A widespread bacterial mobile genetic element encodes weapons against phages, bacteria, and eukaryotes

Tridib Mahata[1], Katarzyna Kanarek[1], Moran G. Goren[1], Eran Bosis[2,*], Udi Qimron[1,*], & Dor Salomon[1,*]

[1] Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[2] Department of Biotechnology Engineering, Braude College of Engineering, Karmiel, Israel

* For correspondence: EB, bosis@braude.ac.il; UQ, ehudq@tauex.tau.ac.il; DS, dorsalomon@mail.tau.ac.il

Keywords: T6SS, anti-phage, toxin, effector, antibacterial, nuclease, DNA mobility, integrase, GMT, Vibrio

Abstract

Bacteria are found in ongoing conflicts with rivals and predators, which lead to an evolutionary arms race and the development of innate and adaptive immune systems. Although diverse bacterial immunity mechanisms have been recently identified, many remain unknown, and their dissemination within bacterial populations is poorly understood. Here, we describe a widespread genetic element, defined by the Gamma-Mobile-Trio (GMT) proteins, that serves as a mobile bacterial weapons armory. We show that GMT islands have cargo comprising various combinations of secreted antibacterial toxins, anti-phage defense systems, and secreted anti-eukaryotic toxins. This finding led us to identify four new anti-phage defense systems encoded within GMT islands and reveal their active domains and mechanisms of action. We also find the phage protein that triggers the activation of one of these systems. Thus, we can identify novel toxins and defense systems by investigating proteins of unknown function encoded within GMT islands. Our findings imply that the concept of 'defense islands' may be broadened to include other types of bacterial innate immunity mechanisms, such as antibacterial and anti-eukaryotic toxins that appear to stockpile with anti-phage defense systems within GMT 'weapon islands'.

Introduction

Competition and predation contribute to bacterial evolution. For example, the type VI secretion system (T6SS), a missile-like dualling apparatus that delivers antibacterial toxins (i.e., effectors) directly into rival bacteria [1–4], was shown to shift the balance of bacterial populations and lead to the emergence of dominant strains [5–10]. Similarly, bacteriophages drive bacterial evolution and population dynamics [11–13]. Their ability to prey on bacteria leads to an arms race in which bacteria evolve or acquire anti-phage defense systems, while bacteriophages counteract these with anti-defense mechanisms [14]. Interestingly, anti-phage defense systems often cluster in so-called 'defense islands' [15,16], a phenomenon that has been used to identify dozens of diverse anti-phage defense systems in recent years [17–19]. Probably, many defense systems are yet to be revealed.

Mobile genetic elements (MGEs), such as plasmids, bacteriophages, and transposable elements, mediate horizontal gene transfer (HGT) and play a significant role in bacterial evolution [20–22]. They contain information that can enhance bacterial fitness [23–26]. For example, the *Vibrio cholerae* Cholera toxin, the Shiga toxin found in certain *Escherichia coli* strains, and a type III secretion system in *E. coli*, are encoded on a prophage [27–29]. Furthermore, anti-phage defense systems and secreted antibacterial toxins had been identified within MGEs and were shown to be horizontally shared [12,19,30–36].

Many MGEs are poorly understood and require further investigation to reveal their distribution and cargo. We previously reported that *Vibrio parahaemolyticus* island-6 (VPaI-6), a genomic island encompassing *vpa1270-vpa1254* in the pathogen *V. parahaemolyticus* RIMD 2210633 [37], contains an antibacterial DNase T6SS effector, VPA1263 (BAC62606.1), and its cognate immunity, Vti2 (WP_005477334.1) [38,39]. Here, we set out to further investigate VPaI-6 and similar genomic islands. We identified a protein trio that defines a widespread mobile genetic element with diverse cargo, including antibacterial T6SS effectors, anti-phage defense systems, and anti-eukaryotic toxins. Examining several genes with unknown function within these islands revealed four new families of anti-phage defense systems. Therefore, the described MGE is a bacterial armory containing diverse offensive and defensive weapons against potential predators, rivals, or hosts, which can be horizontally shared.

Results

Three conserved proteins define a widespread genomic island

Analysis of VPaI-6 revealed that the first three genes in this genomic island, encoding VPA1270, VPA1269, and VPA1268 (WP_005477115.1, WP_005477239.1, and WP_005477284.1, respectively), are annotated in the NCBI protein family models database as a Gamma-Mobile-Trio (GMT) system (**Fig. 1a**). We found similar co-occurring GMT proteins encoded in many assembled bacterial genomes, including those of pathogenic bacteria, such as *Vibrio*, *Pseudomonas*, *Burkholderia*, *Acinetobacter*, *Enterobacter*, and *Aeromonas* (**Fig. 1b** and **Supplementary Dataset S1**). However, very little information is available on the function of these proteins and the genomic islands containing them [37]. Therefore, we set out to further investigate the GMT system.

According to the NCBI Conserved Domain Database [40], the trio's first gene, named *vpa1270* or *gmtY*, encodes a site-specific tyrosine recombinase with a domain belonging to the DNA_BRE_C superfamily. The second gene, *vpa1269*, encodes a phage-like integrase with a domain belonging to the Phage_Integr_2 superfamily, which is predicted to mediate unidirectional site-specific recombination. The third gene, named *vpa1268* or *gmtX*, encodes a protein of unknown function. Sequence analyses revealed a similarity between the N-terminus of GmtX and the DNA binding domain of the partition protein StbA of plasmid R388 (according to HHpred [41]; 93.44% probability between amino acids 3-75 of GmtX and amino acids 1-68 of StbA [PDB: 7PC1_A]). These predicted activities suggest that the GMT proteins play a role in DNA mobility.

Predicted insertion sites of GMT islands comprise inverted or direct repeats

To investigate GMT islands, we focused our analyses on members of the genus *Vibrio* (**Fig. 2**). In all cases, GMT proteins are encoded at the beginning of an apparent genomic island of varying length and gene content: between 7 and 40 kb long, with an average of ~20 kb (**Supplementary Fig. S1**); encoding 5-44 genes, including the GMT genes (**Fig. 2**). By comparing the flanking regions of 32 different types of GMT-containing islands (differing in length and gene content) to closely related *Vibrio* genomes lacking the respective island, we identified their putative insertion sites (**Supplementary Dataset S2**). Most of the GMT islands appear to have been inserted in intergenic regions or within tRNA genes; inverted repeats flank the putative insertion sites and often remain flanking the island after insertion. In contrast, some GMT islands appear to have been inserted inside a protein-encoding gene, such as *greB*, *radC*, a gene encoding DUF1244, or *tauE*; for the first three, the putative insertion sites comprise direct repeat sequences rather than inverted repeats. Phylogenetic analysis of GmtY proteins revealed that they cluster following the type of putative insertion site (**Supplementary Fig. S2**). Interestingly, GMT islands inserted within *greB*, the gene encoding DUF1244, or *tauE* encode a homolog of the disrupted gene, possibly to compensate for its loss (**Supplementary Fig. S3**). Notably, a genome may contain multiple GMT islands (**Supplementary Dataset S1**).

The insertion of a GMT island within a repeat-containing site appears to have diverse outcomes. In some instances, we identified a 'clean insertion', where the recognition site split to become the sequences flanking the GMT island in the genome. In other cases, short spacer sequences between the repeats appear to have been duplicated, resulting in both direct and inverted repeats flanking the GMT island. The nature of the sequences flanking a GMT island is possibly random or dictated by the insertion site rather than by the GMT proteins. This is exemplified in *V. alginlolytics*, in which two identical GMT islands co-occur, one on the chromosome and another on a plasmid. The intergenic chromosomal and plasmid putative insertion sites share inverted repeat sequences with an identical 5' AAGAGC 3' core and a 14 bp spacer. Despite their identity, the chromosomal GMT island is flanked by both direct and inverted repeats, whereas the plasmid GMT island is flanked only by an inverted repeat (i.e., a 'clean insertion') (**Supplementary Fig. S4**). Taken together, our findings suggest that GMT proteins define a mobile genetic element with diverse cargo, which inserts into specific regions comprising repeat sequences.

Many GMT islands encode an antibacterial effector

Next, we investigated the cargo of GMT islands in vibrios. We determined whether T6SS effector and immunity pairs, such as the one found in VPaI-6 (i.e., VPA1263 and Vti2 [39]), are ubiquitous. Interestingly, we identified a T6SS effector and immunity pair in 8 out of 32 GMT island types found in assembled *Vibrio* genomes (**Fig. 2**). They comprise predicted T6SS cargo effectors with T6SS-specific MIX [38,39,42,43] or FIX [44] domains and a specialized effector containing an N-terminal PAAR repeat-containing domain [45]. In some of these T6SS effectors, we identified C-terminal nuclease or phospholipase toxin domain. A known or predicted immunity gene resides immediately downstream of all eight predicted T6SS effectors. We also identified a predicted antibacterial toxin (WP_159066141.1) containing an N-terminal Rearrangement hotspot (Rhs) repeat domain [46] and a C-terminal DUF4258, which is found at C-termini of polymorphic toxins [47]. The protein encoded by the upstream gene (WP_159066140.1) is possibly linked to this predicted toxin, since it is similar to the central region of the RhsA T6SS effector from *Pseudomonas protegens* Pf-5 [48] (according to HHpred [41]; 100% probability between the entire sequence of WP_159066140.1 and amino acids 209-1053 of RhsA [PDB:7Q97_A]). However, since Rhs repeat-containing proteins are not strictly T6SS effectors [46], we cannot determine whether this toxin is associated with the T6SS or another secretion system. Altogether, our findings indicate that ~28% (9 out of 32) of vibrios' GMT island types harbor a secreted antibacterial effector.

Anti-phage defense systems are frequent in GMT islands

VPaI-6 harbors a Septu type I anti-phage defense system [15], encoded by *vpa1261* and *vpa1260* (**Fig. 2**). The presence of an antibacterial effector and an anti-phage defense system within the same GMT island prompted us to investigate whether other GMT islands carry similar cargo. Using the PADLOC [49] and DefenseFinder [50] tools, we identified anti-phage defense systems in ~46% (15 of the 32) of the vibrios' GMT island types (**Fig. 2**). At least six of the GMT island types are so-called 'defense islands' [15,16] containing multiple anti-phage defense systems. Moreover, four GMT island types harbor both an anti-phage defense system and an antibacterial T6SS effector and immunity pair. Therefore, GMT islands appear to be armories stockpiled with diverse weapon arsenals, both antibacterial and anti-phage.

To determine whether GMT islands outside of the *Vibrio* genus also contain weapon arsenals, we analyzed the cargo of randomly selected GMT islands found in other bacteria. We observed that many islands harbor predicted anti-phage defense systems. Remarkably, although we did not find an enrichment of antibacterial T6SS effectors in GMT islands outside the *Vibrio* genus, we did find examples of secreted anti-eukaryotic toxins (e.g., a Tc toxin [51] and a type III secretion system effector, HopAU1 [52]) (**Supplementary Fig. S5**). Taken together, our findings indicate that GMT islands stockpile offensive and defensive weapons against attacking phages or competing bacteria, as well as against eukaryotic hosts or predators.

Identifying four new anti-phage defense systems in GMT islands

Based on the analyses above, we hypothesized that GMT islands harbor yet-unknown anti-phage defense systems. We assembled a list of 13 genes and operons found within *Vibrio* GMT

islands, which we predicted are novel anti-phage defense systems and therefore named them GAPS (GMT-encoded Anti-Phage System) (**Fig. 2**). These candidates were selected based on a wide distribution of their homologs in non-*Vibrio* bacterial species or because we found that they or their homologs reside close to known anti-phage defense systems.

To investigate the anti-phage activities of these candidates, we cloned them into a low copy-number expression plasmid under an arabinose-inducible promoter. *E. coli* K-12 BW25113 strains containing the cloned plasmids were challenged with five coliphages (T7, T4, $P1_{vir}$, T5, and $lambda_{vir}$), and the efficiency of plating (EOP) of the different phages was compared to *E. coli* harboring a control empty plasmid. Remarkably, four candidates: GAPS1, GAPS2, GAPS4, and GAPS6 provided significant protection against at least one phage, as evident by a reduction of at least three orders of magnitude in the number of visible plaques developing on a lawn of bacteria (i.e., the EOP) (**Fig. 3**). Four additional GAPSs (GAPS9, GAPS10, GAPS11, and GAPS13) led to a mild protection phenotype, observed as a low fold decrease in EOP or the appearance of smaller plaques. A phenotype of small plaques may be attributed to a non-specific effect caused by mild toxicity of the expressed proteins and therefore we did not pursue investigations of such systems. Notably, three candidates were considerably toxic to *E. coli* upon inducible over-expression (GAPS3, GAPS5, and GAPS8); therefore, we could not determine their anti-phage activity. GAPS7 and GAPS12 did not provide protection against the tested coliphages. These results prompted us to further investigate three of the newly identified GAPS.

GAPS1 is a predicted nuclease activated by a phage capsid protein

GAPS1 (WP_005477165.1) is a single protein containing a predicted nuclease domain of the PD-(D/E)xK superfamily [53,54] (**Fig. 4a**). Its homologs are widespread in Gram-negative and Gram-positive bacteria (**Supplementary Fig. S6** and **Supplementary Dataset S3**). To determine whether the nuclease domain is required for the anti-phage activity of GAPS1, we substituted D313 and K328 within the conserved PD-(D/E)xK active site with alanines. These substitutions abolished the defense activity against T7 phage, further establishing the role of GAPS1 as an anti-phage defense system and supporting the hypothesis that it uses the nuclease domain to mediate anti-phage defense (**Fig. 4b**).

Most members of the PD-(D/E)xK superfamily are DNases [53,54]. We hypothesized that degradation of the host DNA might be the mechanism by which the infection aborts and the bacterial colony is protected. However, we did not detect a difference in the rate of genomic DNA degradation between T7 phage-infected *E. coli* expressing GAPS1 or containing an empty plasmid (**Supplementary Fig. S7a**). Therefore, we cannot conclude at this time what is the target of GAPS1 activity.

Next, we sought to identify the trigger of GAPS1, and we hypothesized that escape mutants of an attacking T7 phage mutate the protein that triggers GAPS1 to avoid system activation. Therefore, we sequenced the genome of two T7 escape phages that formed plaques in the presence of GAPS1. We identified a mutation in the capsid protein (Gp10)-encoding gene 10 of both isolates; one isolate carried an I217T mutation, and the other a V247A mutation (**Supplementary Table S1**). Notably, I217 and V247 are closely positioned in the folded capsid protein (**Supplementary Fig. S8**). The isolates also contained mutations in other genes, yet these mutations were not common to both isolates. To confirm the hypothesis that the capsid

protein of T7 phage induces GAPS1-mediated defense, we next generated the two mutations identified in gene 10 in naïve T7 phages. We validated that additional mutations that we identified in the escape isolates are not present in these newly constructed phages, and tested the ability of GAPS1 to defend against them. We observed that the EOP of the mutant T7 phages was ~1 on *E. coli* expressing GAPS1 compared to *E. coli* containing an empty plasmid, indicating that GAPS1 could not defend against these mutants (**Fig. 4c**). This finding suggests that the capsid protein, and specifically residues I217 and V247, is the trigger that activates defense mediated by GAPS1.

To investigate whether the capsid protein triggers GAPS1 from outside, i.e., while injecting DNA into the bacterium, we used the T7 phage as a transducing particle. We packaged a plasmid encoding a T7-packaging signal and an antibiotic resistance gene into T7 particles encoding the wild-type gene 10. We then used these transducing particles to transduce bacteria encoding or lacking GAPS1 and selected the transductants on antibiotic-containing LB plates. We expected that if the wild-type capsid protein triggers GAPS1 upon DNA injection, there would be fewer transduced colonies in bacteria encoding GAPS1 than in the control bacteria lacking it. However, we observed a similar number of transduced colonies in the presence or absence of GAPS1 (**Fig. 4d**). This result suggests that the capsid protein does not trigger GAPS1 while injecting the DNA into the bacterial cells. In contrast, over-expression of the wild-type capsid protein (Gp10) from a plasmid inhibited the growth of *E. coli* expressing a chromosomally inserted GAPS1 (**Fig. 4e-f**). These results indicate that GAPS1 is activated after the capsid protein had been expressed during later stages of T7 phage infection (**Fig. 4g**). Our findings imply that GAPS1 is activated as a "last resort" option only after the phage has produced the capsid protein and not upon initial detection of the phage capsid. Notably, we did not detect *E. coli* genomic DNA degradation upon GAPS1 activation by a plasmid-expressed capsid protein (**Supplementary Fig. S7b**), suggesting that the PD-(D/E)xK domain-containing GAPS1 does not function as a non-specific DNase upon activation.

GAPS4 contains a DNase that cleaves bacterial DNA upon phage infection

GAPS4 comprises two genes encoding WP_055466293.1 (GAPS4a) and WP_055466294.1 (GAPS4b). The first protein contains a domain predicted to be a diverged member of the PD-(D/E)xK superfamily of nucleases, in which the D/E residue is substituted with a serine [55] (**Fig. 5a**). Notably, both genes are essential for GAPS4-mediated defense as each gene alone could not protect against the lambda$_{vir}$ phage (**Fig. 5b**). Indeed, these two proteins are encoded together in diverse Gram-negative and Gram-positive bacteria; occasionally, they are fused into a single gene (**Supplementary Fig. S9** and **Supplementary Dataset S4**).

A structural model based on AlphaFold2 [56,57] predicts, with a high degree of confidence, that GAPS4a and GAPS4b interact and form a heterodimer (**Fig. 5c** and **Supplementary Fig. S10**). Two basic lobes, one adjacent to the position of the predicted nuclease domain, are found ~70 Å apart, which is approximately the length of two turns of the double helix (i.e., the distance between two non-adjacent grooves) [58]. This prediction supports a role for GAPS4 in DNA binding.

To determine whether the PD-(D/E)xK nuclease domain plays a role in the defense against phages, we mutated the predicted active site residue D59 to alanine. As expected, the mutation abolished the defense activity of GAPS4, and the EOP of lambda$_{vir}$ phage was

restored (**Fig. 5b**). We hypothesized that degradation of the host DNA by the PD-(D/E)xK nuclease may be the mechanism by which GAPS4 protects the bacterium. To examine this, we monitored the degradation of genomic DNA in *E. coli* expressing GAPS4 and infected with the lambda$_{vir}$ phage. We used *E. coli* expressing an empty plasmid or the inactive GAPS4^{D59A} mutant as controls. We observed gradual genomic DNA degradation in the *E. coli* expressing GAPS4 but not in the control samples. The degradation occurred only in response to lambda$_{vir}$ infection (**Fig. 5d**). These results suggest that GAPS4 is an abortive infection system degrading host cell DNA upon phage infection (**Fig. 5e**). We speculate that the specific activation of the system requires a phage component. However, we could not isolate escape phages from the culture.

GAPS6 contains a predicted PINc domain required for anti-phage defense

GAPS6 comprises two genes encoding WP_248387294.1 (GAPS6a) and WP_248387295.1 (GAPS6b) (**Fig. 6a**). These two proteins are encoded together in diverse Gram-negative bacteria (**Supplementary Fig. S11** and **Supplementary Dataset S5**). The expression of GAPS6a alone was insufficient to provide defense against lambda$_{vir}$, resulting in an EOP comparable to that of an empty plasmid (**Fig. 6b**). This result indicates that GAPS6b is essential for the defense activity of GAPS6. However, we could not determine whether GAPS6b is sufficient since we could not clone it into an expression plasmid alone, possibly due to toxicity issues. GAPS6b is a large protein of 1086 amino acids. It contains predicted TPR repeats, which may play a role in ligand binding [59], at the N-terminus and a predicted PINc domain at the C-terminus (according to HHpred [41]; for TPR, 99% probability between amino acids 18-227 of GAPS6b and amino acids 593-814 of PDB:8F5O_D; for PINc, 97% probability between amino acids 760-913 of GAPS6b and amino acids 2-214 of PIN_5 [PF08745.14]) (**Fig. 6a**). PINc domains are RNases with several conserved acidic residues. They have been implicated as toxins in bacterial toxin-antitoxin modules [60]. To determine whether the PINc domain plays a role in the defense activity of GAPS6, we substituted the conserved D887 of this domain with an alanine. The mutant could not provide protection against lambda$_{vir}$ (**Fig. 6b**), confirming that the PINc domain is required for the anti-phage defense activity of GAPS6.

Discussion

In this work, we expand the notion of anti-phage 'defense islands' [15,16] and propose to consider bacterial 'weapon islands'—genomic islands that contain diverse defensive and offensive weapons against attacking phages, rival bacteria, and eukaryotic predators or hosts. We show that widespread MGEs defined by the GMT proteins are such 'weapon islands'.

Although anti-phage defense systems, secreted antibacterial toxins, and anti-eukaryotic toxins have all been previously described as cargo of MGEs [12,19,35,36,27–34], the GMT system is unique in that it appears to be an armory containing various types of bacterial weapons. It is intriguing that we identified an enrichment of antibacterial T6SS effectors in *Vibrio* GMT islands and not in other bacteria. This enrichment appears to be another example supporting previous observations describing a significant contribution of horizontal gene transfer to the dynamics of T6SS effector repertoires within this genus [32–35,61]. It may also be related to the ecology of these aquatic bacteria, which reside in heterogeneous and complex microhabitats [62,63].

Notably, this work includes only a conservative list of GMT islands. We report and investigate GMT islands found within complete genomes in the NCBI RefSeq database and that are highly similar to those found in the *V. parahaemolyticus* RIMD 2210633 VPaI-6. Thus, GMT islands are probably more ubiquitous and diverse than described here. Moreover, we hypothesize that additional bacterial weapons can be identified upon further investigating the cargo in non-vibrios islands or in islands containing more distant GMT protein homologs.

Our analyses revealed that GMT islands probably insert into specific, repeat-containing sites within bacterial genomes. Several pieces of evidence support the conclusion that the location of GMT island insertion is not random. First, we found that GMT islands that disrupt a protein-coding gene contain a possibly compensatory homolog of this gene in their cargo. Second, we showed that two identical GMT islands co-occur within the same genome at insertion loci that share an identical core of inverted repeats. This observation suggests that each GMT island can only insert into sites containing specific repeat sequences. We will investigate the mechanism governing the mobility of GMT islands and the role of the GMT proteins in this process in future work. We predict that GMT proteins may be utilized as molecular genetic tools to mediate the insertion of large DNA cassettes into specific locations within bacterial genomes.

The various weapons stockpiled in GMT islands led us to hypothesize that many genes of unknown function within these islands are bacterial weapons. To test this hypothesis, we focused on identifying novel anti-phage defense systems. Out of a small pool of 13 selected candidates, we revealed four new anti-phage defense systems and further investigated three of these. Although we could not confirm their role against phages, the nine other candidate systems we investigated may also play a role in anti-phage defense. Our candidates originate from vibrios, yet we examined them in *E. coli* as a surrogate platform against a minimal set of coliphages. Therefore, these additional candidates may protect against a *Vibrio* phage we did not examine. We may have also neglected to include a regulatory or accessory component endogenously found in vibrios when cloning the candidate genes into plasmids under arabinose-inducible promoters.

The first new anti-phage defense system we investigated is GAPS1, which defends against the T7 phage. It is a single protein that requires an intact PD-(D/E)xK nuclease domain for defense activity. However, the PD-(D/E)xK nuclease does not appear to target the host genomic DNA; therefore, its target is yet to be revealed. We show that the expression of the phage capsid protein, Gp10, activates GAPS1-mediated defense. However, this activation does not occur when a phage encoding Gp10 injects its DNA into the host. To our knowledge, this is the first time that the possibility of capsid-mediated anti-phage defense activation during the injection step has been experimentally examined. We propose that this later activation of the defense system is a naturally evolved trait preventing bacteria from triggering a catastrophic mechanism unless all other defense systems have failed. If the defense system could be triggered from outside the cell, bacteria might occasionally trigger an unnecessary detrimental program against phages that do not naturally infect them. Nevertheless, we cannot rule out the possibility that the capsid protein in our assay is not detected when infecting the bacterial cell, and in cases that it does, it can activate the system. Notably, phage capsid proteins appear to be a common trigger of anti-phage defense systems [64–66].

We also identified GAPS4, a defense system comprising two proteins essential for its activity, possibly forming a heterodimer. Identifying GAPS4 homologs in which the two genes

fused into one further supported this prediction. Based on a structural model of the heterodimer, we hypothesized that the two proteins combine to form a DNA-binding complex. Indeed, we show that GAPS4 degrades the bacterial genomic DNA using a PD-(D/E)xK nuclease domain upon lambda phage infection. This activity suggests that GAPS4 uses an abortive infection mechanism of action. Notably, we failed to isolate phage mutants that escape the GAPS4-mediated defense. We hypothesize that several phage components activate the system, and thus simultaneous mutations in all of these components are infrequent, making it virtually impossible to isolate escape mutants. Alternatively, the determinant activating the system may be essential to phage infection, so its mutations are unable to form a plaque.

The third system that we characterized is GAPS6. It also comprises two proteins, although we could not determine whether both are required for its activity. Nevertheless, a single mutation in a predicted PINc domain of GAPS6b abrogated the defense activity of GAPS6, indicating this domain is essential. Although PINc domains are predicted RNases [60], we have yet to identify GAPS6's mechanism of action.

In conclusion, this work demonstrates that GMT islands are bacterial armories. Our results indicate that one should regard 'defense islands' as possible 'weapon islands'. We suggest considering secreted antibacterial toxins and anti-eukaryotic toxins targeted at eukaryotic predators as part of the bacterial innate immunity repertoire. Thus, further examination of cargo genes encoding proteins of unknown function within GMT islands and islands containing more distant GMT homologs will reveal additional immunity mechanisms. It may also shed light on evolutionary constraints for different bacterial genera and bacterial population dynamics.

Materials and Methods

Strains and media: The *E. coli* strains used in this study are listed in **Supplementary Table S2**. The strains were grown in LB (1% [wt/vol] tryptone, 0.5% [wt/vol] yeast extract, and 0.5% [wt/vol] NaCl) or 2xYT (1.6% [wt/vol] tryptone, 1% [wt/vol] yeast extract, 0.5% [wt/vol] NaCl) at 37°C. To prepare solid plates, media were supplemented with 1.5% (wt/vol) agar. When required, media were supplemented with 35 µg/mL chloramphenicol, 50 µg/mL kanamycin, or 100 µg/mL ampicillin to maintain plasmids. To induce the expression from Pbad promoters, 0.04% or 0.2% (wt/vol) L-arabinose was added to the media, as indicated.

Plasmid construction: Plasmids were constructed with standard molecular biology techniques using the Gibson Assembly methods [67]. The Gibson Assembly master mix was obtained from NEB (E2611S). DNA fragments were amplified by PCR from bacterial genomic DNA or from DNA synthesized by TWIST Bioscience, and Gibson Assembly ligations were carried out according to the manufacturer's instructions. Commercially synthesized DNA, plasmids, and primers that were used in this study are listed in **Supplementary Table S3**, **S4**, and **S5**, respectively.

Plaque Assays: The phages used in this study are listed in **Supplementary Table S6**. Phages were propagated on *E. coli* BW25113. To determine the effect of 13 putative defense system (GAPS1-13) against 5 different coliphages (T4,T5,T7, lambda$_{vir}$ and P1$_{vir}$), *E. coli* BW25113 strains harboring the indicated pBAD33.1-based plasmids were grown overnight in LB supplemented with chloramphenicol and 0.2% (wt/vol) D-glucose (to repress expression from

the Pbad promoter) at 37°C. Overnight cultures were washed twice to remove remaining glucose, and then 350 µL of each culture was mixed with 7 mL of 0.7% (wt/vol) molten agar supplemented with 0.2% (wt/vol) L-arabinose, 5 mM $MgCl_2$, 0.1 mM $MnCl_2$, and 5 mM $CaCl_2$. The mixture was poured onto a 1.5% (wt/vol) agar plate supplemented with chloramphenicol and 0.2% (wt/vol) L-arabinose, and the plates were left to dry. Ten-fold serial dilutions of all the phages were prepared, and 7.5 µL of each dilution were spotted on the dried plate. The plates were incubated overnight at 37°C. On the following day, the plaques were counted and plaque forming units (PFU/mL) were calculated. For dilution spots in which no individual plaques were visible but a faint zone of lysis was observed, the dilution we considered as having ten plaques [68]. Plaque assays with *E. coli* containing plasmids for expression of GAPS mutants or other variations were performed in the same manner.

Isolation of escape phages: *E. coli* BW25113 harboring an empty pBAD33.1 or one encoding GAPS1 (pBAD33.1-GAPS1) were grown overnight in LB supplemented with chloramphenicol and 0.2% (wt/vol) D-glucose at 37°C. The cells were washed twice and mixed with 0.7% (wt/vol) molten agar supplemented with 0.2% (wt/vol) L-arabinose, and poured onto a 1.5% (wt/vol) agar plate supplemented with 0.2% (wt/vol) L-arabinose. After the agar dried, 10-fold serially diluted T7 phage was spotted onto the plate, and the plate was incubated overnight at 37°C. Individual plaques growing in the highest dilution on the plates containing *E. coli* expressing GAPS1 were isolated and propagated on naïve *E. coli* BW25113 cells harboring pBAD33.1-GAPS1 in LB supplemented with chloramphenicol and 0.2% (wt/vol) L-arabinose at 37°C. The escape phages were confirmed by spotting them on plates with *E. coli* expressing GAPS1.

For phage genomic DNA isolation, high titers of escape phages and a parental wild-type phage were prepared (~$1.0E^{11}$ PFU/mL). Approximately 40 mL of lysates of each escape phage were mixed with 10% (wt/vol) PEG 8000 and 3 M NaCl, and incubated overnight at 4°C. The lysate-PEG mixture was then centrifuged at 10000 x *g* for 15 minutes to collect the phage pellet. The pellet was re-suspended using resuspension buffer from the Presto™ Mini gDNA isolation kit, and the phage genomic DNA was isolated following the manufacturer's protocol.

Illumina whole genome sequencing was carried out at SeqCenter (Pittsburgh, PA, USA; https://www.seqcenter.com/). Sample libraries were prepared using the Illumina DNA Prep kit and IDT 10 bp UDI indices, and sequenced on an Illumina NextSeq 2000, producing 2 x 151 bp reads. Mutations were identified using variant calling (SeqCenter). Only mutations that were found in escape mutant genomes and not in the sequenced parental T7 phage are reported.

Constructing T7 phage mutants: T7 mutants were constructed using the pORTPHAGE method [69], a MAGE [70,71]-based system for the mutagenesis of bacteriophages. Briefly, *E. coli* K-12 strain harboring the pORTMAGE-Ec1 plasmid was grown to reach early log phase (OD_{600} ~0.3). Then, 1 mM m-toluic acid was added to induce the expression of recombineering proteins expression. The cells were made electrocompetent, and then transformed with mutating oligonucleotides. After electroporation, the culture was infected with a wild-type T7 phage and incubated until complete lysis occurred. The final lysate was cleared using chloroform, diluted, and then plated using the soft agar overlay method to screen for individual mutated plaques. Single plaques were picked, suspended in LB, and used as templates for PCR amplification and sequencing to identify mutants.

Chromosomal integration of GAPS1: GAPS1 was introduced into the chromosome of *E. coli* BW25113 in place of *ydhQ* (*ydhQ*::GAPS1) using the red recombination system, as previously

described [72]. Briefly, *E. coli* BW25113 cells harboring pSim6 were grown overnight in LB supplemented with ampicillin at 30°C. Overnight cultures were diluted 1:100 in 35 mL of fresh LB supplemented with ampicillin, and grown to an OD_{600} of ~0.5. The red recombinase system was then heat induced for 20 minutes in a shaking water bath at 42°C. Immediately after induction, the cells were chilled on ice and pelleted by centrifugation. The cell pellets were washed thrice with ice-cold water and resuspended in 200 µL of ice-cold water. The gene encoding GAPS1 under P_{BAD} promoter control, along with a kanamycin-resistance cassette, was amplified together with flanking sequences identical to 50 bp upstream and downstream of the chromosomal *ydhQ*. The amplified DNA was treated with DpnI, and then ran on an agarose gel and purified. The purified DNA was electroporated into *E. coli*, and bacteria were allowed to recover in 2xYT broth supplemented with 0.2% (wt/vol) D-glucose for 2 hours at 30°C. The transformed cells were then plated onto a 1.5% (wt/vol) agar plate supplemented with 25 µg/mL kanamycin. The integration of GAPS1 was verified by PCR. Bacteria were cured of the pSIM6 plasmid, and the recombinant cells were electroporated with pCP20 plasmid. The kanamycin cassette was flipped out by inducing the pCP20 plasmid at 42°C.

Toxicity Assay: *E. coli* BW25133 Δ*ydhQ* and *E. coli* BW25133 *ydhQ*::GAPS1 cells harboring an empty plasmid or a plasmid for the arabinose-inducible expression of the T7 phage gene10 (encoding the Gp10 capsid protein) were grown overnight at 37°C in LB supplemented with kanamycin and 0.2% (wt/vol) D-glucose. Overnight cultures were diluted to an OD_{600} = 0.02 in fresh LB supplemented with kanamycin, and 200 µL were transferred into 96 well plates in triplicates. Cells were grown under continuous shaking (205 RPM) in a Tecan Infinite M Plex plate reader at 37°C. After 2 hours, the expression of Gp10 and GAPS1 were induced by adding L-arabinose to a final concentration of 0.2% (wt/vol). OD_{600} readings were acquired every 10 minutes.

To determine cell viability after induction, bacteria were collected at the indicated timepoints after arabinose addition. Ten-fold dilutions of each culture were spotted on 1.5% (wt/vol) agar plates supplemented with kanamycin and 0.2% (wt/vol) D-glucose (to repress arabinose-induced expression). The plates were incubated overnight at 37°C, and CFU/mL of each culture were determined the following day.

Transduction assays: Transduction assays were carried out as previously described [73]. Briefly, *E. coli* BW25113 cells harboring the T7-packaging signal-containing pIYPE114 plasmid were grown overnight in LB supplemented with kanamycin at 37°C. The cells were diluted 1:100 in fresh LB supplemented with kanamycin and grown at 37°C. Upon reaching OD_{600} of ~0.5, the cells were infected with a mutant T7 phage Δ(11-12-17) [73] at a multiplicity of infection (MOI) of 2. Infected cells were grown until lysis occurred (1.5-2 hours). Then, chloroform (200 µL) was added and the lysates were centrifuged to clear cell debris. The resulting supernatants comprised a mixture of T7 phage particle with either a T7 genome or a pIYPE114 plasmid containing a T7-packaging signal. For the transduction assay, a *E. coli* BW25113 Δ*trxA* strain harboring either pBAD33.1 (empty vector) or pBAD33.1-GAPS1 were grown overnight in LB supplemented with chloramphenicol and 0.2% (wt/vol) D-glucose at 37°C. Overnight cultures were washed twice and diluted 1:100 in 3 mL of fresh LB supplemented with chloramphenicol. Cells were grown for an hour at 37°C, and GAPS1 expression was induced with the addition of 0.2% (wt/vol) L-arabinose and an additional hour of growth at 37°C. The induced cells were mixed at a 1:1 ratio (vol/vol) with 10-fold serial dilutions of the prepared lysates described above. Transduction was carried out for 1 hour at 37°C with shaking. Then, the cells were

spotted onto 1.5% (wt/vol) agar plate supplemented with kanamycin, and incubated overnight at 37°C. On the next day, the transduced colonies were counted and the CFU/mL was calculated.

In vivo genomic DNA degradation upon expression of Gp10 from a plasmid: *E. coli* BW25133 Δ*ydhQ* and *E. coli* BW25133 *ydhQ*::GAPS1 cells containing an empty plasmid or a plasmid for the arabinose-inducible expression of the T7 phage Gp10 were grown overnight at 37°C in LB supplemented with kanamycin and 0.2% (wt/vol) D-glucose. Overnight cultures were washed twice and diluted 1:100 in fresh LB supplemented with kanamycin. The cells were grown to an OD_{600} of ~0.5, and then the media were supplemented with 0.2% (wt/vol) L-arabinose to induce expression from Pbad promoters. To monitor the integrity of genomic DNA, 1 OD_{600} of cells were taken out at the 0 and 2-hour timepoints after arabinose induction, and genomic DNA was isolated using the Presto™ Mini gDNA isolation kit. Equal volumes of genomic DNA extracts were resolved on 0.8% (wt/vol) agarose gel and the DNA was visualized using a Fusion Solo X (Vilber).

In vivo genomic DNA degradation upon T7 phage infection: *E. coli* BW25113 cells containing an empty plasmid (pBAD33.1) or a plasmid for the arabinose-inducible expression of GAPS1 (pBAD33.1-GAPS1) were grown overnight in LB supplemented with chloramphenicol and 0.2% (wt/vol) D-glucose at 37°C. Overnight cultures were washed twice, diluted 1:100 in 3 mL of fresh LB supplemented with chloramphenicol, and grown for an hour at 37°C. L-arabinose (0.2% [wt/vol]) was added to the cultures to induce expression from Pbad promoters, and cells were grown for an additional hour at 37°C. The cultures were transferred to ice for 30 minutes, and then the bacteria were infected with T7 phage at an MOI of 1.5. The cultures were then incubated with shaking at 30°C, and 0.5 mL of the cultures were collected at 0, 20, and 30 minutes after T7 phage addition. The cells were pelleted and frozen in liquid nitrogen. Genomic DNA was isolated using the Presto™ Mini gDNA isolation kit. Equal volumes of genomic DNA were then resolved on a 0.8% (wt/vol) agarose gel and visualized as detailed above.

In vivo genomic DNA degradation upon lambda$_{vir}$ infection*:* *E. coli* BW25113 cells containing an empty plasmid (pBAD33.1) or a plasmid for the arabinose-inducible expression of GAPS4 (pBAD33.1-GAPS4) or its inactive mutant (pBAD33.1-GAPS4^{D59A}) were grown overnight in LB supplemented with chloramphenicol and 0.2% (wt/vol) D-glucose at 37°C. Overnight cultures were washed twice, diluted 1:100 in 3.5 mL of fresh LB supplemented with chloramphenicol, and grown for an hour at 37°C. L-arabinose (0.04% [wt/vol]) was added to the cultures to induce expression from Pbad promoters, and cells were grown for an additional hour at 37°C. The cells were then infected with lambda$_{vir}$ at an MOI of ~10, and the cultures were incubated at 37°C with shaking. At the 0, 45, and 90 minutes timepoints after infection, 1 mL of were collected from each culture, washed with PBS, and frozen in liquid nitrogen. Genomic DNA was isolated using the Presto™ Mini gDNA isolation kit. Equal volumes of genomic DNA were then resolved on a 0.8% (wt/vol) agarose gel and visualized as detailed above.

Identifying GMT proteins in bacterial genomes: First, the position-specific scoring matrices (PSSMs) of VPA1270, VPA1269, and VPA1268 were constructed using the full-length sequences from *Vibrio parahaemolyticus* RIMD 2210633 (WP_005477115.1, WP_005477239.1, and WP_005477284.1, respectively). Five iterations of PSI-BLAST were performed against the RefSeq protein database. A maximum of 500 hits with an expect value threshold of 10^{-6} and a query coverage of 70% were used in each iteration. A local database containing the RefSeq bacterial nucleotide and protein sequences was generated (last updated on June 11, 2022). RPS-BLAST was used to identify VPA1270 homologs in the local database. The results were

filtered using an expect value threshold of 10^{-6} and a query coverage of 70%. Analysis was limited to complete genomes. Subsequently, the genomic neighborhood was analyzed as described before [43,74,75]. The results were further analyzed to identify bacterial sequences containing the three GMT proteins in tandem (**Supplementary Dataset S1**). The Conserved Domain Database (CDD) and related information were downloaded from NCBI on June 15, 2022 [76]. RPS-BLAST was employed to identify conserved domains and the output was processed using the Post-RPS-BLAST Processing Utility v0.1. The expect value threshold was set to 10^{-5}.

Identifying the boundaries and insertion sites of GMT islands: For each GMT island, BLASTN was performed against the NR nucleotide collection to identify closely-related genomes that lack it and contain no insertion between the sequences flanking it. Alignment of the two related sequences, one with the GMT island and one without, allowed us to determine the island boundaries and to identify the putative "naïve" insertion site within the genome lacking the GMT island. Inverted or direct repeat sequences within the putative insertion sites were identified manually.

Characterizing the cargo of GMT islands: The gene cargo, gene number, and the length of representative GMT islands was manually curated and annotated. T6SS effectors were identified by the presence of T6SS effector-specific domains (e.g., MIX [38], FIX [44], PAAR and PAAR-like [45]), as determined by the NCBI CDD [40]. Anti-eukaryotic toxins were identified in a similar manner. Predicted toxin domains of T6SS effectors were identified using the CDD or by similarity detected using hidden Markov modeling (HHpred [41]). Small genes downstream of T6SS effectors were annotated as putative immunity genes. Anti-phage defense systems were identified using the PADLOC [49] and DefenseFinder [50] tools.

Identifying GAPS1, GAPS4, and GAPS6 homologs: PSSMs of the GAPS1, GAPS4a, GAPS4b, GAPS6a, and GAPS6b were reconstructed using full-length sequences (WP_005477165.1, WP_055466293.1, WP_055466294.1, WP_248387294.1, and WP_248387295.1, respectively). PSI-BLAST was performed as described above for VPA1270. RPS-BLAST was then performed to identify GAPS1, GAPS4a and GAPS6b homologs. The results were filtered using an expect value threshold of 10^{-15} and a minimal coverage of 70%. The genomic neighborhoods of identified proteins were analyzed as described previously [43,74,75]. Duplicated protein accessions appearing in the same genome in more than one genomic accession were removed if the same downstream protein existed at the same distance. Protein accessions located at the ends of contigs were removed. In the case of GAPS4a, bacterial sequences lacking GAPS4b downstream of GAPS4a were removed unless the two proteins were fused together. In the case of GAPS6b, bacterial sequences lacking GAPS6a upstream of GAPS6b were removed.

Construction of phylogenetic trees: Phylogenetic analysis of bacterial strains was conducted using the MAFFT server (mafft.cbrc.jp/alignment/server/) [77]. The nucleotide sequences of *rpoB* gene coding for DNA-directed RNA polymerase subunit beta were aligned using MAFFT version 7 (FFT-NS-i was used for bacterial strains containing VPA1270 and GAPS6, and FFT-NS-2 was used for bacterial strains containing GAPS1 and GAPS4) [78]. Partial and pseudo-gene sequences were not included in the analysis. The evolutionary history was inferred using the neighbor-joining method [79] with the Jukes-Cantor substitution model (JC69). The evolutionary distances are in the units of the number of base substitutions per site.

For *Vibrio* GmtY (VPA1270) homologs, the amino acid sequences were aligned using MUSCLE [80] and the evolutionary history was inferred using the neighbor-joining method [79]. The evolutionary distances were computed using the Poisson correction method, and they represent the number of amino acid substitutions per site. All positions with less than 95% site coverage were eliminated. The evolutionary analyses were conducted in MEGA11 [81].

Illustration of conserved residues using Weblogo: The protein sequences of GAPS1, GAPS4a, and GAPS6b homologs were aligned using CLUSTAL Omega [82]. Aligned columns not found in representative proteins were discarded. The conserved residues were illustrated using the WebLogo server (weblogo.berkeley.edu) [83].

GAPS4 structure prediction: A model structure of the GAPS4 complex (WP_055466293.1 and WP_055466294.1) was predicted in ColabFold: AlphaFold2 using MMseqs2 [56]. The protein structure of the highest ranking model was visualized using ChimeraX 1.5 [84].

Acknowledgments

We thank members of the Salomon, Qimron and Bosis labs for helpful discussions and suggestions. DS and EB received funding from the Israel Science Foundation (ISF grant number 1362/21). UQ is supported by the European Research Council – Horizon 2020 research and innovation program, grant no. 818878. UQ has also received funding from the Israeli Ministry of Health in the framework of the ERANET-JPI-AMR, grant no. 15370.

References

1. Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc. Natl. Acad. Sci.* **103**, 1528–1533 (2006).

2. Hood, R. D. *et al.* A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell Host Microbe* **7**, 25–37 (2010).

3. Jana, B. & Salomon, D. Type VI secretion system: a modular toolkit for bacterial dominance. *Future Microbiol.* **14**, fmb-2019-0194 (2019).

4. MacIntyre, D. L., Miyata, S. T., Kitaoka, M. & Pukatzki, S. The Vibrio cholerae type VI secretion system displays antimicrobial properties. *Proc. Natl. Acad. Sci.* **107**, 19520–19524 (2010).

5. Speare, L. *et al.* Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8528–E8537 (2018).

6. Ma, L. S., Hachani, A., Lin, J. S., Filloux, A. & Lai, E. M. Agrobacterium tumefaciens deploys a superfamily of type VI secretion DNase effectors as weapons for interbacterial competition in planta. *Cell Host Microbe* **16**, 94–104 (2014).

7. Borgeaud, S., Metzger, L. C., Scrignari, T. & Blokesch, M. The type VI secretion system of Vibrio cholerae fosters horizontal gene transfer. *Science* **347**, 63–7 (2015).

8. Sana, T. G. *et al. Salmonella* Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. *Proc. Natl. Acad. Sci.* **113**, E5044–E5051 (2016).

9. Verster, A. J. *et al.* The Landscape of Type VI Secretion across Human Gut Microbiomes Reveals Its Role in Community Composition. *Cell Host Microbe* **22**, 411-419.e4 (2017).

10. Unterweger, D. *et al.* The Vibrio cholerae type VI secretion system employs diverse effector modules for intraspecific competition. *Nat. Commun.* **5**, 3549 (2014).

11. Pleška, M., Lang, M., Refardt, D., Levin, B. R. & Guet, C. C. Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nat. Ecol. Evol.* **2**, (2018).

12. Hussain, F. A. *et al.* Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science (80-.).* **374**, 488–492 (2021).

13. Piel, D. *et al.* Phage–host coevolution in natural populations. *Nat. Microbiol. 2022 77* **7**, 1075–1086 (2022).

14. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* vol. 577 (2020).

15. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).

16. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J. Bacteriol.* **193**, (2011).

17. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol. 2019 182* **18**, 113–119 (2019).

18. Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556-1569.e5 (2022).

19. Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. *Nat. Microbiol. 2022 710* **7**, 1568–1579 (2022).

20. Le Roux, F. & Blokesch, M. Eco-evolutionary dynamics linked to horizontal gene transfer in vibrios. *Annu. Rev. Microbiol.* **72**, annurev-micro-090817-062148 (2018).

21. Arnold, B. J., Huang, I. T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology* vol. 20 (2022).

22. Horne, T., Orr, V. T. & Hall, J. P. How do interactions between mobile genetic elements affect horizontal gene transfer? *Curr. Opin. Microbiol.* **73**, 102282 (2023).

23. Bellanger, X., Payot, S., Leblond-Bourget, N. & Guédon, G. Conjugative and mobilizable genomic islands in bacteria: Evolution and diversity. *FEMS Microbiology Reviews* vol. 38 (2014).

24. Khedkar, S. *et al.* Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Res.* **50**, 3155–3168 (2022).

25. Stephens, C. *et al.* F Plasmids Are the Major Carriers of Antibiotic Resistance Genes in Human-Associated Commensal Escherichia coli. *mSphere* **5**, (2020).

26. Alekshun, M. N. & Levy, S. B. Molecular Mechanisms of Antibacterial Multidrug Resistance. *Cell* vol. 128 (2007).

27. Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science (80-.).* **272**, (1996).

28. Asakura, M. *et al.* An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **104**, (2007).

29. Schmidt, H. Shiga-toxin-converting bacteriophages. *Research in Microbiology* vol. 152 (2001).

30. Fillol-Salom, A. *et al.* Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. *Cell* **185**, 3248-3262.e20 (2022).

31. Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740-753.e5 (2022).

32. Salomon, D. *et al.* Type VI secretion system toxins horizontally shared between marine bacteria. *PLoS Pathog.* **11**, 1–20 (2015).

33. Jana, B., Keppel, K., Fridman, C. M., Bosis, E. & Salomon, D. Multiple T6SSs, mobile auxiliary modules, and effectors revealed in a systematic analysis of the Vibrio parahaemolyticus pan-genome. *mSystems* e00723-22 (2022) doi:10.1128/MSYSTEMS.00723-22.

34. Thomas, J., Watve, S. S., Ratcliff, W. C. & Hammer, B. K. Horizontal gene transfer of functional type VI killing genes by natural transformation. *MBio* **8**, 1–11 (2017).

35. Santoriello, F. J., Kirchberger, P. C., Boucher, Y. & Pukatzki, S. Pandemic Vibrio cholerae acquired competitive traits from an environmental Vibrio species. *Life Sci. Alliance* **6**, (2023).

36. Ruhe, Z. C., Low, D. A. & Hayes, C. S. Polymorphic toxins and their immunity proteins: Diversity, evolution, and mechanisms of delivery. *Annu. Rev. Microbiol.* **74**, 497–520 (2020).

37. Hurley, C. C., Quirke, A. M., Reen, F. J. & Boyd, E. F. Four genomic islands that mark post-1995 pandemic Vibrio parahaemolyticus isolates. *BMC Genomics* **7**, 104 (2006).

38. Salomon, D. *et al.* Marker for type VI secretion system effectors. *Proc. Natl. Acad. Sci.* **111**, 9271–9276 (2014).

39. Fridman, C. M., Jana, B., Ben-Yaakov, R., Bosis, E. & Salomon, D. A DNase Type VI Secretion System Effector Requires Its MIX Domain for Secretion. *Microbiol. Spectr.* (2022) doi:10.1128/SPECTRUM.02465-22/SUPPL_FILE/REVIEWER-COMMENTS.PDF.

40. Marchler-Bauer, A. *et al.* CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* **35**, D237-40 (2007).

41. Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

42. Ray, A. *et al.* Type VI secretion system MIX-effectors carry both antibacterial and anti-eukaryotic activities. *EMBO Rep.* **18**, e201744226 (2017).

43. Dar, Y., Salomon, D. & Bosis, E. The antibacterial and anti-eukaryotic Type VI secretion system MIX-effector repertoire in Vibrionaceae. *Mar. Drugs* **16**, 433 (2018).

44. Jana, B., Fridman, C. M., Bosis, E. & Salomon, D. A modular effector with a DNase domain and a marker for T6SS substrates. *Nat. Commun.* **10**, 3595 (2019).

45. Shneider, M. M. *et al.* PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**, 350–353 (2013).

46. Koskiniemi, S. *et al.* Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7032–7 (2013).

47. Willett, J. L. E., Ruhe, Z. C., Goulding, C. W., Low, D. A. & Hayes, C. S. Contact-Dependent Growth Inhibition (CDI) and CdiB/CdiA Two-Partner Secretion Proteins. *J. Mol. Biol.* **427**, 3754–65 (2015).

48. Günther, P. *et al.* Structure of a bacterial Rhs effector exported by the type VI secretion system. *PLoS Pathog.* **18**, (2022).

49. Payne, L. J. *et al.* Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res.* **49**, 10868–10878 (2021).

50. Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, (2022).

51. Roderer, D. & Raunser, S. Tc toxin complexes: Assembly, membrane permeation, and protein translocation. *Annual Review of Microbiology* vol. 73 (2019).

52. Zhang, J., Zhou, M., Liu, W., Nie, J. & Huang, L. Pseudomonas syringae pv. actinidiae Effector HopAU1 Interacts with Calcium-Sensing Receptor to Activate Plant Immunity. *Int. J. Mol. Sci.* **23**, (2022).

53. Knizewski, L., Kinch, L. N., Grishin, N. V, Rychlewski, L. & Ginalski, K. Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* **7**, 40 (2007).

54. Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L. & Ginalski, K. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* **40**, 7016–7045 (2012).

55. Johnson, P. M. *et al.* Functional diversity of cytotoxic tRNase/Immunity protein complexes from burkholderia pseudomallei. *J. Biol. Chem.* **291**, (2016).

56. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods 2022 196* **19**, 679–682 (2022).

57. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* **596**, 583–589 (2021).

58. The complementary structure of deoxyribonucleic acid. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **223**, (1954).

59. Grove, T. Z., Cortajarena, A. L. & Regan, L. Ligand binding by repeat proteins: natural

and designed. *Current Opinion in Structural Biology* vol. 18 (2008).

60. Arcus, V. L., Mckenzie, J. L., Robson, J. & Cook, G. M. The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. *Protein Engineering, Design and Selection* vol. 24 (2011).

61. Kirchberger, P. C., Unterweger, D., Provenzano, D., Pukatzki, S. & Boucher, Y. Sequential displacement of type VI secretion system effector genes leads to evolution of diverse immunity gene arrays in Vibrio cholerae. *Sci. Rep.* **7**, 1–12 (2017).

62. Hurst, C. J. Understanding Aquatic Microbial Communities. 1–12 (2019) doi:10.1007/978-3-030-16775-2_1.

63. Zoccarato, L. & Grossart, H. P. Relationship Between Lifestyle and Structure of Bacterial Communities and Their Functionality in Aquatic Systems. 13–52 (2019) doi:10.1007/978-3-030-16775-2_2.

64. Zhang, T. *et al.* Direct activation of a bacterial innate immune system by a viral capsid protein. *Nat. 2022 6127938* **612**, 132–140 (2022).

65. Molineux, I. J., Schmitt, C. K. & Condreay, J. P. Mutants of bacteriophage T7 that escape F restriction. *J. Mol. Biol.* **207**, 563–574 (1989).

66. Tal, N. *et al.* Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, (2021).

67. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

68. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, (2019).

69. Goren, M. G., Mahata, T. & Qimron, U. An efficient, scarless, selection-free technology for phage engineering. *bioRxiv* 2023.03.26.534254 (2023) doi:10.1101/2023.03.26.534254.

70. Wannier, T. M. *et al.* Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci. U. S. A.* **117**, (2020).

71. Nyerges, Á. *et al.* A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl. Acad. Sci.* **113**, 2502–2507 (2016).

72. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc. 2009 42* **4**, 206–223 (2009).

73. Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S. & Qimron, U. Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Mol. Cell* **66**, 721-728.e3 (2017).

74. Dar, Y., Jana, B., Bosis, E. & Salomon, D. A binary effector module secreted by a type VI secretion system. *EMBO Rep.* **23**, e53981 (2022).

75. Fridman, C. M., Keppel, K., Gerlic, M., Bosis, E. & Salomon, D. A comparative genomics methodology reveals a widespread family of membrane-disrupting T6SS effectors. *Nat. Commun.* **11**, 1085 (2020).

76. Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327-31 (2004).

77. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2018).

78. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–66 (2002).

79. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

80. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

81. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, (2021).

82. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, (2022).

83. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–90 (2004).

84. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, (2021).

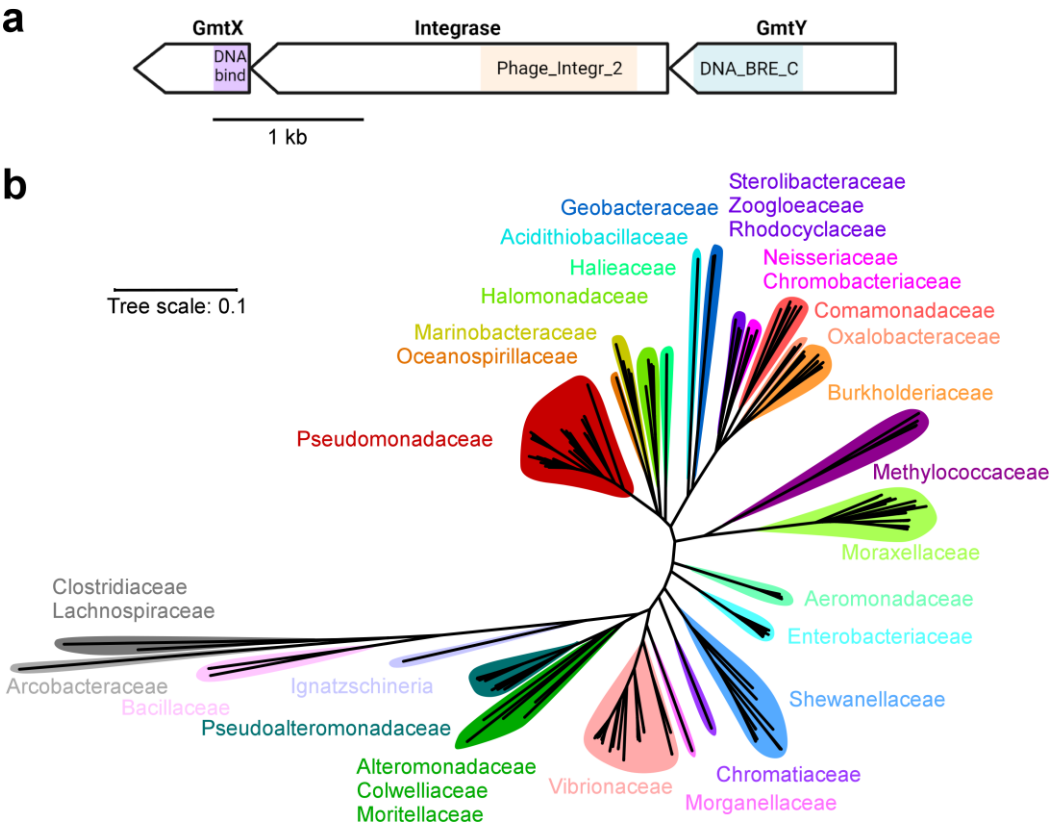
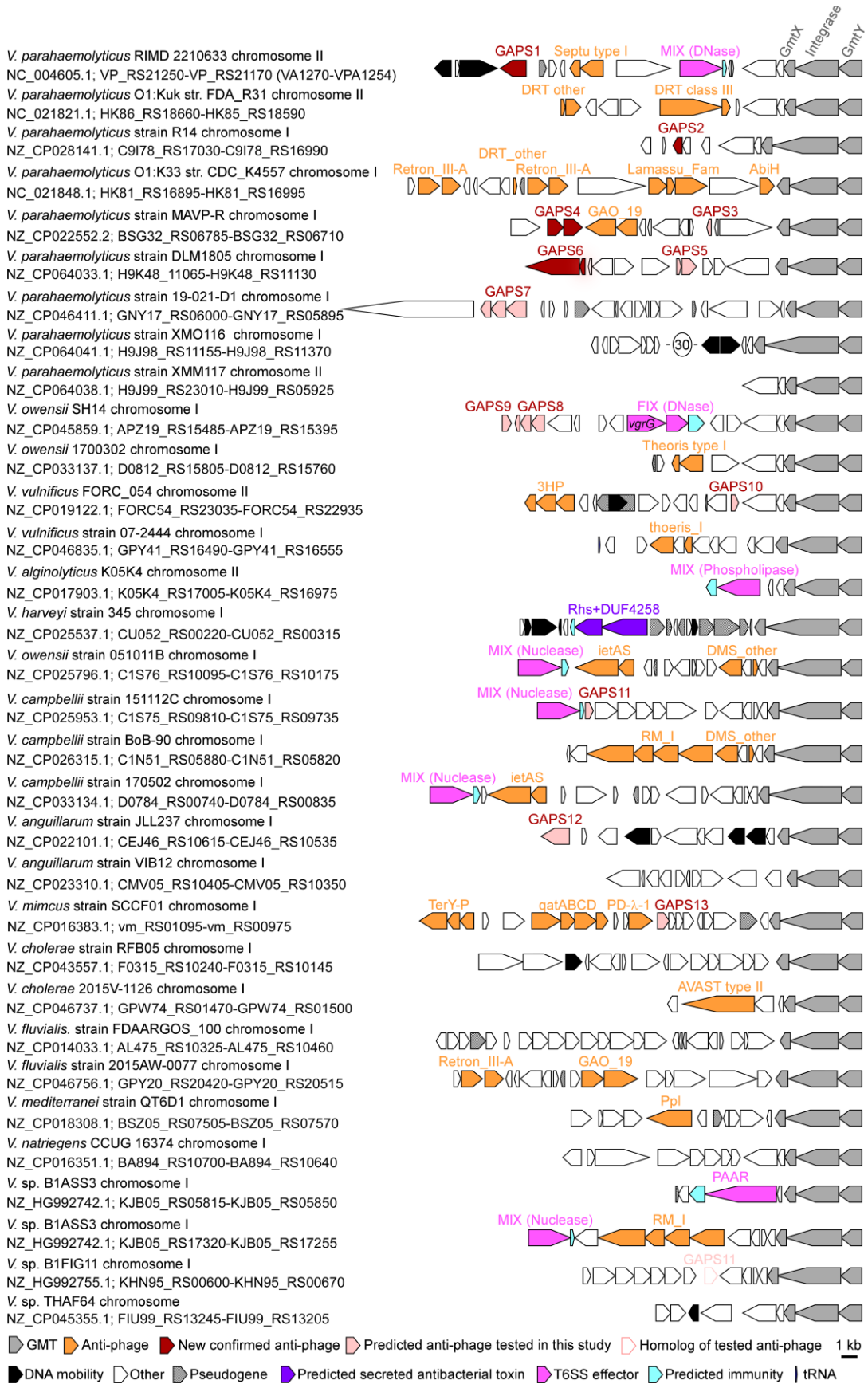


Fig. 1. GMT proteins are widespread in bacteria. (a) Schematic representation of the GMT genes and predicted domains and activities. **(b)** Phylogenetic distribution of bacteria containing GMT proteins, based on the DNA sequence of *rpoB*. Bacterial families are dented. The evolutionary history was inferred using the neighbor-joining method.


V. parahaemolyticus RIMD 2210633 chromosome II
NC_004605.1; VP_RS21250-VP_RS21170 (VA1270-VPA1254)
V. parahaemolyticus O1:Kuk str. FDA_R31 chromosome II
NC_021821.1; HK86_RS18660-HK85_RS18590
V. parahaemolyticus strain R14 chromosome I
NZ_CP028141.1; C9I78_RS17030-C9I78_RS16990
V. parahaemolyticus O1:K33 str. CDC_K4557 chromosome I
NC_021848.1; HK81_RS16895-HK81_RS16995
V. parahaemolyticus strain MAVP-R chromosome I
NZ_CP022552.2; BSG32_RS06785-BSG32_RS06710
V. parahaemolyticus strain DLM1805 chromosome I
NZ_CP064033.1; H9K48_11065-H9K48_RS11130
V. parahaemolyticus strain 19-021-D1 chromosome I
NZ_CP046411.1; GNY17_RS06000-GNY17_RS05895
V. parahaemolyticus strain XMO116 chromosome I
NZ_CP064041.1; H9J98_RS11155-H9J98_RS11370
V. parahaemolyticus strain XMM117 chromosome II
NZ_CP064038.1; H9J99_RS23010-H9J99_RS05925
V. owensii SH14 chromosome I
NZ_CP045859.1; APZ19_RS15485-APZ19_RS15395
V. owensii 1700302 chromosome I
NZ_CP033137.1; D0812_RS15805-D0812_RS15760
V. vulnificus FORC_054 chromosome II
NZ_CP019122.1; FORC54_RS23035-FORC54_RS22935
V. vulnificus strain 07-2444 chromosome I
NZ_CP046835.1; GPY41_RS16490-GPY41_RS16555
V. alginolyticus K05K4 chromosome II
NZ_CP017903.1; K05K4_RS17005-K05K4_RS16975
V. harveyi strain 345 chromosome I
NZ_CP025537.1; CU052_RS00220-CU052_RS00315
V. owensii strain 051011B chromosome I
NZ_CP025796.1; C1S76_RS10095-C1S76_RS10175
V. campbellii strain 151112C chromosome I
NZ_CP025953.1; C1S75_RS09810-C1S75_RS09735
V. campbellii strain BoB-90 chromosome I
NZ_CP026315.1; C1N51_RS05880-C1N51_RS05820
V. campbellii strain 170502 chromosome I
NZ_CP033134.1; D0784_RS00740-D0784_RS00835
V. anguillarum strain JLL237 chromosome I
NZ_CP022101.1; CEJ46_RS10615-CEJ46_RS10535
V. anguillarum strain VIB12 chromosome I
NZ_CP023310.1; CMV05_RS10405-CMV05_RS10350
V. mimcus strain SCCF01 chromosome I
NZ_CP016383.1; vm_RS01095-vm_RS00975
V. cholerae strain RFB05 chromosome I
NZ_CP043557.1; F0315_RS10240-F0315_RS10145
V. cholerae 2015V-1126 chromosome I
NZ_CP046737.1; GPW74_RS01470-GPW74_RS01500
V. fluvialis. strain FDAARGOS_100 chromosome I
NZ_CP014033.1; AL475_RS10325-AL475_RS10460
V. fluvialis strain 2015AW-0077 chromosome I
NZ_CP046756.1; GPY20_RS20420-GPY20_RS20515
V. mediterranei strain QT6D1 chromosome I
NZ_CP018308.1; BSZ05_RS07505-BSZ05_RS07570
V. natriegens CCUG 16374 chromosome I
NZ_CP016351.1; BA894_RS10700-BA894_RS10640
V. sp. B1ASS3 chromosome I
NZ_HG992742.1; KJB05_RS05815-KJB05_RS05850
V. sp. B1ASS3 chromosome I
NZ_HG992742.1; KJB05_RS17320-KJB05_RS17255
V. sp. B1FIG11 chromosome I
NZ_HG992755.1; KHN95_RS00600-KHN95_RS00670
V. sp. THAF64 chromosome
NZ_CP045355.1; FIU99_RS13245-FIU99_RS13205
GAPS1
Septu type I
MIX (DNase)
GmtX
Integrase
GmtY
DRT other
DRT class III
GAPS2
DRT_other
Retron_III-A
Retron_III-A
Lamassu_Fam
AbiH
GAPS4
GAO_19
GAPS3
GAPS6
GAPS5
GAPS7
30
GAPS9
GAPS8
FIX (DNase)
vgrG
Theoris type I
3HP
GAPS10
thoeris_I
MIX (Phospholipase)
Rhs+DUF4258
MIX (Nuclease)
ietAS
DMS_other
MIX (Nuclease)
GAPS11
RM_I
DMS_other
MIX (Nuclease)
ietAS
GAPS12
TerY-P
qatABCD
PD-λ-1
GAPS13
AVAST type II
Retron_III-A
GAO_19
Ppl
PAAR
MIX (Nuclease)
RM_I
GAPS11
GMT
Anti-phage
New confirmed anti-phage
Predicted anti-phage tested in this study
Homolog of tested anti-phage
1 kb
DNA mobility
Other
Pseudogene
Predicted secreted antibacterial toxin
T6SS effector
Predicted immunity
tRNA

783

Fig. 2. GMT islands in vibrios contain diverse cargo. The gene structure of representative GMT island types found in *Vibrio* genomes. The strain names, the GenBank accession numbers, and the locus tags of the island edges are provided. The names of encoded proteins and predicted activities or domains are denoted above.

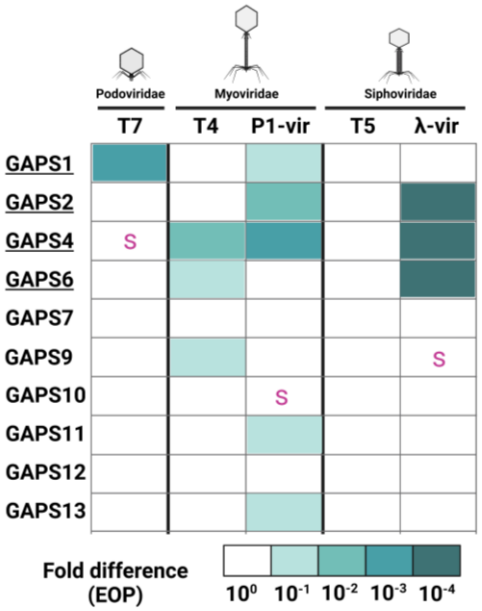


Fig. 3. Four new anti-phage defense systems identified within GMT islands. The efficiency of plating (EOP) was tested for five coliphages against *E. coli* expressing the indicated candidate defense systems. “S” denoted a smaller plaque size phenotype. GAPS3, GAPS5, and GAPS8 are not shown because they were toxic to *E. coli* upon expression. The four confirmed systems are underlined.

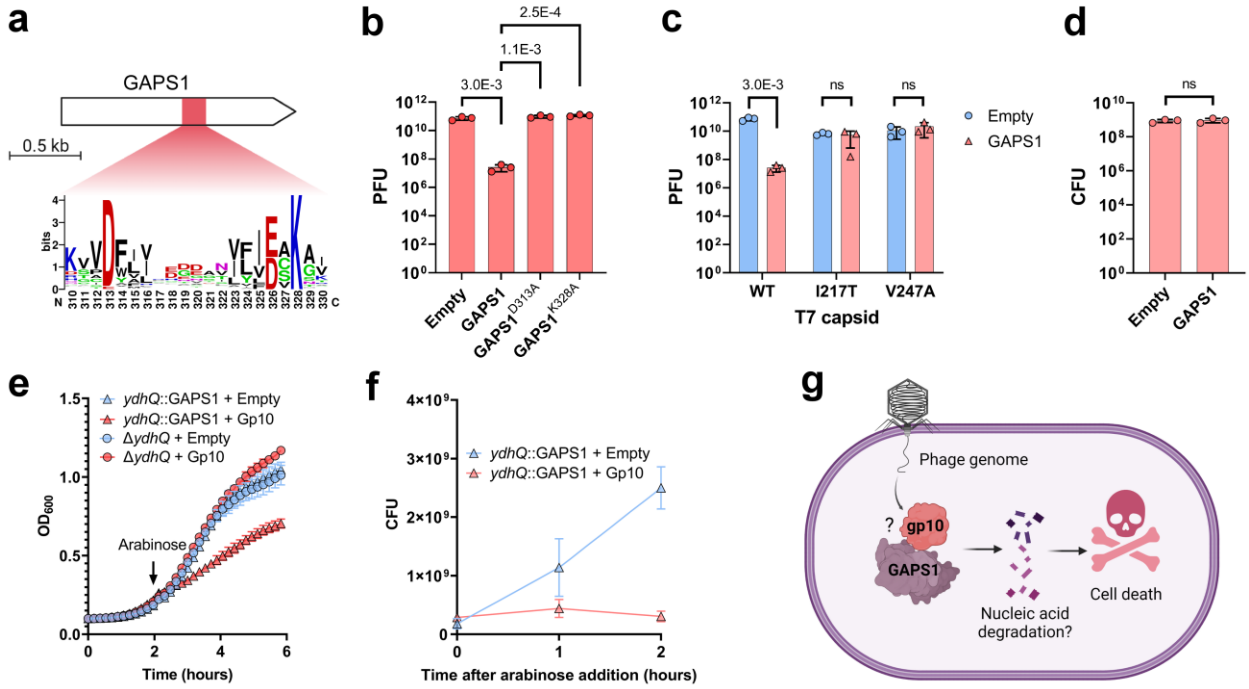


Fig. 4. GAPS1 is a predicted nuclease activated by a phage capsid protein. (a) Schematic representation of GAPS1 (WP_005477165.1) and the location of the predicted PD-(D/E)xK nuclease domain. The conservation logo of GAPS1 homologs is shown below. The residue numbers correspond to the positions in WP_005477165.1. **(b)** Plaque forming units (PFU) of T7 phage upon infection of *E. coli* strains containing an empty plasmid or a plasmid for the arabinose-inducible expression of the indicated GAPS1 version. **(c)** PFU of T7 phage, either wild-type (WT) or containing the indicated mutation in the Gp10 capsid protein, upon infection of *E. coli* strains containing an empty plasmid or a plasmid for the arabinose-inducible expression of GAPS1. **(d)** Colony forming units (CFU) of *E. coli* BW25113 Δ*trxA* containing an empty plasmid or a plasmid for the arabinose-inducible expression of GAPS1, transduced with a T7 phage-based particle containing a plasmid conferring kanamycin resistance. **(e)** Growth of *E. coli* strains in which the chromosomal *ydhQ* gene was deleted (Δ*ydhQ*) or replaced with an arabinose-inducible GAPS1 (*ydhQ*::GAPS1), containing either an empty plasmid or a plasmid for the arabinose-inducible expression of the T7 phage Gp10. An arrow denotes the time at which arabinose was added. **(f)** Viability (measured as CFU) of the indicated *E. coli* strains after arabinose induction, as described in (e). In (b-f), the data are shown as the mean ± SD. Statistical significance between samples was determined by an unpaired, two-tailed Student's *t*-test; ns, no significant difference ($P > 0.05$). **(g)** A model for activation of GAPS1-mediated defense by T7 phage infection. After injection of the phage genome, the expression of the capsid protein Gp10 is sensed by GAPS1. This activates the PD-(D/E)xK nuclease domain of GAPS1, leading to cell death.

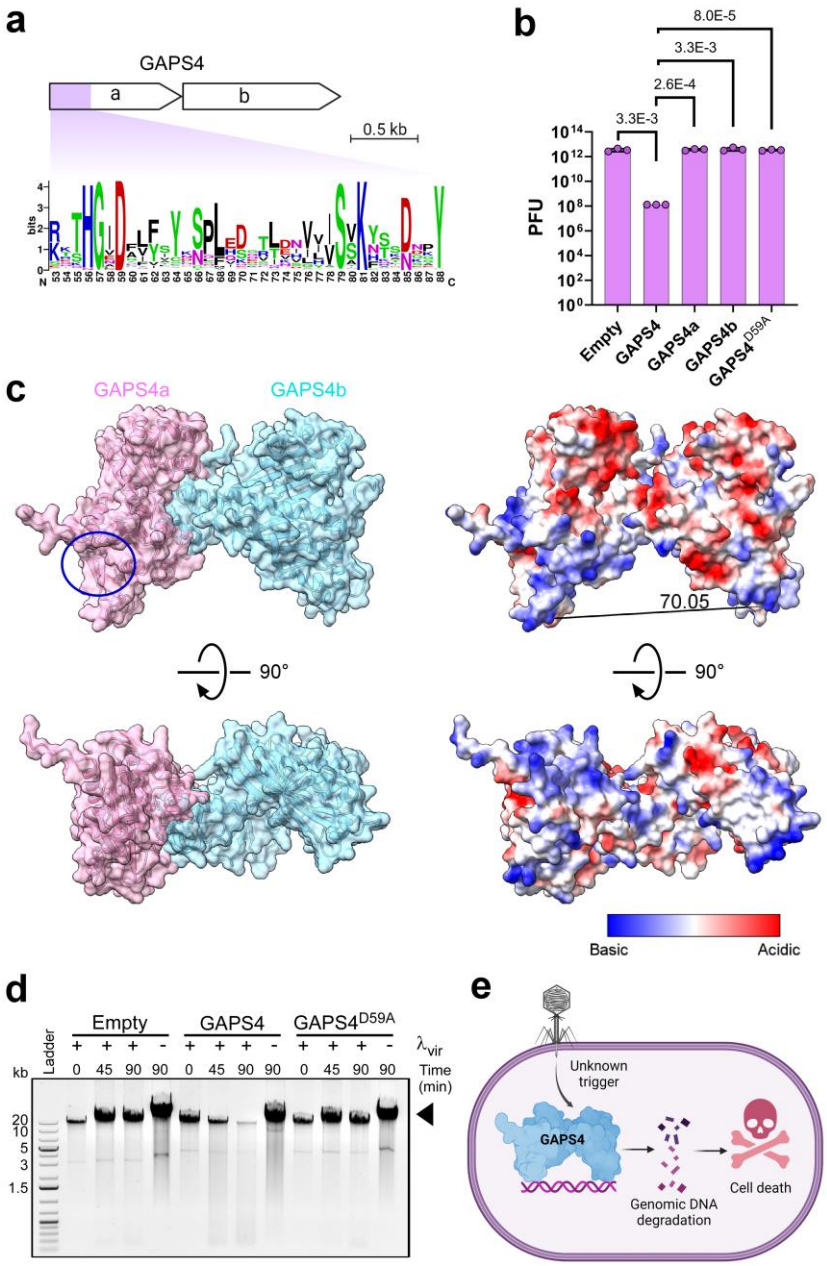


Fig. 5. GAPS4 is a DNase protecting against lambda phage. (a) Schematic representation of GAPS4 (a - WP_055466293.1 and b - WP_055466294.1) and the location of the predicted PD-(D/E)xK nuclease domain in GAPS4a. The conservation logo of GAPS4a homologs is shown below. The residue numbers correspond to the positions in WP_055466293.1. **(b)** PFU of lambda phage upon infection of *E. coli* strains containing an empty plasmid or a plasmid for the arabinose-inducible expression of the indicated GAPS4 version. Data are shown as the mean ± SD. Statistical significance between samples was determined by an unpaired, two-tailed Student's *t*-test. **(c)** An alphaFold2 structure prediction of GAPS4. The surface and backbone are shown on the left; the electrostatic potential of the surface is shown on the right; a dark blue circle denotes the predicted nuclease active site in GAPS4a; the distance (Å) between the predicted DNA-binding lobes of the two proteins is denoted. **(d)** Stability of *E. coli* genomic DNA

(denoted with an arrowhead) in cells containing an empty plasmid or a plasmid for the arabinose-inducible expression of the indicated GAPS4 in the presence or absence of infection with lambda phage (λ_{vir}) for the indicated duration of time. **(e)** A model for activation of GAPS4-mediated defense by lambda phage. After injection, an unknown protein or signal triggers the activation of GAPS4 that uses its PD-(D/E)xK nuclease domain to cleave bacterial genomic DNA, leading to cell death.

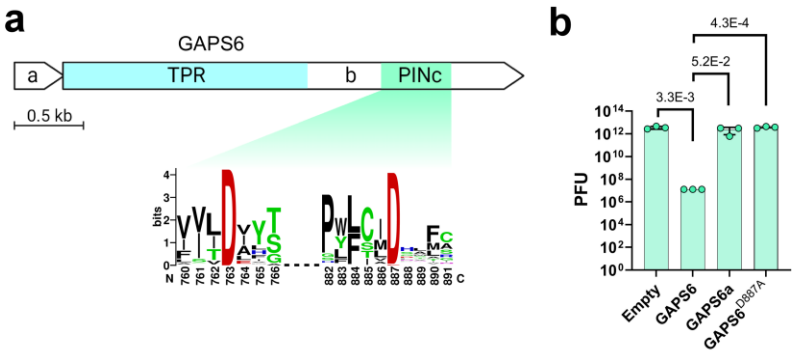


Fig. 6. GAPS6 protects against lambda phage using a predicted PINc domain. (a) Schematic representation of GAPS6 (a - WP_248387294.1 and b - WP_248387295.1) and the location of the predicted TPR repeats and PINc domain in GAPS6b. The conservation logo of the predicted catalytic PINc domain aspartic acid residues in GAPS6b homologs is shown below. The residue numbers correspond to the positions in WP_248387295.1. **(b)** PFU of lambda$_{vir}$ phage upon infection of *E. coli* strains containing an empty plasmid or a plasmid for the arabinose-inducible expression of the indicated GAPS6 version. Data are shown as the mean ± SD. Statistical significance between samples was determined by an unpaired, two-tailed Student's *t*-test.