	*avs4*	WP_044068927.1
	F	*brxL*	WP_001193074.1	16	A	*avs5*	WP_001515187.1
Druantia type I	A	*druA*	WP_000549798.1	17	A	*dsr1*	WP_029488749.1
	B	*druB*	WP_001315973.1	18	A	*dsr2*	WP_015387030.1*
	C	*druC*	WP_021520530.1	19	A	–	WP_021577683.1
	D	*druD*	WP_000455180.1		B	*herA*	WP_021577682.1
	E	*druE*	WP_000608843.1	20	A	–	WP_016239654.1
RT-Abi-P2	A	–	WP_047657908.1		B	*herA*	WP_016239655.1
1	A	–	WP_005025120.1*	21	A	*tmn*	WP_001683567.1
2	A	Ec67	WP_000169432.1	22	A	*qatA*	STG85056.1
3	A	–	WP_001034589.1		B	*qatB*	STG85057.1
	B	Ec86	WP_001320043.1		C	*qatC*	STG85058.1
4	A	Ec78	WP_001549208.1		D	*qatD*	STG85059.1
	B	*ptuA*	WP_001549209.1	23	A	*hhe*	WP_032200272.1
	C	*ptuB*	WP_001549210.1	24	A	*mzaA*	VEA06816.1*
5	A	*drt1a*	WP_115196278.1		B	*mzaB*	VEA06814.1
	B	*drt1b*	WP_040189938.1		C	*mzaC*	VEA06812.1
6	A	*drt2*	WP_012737279.1		D	*mzaD*	VEA06810.1
7	A	*drt3a*	WP_087902017.1		E	*mzaE*	VEA06808.1
	B	*drt3b*	WP_062891751.1	25	A	*terY*	WP_115257868.1
8	A	*drt4*	GCK53192.1		B	–	WP_115257869.1
9	A	*drt5*	WP_001524904.1		C	–	WP_115257870.1
10.A	A	*rdrA*	WP_012906049.1	26	A	*upx*	WP_060647174.1
	B	*rdrB*	WP_012906048.1	27	A	*ppl*	STM52149.1
10.B	A	*rdrA*	WP_155731552.1	28	A	*ietA*	WP_000385105.1
	B	*rdrB*	WP_064360593.1		B	*ietS*	WP_001551050.1
	C	*rdrD*	WP_064360592.1	29	A	–	WP_000860009.1
11	A	*apeA*	WP_000706972.1		B	–	WP_001044652.1
12	A	*avs1a*	WP_023654314.1		C	–	WP_001207938.1
	B	*avs1b*	WP_084007836.1*		D	–	WP_000985714.1

* Probable error in annotated protein start position corrected.

Table S8. Predicted protein domains within validated defense systems and controls. Transmembrane helices were predicted using TMHMM, and all other domains were predicted using HHpred.

ID	Gene	Residues	Domain	Representative HHpred Hit	Probability	Start	End
BREX type I	A	201	DUF1819	PF08849.11	100	6	189
	B	200	DUF1788	PF08747.11	100	65	187
	C	1213	ATPase	PF07693.14	96.66	43	348
			DUF499	PF04465.12	99.88	247	846
	D	1201	Methyltransferase	PF02384.16	99.7	210	622
	E	865	PglZ	PF08665.12	99.12	474	650
	F	694	Lon protease	PF13337.6	100	30	484
			Lon protease	PF05362.13	99.9	486	693
Druantia type I	A	404	DUF4338	PF14236.6	99.92	45	339
	B	548	CoiA	PF06054.11	99.77	1	182
	C	627	Macoilin	PF09726.9	96.72	167	323
	D	347	(none)	--	--	--	--
	E	1836	Helicase	PF00270.29	98.45	99	388
			Helicase	5V9X_A	97.55	1071	1208
			DUF1998	PF09369.10	98.92	1626	1710
RT-Abi-P2	A	515	RT	PF00078.27	99.09	68	291
1	A	542	RT	PF00078.27	99.43	105	309
			TIR	PF13676.6	97.91	411	536
2	A	586	RT	PF00078.27	99.45	48	262
			TOPRIM	cd01026	96.88	367	465
3	A	307	Nuc_deoxy	PF15891.5	96.04	29	128
	B	320	RT	PF00078.27	99.52	53	248
4	A	311	RT	PF00078.27	99.37	34	241
	B	550	ATPase	PF13175.6	99.8	64	432
	C	216	HNH	PF01844.23	97.57	43	85
5	A	1232	RT	PF00078.27	99.06	80	382
			Nitrilase	PF00795.22	98.89	953	1216
	B	144	Transmembrane	--	--	4	26
6	A	425	RT	PF00078.27	99.63	54	328
7	A	398	RT	PF00078.27	99.39	53	251
	B	667	RT	PF00078.27	98.96	63	323
8	A	540	RT	PF00078.27	99.12	67	296
9	A	494	RT	PF00078.27	99.14	59	263
10.A	A	851	ATPase	PF07693.14	99.6	33	364
	B	856	Adenosine deaminase	PF00962.22	99.52	166	831
10.B	A	907	ATPase	PF07693.14	99.48	29	349
	B	914	Adenosine deaminase	PF00962.22	97.63	789	901

	C	245	SLATT	PF18183.1	96.01	120	241
			Transmembrane	--	--	44	63
			Transmembrane	--	--	78	100
			Transmembrane	--	--	127	146
			Transmembrane	--	--	151	168
11	A	601	HEPN	PF18739.1	86.57	507	532
12	A	386	MBL-fold hydrolase	PF00753.27	98.79	8	324
	B	1935	Protease	PF02122.15	98.23	2	187
			ATPase	PF14516.6	99.36	204	535
	C	93	(none)	--	--	--	--
13	A	1484	ATPase	PF14516.6	98.93	316	643
14	A	2092	DUF4297	PF14130.6	98.41	8	223
			ATPase	PF14516.6	99.44	250	597
	B	207	(none)	--	--	--	--
15	A	1587	Mrr	PF13156.6	97.05	17	162
			ATPase	PF14516.6	99.07	204	476
16	A	769	SIR2	cd00296	99.26	22	244
			ATPase	PF14516.6	97.6	312	464
17	A	1275	SIR2	cd00296	99.44	21	253
			DUF4020	PF13212.6	98.39	1114	1268
18	A	1207	SIR2	cd00296	99.47	21	240
19	A	415	SIR2	cd00296	99.59	26	338
	B	610	HerA helicase	4D2I_B	100	10	608
20	A	394	DUF4297	PF14130.6	99.05	1	191
	B	571	HerA helicase	4D2I_B	100	7	568
21	A	1273	ATPase	PF07693.14	97.62	39	390
			Transmembrane	--	--	160	177
			Transmembrane	--	--	199	218
22	A	643	ATPase	PF07693.14	99.8	15	385
	B	274	(none)	--	--	--	--
	C	457	QueC	PF06508.13	99.67	150	369
	D	263	TatD DNase	PF01026.21	99.94	13	254
23	A	1911	DUF4011	PF13195.6	99.81	33	308
			ATPase	PF13086.6	97.93	427	552
			Helicase	PF01443.18	97.82	1379	1636
			Endonuclease	PF18741.1	98.7	1683	1780
			DUF3320	PF11784.8	98.1	1841	1885
24	A	679	Ankyrin repeat	COG0666	99.52	10	188
			Sigma	COG1191	99.81	411	657
	B	500	MutL	COG0323	99.81	1	352
	C	952	ATPase	PF13872.6	97.51	117	349
			Z1	PF10593.9	100	437	672
	D	342	DUF4420	PF14390.6	100	9	317

	E	601	AIPR	PF10592.9	100	245	562
25	A	277	vWA	PF00092.28	98.93	14	203
	B	239	Phosphatase	PF00481.21	99.74	5	232
	C	561	Kinase	PF00069.25	100	34	296
			ssDNA-binding	PF01336.25	96.18	344	435
26	A	1272	DUF1887	PF09002.11	92.5	1105	1272
27	A	891	PHP	cd07436	99.36	4	238
			ATPase	PF13166.6	99.74	266	836
28	A	384	ATPase	PF13654.6	97.36	5	349
	B	754	Protease	PF00082.22	99.87	264	561
29	A	1022	ATPase	PF07693.14	96.47	49	312
			DUF499	PF04465.12	100	79	745
	B	195	DUF3780	PF12635.7	100	1	187
	C	945	DUF1156	PF06634.12	99	18	81
			Methyltransferase	PF01555.18	96.08	150	202
			Methyltransferase	PF01555.18	97.76	548	682
	D	907	PLD	cd09179	99.17	4	177
			Helicase	6BOG_B	100	218	865

Table S9. Sequence of vector backbone. Inserts were cloned between the HindIII and EcoRI restriction sites (underlined).

CCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC
CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGC**AAGCTT**CTGCA**GAATTC**GC
AATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAACCT
CAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGC
AATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTT
TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCACCAGGCGTTTAAGG
GCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAATACTGTTGTA
ATTCATTTAACATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCC
AGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATAGTGAAAACGGGGGCGAA
GAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTG
AGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACAC
GCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAG
CGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA
TCACCAGCTCACCGTCTTTCATCGCCATACGGAACTCTGGATGAGCATTCATCAGGCGGGCA
AGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGC
CGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAA
AATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCA
TTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTA
TTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGT
TGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCT
TCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTA
GTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTC
AGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGT
GTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGA
GAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCG
CTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGG
CGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAA
AGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAG
TGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGT
GCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCT
CATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACG
AACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG
GAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTC
TTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCC
AAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCA
AGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGAT
CATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCA
CCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGC

Table S10 (available online). Sequences of validated defense systems.

Table S11 (available online). Additional tested homologs of predicted defense systems.

Table S12 (available online). Genome coordinates of RADAR editing sites in Figure S7.

Table S13 (available online). Description of phage T2 fragments in Figure S8C-E.

References and Notes

1. C. A. Suttle, Viruses: Unlocking the greatest biodiversity on Earth. *Genome* **56**, 542–544 (2013). doi:10.1139/gen-2013-0152 Medline

2. A. G. Cobián Güemes, M. Youle, V. A. Cantú, B. Felts, J. Nulton, F. Rohwer, Viruses as Winners in the Game of Life. *Annu. Rev. Virol.* **3**, 197–214 (2016). doi:10.1146/annurev-virology-100114-054952 Medline

3. F. Hille, H. Richter, S. P. Wong, M. Bratovič, S. Ressel, E. Charpentier, The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018). doi:10.1016/j.cell.2017.11.032 Medline

4. S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018). doi:10.1126/science.aar4120 Medline

5. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013). doi:10.1038/nrmicro3096 Medline

6. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013). doi:10.1038/nature11723 Medline

7. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41**, 4360–4377 (2013). doi:10.1093/nar/gkt157 Medline

8. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: Guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020). doi:10.1038/s41576-019-0172-9 Medline

9. G. Faure, S. A. Shmakov, W. X. Yan, D. R. Cheng, D. A. Scott, J. E. Peters, K. S. Makarova, E. V. Koonin, CRISPR-Cas in mobile genetic elements: Counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019). doi:10.1038/s41579-019-0204-7 Medline

10. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011). doi:10.1128/JB.05535-11 Medline

11. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5307–E5316 (2018). doi:10.1073/pnas.1803440115 Medline

12. S. A. Shmakov, G. Faure, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of functionally linked genes in bacterial and archaeal genomes. *Nat. Protoc.* **14**, 3013–3031 (2019). doi:10.1038/s41596-019-0211-1 Medline

13. J. Gordeeva, N. Morozova, N. Sierro, A. Isaev, T. Sinkunas, K. Tsvetkova, M. Matlashov, L. Truncaitė, R. D. Morgan, N. V. Ivanov, V. Siksnys, L. Zeng, K. Severinov, BREX

system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res.* **47**, 253–265 (2019). doi:10.1093/nar/gky1125 Medline

14. T. Goldfarb, H. Sberro, E. Weinstock, O. Cohen, S. Doron, Y. Charpak-Amikam, S. Afik, G. Ofir, R. Sorek, BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015). doi:10.15252/embj.201489455 Medline

15. R. Odegrip, A. S. Nilsson, E. Haggård-Ljungquist, Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. *J. Bacteriol.* **188**, 1643–1647 (2006). doi:10.1128/JB.188.4.1643-1647.2006 Medline

16. A. M. Burroughs, D. Zhang, D. E. Schäffer, L. M. Iyer, L. Aravind, Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015). doi:10.1093/nar/gkv1267 Medline

17. K. S. Makarova, L. Gao, F. Zhang, E. V. Koonin, Unexpected connections between type VI-B CRISPR-Cas systems, bacterial natural competence, ubiquitin signaling network and DNA modification through a distinct family of membrane proteins. *FEMS Microbiol. Lett.* **366**, fnz088 (2019). doi:10.1093/femsle/fnz088 Medline

18. C. D. Rae, Y. Gordiyenko, V. Ramakrishnan, How a circularized tmRNA moves through the ribosome. *Science* **363**, 740–744 (2019). doi:10.1126/science.aav9370 Medline

19. S. Zimmerly, L. Wu, An Unexplored Diversity of Reverse Transcriptases in Bacteria. *Microbiol. Spectr.* **3**, A3–A0058, 2014 (2015). 10.1128/microbiolspec.MDNA3-0058-2014 Medline

20. N. Toro, R. Nisa-Martínez, Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLOS ONE* **9**, e114083 (2014). doi:10.1371/journal.pone.0114083 Medline

21. K. K. Kojima, M. Kanehisa, Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. *Mol. Biol. Evol.* **25**, 1395–1404 (2008). doi:10.1093/molbev/msn081 Medline

22. D. M. Simon, S. Zimmerly, A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res.* **36**, 7219–7229 (2008). doi:10.1093/nar/gkn867 Medline

23. H. C. Pace, C. Brenner, The nitrilase superfamily: Classification, structure and function. *Genome Biol.* **2**, S0001 (2001). doi:10.1186/gb-2001-2-1-reviews0001 Medline

24. A. J. Simon, A. D. Ellington, I. J. Finkelstein, Retrons and their applications in genome engineering. *Nucleic Acids Res.* **47**, 11007–11019 (2019). doi:10.1093/nar/gkz865 Medline

25. F. Farzadfard, T. K. Lu, Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014). doi:10.1126/science.1256272 Medline

26. B. C. Lampson, J. Sun, M. Y. Hsu, J. Vallejo-Ramirez, S. Inouye, M. Inouye, Reverse transcriptase in a clinical strain of Escherichia coli: Production of branched RNA-linked msDNA. *Science* **243**, 1033–1038 (1989). doi:10.1126/science.2466332 Medline

27. D. Lim, W. K. Maas, Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in E. coli B. *Cell* **56**, 891–904 (1989). doi:10.1016/0092-8674(89)90693-4 Medline

28. T. M. Lima, D. Lim, A novel retron that produces RNA-less msDNA in Escherichia coli using reverse transcriptase. *Plasmid* **38**, 25–33 (1997). doi:10.1006/plas.1997.1298 Medline

29. L. Aravind, D. D. Leipe, E. V. Koonin, Toprim—A conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205–4213 (1998). doi:10.1093/nar/26.18.4205 Medline

30. S. Horsefield, H. Burdett, X. Zhang, M. K. Manik, Y. Shi, J. Chen, T. Qi, J. Gilley, J.-S. Lai, M. X. Rank, L. W. Casey, W. Gu, D. J. Ericsson, G. Foley, R. O. Hughes, T. Bosanac, M. von Itzstein, J. P. Rathjen, J. D. Nanson, M. Boden, I. B. Dry, S. J. Williams, B. J. Staskawicz, M. P. Coleman, T. Ve, P. N. Dodds, B. Kobe, NAD^+ cleavage activity by animal and plant TIR domains in cell death pathways. *Science* **365**, 793–799 (2019). doi:10.1126/science.aax1911 Medline

31. V. Anantharaman, L. M. Iyer, L. Aravind, Ter-dependent stress response systems: Novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. *Mol. Biosyst.* **8**, 3142–3165 (2012). doi:10.1039/c2mb25239b Medline

32. K. S. Makarova, Y. I. Wolf, J. van der Oost, E. V. Koonin, Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29 (2009). doi:10.1186/1745-6150-4-29 Medline

33. V. Anantharaman, K. S. Makarova, A. M. Burroughs, E. V. Koonin, L. Aravind, Comprehensive analysis of the HEPN superfamily: Identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol. Direct* **8**, 15 (2013). doi:10.1186/1745-6150-8-15 Medline

34. L. Aravind, L. M. Iyer, D. D. Leipe, E. V. Koonin, A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. *Genome Biol.* **5**, R30 (2004). doi:10.1186/gb-2004-5-5-r30 Medline

35. D. D. Leipe, E. V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28 (2004). doi:10.1016/j.jmb.2004.08.023 Medline

36. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. *Structure* **17**, 172–182 (2009). doi:10.1016/j.str.2009.01.001 Medline

37. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: The bacterial connection. *Cell Death Differ.* **9**, 394–404 (2002). doi:10.1038/sj.cdd.4400991 Medline

38. A. Bernheim, R. Sorek, The pan-immune system of bacteria: Antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020). doi:10.1038/s41579-019-0278-2 Medline

39. D. Hyatt, G.-L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010). doi:10.1186/1471-2105-11-119 Medline

40. M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, R. D. Finn, The Pfam protein families database. *Nucleic Acids Res.* **40** (D1), D290–D301 (2012). doi:10.1093/nar/gkr1065 Medline

41. A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, L. Y. Geer, S. H. Bryant, CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45** (D1), D200–D203 (2017). doi:10.1093/nar/gkw1129 Medline

42. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017). doi:10.1038/nbt.3988 Medline

43. M. Steinegger, J. Söding, Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9**, 2542 (2018). doi:10.1038/s41467-018-04964-5 Medline

44. R. J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* **43** (D1), D298–D299 (2015). doi:10.1093/nar/gku1046 Medline

45. D. Cohen, S. Melamed, A. Millman, G. Shulman, Y. Oppenheimer-Shaanan, A. Kacen, S. Doron, G. Amitai, R. Sorek, Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019). doi:10.1038/s41586-019-1605-5 Medline

46. G. Ofir, S. Melamed, H. Sberro, Z. Mukamel, S. Silverman, G. Yaakov, S. Doron, R. Sorek, DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018). doi:10.1038/s41564-017-0051-0 Medline

47. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline

48. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018). doi:10.1016/j.jmb.2017.12.007 Medline

49. J. C. Petricciani, F. C. Chu, J. B. Johnson, H. M. Meyer Jr., Bacteriophages in live virus vaccines. *Proc. Soc. Exp. Biol. Med.* **144**, 789–791 (1973). doi:10.3181/00379727-144-37683 Medline

50. J. B. Milstien, J. R. Walker, J. C. Petricciani, Bacteriophages in live virus vaccines: Lack of evidence for effects on the genome of rhesus monkeys. *Science* **197**, 469–470 (1977). doi:10.1126/science.406673 Medline

51. B. Xu, X. Ma, H. Xiong, Y. Li, Complete genome sequence of 285P, a novel T7-like polyvalent E. coli bacteriophage. *Virus Genes* **48**, 528–533 (2014). doi:10.1007/s11262-014-1059-7 Medline

52. S. Picelli, Å. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014). doi:10.1101/gr.177881.114 Medline

53. E. S. Miller, E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, W. Rüger, Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**, 86–156 (2003). doi:10.1128/MMBR.67.1.86-156.2003 Medline

54. D. H. Turner, D. H. Mathews, NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* **38** (suppl_1), D280–D282 (2010). doi:10.1093/nar/gkp892 Medline

55. Y. Zhou, Y. Liang, K. H. Lynch, J. J. Dennis, D. S. Wishart, PHAST: A fast phage search tool. *Nucleic Acids Res.* **39** (suppl.), W347–W352 (2011). doi:10.1093/nar/gkr485 Medline

56. D. Arndt, J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, D. S. Wishart, PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44** , W16–W21 (2016). doi:10.1093/nar/gkw387 Medline

57. J. Strecker, A. Ladha, Z. Gardner, J. L. Schmid-Burgk, K. S. Makarova, E. V. Koonin, F. Zhang, RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019). doi:10.1126/science.aax9181 Medline

58. S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019). doi:10.1038/s41586-019-1323-z Medline

59. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017). doi:10.1146/annurev-micro-090816-093830 Medline

60. S. Yamamoto, K. Kiyokawa, K. Tanaka, K. Moriguchi, K. Suzuki, Novel toxin-antitoxin system composed of serine protease and AAA-ATPase homologues determines the high level of stability and incompatibility of the tumor-inducing plasmid pTiC58. *J. Bacteriol.* **191**, 4656–4666 (2009). doi:10.1128/JB.00124-09 Medline