Nucleic Acids Research, 2025, **53**, gkae1171
https://doi.org/10.1093/nar/gkae1171
Advance access publication date: 9 December 2024
Genomics

Exploring the diversity of anti-defense systems across prokaryotes, phages and mobile genetic elements

Florian Tesson[1,†], Erin Huiting[2,†], Linlin Wei[3], Jie Ren[4], Matthew Johnson[2], Rémi Planel[5], Jean Cury[1], Yue Feng[3], Joseph Bondy-Denomy[2,6,*] and Aude Bernheim[1,*]

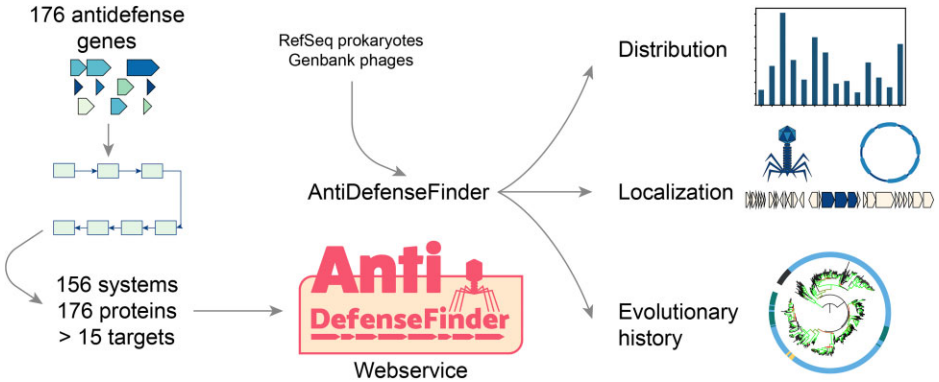
[1]Institut Pasteur, CNRS UMR3525, Molecular Diversity of Microbes Lab, 25-28 rue du Docteur Roux, 75015, Paris, France
[2]Department of Microbiology and Immunology, University of California San Francisco, Genentech Hall Room N372E UCSF Mail Code 2200 600 16th Street San Francisco, San Francisco, CA 94158, USA
[3]State Key Laboratory of Chemical Resource Engineering, Beijing Key Laboratory of Bioprocess, College of Life Science and Technology, Beijing University of Chemical Technology, Beisanhuan EastRoad 15, Chaoyang Distract, 100029 Beijing, China
[4]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Ministry of Agriculture, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, 2 Yuanmingyuan W Rd, Haidian District, 100091 Beijing, China
[5]Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, 25-28 rue du Docteur Roux, 75015, Paris, France
[6]Quantitative Biosciences Institute, University of California San Francisco, 1700 4th St, San Francisco, CA 94158, USA

[*]To whom correspondence should be addressed. Tel: +33 695574972; Email: aude.bernheim@pasteur.fr
Correspondence may also be addressed to Joseph Bondy-Denomy. Tel: +1 415 514 1381; Email: joseph.bondy-denomy@ucsf.edu
[†]The first two authors should be regarded as Joint First Authors.

Abstract

The co-evolution of prokaryotes, phages and mobile genetic elements (MGEs) has driven the diversification of defense and anti-defense systems alike. Anti-defense proteins have diverse functional domains, sequences and are typically small, creating a challenge to detect anti-defense homologs across prokaryotic and phage genomes. To date, no tools comprehensively annotate anti-defense proteins within a desired sequence. Here, we developed 'AntiDefenseFinder'—a free open-source tool and web service that detects 156 anti-defense systems of one or more proteins in any genomic sequence. Using this dataset, we identified 47 981 anti-defense systems distributed across prokaryotes and their viruses. We found that some genes co-localize in 'anti-defense islands', including *Escherichia coli* T4 and Lambda phages, although many appear standalone. Eighty-nine per cent anti-defense systems localize only or preferentially in MGE. However, >80% of anti-Pycsar protein 1 (Apyc1) resides in nonmobile regions of bacterial genomes. Evolutionary analysis and biochemical experiments revealed that Apyc1 likely originated in bacteria to regulate cyclic nucleotide (cNMP) signaling, but phage co-opted Apyc1 to overcome cNMP-utilizing defenses. With the AntiDefenseFinder tool, we hope to facilitate the identification of the full repertoire of anti-defense systems in MGEs, the discovery of new protein functions and a deeper understanding of host–pathogen arms race.

Graphical abstract



Introduction

In the past 6 years, there have been >100 newly identified systems in prokaryotes that defend against phages (1). Several studies have revealed mechanistic diversity of defense systems, spanning nucleic acid (2–9) or metabolite (10–16) depletion, signaling molecule cascades (17–20) membrane disruption (21–23) and many more (24–26). To counteract these systems, phages evolved a diversity of anti-defense systems

Received: August 9, 2024. **Revised:** October 30, 2024. **Editorial Decision:** November 6, 2024. **Accepted:** December 6, 2024

that directly inhibit individual defense proteins (7,10,26–29) or signaling molecules (29–36) or indirectly inhibit these systems through reversing the depletion of host metabolites (37).

To date, the most well-studied anti-defense strategies are anti-Restriction-Modification (anti-RM) and anti-CRISPR (Acr) proteins that provide protection against nucleic acid targeting systems. These proteins have been extensively studied in phages, prophages (38,39), plasmids (40) and conjugative elements (39). In certain cases, mobile genetic element (MGE)-encoding Acr proteins that inhibit Type III CRISPR-Cas systems (41) have been co-opted by the bacterial host to regulate the Type III CRISPR-Cas activity (42). Beyond inhibitors of CRISPR-Cas and RM, the distribution and localization of other anti-defense systems remain vastly understudied. The main challenge in identifying anti-defense proteins is due to the vast diversity of the functional domains and the often small protein size (i.e. 80% of anti-defense proteins are smaller than 200 amino acids), a bottleneck for both sequence and structure-based detection.

To address this, we built upon the established DefenseFinder (1,43,44) search tool and web service to detect all known anti-defense systems in prokaryotic and phage genomes. Since the discovery of the first anti-restriction protein (45), there have been at least 180 proteins identified to inhibit prokaryotic defense systems. A pre-computed database of 41 experimentally validated anti-defense systems (i.e. dbAPIS) was recently published that identified 4428 homologs of anti-defense systems in phages (46). Our newly developed AntiDefenseFinder tool can detect 156 anti-defense systems, in which some are composed of multiple proteins, like NARP1 and NARP2. We also grouped some independently identified anti-defense genes that have proven to be close homologs, like Thoeris anti-defense 2 (Tad2) and AcrIIA7. When applied to the RefSeq database of 21 855 prokaryotic complete genomes and from the GenBank database of 13 487 phage sequences, it detects 41 972 and 6009 anti-defense systems in prokaryotic and phage genomes, respectively. Alongside this comprehensive dataset, the search tool is available on a freely accessible web service and via command line, which we hope will facilitate the identification of anti-defense genes within any DNA or protein sequences.

We found that most anti-defense systems are variable in frequency and distribution across prokaryotic species. We observed several instances of anti-defense genes co-localizing into 'anti-defense islands', including the model *Escherichia coli* T4 and Lambda phages. In some cases, these anti-defense islands contain only anti-defense genes from a single family, such as Acrs, anti-Gabija or anti-Thoeris. We also observed that many anti-defense genes tend to be encoded alone across a combination of prophages, plasmids, phage satellites, integrons and integrative and conjugative elements. We also identified that NAD^+ reconstitution pathway 1 and 2 (NARP1/2) and anti-Pycsar gene 1 (*apyc1*) genes are predominantly identified in sequences within the bacterial chromosome. Based on our evolutionary and functional analyses, we propose that Apyc1 homologs are common in prokaryotic genomes to regulate housekeeping signals, such as cyclic AMP (cAMP), but this enzyme family can cleave any cyclic nucleotide (cNMP). Therefore, this cNMP-cleaving protein was co-opted by phages to counteract defense systems using cyclic CMP (cCMP) and cyclic UMP (cUMP). We anticipate that AntiDefenseFinder can be used by groups in the future to discover new inhibitor proteins and dissect evolutionary and functional relationships between inhibitors, their hosts and the defense they antagonize, as executed here with Apyc1 as an exemplar.

Materials and methods

Databases used in the study

Two databases were utilized in this study. First, we used the RefSeq complete genome database for bacteria and archaea, which was downloaded in July 2022 and contains 21 855 genomes. For phage genomes, we utilized the GenBank database, which was downloaded in December 2023 and includes 13 487 genomes.

Protein sequence models

All experimentally validated protein sequences were retrieved from the literature (Supplementary Table S4). All proteins were blasted using BLASTp against the NCBI nonredundant database with an *E*-value threshold of 1e−5. The resulting hits were then compared with the original protein sequence to ensure a minimum of 30% identity. Additionally, a coverage threshold was applied: 80% of coverage of the original protein and 70% of coverage of the hit (i.e. the 70% of the hit protein corresponds to the original protein). All conserved hits were then clustered at 95% identity and 95% coverage using MMseqs2 v13.45111 (47) easy-cluster. If the number of representative sequences was >200, the sequences were clustered at 80% coverage and 80% identity. All representative sequences were then aligned using MAFFT v7.505 (48) (default settings) and Hidden Markov Model (HMM) profiles were built using hmmbuild (HMMER v3.3.2) (49).

MGE and defense system detection

RefSeq annotation was used to determine if a given replicon was a plasmid. Prophages were detected using VirSorter2 v2.2.3 (50). An anti-defense system was classified as inside a prophage if it was present in the boundaries of the prophage (±2 kb). Satellites were detected using SatelliteFinder v0.9.1. An anti-defense system was classified as inside a satellite if it was present in the boundaries of the prophage (±2 kb). Integrons were detected using IntegronFinder v2.0.2 (51). An anti-defense system was classified as inside an integron if the protein was detected as part of an integron cassette by IntegronFinder. Integrative conjugative elements (ICEs) were detected using CONJscan MacSyFinder models v2.0.1 (52) to detect conjugative systems on chromosomal replicon (not annotated as plasmid). An anti-defense system was classified as inside an ICE if it was present between the extremities of the detected proteins (±10 kb). All integrases were detected using 108 Pfam (53) with the Pfam description containing 'Transposase', 'Recombinase', 'Integrase' and 'Resolvase' using gathering bit score (GA) thresholds with hmmsearch (HMMER v3.3.2) (49). The genomic localization was defined as non-MGE if it was not found in any previously described MGEs and without any integrase in 10 kb downstream and upstream. For the first detection threshold determination, only plasmid and prophages (most abundant MGE) were detected and hit localization were assigned as non-MGE if not encoded inside prophage, plasmids or detected in the GenBank phage database.

First detection of anti-defense system and threshold choice

All profile HMMs' detection was done using hmmsearch (HMMER v3.3.2) (49) on both the prokaryotic RefSeq database and GenBank phage database with GA cut threshold at 20 and profile coverage of 40%. All hits were then classified into four categories based on their localization: phage (GenBank database), plasmid, prophage or other. All GA thresholds were manually chosen. Those thresholds were defined using three main factors: hit score, coverage distribution and hit localization in the genome. These criteria were combined in a single graph illustrated in Figure 1B and available for genes with more than 1000 hits (Supplementary Figure S1) and for all profiles on GitHub (https://github.com/mdmparis/antidefensefinder_2024) and on Figshare under the DOI: 10.6084/m9.figshare.26526487.

Anti-defense system and defense system detection

Anti-defense systems and defense systems were detected using DefenseFinder v1.3.0 with the argument –antidefense on the two databases.

Apyc1 phylogenetic tree

All Apyc1 homologs detected by AntiDefenseFinder were retrieved. Bacterial homologs were clustered together at 80% identity and 80% coverage with MMseqs2 v13.45111 (47). Phage homologs were clustered with MMseqs2 at 95% identity and 95% coverage. All representative sequences were used for the alignment. Eighteen sequences of Metallo Beta Lactamase (MβL) fold protein known to be antimicrobial resistance genes were used as an outgroup of the tree. The alignment used for the tree construction was made using MUSCLE v5.1 (54) with the -super5 option. The alignment was trimmed using ClipKIT v1.3.0 (55) in smart gap mode. The tree was built using IQ-TREE v2.2.3 (56) with models finder and 2000 ultrafast bootstrap.

Apyc1 multiple sequence alignment

Apyc1 protein sequences in Figure 4 were aligned using EMBL-EBI MUSCLE and then visualized using Jalview v2.11.3.3. These Apyc1 sequences included *Thalassospira* WP_223304948.1 (THSP027), *Archangium violaceum* WP_204220610.1 (ARVI001), *Bacillus* phage SBSphiJ (31), *Paenibacillus sp. J14* WP_028539944.1 (PASP001), *ohnella* WP_174887610.1 (COSP018), *Legionella sp. MW5194* WP_203455517.1 (LESP016), *Synechocystis* WP_010871596.1 (SYSP007), *Staphylococcus* phage Madawaska QQO92874.1 (MW349129) and *Caldicellulosiruptor bescii* WP_041727399.1 (CABE001).

Apyc1 protein structure predictions

Apyc1 protein sequences in Figure 4 (listed above) were predicted using AlphaFold2 (AF2) ColabFold v1.5.5 (57). Structural comparison of the Apyc1 proteins was performed using the super function in PyMOL v2.1.

Apyc1 protein purification

The *apyc1* genes were synthesized and cloned into pET28a vectors in which the expressed protein contains an N-terminal His_6 tag. All the proteins were expressed in *E. coli* strain BL21(DE3) in lysogeny broth medium. After growth at 37°C, the cells were induced by 0.2 mM isopropyl-β-d-thiogalactopyranoside when the cell density reached an optical density at 600 nm of 0.8. After growth at 18°C for 12 h, the cells were harvested, resuspended in lysis buffer (50 mM Tris–HCl pH 8.0, 300 mM NaCl, 30 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF)) and lysed by sonication. The cell lysate was centrifuged at 20 000 × *g* for 50 min at 4°C to remove cell debris. The supernatant was applied onto a self-packaged Ni-affinity column (2 mL Ni-NTA, GenScript) and contaminant proteins were removed with washing buffer (50 mM Tris–HCl pH 8.0, 300 mM NaCl, 30 mM imidazole). Then the protein was eluted with an elution buffer (50 mM Tris pH 8.0, 300 mM NaCl, 300 mM imidazole). The protein eluent was concentrated and further purified using a Superdex-200 increase 10/300 GL (Cytiva) column equilibrated with a buffer containing 10 mM Tris–HCl pH 8.0, 200 mM NaCl and 5 mM DTT. For the LESP016-Apyc1 and MW349129-Apyc1, buffers contained 500 mM NaCl along with an additional 5% glycerol throughout the purification process.

Apyc1 *in vitro* cleavage assays

Reactions of the assay consisted of 50 mM Tris–HCl pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 100 μM cNMP and 1 μM recombinant protein in a 100 μl volume. The reaction mix was incubated at 37°C for 20 min and then filtered using a 3-kDa cutoff filter (Millipore) at 4°C. Filtered nucleotide products were analyzed using a C18 column (Agilent ZORBAX Bonus-RP 4.6 mm × 150 mm) heated to 30°C and run at 1 ml/min in a buffer of 50 mM NaH_2PO_4 adjusted to pH 6.8, supplemented with 3% acetonitrile and 0.1% trifluoroacetic acid. Raw data are provided in Supplementary Figure S2.

Apyc1 enzymatic kinetics assays

The kinetic experiments were conducted at 37°C with a total reaction volume of 100 μl, in a buffer containing 50 mM Tris–HCl pH 7.5, 100 mM KCl, 1 mM DTT and 1 mM $MgCl_2$. Reactions were initiated by adding protein and proceeded for 20 s, then they were terminated with 0.1 M NaOH. Subsequently, the reaction samples were placed into the High-Performance Liquid Chromatography (HPLC) autosampler. Each reaction mix was analyzed using the C18 column under the above conditions. The area of the substrate peak at 254 nm was integrated to determine the substrate consumption at each substrate concentration. The data were converted into reaction rates and plotted against substrate concentrations. Curve fitting and kinetics parameter determination were performed using the Origin software. Raw data are provided in Supplementary Figure S3.

Results

AntiDefenseFinder: a search tool to detect known inhibitors of prokaryotic defense systems

To systematically detect anti-defense systems, we developed and added an AntiDefenseFinder option, to DefenseFinder (1,43,44), a program that already detects defense systems. We first conducted a comprehensive literature review of all known anti-defense proteins and retrieved experimentally validated sequences of 180 anti-defense genes. DefenseFinder relies on HMM profiles for sensitive homology search. We thus needed to build one HMM profile per anti-defense protein.

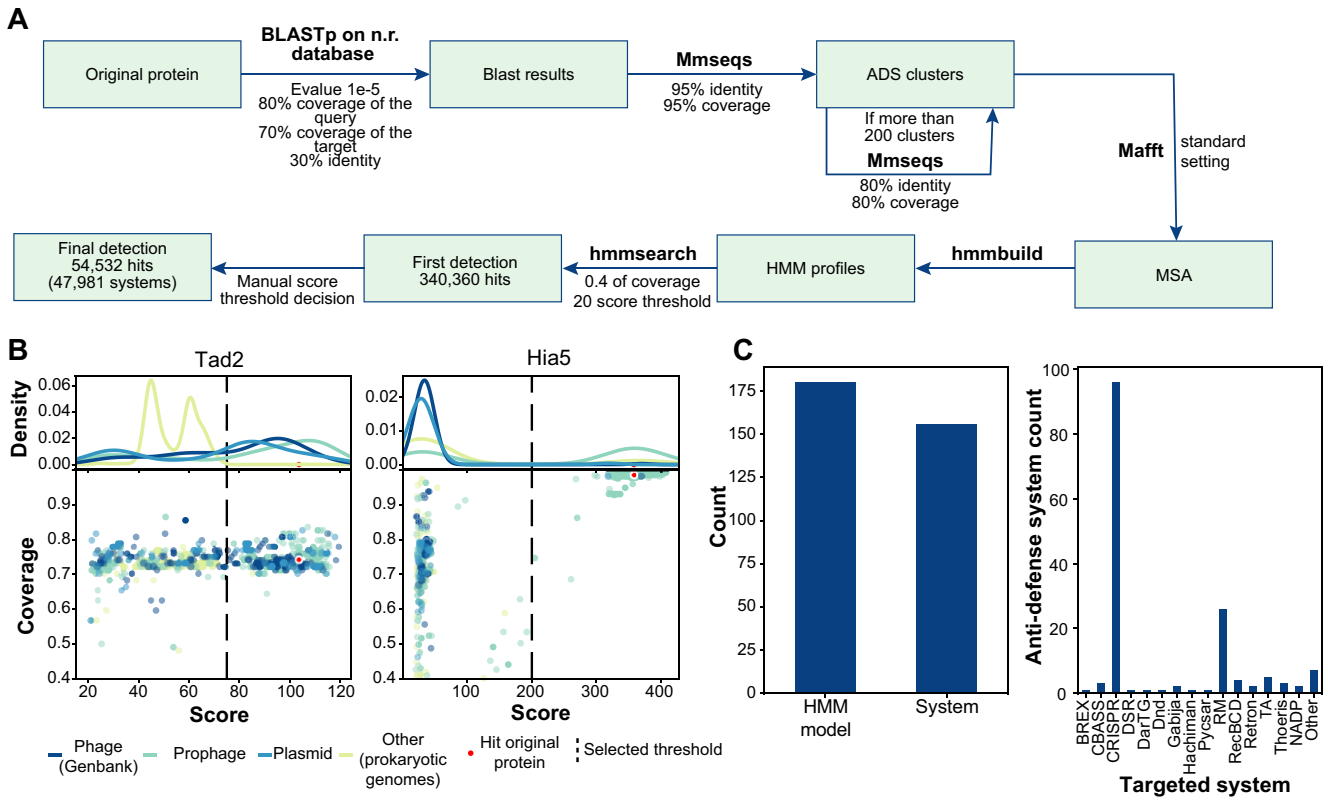


Figure 1. AntiDefenseFinder is a tool to systematically detect known inhibitors of prokaryotic defense systems. (**A**) Pipeline of creating HMM profiles for AntiDefenseFinder. (**B**) Filtering of positive hits based on selected threshold and protein sequence coverage (≥40%). The selection threshold for each anti-defense protein was manually analyzed and chosen based on the distribution of hits relative to the originally discovered protein. (**C**) Total number of HMM models developed relative to the total number of anti-defense proteins, and total number of anti-defense systems detected across prokaryote and phage sequences that inhibit a specific type or family of defense systems.

To automate the creation of the HMM profiles, we started with a homology search using BLASTp on the RefSeq nonredundant database to capture sequence diversity (Figure 1A). BLASTp results were filtered using a minimal coverage and sequence identity. Next, the sequences were clustered to reduce the weight of closely related homologs (e.g. *E. coli* proteins) in the multiple sequence alignment. Cluster representatives were then aligned and an HMM profile was constructed for 156 anti-defense systems because several systems may be composed of two or more proteins. Specifically, four systems are exclusively multiple gene systems – NARP1, NARP2, vcrx089_090 and vrcx091_093 – and the Acr genes can be detected alone or with an anti-CRISPR associated gene. There are also several independently identified anti-defense genes that are closely related homologs and we therefore classified them as a single anti-defense family, including (i) Tad2 and AcrIIA7, (ii) ArdB, KlcA and AcrIC11, (iii) ArdA and ArdU and (iv) Hin/Hia/Nma (Supplementary Figure S4).

We performed initial detections on two databases: RefSeq prokaryotic complete genomes and GenBank phage genomes (see 'Materials and methods' section). Using a low threshold (GA: 20 and coverage >40%), we identified 340 360 hits (Supplementary Table S1). These hits were used to refine each HMM profile's GA threshold based on the distribution of both hit scores and profile coverage (Supplementary Figure S1). As anti-defense genes are often encoded inside MGEs, we also took into consideration the localization of hits within genomes or MGEs to further define a true positive hit (Figure 1B). Hits within MGEs (e.g. plasmids, prophages or phage databases) were considered more likely to be true positives. This approach allowed us to manually set a threshold for each profile (Supplementary Figure S1). Overall, AntiDefenseFinder detects 156 anti-defense systems with HMM profiles encompassing 180 proteins (Figure 1C). The majority of known anti-defense systems are Acr ($n = 96$) and anti-RM ($n = 26$); however, AntiDefenseFinder also identifies a variety of other anti-defense systems that target the expanding diversity of prokaryotic defense systems. AntiDefenseFinder is now integrated into DefenseFinder version v1.3.0 available in command line and as a web service. It can be executed alongside DefenseFinder (–antidefensefinder) or using only AntiDefenseFinder models (–antidefensefinder-only).

Anti-defense systems are variably distributed across genomes and genetic elements

We initially sought a comprehensive view of anti-defense system distribution across prokaryotes and phages. We applied AntiDefenseFinder to a database of 21 855 prokaryotic and 13 487 phage genomes and detected a total of 47 981 anti-defense systems. In bacteria, 41 946 total anti-defense systems were identified and were predominantly identified in the genera *Escherichia* (12 544 total, ~25%), *Klebsiella* (9108), *Staphylococcus* (1781), *Enterococcus* (1242), *Pseudomonas* (579) and *Bacillus* (320) (Figure 2A and C; Supplementary Table S2). We also found that anti-RM and Acr are the most abundant anti-defense systems in bacteria in our detection with a total count of 22 708 and 6880, respectively

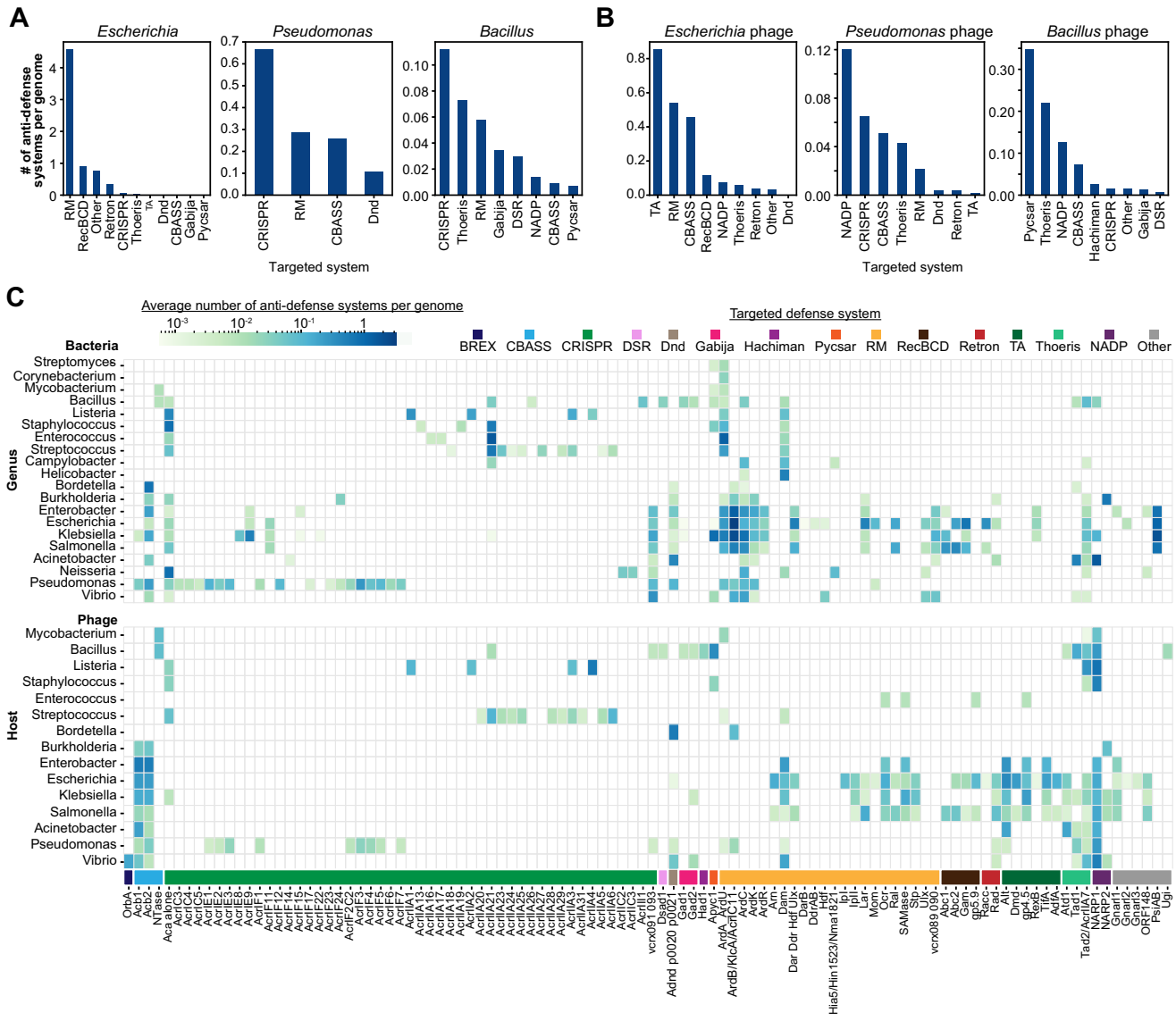


Figure 2. Anti-defense system distribution across different bacterial genera and phage host. Total anti-defense systems found across (**A**) bacteria or (**B**) phages in *Escherichia, Pseudomonas* or *Bacillus* genera. (**C**) Average number of anti-defense systems found per genome organized by genus. Some systems discovered separately were grouped based on homology into a single system where they are separated by '/'. Anti-defense system activity refers to the activity observed at the time of its initial discovery although some anti-defense systems may act against multiple bacterial defense systems (e.g. Ocr, which targets both RM and BREX).

(Figure 2A and C). In *Escherichia*, anti-RM systems are the most detected anti-defense systems and are notably abundant with 3132 instances of ArdB/KlcA. This may have occurred because ArdB was discovered in *E. coli* in 1993 (58), and has henceforth been studied in-depth in the same bacteria host. In *Pseudomonas*, Acr systems are the most abundant in our detection especially with Type I and II CRISPR-Cas Acr proteins, but most notably 114 instances of AcrIF3. This again may be due to the discovery of Acrs in *Pseudomonas aeruginosa* (38,59). Apart from anti-RM and Acrs, 43% (10/23) of anti-defense systems with >10 instances are only detected in the phylogenetic order where the system was originally discovered. Otherwise, anti-defense systems are variable between bacterial species. For instance, in *Klebsiella*, the anti-Pycsar protein 1 (Apyc1) is the most detected system with 1233 homologs detected, and in *Acinetobacter,* the newly identified NARP1 is the most abundant with 469 occurrences. In the 383 genomes of archaea, only 26 anti-defense systems were detected and 65% of those systems were Acrs (AcrIII1 $n = 7$, AcrIIA26 $n = 7$ and AcrIA1 $n = 3$). Only five anti-defense systems detected were not anti-RM or Acrs. This limited detection of anti-defense systems in archaea is likely due to the low number of complete archaeal genomes in the RefSeq database (383 complete genomes), and in turn, we anticipate more anti-defense systems may be identified using RefSeq contig assembly or the broader GTDB database. Moreover, as previously described, most anti-defense systems were discovered in bacteria.

In phages, 6009 total anti-defense systems were identified and most were detected in phages that infect the genera *Escherichia* (2124 total, ~35%), *Klebsiella* (453), *Vibrio* (321), *Salmonella* (299), *Pseudomonas* (158) and *Bacillus* (254) (Figure 2B and C; Supplementary Figure S5 and Supplementary Table S3). Similarly to bacteria, anti-RM is

the most abundant anti-defense system. We suspect this is due to a bias in the available genomic sequences and the early discovery and the prevalence of RM in bacteria. Aside from anti-RM, 56% (14/25) of anti-defense systems with >10 instances are only detected in the phylogenetic order where they were discovered (e.g. ArdB in *Enterobacterales*, AcrIIA1 in *Bacillales* or AcrIIA23 in *Lactobacillales*). For example, anti-CBASS protein 1 (Acb1) and Acb2 are predominantly identified in phage genomes infecting eight related genera (Figure 2C). There are also instances when anti-defense systems are only found in phage (e.g. Had1, Ocr, etc.) or only in bacterial genomes (e.g. AcrIIA13, PsiAB, etc). Furthermore, out of the 13 487 phage genomes, we detected at least one anti-defense system in 2796 genomes (20%) with notable disparities between species (Figure 3D; Supplementary Figure S5). In any case, many anti-defense systems are very rare and present in <1% of prokaryotic and phage genomes. Overall, these results demonstrate that anti-defense system distribution is variable across distinct prokaryotic and phage genomes with a bias toward model organisms where they were originally identified. This suggests that discovery of anti-defense in new species is important for a better understanding of the anti-defense diversity.

We then set out to understand how anti-defense systems are localized across the prokaryotic pan-genome and MGEs. Acr genes typically co-localize or are encoded in specific genomic loci of prophages (38,39). Anti-RM, anti-SOS systems and Acrs can co-localize on the leading strand of conjugative plasmids (40), which has been collectively referred to as 'anti-defense islands' (39). We therefore evaluated whether this observation could extend to other anti-defense systems and observed that anti-defense systems co-localize within 10 kb of one another in 31.7% and 32.9% of bacterial and phage genomes, respectively (Supplementary Figure S6). However, given the proportionally smaller size of phage, we also examined anti-defense systems co-localizing within 1 kb and observed 17.8% of systems co-localize within this 1-kb range (Supplementary Figure S6). The well-studied *E. coli* T4 phage has at least three independent instances of anti-defense genes co-localizing together in an anti-defense island while the *E. coli* phage Lambda has one instance (Figure 3A). Other co-localization of anti-defense systems occurs in phages from the BASEL collection, such as Bas31 and Bas35 (Supplementary Figure S7). In all these cases, these anti-defense islands include genes that have been shown to inhibit distinct bacterial defense systems. In other phages, anti-defense genes that target the same bacterial defense system co-localize in the genome, such as Acrs, Gabija anti-defense and Tad genes across *Pseudomonas*, *Bacillus* and *Blautia* bacterial genomes, respectively (Figure 3B). We anticipate that more anti-defense islands are present in MGEs due to the increasing identification and diversity of anti-defense systems. Furthermore, these results demonstrate that ~66% of all known anti-defense genes are not encoded in the same genetic loci, but are rather encoded alone (Supplementary Figure S7). As an example, applying AntiDefenseFinder to the well-studied *P. aeruginosa* model jumbo phage phiKZ revealed only one anti-defense gene, Tad1, despite encoding dozens of small genes of unknown function (34). These collective results reflect a need for discovering new anti-defense genes.

We next evaluated whether anti-defense systems are encoded in the same genome as the defense system they were originally identified to inhibit. We found that most anti-defense genes do not appear to co-occur in the same genome as its targeted defense system (Figure 3C). However, AcrIE8 is a unique example that is often encoded in genomes that also encode Type I CRISPR-Cas (Figure 3C). Nearly all instances of AcrIE8 are encoded on prophages (Figure 3D), with previous work suggesting that Acrs are expressed to neutralize CRISPR and prevent self-targeting (38,39). Some other Acr genes (i.e. *acrIF11*, *acrIIA1* and *acrIIA23*) are often found in the same genome with the CRISPR-Cas system they inhibit. Expanding upon this analysis revealed that many anti-defense genes are encoded in MGEs, including satellites, prophages, ICEs, plasmids and nearby integrases, and fewer anti-defense genes are encoded in nonmobile regions (Figure 3D). In many cases, ≥80% of instances of the detected anti-defense gene are encoded within a single type of MGE (Figure 3D), suggesting that the inhibited defense system may predominantly target that type of MGE.

We hypothesized that identifying anti-defense genes encoded on a distinct type of MGE would reveal an unexpected target of the defense system. However, our findings generally align with the known defense system mechanism. For example, we observed that anti-RM and Acr genes are encoded on diverse types of MGEs (Figure 3D), and it is known that RM and CRISPR-Cas systems target various MGEs (60,61). By comparison, in bacterial genomes, 97% of detectable Acb2 are encoded in prophages, and nearly 100% of Tad1 homologs are encoded in prophages (Figure 3D and E). To date, both CBASS and Thoeris have been demonstrated to only target phages (17,20,32). Other anti-defense genes are only encoded in virulent phages, including Ocr (anti-RM; *Teseptimavirus* and *Kayfunavirus* phage) (62), Had1 (anti-Hachiman; *Bastillevirinae* phage) (63), Atd1 (anti-TIR; *Phapecoctavirus*, *Justusliebigvirus* and *Lazarusvirus* phage) (36) and AdfA (anti-TA; *Tequatrovirus* and *Mosigvirus* phage) (26) (Figure 3E). In several of these cases, the cognate defense system has been demonstrated to solely target phages. Surprisingly, however, our final analyses demonstrated that a limited number of anti-defense genes that inhibit phage-targeting systems – Apyc1 (31), NARP1/2 (37) and NTases (anti-CBASS) (36) – were mostly encoded in the bacterial genome outside of prophages (Figure 3D), suggesting a nondefense function for these proteins. We further investigate bacterial and phage Apyc1 below.

Apyc1 is common in bacterial chromosome and co-opted by phages

The pyrimidine cyclase system for anti-phage resistance (Pycsar) uses cCMP or cUMP signaling molecules to activate a downstream effector that limits phage replication (18). In response, phage encodes the Apyc1 that counteracts this system through cleavage of cyclic mononucleotides [cCMP, cUMP, cyclic GMP, cAMP] (31). Hobbs *et al.* (31) also identified 107 Apyc1 homologs in distinct phages and bacterial chromosomes in two predominant *Bacillus* and *Staphylococcus* clades and then 10 homologs were experimentally validated to cleave cCMP and cUMP (31). Using the AntiDefenseFinder tool, we detected 2301 total instances of Apyc1 with 80.7% encoded in the bacterial chromosome outside of an obvious MGE (Figure 3C). To determine the evolutionary history of Apyc1 homologs, we built a phylogenetic tree of bacterial and phage Apyc1 and used an antimicrobial resistance MβL-fold protein as an outgroup to root the tree. We observed three

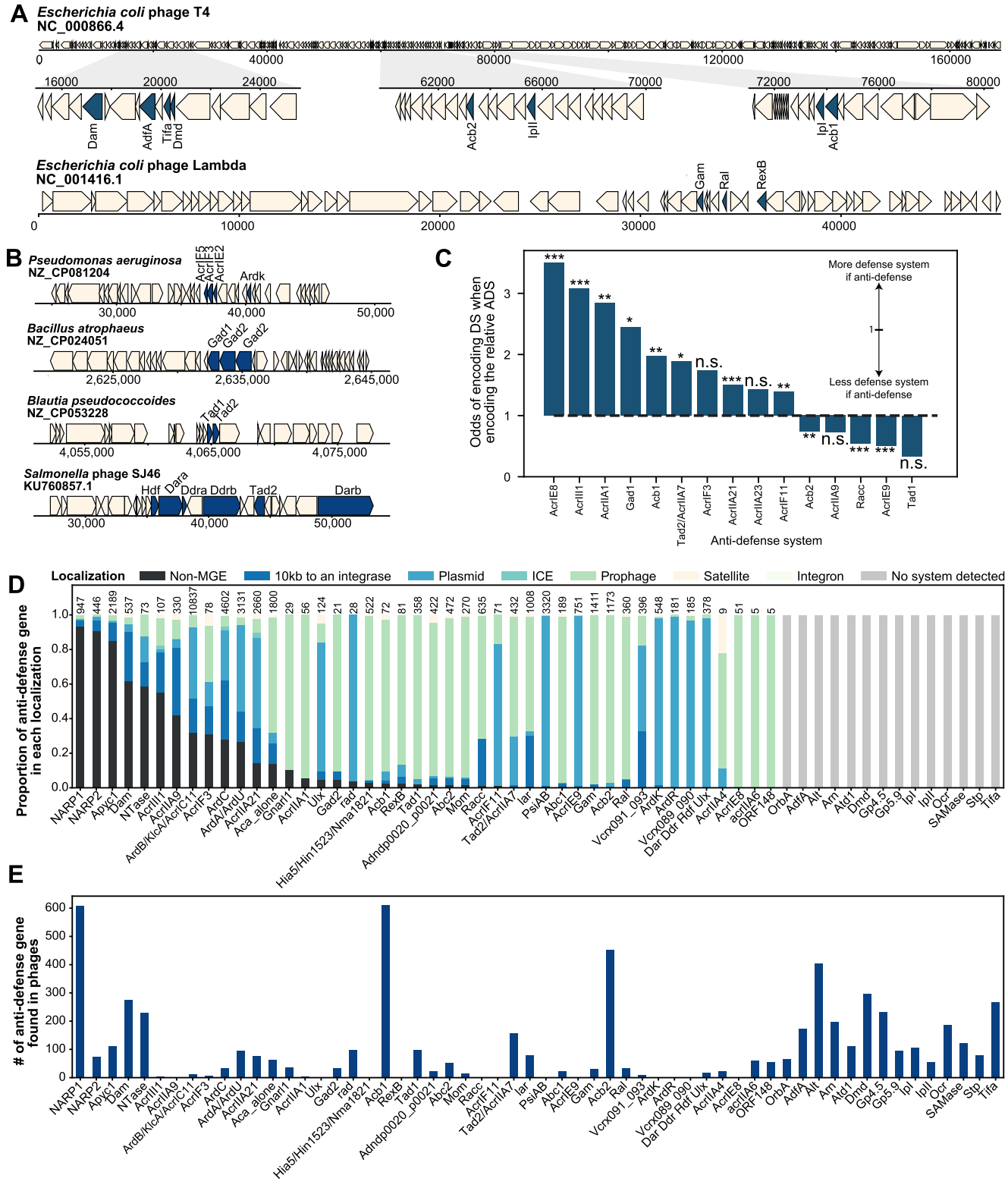


Figure 3. Localization of anti-defense systems in genomes and MGEs. (**A**) Examples of anti-defense genes co-localized in an anti-defense island within the well-studied *E. coli* phages T4 and Lambda, and (**B**) diverse bacterial and phage genomes. (**C**) Odds ratio of co-encoding a defense system in a genome that encodes its corresponding anti-defense system. DS: defense system; ADS: anti-defense system. Odds ratio is calculated by dividing the proportion of genomes encoding the defense system and its corresponding anti-defense system by the proportion of genomes encoding the defense system without its corresponding anti-defense system. Statistical test: Chi-squared test corrected by Bonferonni (*** < 10e−5, ** < 10e−3, * < 0.05 and n.s.: not significant). (**D**) Relative proportion of a single anti-defense gene localized in distinct genomic localizations, including satellites, prophages, ICEs, plasmids, nearby integrases and not in MGEs like bacterial chromosomes. The number of detected systems in the RefSeq database is indicated at the tip of the bars. (**E**) The total number of anti-defense genes localized in phage genomes. Systems represented in Figure 3D and E are anti-defense systems detected at least 50 times in the RefSeq database or 20 times in the GenBank phage database.

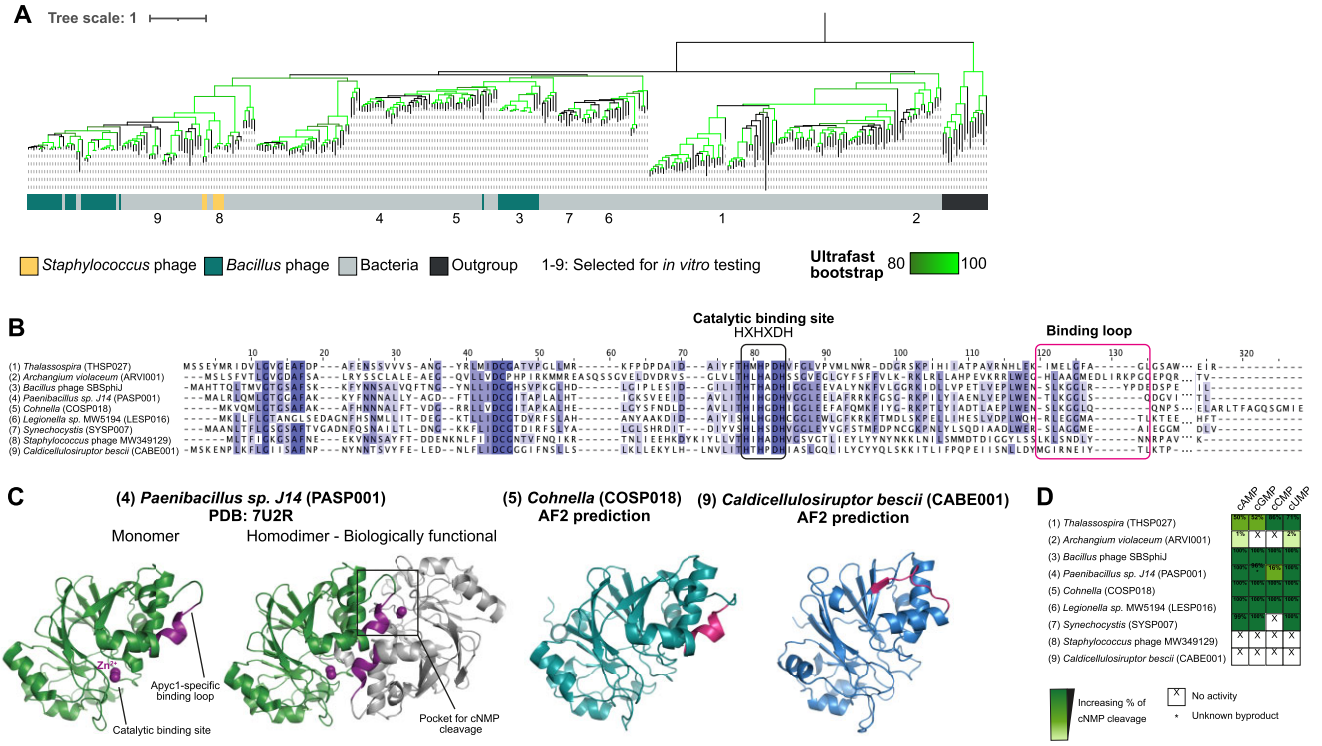


Figure 4. Apyc1 is abundant and functionally conserved across bacteria. (**A**) Phylogenetic tree of SBSphiJ Apyc1 and >350 homologs from bacteria and phage. Colors represent bacterial genus, highlighting the most abundant *Bacillus* and *Staphylococcus*. Black circles indicate Apyc1 homologs tested for *in vitro* cleavage of cNMPs. (**B**) Multiple sequence alignment of Apyc1 homologs (see Supplementary Figure S6 for full alignment). Residues that are >80% conserved, >60% conserved and > 40% conserved are shaded in dark purple, light purple and light gray, respectively. Residues involved in catalysis and binding are circled in black (left) and pink (right), respectively. (**C**) Structures of *Paenibacillus sp. J14* (PASP001), *Cohnella* (COSP018) and *C. bescii* (CABE001) Apyc1 homologs. PASP001 was experimentally solved and deposited on the RCSB Protein Data Bank (PBD: 7U2R), and COSP018 and CABE011 were computationally predicted using AF2. Zn^{2+} ions that coordinate cNMP cleavage in the catalytic binding site, as well as the Apyc1-specific loop that extends into the cNMP catalytic binding site, are labeled and highlighted in pink. (**D**) Summary of the *in vitro* cleavage assay data ($n = 3$), with corresponding conversion rates as defined by the percentage of 100 μM cNMPs that reacted with 1 μM Apyc1 protein within 20 min.

independent monophyletic clades of phage Apyc1 branching in bacterial Apyc1 (Figure 4A), suggesting that bacterial Apyc1 represents the ancestral form that phage likely acquired Apyc1 from a bacterial host. Upon further investigation, we observed that bacterial Apyc1 is encoded in genomes that also include Pycsar, CBASS and occasionally, Apyc1 is adjacent to a cyclase with no obvious effector nearby (Supplementary Figure S8).

To determine whether bacterial Apyc1 is an active enzyme, we initially examined the sequence and structure of evolutionarily diverged Apyc1 homologs in bacteria and phage. We observed that the Apyc1 sequences all retain the catalytic site, including the versions encoded next to cyclases, but exhibit diversity in the nucleotide binding loop (Figure 4B), which is proposed to extend into the nucleotide-binding pocket and stabilize the small cyclic mononucleotide substrates (31). For the *Paenibacillus sp. J14* Apyc1 homolog (PASP001), the structure was previously solved and demonstrated that the loop from one monomer interacts with the catalytic binding pocket of the opposing monomer and subsequently enables cCMP hydrolysis (31) (Figure 4). For the bacterial homologs we examined, such as *Cohnella* (COSP018), the nucleotide binding loop is intact and overlays well with PASP001 loop (Figure 4C; Supplementary Figure S9), suggesting that it also retains the nucleotide cleavage function. Some bacterial homologs like *C. bescii* (CABE001) exhibit a shortened loop (Figure 4C) while others exhibit a lengthened loop (Supplementary Figure S9), and in turn, may not effectively interact with the catalytic binding pocket.

To examine the function of these Apyc1 homologs, we performed *in vitro* cleavage assays and observed that bacterial Apyc1 homologs with structurally conserved nucleotide binding loops were able to strongly cleave cAMP, cGMP, cCMP, and cUMP signals (Figure 4D; Supplementary Figure S2). The PASP011 and SBSphiJ Apyc1 homologs examined in Hobbs *et al.* (31) also demonstrated cleavage of all cNMP signals. By comparison, CABE001, *A. violaceum* (ARVI001) and *Staphylococcus* phage (MW349129) homologs with shortened or lengthened Apyc1-specific nucleotide binding loops showed weak or no cleavage of cNMPs (Figure 4D). These data suggest that the bacterial Apyc1 with intact, shortened nucleotide-binding loops can degrade cNMPs. Finally, we investigated whether the phage versions of the enzyme displayed faster turnover compared with the host version. To do so, we examined enzymatic kinetics from Apyc1 homologs in the *Bacillales* order – bacterial PASP011 and COSP018 and phage SBSphiJ Apyc1 – with the Pycsar signals cCMP and cUMP. We observed that the bacterial COSP0018 Apyc1 homolog cleaves cCMP and cUMP with nearly identical kinetics compared with phage Apyc1 while the bacterial PASP011 Apyc1 demonstrated ∼6-fold and ∼2-fold slower kinetics with cCMP and cUMP, respectively (Supplementary Figure S3). In addition, we observed that all Apyc1 homologs exhibit high K_{cat} values (275–1581 per s) (Supplementary Figure S3).

These findings suggest that bacterial and phage Apyc1 have generally similar enzyme kinetics without specialization or adaptation by the phage homologs. Altogether, we conclude that the Apyc1 family functions in rapid cleavage of cNMPs that are likely utilized in both regulatory and defense systems.

Discussion

We developed the AntiDefenseFinder tool and web service (https://defensefinder.mdmlab.fr) that detects all known anti-defense systems across prokaryotic and phage genomes, as well as MGEs. In doing so, we provided a quantitative overview of 156 anti-defense systems families and 47 981 homologs that span a diversity of bacterial genera, genomic localizations, and functional strategies. A recently developed pre-computed database, dbAPIS, detects 41 anti-defense systems and 4428 total homologs encoded in phage genomes (46). We hope that the free and open-source, searchable nature of AntiDefenseFinder will enable the field to identify the full repertoire of anti-defense systems, especially in understudied MGEs. AntiDefenseFinder is also easily adaptable to add new anti-defense systems given that we built upon the pre-existing framework of the DefenseFinder tool (1,43,44). Over time, we will continue building new profiles of anti-defense systems.

Many gaps of knowledge remain regarding anti-defense systems, such as species diversity and anti-defense island abundance. Although we observed anti-defense genes widespread across many distinct bacterial species, there is biased enrichment in *Escherichia* (14 668 detected) and related species, likely because these model organisms were used to discover the first instance of the anti-defense gene and the 1861 *Escherichia* bacteria and 971 *Escherichia* phage genomes in our databases. We also speculate that there may be biases in anti-defense system detection, in which some are highly sequenced, but have low overall abundance in nature and vice versa. This may be particularly relevant for anti-defense systems in archaea and phages, in which we identified ∼0.1% and ∼20% anti-defense systems in their genomes, respectively. Furthermore, in both bacterial and phage genomes, we also observed that over 30% of detectable anti-defense genes co-localize within 10 kb of one another in bacteria and 17.8% within 1 kb in bacteriophages, which is a defining feature of anti-defense islands (64). The model *E. coli* T4 phage notably encoded three independent instances of anti-defense islands; however, many bacteria and phage still lack these islands. Conversely, there are over 60% of anti-defense genes that are standalone. It is possible that applying a 'guilt-by-association' analysis may reveal entirely new anti-defense genes as it did with Acrs (39,65). We anticipate that an abundance of anti-defense systems await discovery in prokaryotic host species that currently lack known anti-defense genes or islands.

Challenges remain in the detection of distantly related anti-defense proteins due to their small size and vast sequence divergence. In some cases, the functional domains of enzymatic proteins are widely conserved, like the phosphodiesterase domain of Acb1 (31). Enzymatic domains have been found to retain conserved structural folds, enabling the discovery of an Acb1 homolog in eukaryotic viruses (66). With advances in structural predictions and comparative analyses, the field is pivoting toward structure-guided discovery of new anti-defense systems and has been applied to discover new Acrs in phage (67). The recent release of the Big Fantastic Viral Database (BFVD) provides a collection of 350K + predicted structures of viral proteins of eukaryotic and prokaryotic origin (68). Therefore, to address the challenge of identifying highly divergent anti-defense systems, future iterations of AntiDefense Finder could include a database of experimental and predicted protein structures of all known anti-defense systems that may be then used to query the BFVD via the open-source FoldSeek protein search tool (69). These findings will also enable the creation of new HMM profiles to improve overall detection and eventually support machine learning algorithms to open a new frontier of anti-defense system discovery.

Despite these limitations, our quantitative detection and analysis of known anti-defense systems revealed fundamental insights into bacterial and phage biology. We observed that over 80% of detected instances of Apyc1 were encoded in nonmobile regions of the bacterial genome. Apyc1 was previously identified in phages and prophages and functioned in the degradation of cyclic mononucleotides (cAMP, cGMP, cCMP and cUMP) (31). Pycsar defense solely relies on cCMP and cUMP (18) whereas cAMP and cGMP are involved in housekeeping functions (70,71). However, our evolutionary and functional analyses suggest that Apyc1 originated in bacteria and then phage co-opted Apyc1 to counteract Pycsar defense. An alternative scenario has been observed with Type III CRISPR-Cas defense: Phage encodes a ring nuclease (AcrIII-1) that degrades cA_4 and inhibits CRISPR effector activity (41) and then bacteria co-opted this inhibitor (Crn2) to regulate CRISPR (42). Lastly, the recently identified NARP1, which inhibits defense systems metabolizing NAD^+, was also found in nonmobile regions of bacteria (37), aligning with our findings and likely plays housekeeping functions in bacteria. Altogether, the AntiDefenseFinder tool has enabled us to explore diverse anti-defense proteins across prokaryotes, phages and MGEs and we hope that we have given others the agency to do the same.

Data availability

The AntiDefenseFinder web service can be found at https://defensefinder.mdmlab.fr/. The command line tool is available on GitHub at https://github.com/mdmparis/defense-finder, and its associate MacSyFinder models are also available on GitHub at https://github.com/mdmparis/defense-finder-models. Code and supplementary information are available on GitHub: https://github.com/mdmparis/antidefensefinder_2024 and on Figshare under the DOI: 10.6084/m9.figshare.26526487.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Sukrit Silas from the Bondy-Denomy Lab for his insight into the genomics of phage anti-defense systems. We thank the IT Department of the Institut Pasteur for providing support for the AntiDefenseFinder web service. To promote gender equality and inclusivity in research, we are convinced of the importance of acknowledging gender bias in research article citations. We estimated that among the 71 references cited in the main text, approximately 32% (23) have a female first author and approximately 16% (11) have a female last author.

Author contributions: E.H. and F.T. conceived the project; A.B. and J.B.-D. supervised the project and provided feedback; F.T. and J.C. led the development of the AntiDefenseFinder pipeline and HMM profiles, as well as detection and bioinformatic analyses of anti-defense systems; J.C. and M.J. provided support in developing the pipeline and HMM profiles; R. P. developed the web service; E.H. performed the in-depth sequence and structural analyses of Apyc1 homologs; L.W. and J.R. conducted the protein purification and *in vitro* cleavage and kinetics assays of Apyc1 homologs supervised by Y.F.; E.H. and F.T. wrote the manuscript and created the figures; and all authors provided editing and feedback.

Funding

Institut National de la Santé et de la Recherche Médicale (Inserm) [ATIP-Avenir R21042KS/RSE22002KSA to F.T.]; MSDAVENIR [to F.T., to A.B.]; National Institutes of Health [5T32AI060537-20 to E.H., R21AI168811 to J.B.-D.]; National Key Research and Development Program of China [2022YFC3401500 and 2022YFC2104800 to Y.F.]; Kleberg Foundation [to J.B.-D.]; Vallee Foundation [to J.B.-D.]; Searle Scholars Program [to J.B.-D.]; Bettencourt Schueller Foundation [to A.B.]; ERC [PECAN 101040529 to A.B.]; Institut Pasteur [to A.B.]. Funding for open access charge: ERC [PECAN 101040529].

Conflict of interest statement

J.B.-D. is a scientific advisory board member of SNIPR Biome and Excision Biotherapeutics, a consultant to LeapFrog Bio and BiomX, and a scientific advisory board member and co-founder of Acrigen Biosciences and ePhective Therapeutics.

References

1. Tesson,F., Planel,R., Egorov,A.A., Georjon,H., Vaysset,H., Brancotte,B., Néron,B., Mordret,E., Atkinson,G.C., Bernheim,A., *et al.* (2024) A comprehensive resource for exploring antiphage defense: DefenseFinder webservice, Wiki and databases. *Peer Community J.*, **4**, 91.
2. Cheng,R., Huang,F., Wu,H., Lu,X., Yan,Y., Yu,B., Wang,X. and Zhu,B. (2021) A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.*, **49**, 5216–5229.
3. Gao,L.A., Wilkinson,M.E., Strecker,J., Makarova,K.S., Macrae,R.K., Koonin,E.V. and Zhang,F. (2022) Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science*, **377**, eabm4096.
4. Bari,S.M.N., Chou-Zheng,L., Howell,O., Hossain,M., Hill,C.M., Boyle,T.A., Cater,K., Dandu,V.S., Thomas,A., Aslan,B., *et al.* (2022) A unique mode of nucleic acid immunity performed by a multifunctional bacterial enzyme. *Cell Host Microbe*, **30**, 570–582.
5. Hsueh,B.Y., Severin,G.B., Elg,C.A., Waldron,E.J., Kant,A., Wessel,A.J., Dover,J.A., Rhoades,C.R., Ridenhour,B.J., Parent,K.N., *et al.* (2022) Phage defence by deaminase-mediated depletion of deoxynucleotides in bacteria. *Nat. Microbiol.*, **7**, 1210–1220.
6. Tal,N., Millman,A., Stokar-Avihail,A., Fedorenko,T., Leavitt,A., Melamed,S., Yirmiya,E., Avraham,C., Brandis,A., Mehlman,T., *et al.* (2022) Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nat. Microbiol.*, **7**, 1200–1209.
7. Antine,S.P., Johnson,A.G., Mooney,S.E., Leavitt,A., Mayer,M.L., Yirmiya,E., Amitai,G., Sorek,R. and Kranzusch,P.J. (2024) Structural basis of Gabija anti-phage defence and viral immune evasion. *Nature*, **625**, 360–365.
8. Tuck,O.T., Adler,B.A., Armbruster,E.G., Lahiri,A., Hu,J.J., Zhou,J., Pogliano,J. and Doudna,J.A. (2024) Genome integrity sensing by the broad-spectrum Hachiman antiphage defense complex. *Cell*, **184**, 6914–6920.
9. Bobonis,J., Mitosch,K., Mateus,A., Karcher,N., Kritikos,G., Selkrig,J., Zietek,M., Monzon,V., Pfalz,B., Garcia-Santamarina,S., *et al.* (2022) Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems. *Nature*, **609**, 144–150.
10. Garb,J., Lopatina,A., Bernheim,A., Zaremba,M., Siksnys,V., Melamed,S., Leavitt,A., Millman,A., Amitai,G. and Sorek,R. (2022) Multiple phage resistance systems inhibit infection via SIR2-dependent NAD$^+$ depletion. *Nat. Microbiol.*, **7**, 1849–1856.
11. Duncan-Lowey,B., Tal,N., Johnson,A.G., Rawson,S., Mayer,M.L., Doron,S., Millman,A., Melamed,S., Fedorenko,T., Kacen,A., *et al.* (2023) Cryo-EM structure of the RADAR supramolecular anti-phage defense complex. *Cell*, **186**, 987–998.
12. Gao,Y., Luo,X., Li,P., Li,Z., Ye,F., Liu,S. and Gao,P. (2023) Molecular basis of RADAR anti-phage supramolecular assemblies. *Cell*, **186**, 999–1012.
13. Rousset,F., Yirmiya,E., Nesher,S., Brandis,A., Mehlman,T., Itkin,M., Malitsky,S., Millman,A., Melamed,S. and Sorek,R. (2023) A conserved family of immune effectors cleaves cellular ATP upon viral infection. *Cell*, **186**, 3619–3631.
14. Shen,Z., Lin,Q., Yang,X.-Y., Fosuah,E. and Fu,T.-M. (2023) Assembly-mediated activation of the SIR2-HerA supramolecular complex for anti-phage defense. *Mol. Cell*, **83**, 4586–4599.
15. Ka,D., Oh,H., Park,E., Kim,J.-H. and Bae,E. (2020) Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD$^+$ degradation. *Nat. Commun.*, **11**, 2816.
16. Tang,D., Chen,Y., Chen,H., Jia,T., Chen,Q. and Yu,Y. (2023) Multiple enzymatic activities of a Sir2-HerA system cooperate for anti-phage defense. *Mol. Cell*, **83**, 4600–4613.
17. Cohen,D., Melamed,S., Millman,A., Shulman,G., Oppenheimer-Shaanan,Y., Kacen,A., Doron,S., Amitai,G. and Sorek,R. (2019) Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature*, **574**, 691–695.
18. Tal,N., Morehouse,B.R., Millman,A., Stokar-Avihail,A., Avraham,C., Fedorenko,T., Yirmiya,E., Herbst,E., Brandis,A., Mehlman,T., *et al.* (2021) Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell*, **184**, 5728–5739.
19. Sabonis,D., Avraham,C., Lu,A., Herbst,E., Silanskas,A., Leavitt,A., Yirmiya,E., Zaremba,M., Amitai,G., Kranzusch,P.J., *et al.* (2024) TIR domains produce histidine-ADPR conjugates as immune signaling molecules in bacteria. bioRxiv doi: https://doi.org/10.1101/2024.01.03.573942, 03 January 2024, preprint: not peer reviewed.
20. Ofir,G., Herbst,E., Baroz,M., Cohen,D., Millman,A., Doron,S., Tal,N., Malheiro,D.B.A., Malitsky,S., Amitai,G., *et al.* (2021) Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature*, **600**, 116–120.
21. Millman,A., Bernheim,A., Stokar-Avihail,A., Fedorenko,T., Voichek,M., Leavitt,A., Oppenheimer-Shaanan,Y. and Sorek,R. (2020) Bacterial retrons function in anti-phage defense. *Cell*, **183**, 1551–1561.
22. Duncan-Lowey,B., McNamara-Bordewick,N.K., Tal,N., Sorek,R. and Kranzusch,P.J. (2021) Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. *Mol. Cell*, **81**, 5039–5051.
23. Johnson,A.G., Wein,T., Mayer,M.L., Duncan-Lowey,B., Yirmiya,E., Oppenheimer-Shaanan,Y., Amitai,G., Sorek,R. and Kranzusch,P.J. (2022) Bacterial gasdermins reveal an ancient mechanism of cell death. *Science*, **375**, 221–225.
24. Bernheim,A., Millman,A., Ofir,G., Meitav,G., Avraham,C., Shomar,H., Rosenberg,M.M., Tal,N., Melamed,S., Amitai,G., *et al.* (2021) Prokaryotic viperins produce diverse antiviral molecules. *Nature*, **589**, 120–124.
25. Zhang,T., Tamman,H., Coppieters 't Wallant,K., Kurata,T., LeRoux,M., Srikant,S., Brodiazhenko,T., Cepauskas,A., Talavera,A., Martens,C., *et al.* (2022) Direct activation of a

bacterial innate immune system by a viral capsid protein. *Nature*, **612**, 132–140.
26. LeRoux,M., Srikant,S., Teodoro,G.I.C., Zhang,T., Littlehale,M.L., Doron,S., Badiee,M., Leung,A.K.L., Sorek,R. and Laub,M.T. (2022) The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. *Nat. Microbiol.*, **7**, 1028–1040.
27. Wilkinson,M., Wilkinson,O.J., Feyerherm,C., Fletcher,E.E., Wigley,D.B. and Dillingham,M.S. (2022) Structures of RecBCD in complex with phage-encoded inhibitor proteins reveal distinctive strategies for evasion of a bacterial immunity hub. *eLife*, **11**, e83409.
28. Azam,A.H., Chihara,K., Kondo,K., Nakamura,T., Ojima,S., Tamura,A., Yamashita,W., Cui,L., Takahashi,Y., Watashi,K., *et al.* (2023) Viruses encode tRNA and anti-retron to evade bacterial immunity. bioRxiv doi: https://doi.org/10.1101/2023.03.15.532788, 15 March 2023, preprint: not peer reviewed.
29. Yirmiya,E., Leavitt,A., Lu,A., Ragucci,A.E., Avraham,C., Osterman,I., Garb,J., Antine,S.P., Mooney,S.E., Hobbs,S.J., *et al.* (2024) Phages overcome bacterial immunity via diverse anti-defence proteins. *Nature*, **625**, 352–359.
30. Leavitt,A., Yirmiya,E., Amitai,G., Lu,A., Garb,J., Herbst,E., Morehouse,B.R., Hobbs,S.J., Antine,S.P., Sun,Z.-Y.J., *et al.* (2022) Viruses inhibit TIR gcADPR signalling to overcome bacterial defence. *Nature*, **611**, 326–331.
31. Hobbs,S.J., Wein,T., Lu,A., Morehouse,B.R., Schnabel,J., Leavitt,A., Yirmiya,E., Sorek,R. and Kranzusch,P.J. (2022) Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature*, **605**, 522–526.
32. Huiting,E., Cao,X., Ren,J., Athukoralage,J.S., Luo,Z., Silas,S., An,N., Carion,H., Zhou,Y., Fraser,J.S., *et al.* (2023) Bacteriophages inhibit and evade cGAS-like immune function in bacteria. *Cell*, **186**, 864–876.
33. Cao,X., Xiao,Y., Huiting,E., Cao,X., Li,D., Ren,J., Fedorova,I., Wang,H., Guan,L., Wang,Y., *et al.* (2024) Phage anti-CBASS protein simultaneously sequesters cyclic trinucleotides and dinucleotides. *Mol. Cell*, **84**, 375–385..
34. Li,D., Xiao,Y., Fedorova,I., Xiong,W., Wang,Y., Liu,X., Huiting,E., Ren,J., Gao,Z., Zhao,X., *et al.* (2024) Single phage proteins sequester signals from TIR and cGAS-like enzymes. *Nature*, **635**, 719–727.
35. Jenson,J.M., Li,T., Du,F., Ea,C.-K. and Chen,Z.J. (2023) Ubiquitin-like conjugation by bacterial cGAS enhances anti-phage defence. *Nature*, **616**, 326–331.
36. Ho,P., Chen,Y., Biswas,S., Canfield,E., Abdolvahabi,A. and Feldman,D.E. (2023) Bacteriophage antidefense genes that neutralize TIR and STING immune responses. *Cell Rep.*, **42**, 112305.
37. Osterman,I., Samra,H., Rousset,F., Loseva,E., Itkin,M., Malitsky,S., Yirmiya,E., Millman,A. and Sorek,R. (2024) Phages reconstitute NAD^+ to counter bacterial immunity. *Nature*, **634**, 1160–1167.
38. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
39. Pinilla-Redondo,R., Shehreen,S., Marino,N.D., Fagerlund,R.D., Brown,C.M., Sørensen,S.J., Fineran,P.C. and Bondy-Denomy,J. (2020) Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat. Commun.*, **11**, 5652.
40. Samuel,B., Mittelman,K., Croitoru,S.Y., Ben Haim,M. and Burstein,D. (2024) Diverse anti-defence systems are encoded in the leading region of plasmids. *Nature*, **635**, 186–192.
41. Athukoralage,J.S., McMahon,S.A., Zhang,C., Grüschow,S., Graham,S., Krupovic,M., Whitaker,R.J., Gloster,T.M. and White,M.F. (2020) An anti-CRISPR viral ring nuclease subverts Type III CRISPR immunity. *Nature*, **577**, 572–575.
42. Samolygo,A., Athukoralage,J.S., Graham,S. and White,M.F. (2020) Fuse to defuse: a self-limiting ribonuclease-ring nuclease fusion for Type III CRISPR defence. *Nucleic Acids Res.*, **48**, 6149–6156.
43. Tesson,F., Hervé,A., Mordret,E., Touchon,M., Humières,C., Cury,J. and Bernheim,A. (2022) Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.*, **13**, 2561.
44. Néron,B., Denise,R., Coluzzi,C., Touchon,M., Rocha,E.P.C. and Abby,S.S. (2023) MacSyFinder v2: improved modelling and search engine to identify molecular systems in genomes. *Peer Community J.*, **3**, e28.
45. Krüger,D.H., Gola,G., Weisshuhn,I. and Hansen,S. (1978) The ocr gene function of bacterial viruses T3 and T7 prevents host-controlled modification. *J. Gen. Virol.*, **41**, 189–192.
46. Yan,Y., Zheng,J., Zhang,X. and Yin,Y. (2024) dbAPIS: a database of anti-prokaryotic immune system genes. *Nucleic Acids Res.*, **52**, D419–D425.
47. Mirdita,M., Steinegger,M. and Söding,J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*, **35**, 2856–2858.
48. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
49. Eddy,S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.*, **7**, e1002195.
50. Guo,J., Bolduc,B., Zayed,A.A., Varsani,A., Dominguez-Huerta,G., Delmont,T.O., Pratama,A.A., Gazitúa,M.C., Vik,D., Sullivan,M.B., *et al.* (2021) VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, **9**, 37 .
51. Néron,B., Littner,E., Haudiquet,M., Perrin,A., Cury,J. and Rocha,E.P.C. (2022) IntegronFinder 2.0: identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in *Klebsiella*. *Microorganisms*, **10**, 700.
52. Cury,J., Abby,S.S., Doppelt-Azeroual,O., Néron,B. and Rocha,E.P.C. (2020) Identifying conjugative plasmids and integrative conjugative elements with CONJscan. *Methods Mol. Biol.*, **2075**, 265–283.
53. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J., *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
54. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
55. Steenwyk,J.L., Iii,T.J.B., Li,Y., Shen,X.-X. and Rokas,A. (2020) ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.*, **18**, e3001007.
56. Nguyen,L.-T., Schmidt,H.A., von Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.
57. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.
58. Belogurov,A.A., Delver,E.P. and Rodzevich,O.V. (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J. Bacteriol.*, **175**, 4843–4850.
59. Pawluk,A., Bondy-Denomy,J., Cheung,V.H.W., Maxwell,K.L. and Davidson,A.R. (2014) A new group of phage anti-CRISPR genes inhibits the Type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *mBio*, **5**, e00896.
60. Tock,M.R. and Dryden,D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
61. Hille,F., Richter,H., Wong,S.P., Bratovič,M., Ressel,S. and Charpentier,E. (2018) The biology of CRISPR-Cas: backward and forward. *Cell*, **172**, 1239–1259.
62. Maffei,E., Shaidullina,A., Burkolter,M., Heyer,Y., Estermann,F., Druelle,V., Sauer,P., Willi,L., Michaelis,S., Hilbi,H., *et al.* (2021) Systematic exploration of *Escherichia coli* phage–host interactions with the BASEL phage collection. *PLoS Biol.*, **19**, e3001424.
63. Węglewska,M., Barylski,J., Wojnarowski,F., Nowicki,G. and Łukaszewicz,M. (2023) Genome, biology and stability of the

Thurquoise phage—a new virus from the *Bastillevirinae* subfamily. *Front. Microbiol.*, **14**, 1120147.
64. Doron,S., Melamed,S., Ofir,G., Leavitt,A., Lopatina,A., Keren,M., Amitai,G. and Sorek,R. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, **359**, eaar4120.
65. Mahendra,C., Christie,K.A., Osuna,B.A., Pinilla-Redondo,R., Kleinstiver,B.P. and Bondy-Denomy,J. (2020) Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat. Microbiol.*, **5**, 620–629.
66. Nomburg,J., Doherty,E.E., Price,N., Bellieny-Rabelo,D., Zhu,Y.K. and Doudna,J.A. (2024) Birth of protein folds and functions in the virome. *Nature*, **633**, 710–717.
67. Duan,N., Hand,E., Pheko,M., Sharma,S. and Emiola,A. (2024) Structure-guided discovery of anti-CRISPR and anti-phage defense proteins. *Nat. Commun.*, **15**, 649.
68. Kim,R.S., Karin,E.L. and Steinegger,M. (2024) BFVD—a large repository of predicted viral protein structures. bioRxiv doi: https://doi.org/10.1101/2024.09.08.611582, 09 September 2024, preprint: not peer reviewed.
69. van Kempen,M., Kim,S.S., Tumescheit,C., Mirdita,M., Lee,J., Gilchrist,C.L.M., Söding,J. and Steinegger,M. (2024) Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.*, **42**, 243–246.
70. Green,J., Stapleton,M.R., Smith,L.J., Artymiuk,P.J., Kahramanoglou,C., Hunt,D.M. and Buxton,R.S. (2014) Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. *Curr. Opin. Microbiol.*, **18**, 1–7.
71. Linder,J.U. (2010) cGMP production in bacteria. *Mol. Cell. Biochem.*, **334**, 215–219.

Received: August 9, 2024. **Revised:** October 30, 2024. **Editorial Decision:** November 6, 2024. **Accepted:** December 6, 2024