



HAL
open science

Sedentary chromosomal integrons as biobanks of bacterial anti-phage defence systems

Baptiste Darracq, Eloi Littner, Manon Brunie, Julia Bos, Pierre Alexandre Kaminski, Florence Depardieu, Weronika Slesak, Kevin Debatisse, Marie Touchon, Frederique Le Roux, et al.

► **To cite this version:**

Baptiste Darracq, Eloi Littner, Manon Brunie, Julia Bos, Pierre Alexandre Kaminski, et al.. Sedentary chromosomal integrons as biobanks of bacterial anti-phage defence systems. 2024. hal-04783441

HAL Id: hal-04783441

<https://hal.science/hal-04783441v1>

Preprint submitted on 18 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Sedentary chromosomal integrons as biobanks of bacterial anti-phage defence systems

Baptiste Darracq^{1,2,*}, Eloi Littner^{2,3,4,*}, Manon Brunie¹, Julia Bos¹, Pierre-Alexandre Kaminski¹, Florence Depardieu⁵, Weronika Slesak¹, Kevin Debatisse¹, Marie Touchon³, Aude Bernheim⁶, David Bikard⁵, Frédérique Le Roux^{7,†}, Didier Mazel^{1,†}, Eduardo P.C. Rocha^{3,#,†} and Céline Loot^{1,#,†}

¹Institut Pasteur, Université Paris Cité, Unité Plasticité du Génome Bactérien, CNRS UMR3525, 75724 Paris, France.

²Sorbonne Université, Collège doctoral, F-75005, Paris, France.

³Institut Pasteur, Université Paris Cité, Microbial Evolutionary Genomics, CNRS UMR3525, 75724 Paris, France.

⁴DGA CBRN Defence, 91710 Vert-le-Petit, France.

⁵Institut Pasteur, Université Paris Cité, Synthetic Biology, CNRS UMR6047, 75724 Paris, France.

⁶Institut Pasteur, Université Paris Cité, Molecular Diversity of Microbes lab, INSERM U1284, 75724 Paris, France.

⁷University of Montréal, Faculty of medicine, Department of microbiology, infectiology and immunology, Montreal, QC H3T 1J4, Canada.

* co-first authors

† co-corresponding authors

co-last authors

Corresponding author email: celine.loot@pasteur.fr; eduardo.rocha@pasteur.fr; didier.mazel@pasteur.fr; frederique.le.roux@umontreal.ca;

ABSTRACT

Integrons are genetic systems that accelerate bacterial adaptation by acquiring and shuffling gene cassettes. Mobile integrons spread antibiotic resistance genes among bacteria, while the sedentary chromosomal integrons contain up to hundreds of cassettes of unknown function. Here, we show that many of these cassettes encode anti-phage defence systems. We found numerous streamlined variants of known systems, which have presumably evolved to fit the small size constraints of integron cassettes' recombination and genesis. Intrigued by the rarity of known systems in the sedentary chromosomal integron of the *Vibrio cholerae* 7th cholera pandemic strain, we tested the presence of anti-phage functions in all its cassettes of unknown function. We found that at least 16 of the strain cassettes have an anti-phage activity in *V. cholerae* or *E. coli*. This represents 18% of the tested cassettes and almost 10% of all the integron cassettes, providing at long last a key adaptive role for a significant fraction of the

sedentary integrons. Most of the newly discovered systems have little or no similarity to previously known ones and our experiments show that several mediate defence through cell lysis or growth arrest. One of these systems encodes a 64 amino acids protein, which represents the smallest known protein providing autonomous phage resistance. Given the thousands of uncharacterized integron cassette families, integrons could represent an untapped treasure trove of streamlined anti-phage systems.

Keywords: genetic adaptation; bacterial immunity; streamlined anti-phage system; *Vibrio* spp.

INTRODUCTION

Unexpected challenges require fast responses that may be difficult to prepare. This is especially true for antagonistic biotic interactions, such as the attacks bacteria endure by bacteriophages (phages) (1). Many anti-phage systems have evolved to cope with this challenge (2). They constitute a fast-changing pangenome of defence systems extensively exchanged within bacterial communities by horizontal transfer (3-5). The high turnover of these functions is caused by the huge genetic diversity of phages and other MGEs, and by their ability to evolve anti-defence systems (6). Defence systems are under positive selection when targeting co-occurring infectious phages but may be counter-selected when their costs are not sufficiently compensated by their protective role. The latter tends to fade away as phages evolve to become resistant, which is thought to result in cycles of gain and loss of defence functions. Yet, gene gain by horizontal transfer evolves in a much longer time scale than that of phage infections. If a bacterial genetic system could integrate and express adaptive functions at a given moment while storing them at low cost when not needed, then bacteria could rapidly respond to phage pressure without relying on hazardous horizontal gene transfer. We postulated that integrons could perform these multiple roles: facilitate acquisition defence systems, archive them as a biobank and orchestrate their stochastic expression or silencing. They could thus facilitate bacterial's response to the challenges posed by phages, which will inevitably infect bacteria, but whose diversity and rapid evolution complicate the evolution of a fixed set of defence systems.

Integrons are genetic structures that capture, stockpile, shuffle and express promoter-less gene cassettes terminated by recombination sites (*attC*) (7). These events are mediated by a peculiar recombinase, the IntI integrase that acts in two ways: it excises cassettes from the array, and it integrates them at its beginning (*attI* site) where they become expressed under the action of the integron promoter (P_C) (Figure 1A). Cassette integrations gradually shift the other

cassettes away from the promoter. The cassettes may thus remain silent and costless for large periods of time until they are mobilized to the expression site (7-10). Recombination between *attC* sites decreases as the length of the cassettes increases (11). This is thought to constrain the functions of genes present in integron cassettes, since the latter are significantly smaller than the average gene (11). The integron integrase is expressed under stress conditions allowing combinatorial sampling of the phenotypic diversity (12). The configurations of the array of cassettes that increase bacterial fitness in the local conditions will then be selected and rise in frequency in the population. This stochastic system of adaptation on demand could fit exquisitely the bacterial requirement for a reservoir of anti-phage systems. Integrons are commonly divided in two types (Figure 1B). Mobile integrons (MIs) have typically less than a dozen cassettes, often encoding antibiotic resistance genes, and are associated with MGEs (13). Sedentary chromosomal integrons (SCIs) are not mobile. Their presence is conserved across bacterial species, even though their repertoire of cassettes may change rapidly. SCIs encode large arrays of gene cassettes (up to more than 200), most of which of unknown function (14, 15). Yet, it was previously shown that integrons over-represent genes matching the COG categories including toxin-antitoxin, restriction-modification, and other defence systems (14-16). The toxin-antitoxin systems were experimentally shown to stabilize the integron by their addictive behavior (17), but they could also be anti-phage systems (18).

Here, we explore the idea that integrons are biobanks of anti-phage systems. We start by identifying many known or putative defence systems in integrons, especially in the SCIs. Intriguingly, only one was identified in the large (179 cassettes, 130 kb) model SCI of the seventh pandemic *Vibrio cholerae* O1 El Tor strain, N16961. We therefore hypothesized that this SCI might harbor unknown anti-phage functions. We independently cloned all non-redundant cassettes of unknown function and challenged them against phages of *V. cholerae* and *E. coli*. We found that many of these cassettes provide resistance to phages. We performed computational annotations on these cassettes and subsequently conducted experiments demonstrating that several of them induce cell lysis or growth arrest, preventing phage amplification. We postulated that the significance of integrons hinges on their capability to position cassettes downstream of an active promoter and validated that relocating cassettes to the first position in the array indeed confers protection. Our results confirm that integrons are cost-effective biobanks and orchestrators of anti-bacteriophage defences, highlighting their potential as untapped treasure troves of defence systems.

RESULTS

Sedentary chromosomal integrons have numerous streamlined anti-phages systems

We searched the complete reference genome database RefSeq (Table S1) for known or putative defence systems using DefenseFinder (19) and identified 218,635 systems in 32,798 genomes (Table S2). In parallel, we used IntegronFinder on the same dataset and identified 85,204 integron cassettes comprising 121,711 protein-coding genes (Table S3). The intersection of both analyses revealed that 783 complete defence systems are encoded within integron cassettes. The frequency of defence systems in integrons was 2.2 times higher than expected given the distribution of these systems in genomes ($P < 0.001$, χ^2 test). The defence systems in integrons show a peculiar taxonomic distribution, since the vast majority were found in *Vibrio* (73%, Figure 1C). This is largely because this clade has the largest number of cassettes (60%, Figure S1). One could have thought that such systems were present in MIs which are associated with MGEs. Instead, SCIs, representing 86% of the cassettes, account for over 99% of the identified defence systems. This may be the consequence of the over-representation of MIs carrying antibiotic resistance in current databases (leaving the other MIs under-represented). We conclude that recognizable defence systems are abundant in integrons, especially in the large chromosomal ones.

We then investigated the types of defence systems identified in integron cassettes (Figure 1D, Table S4). By far the most common systems are toxin-antitoxin (TA) systems (a third of all systems) of the PfiAT family (20). TAs also have properties of poison-antidote systems making them addictive and thereby capable of stabilizing the repertoire of cassettes of SCIs (17). The SCI of *Vibrio cholerae* El Tor str N16961 - belonging to the current 7th pandemic lineage and chosen as the focal strain of this study, see below - harbors 19 TAs (21). Among them, the relB/parE-like TA (VCA0488/0489) was detected as a PfiAT defence system. More generally, all the SCIs carried by 7th pandemics *V. cholerae* harbor a PfiAT and no other defence systems (Figure S1). The second most frequent defence systems are Shedu systems which are proposed to act as nucleases (2, 22), followed by an abortive infection system (AbiV) and by PD-T7-4 whose function is yet poorly understood.

Interestingly, the genetic loci encoding these systems are small, TA systems are usually composed of two short genes and the remaining systems are composed of a single gene. Since the integron cassettes contain smaller CDSs than replicons probably due to the cassette genesis

and recombination constraints (*II*), we examined the hypothesis that the pool of defence systems in integrons comprises systems with a smaller-than-average number of components. Indeed, the types of systems encoded in integrons tend to be smaller than the average system identified across the genome database and never exceeds 2 components (Figure 1E). We then enquired if across the types of systems identified in integrons, they tended to be smaller than the same types of systems found at other locations in the same genomes. For this, we built a mixed model describing the variation in the length of the system where the presence in or outside the integron was the fixed effect and the type of system was the random effect. The results show that length is significantly smaller for systems in integrons when controlling in this way for the effect of the type of system ($R^2=0.87$, $P<0.0001$, F Test, Table S5 and S6). This trend becomes particularly evident when one compares the average size of systems present in more than 10 occurrences both within and outside integrons (Figure 1F). Hence, these results show that the defences encoded in cassettes are smaller than expected, both when looking at systems (reduced number of components) and at featured genes themselves (shorter CDS). The need for gene compaction to facilitate recombination within the integron may explain why smaller restriction systems (such as type II and IV) are more prevalent, while defence systems featuring lengthy and intricate loci encoding anti-phage mechanisms like CRISPR-Cas or BREX are entirely absent from integrons.

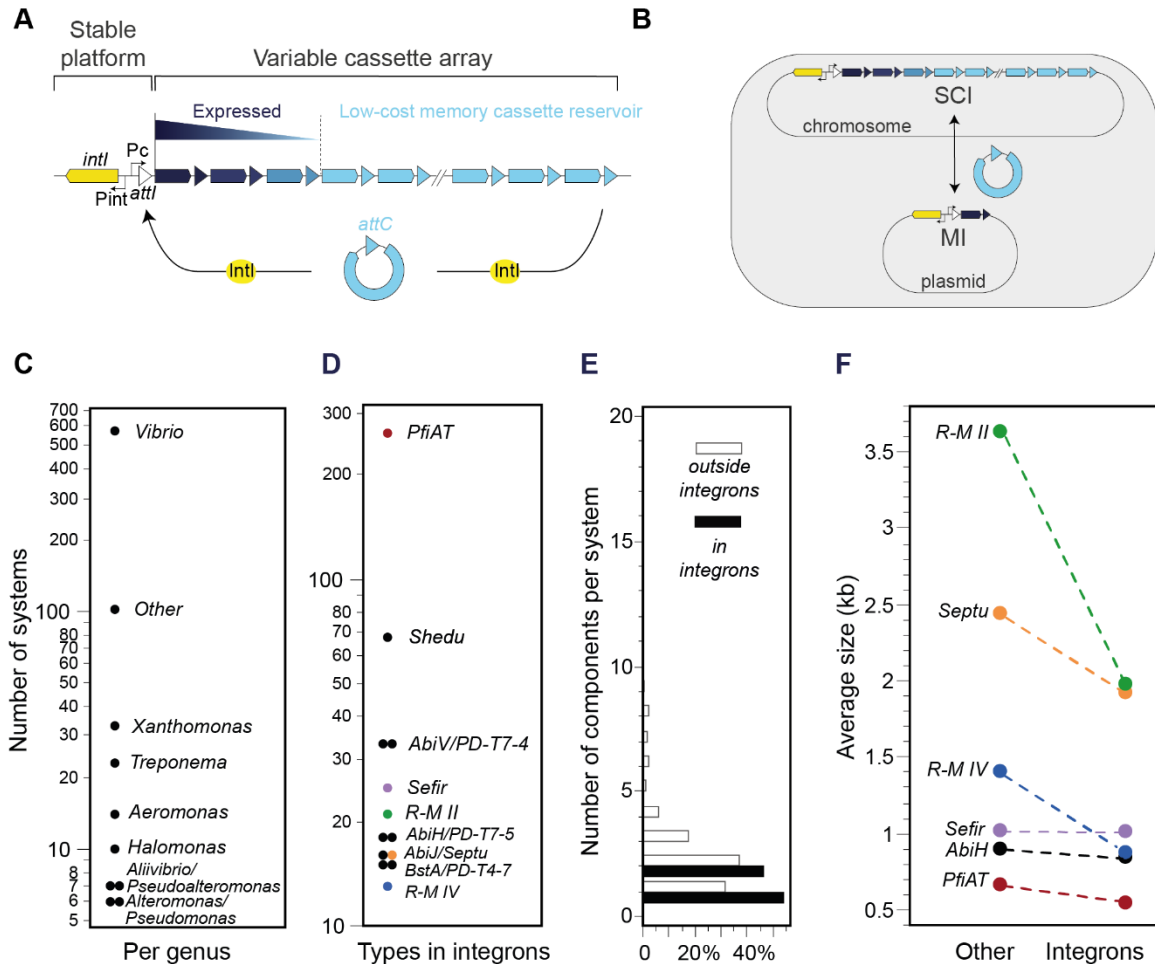


FIGURE 1: Integrons and their known anti-phage systems.

A. Representation of the integron system and function. The stable platform consists of a gene encoding the integrase (*intI*, yellow) and its P_{int} promoter, the *attI* cassette insertion site (white) and the P_C cassette promoter which drives the expression of downstream cassettes along an expression gradient. The cassettes are represented by blue arrows and triangles: arrows represent the CDS and triangles, the *attC* sites. The intensity of their color means their level of expression, from dark blue (highly expressed) to light blue (not expressed). Cassettes can be excised by an intramolecular $attC \times attC$ reaction and reintegrated into the 1st position near the P_C promoter by an intermolecular $attC \times attI$ reaction and become expressed.

B. This process may also result in cassette exchange between Sedimentary chromosomal and mobile integrons (SCIs and MIs).

C-D. Number of systems identified in integron cassettes in function of the taxonomy and in function of the type of system.

E. Number of protein coding genes (components) identified per anti-phage system when comparing systems identified in integrons with systems identified in other regions of the bacterial genomes (all genomes of the database).

F. Average size of anti-phage systems identified in replicons with integrons, separating the systems encoded in integrons from the others. To address taxonomic biases, we compared systems within similar genetic backgrounds. Thus, for the analysis displayed in this panel, we only included systems present in replicons with at least one integron-located defence system. Only system types with more than 10 occurrences in integrons and more than 10 occurrences outside integrons are represented.

Discovery of unknown anti-phage systems in the *Vibrio cholerae* SCI

The previous analyses revealed few defence systems in the integrons of the 7th pandemic strains, which we found intriguing. This led us to investigate the anti-phage activity of the 88 non-redundant function-unknown cassettes of the SCI carried by the *V. cholerae* El Tor strain N16961, the model SCI for *V. cholerae*'s 7th pandemic. Since these cassettes are poorly or not expressed under our laboratory conditions (23), we cloned them under an inducible promoter in a plasmid (Methods). Each construct was transferred to a *V. cholerae* derivative lacking both the SCI (to avoid interference with other cassettes) (24) and the anti-phage system DdmABC (to recover the phage X29 activity, Figure S2) (25). We then challenged the bacteria with our collection of five phages: ICP1, ICP2, ICP3, 24 and X29. We observed a reduction in plaque number and/or a marked reduction in plaque size, both phenotypes indicative of altered susceptibility to phages. Eight cassettes provided robust and reproducible resistance to one phage. One cassette alters the susceptibility to three phages out of five (Figure 2A, Figure S3).

The exploration of the anti-phage function of 88 cassettes remains limited in *V. cholerae* due to the scarcity of available phages and their potential previous adaptation to the strain. It is also useful to note that integron cassettes are commonly transferred across bacterial species, via MIs, and could confer anti-phage protection to a wide range of hosts. Consequently, we sought to re-evaluate cassette function using a "naive" heterologous system, *E. coli* MG1655, for which a much broader panel of phages is available. Each cassette was expressed in *E. coli* MG1655 and challenged with a collection of 23 phages. Five cassettes offer protection against 3 to 5 phages, whereas 4 cassettes seem to be more phage-specific (Figure 2A, Figure S3). Anti-phage activities were observed in both hosts for two cassettes. We named these 16 verified unknown anti-phage systems after deities from Celtic mythologies (Figure S4). Overall, 16 out of the 88 cassettes (18%), representing 10% of all integron cassettes, worked as anti-phage systems in either the *E. coli* or the *V. cholerae* strains (Figure 2B).

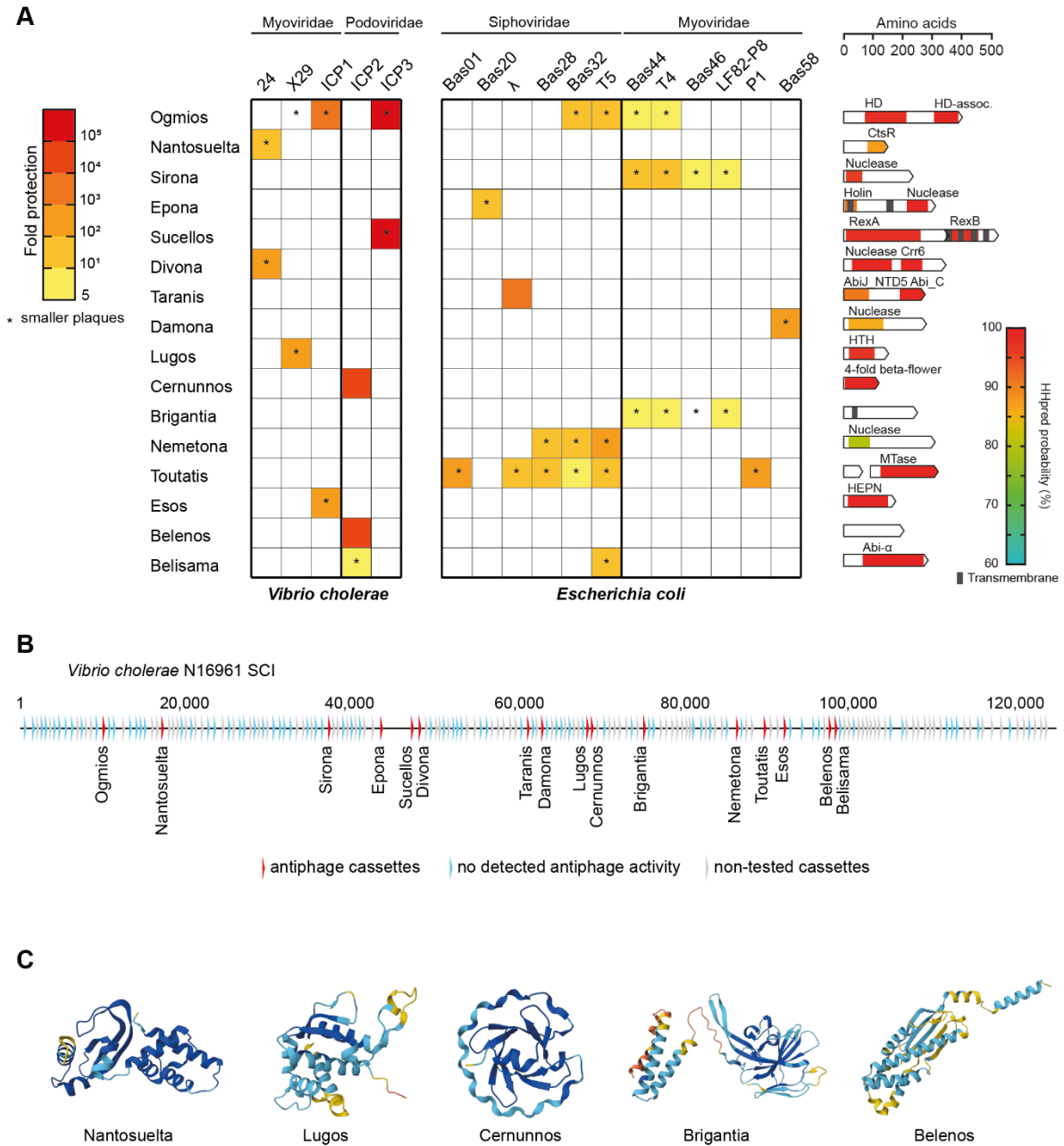


FIGURE 2: *V. cholerae* SCI cassettes protect against phages

A. Heatmap of the phage resistance phenotype representing the mean fold resistance of three independent replicates against a panel of 17 *V. cholerae* or *E. coli* phages. Systems are represented in lines and homology to known protein domains is displayed on the right. Phages are represented in columns and classed according to morphology. In two cases – Sucellos and Toutatis – two genes are in the same integron cassette. * stands for a marked reduction in plaque size.

B. Position of the identified anti-phage cassettes in the *V. cholerae* SCI (red), those that were tested but did not reveal activity (blue), and those not tested (grey).

C. Predicted structures (AlphaFold2) of five anti-phage cassettes in the *V. cholerae* SCI. The predicted structures of all anti-phage cassettes are depicted in Figure S4.

Description of the anti-phage systems found in the *Vibrio cholerae* SCI

None of the 16 anti-phage systems that we identified in *Vibrio cholerae* SCI cassettes was detected by DefenseFinder using default parameters. Defence systems are often a modular combination of domains (*e.g.* a sensor and an effector) (26-28). We thus investigated whether the newly identified systems had protein domains found in known systems using available data (Uniprot, (29)), sensitive sequence similarity detection programs (HHPred (30)), and structural similarity (AlphaFold (31) and FoldSeek (32)) (Figure 2C, Figure S4). Of note, as expected given our observation above that known defence systems in integrons tend to be small, all the discovered systems are small: two have two genes, all others have a single one.

Five systems have domains also found in systems previously described causing growth arrest or cell death upon infection (33). Ogmios encodes a putative deoxyguanosine triphosphatase (dGTPase) enzyme, which degrades dGTP into phosphate-free deoxyguanosine. A dGTPase was previously shown to deplete the nucleotide pool available for phage assembly (34). A homolog of Ogmios, classified as antiviral by DefenseFinder, was actually detected elsewhere in the genome of *V. cholerae* El Tor str N16961. Ogmios may hence be part of a larger family of antiviral dGTPases.

Sucellos has high predicted structural similarity to RexAB encoded by phage λ (35), even though it is not matched by the respective PFAM domain. RexA senses phage infection and triggers the membrane-anchored protein RexB, an ion channel which destroys the membrane potential resulting in cell death (36). Since the RexAB function requires the coordinated action of RexA and RexB, we tested separately the anti-phage function of the two genes of Sucellos (that we named *scIA* and *scIB*) and, as expected, found that both are necessary to provide anti-phage activity (Figure S5A and B).

Taranis may have a function related to that of AbiJ since it matches the PFAM domain Abi_C, corresponding to the C-terminus of the *Lactococcus lactis* abortive infection protein Abi-859 (37), a system later renamed AbiJ (38). The C-terminus of AbiJ is conserved and belongs to the superfamily of HEPN domains that is common in defence systems (39), especially those involved in growth arrest or cell death upon infection (40). Conversely, its N-terminus can vary. The N-terminus of Taranis is distantly related to AbiJ_NTD5. Several systems containing an Abi_C or an AbiJ_NTD domain were recently identified in phage-inducible chromosomal islands (PICIs) of *Klebsiella* (41).

Esos has only a recognizable HEPN-like domain and no other detectable feature. Finally, Belisama has a DUF2806 that has detectable homology with an Abi- α domain, an abortive system often found in prophages (42).

Five anti-phage defence systems harbor diverse fusions of nuclease domains that might act as effectors and other protein domains. Epona harbors a catalytic domain (Mrr_cat) found in type II and type IV restriction enzymes (43), fused to a holin domain with 2 transmembrane domains. Of note, an Mrr_cat domain can also be found in other recently described defence systems' effectors such as Lamassu, Avs and NLR-related systems (44-46). Divona has an unknown function domain (DUF4365) in the N-terminus and a RecB-like nuclease of the Mrr subfamily (47), fused to a chlororespiratory reduction 6 (crr6, PF08847) domain. To our knowledge, the latter was not previously found in anti-phage defence. Damona harbors a domain DUF6602 which is distantly related to RmuC. The latter is a key domain involved in the defence system Shield (48), where it has non-specific (endo)nuclease activity for phage, plasmid, and chromosomal DNA. Damona is further matched by two domains involved in more complex defence systems relying on nuclease activity: CRISPR-Cas (Cas_PD_DExK) and CBASS (CBASS_Endonuc_small). Sirona and Nemetona are both distantly related with His-Me finger endonucleases (HNH), notably to a type II restriction endonuclease domain and to phage λ NinG protein (PF05766). NinG (or Rap) increases the λ phage recombination catalysed by *Escherichia coli*'s RecBCD pathway (49).

Toutatis has two genes, *tutA* and *tutB*, but was considered as a single system because the two genes co-localize in the same cassette in all 61 identified occurrences in *V. cholerae* genomes. *tutB* is a putative methyltransferase. *tutA* is homologous to one domain of the restriction endonuclease Hpy99I involved in cleavage site recognition (50), but does not contain catalytic residues. We tested separately the anti-phage function of the two genes and found that *tutA* is necessary and sufficient to provide anti-phage activity (Figure S5C and D). To our knowledge, the small protein (64 amino acids) encoded by *tutA* is the smallest known to provide autonomous phage resistance.

Five systems have no homology to domains previously associated with defence systems (Figure 2C). Nantosuelta is distantly related to the C-terminal dimerization domain of CtsR (PF17727), a protein that controls heat shock gene expression in *Listeria monocytogenes* (51). Lugos is distantly related to helix-turn-helix DNA-binding proteins. Cernunnos harbors a 4-fold beta flower domain (PF21784) recently described (52). Brigantia has a transmembrane

domain and does not match any other domain of known function. Finally, Belenos lacks known functional domains and homologs of known function. In summary, 11 systems have protein domains or homologous protein domains that can be found in known defence systems, whereas 5 are completely unknown. Their molecular mechanisms will have to be determined in the future and we are currently concentrating our attention on Toutatis system.

***V. cholerae* anti-phage systems trigger cell lysis or growth arrest**

We investigated the function of the systems that are effective in their natural *V. cholerae* host. We monitored growth of cells carrying each of the nine systems or a control plasmid during infection at a low (0.01) or high (10) MOI. Bacterial populations resisted well to phages at low MOI (Figure 3A). Interestingly, at a high MOI, all but one (Nantosuelta) display cell lysis or halted growth after infection. While Lugos and Belisama elicit a strong and unrecoverable culture lysis, most of the others reveal short growth arrest followed by regrowth, which could be associated to post-infection dormancy (Figure 3A) (33). Phage quantifications after lysis when infected at high MOI show an absence of amplification (Figure 3B). Therefore, all systems tested here, besides Nantosuelta, seem to halt growth or kill infected cells preventing completion of the phage cycle. This is consistent with our observation that four systems (Ogmios, Sucellos, Esos and Belisama) present domains common or homologous in systems causing growth arrest or cell death upon infection (Figure 2).

We analyzed in detail the Sucellos system with Alphafold (53). Our prediction indicates that the SclA and SclB proteins each make a homomultimer with the SclB protein forming a transmembrane pore (Figure S6A). Transmembrane proteins are known to mediate cell suicide in other defence systems (33, 54). This cassette contains its own promoter that we named P_{sucellos} . To get closer to the natural conditions of expression, we cloned and expressed this system under the control of P_{sucellos} . While Sucellos induces a strong resistance against ICP3, we observed smaller plaques using phages 24 and ICP1 but no significant reduction of the number of plaques. This was not previously seen using the inducible P_{BAD} promoter (Figure S6B). Infection in liquid culture and phage amplification assays confirm the suicide mechanism triggered by this cassette this time induced by the phage 24 (Figure S6C). We also performed real time microscopy to visualize cells during phage 24 infection (Figure S6D). Cells devoid of the anti-phage cassette were not protected against the phage and exhibited phage-mediated cell lysis around 30 min after initial infection. However, several cells containing the Sucellos system exhibited lysis right after infection (at 0 min), indicating that in these conditions the

Sucellos system induces cell death in infected cells, preventing the completion of the phage cycle and the phage mediated lysis.

To gain insight into the function of the Sucellos system, we performed a phage bacteria co-evolution experiment, and we isolated escapers of the phage 24 able to overcome this system. Whole genome sequencing of 6 escaper phages revealed that all of them carried common mutations (Y112R and N110N/N115D) in the gene *24_60* (Figure S7A), which encodes a pentapeptide repeats containing-protein (Figure S7B). We hypothesized that *24_60* protein might counter the anti-phage activity of Sucellos. We deleted the *24_60* gene of the phage 24 and infected *V. cholerae* with this edited phage. In the presence of the Sucellos system, we obtained a very high fold resistance compared to the infection with the *wt* phage (Figure S7C). We can therefore postulate that the *24_60* protein works as an anti-defence system and that obtained mutations increased its effect against Sucellos.

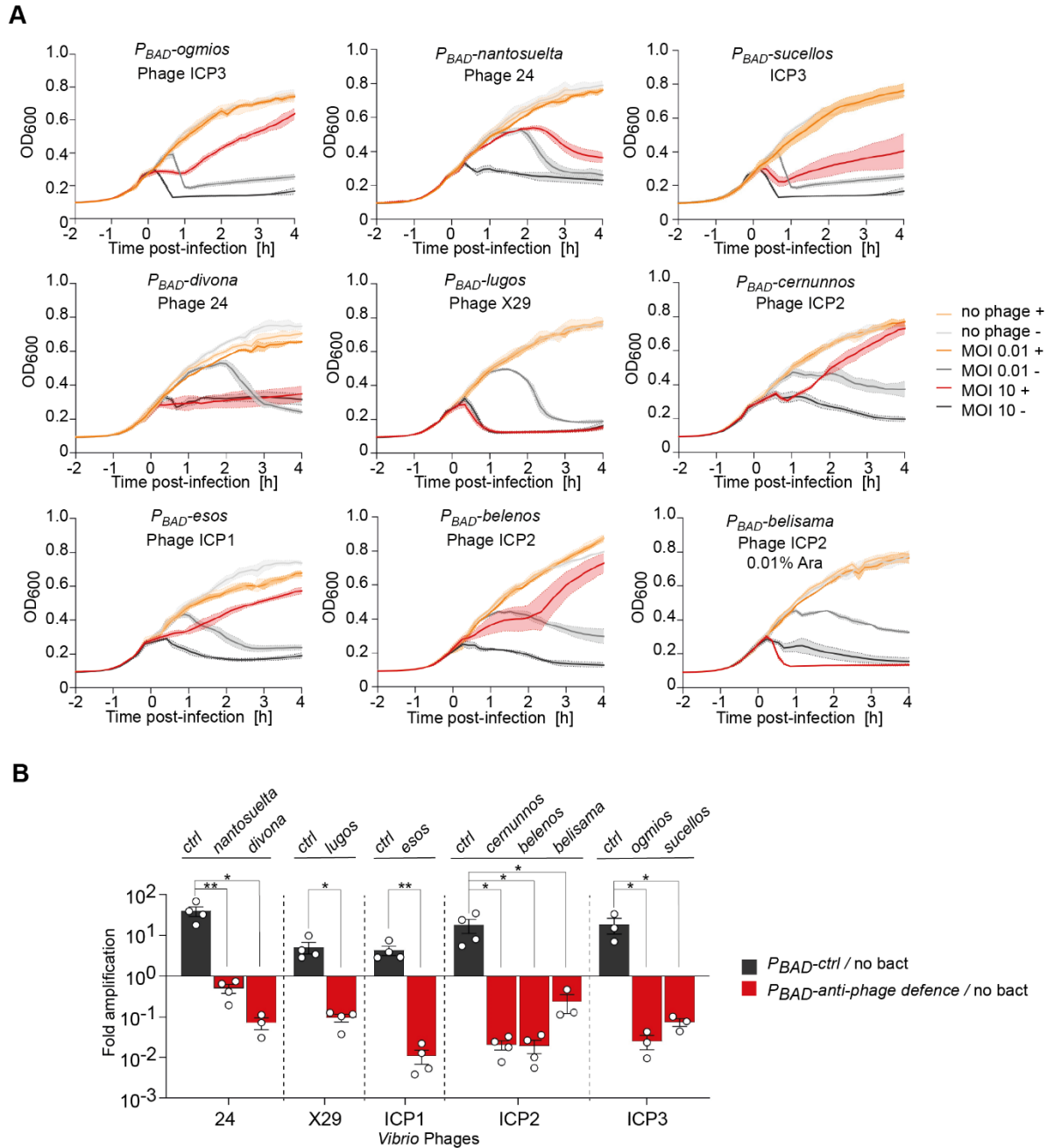


FIGURE 3: Dynamics of phage infection in presence of anti-phage cassettes

A. Growth curves in liquid culture for *V. cholerae* N16961 bacteria that contain the anti-phage system (+) or not (-), infected by phages (MOI 0.01 or 10) or not (no phage). Bacteria were infected at time 0, and measurement of the optical density (OD 600nm) was taken every 10 min. For each experiment, three biological replicates are presented as individual curves.

B. Amplification of phages in *V. cholerae* strains harbouring a plasmid expressing the anti-phage defences or not (ctrl) or without bacteria (no bact). Fold amplification measured as the ratio between *V. cholerae* strains and no bacteria are shown. Bar charts show the mean fold amplification of at least three independent experiments ($n \geq 3$, individual plots) and error bars show the standard error of mean (SEM). Statistical comparisons (Student's t-test, one-tailed) are as follow: **Pvalue < 0.01 and *Pvalue < 0.05.

Anti-phage cassettes can efficiently recombine and be expressed in the *V. cholerae* SCI

As the majority of integron cassettes are devoid of promoter, their distance from the P_C promoter is important for their expression (55-57). All the anti-phage cassettes discovered here were found to be located far from the P_C promoter in the *V. cholerae* SCI and probably not sufficiently expressed to confer phage resistance. We decided to test the effect of translocating one system, Divona, to the first position of the SCI array. We first deleted this cassette (SCI with Divona deleted, Figure 4A). We then inserted it at the first position in the *attI* site, between the G and T nucleotides of the GTTA cleavage site, exactly where recombination is expected to occur (SCI with Divona relocated). This shuffling event should ensure cassette expression, thanks to the presence of the P_C promoter. We then confirmed that integration of the Divona-containing cassette in the first position activated the anti-phage function against phage 24 compared with *wt* or the deleted Divona strains (Figure 4A). As there are a few weak promoters in the integron array (such as those of the TA cassettes, (23, 58)), we mimicked an excision event of cassettes located just upstream of the Divona-containing cassette (SCI with Sucellos to VCA0373 deleted) and demonstrated a slight activation of the Divona anti-phage defence, as we observed a decrease in plaque size (Figure 4A).

The frequency of cassette insertion depends on the properties of *attC* sites involved in the reaction (59, 60), which must adopt a folded single-stranded recombinogenic structure to bind the integrase (59, 61). We calculated the folding probability (pfold (62)) of the 179 *attC* sites of the *V. cholerae* SCI. We found no significant differences between the pfold distribution of the 16 anti-phage cassettes and the others (Figure 4B). This analysis suggests that anti-phage cassettes can be mobilized like the other cassettes. This was confirmed experimentally using a previously constructed inverted SCI strain, which achieves a detectable recombination rate along the entire integron (17). We constructed a library of shuffled cassettes by inducing the integrase and used PCR primer pairs specific to each anti-phage cassettes to identify them. We were able to locate all anti-phage cassettes at the first position, even those with a low pfold (that is, Damona-containing cassette, Figure 4B and 4C) meaning that the *attC* sites of anti-phage cassettes are functional. We conclude that *V. cholerae* anti-phage cassettes can recombine and be expressed when positioned in first position in the integron array.

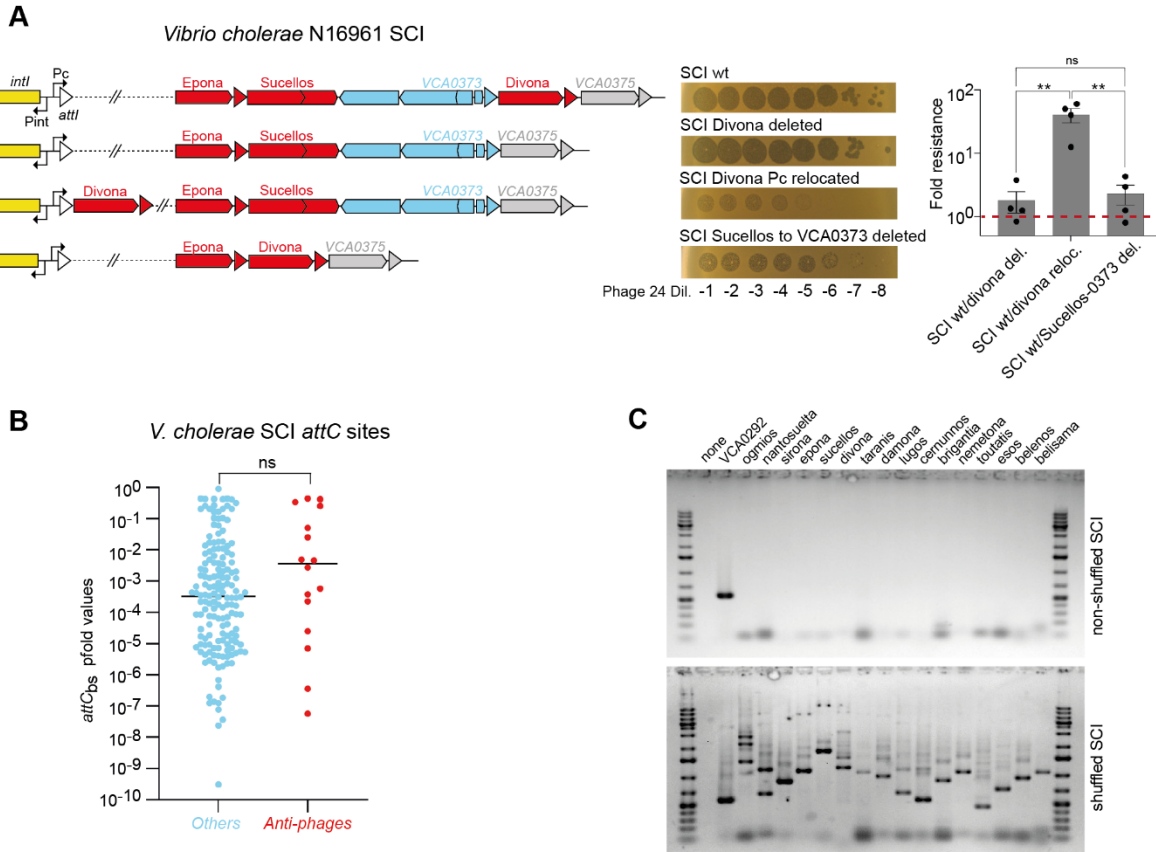


FIGURE 4: Efficient recombination and expression of anti-phage cassettes in the *V. cholerae* SCI

A. Schematic representation of the *V. cholerae* constructed strains is shown on the left panel. The 10-fold serial dilutions of a high titer lysate of phage 24 spotted on these constructed strains are shown in the middle panel. Representative images of a single replicate out of four independent replicates are shown. Fold resistances measured as the ratio between WT and constructed strain values are shown on the right panel. Bar charts show the mean fold resistance of four independent experiments (n=4, individual plots) and error bars show the standard error of mean (SEM). Statistical comparisons (Student's *t*-test, two-tailed) are as follow: ns, not significant; ***P*value < 0.01.

B. Box plot showing the distribution of *attC_{bs}* pfold values of the 179 *attC* sites (bs: bottom strand) of the *V. cholerae* SCI. In red, the 16 *attC* sites associated with the 16 anti-phage cassettes and in blue the remaining (others). Statistical comparison (Mann-Whitney test) is as follow: ns, not significant, *P*=0.1308.

C. Results of the PCR made on the library of shuffled cassettes to identify the anti-phage cassette integration in the first position in the array. The non-shuffled SCI strain is used as control. PCRs were performed using a forward primer specific for the beginning of the integron (non-variable part) and a reverse primer specific for each cassette tested. The word "none" means that only the primer specific for the non-variable part is used, and that no cassette-specific primers are used. The multiple bands, seen just above the band of interest only in the shuffled SCI, correspond to the specific shuffled cassette but with one or more other cassettes inserted before it.

Frequency and locations of the newly discovered anti-phage defence systems

We searched for homologs of the 16 anti-phage systems across Bacteria to understand their frequency and location. Since most systems have one single gene one cannot use co-localisation between genes to separate true from false positives. Hence, we used blastp (except for *tutA*, searched with tblastn, see Methods) instead of HMM models or protein fold predictions to identify homologs, which may have rendered our search conservative. Among the 2505 identified systems (Table S7), around a third (807) were found in *V. cholerae*. These are usually more than 90% identical to the ones found in *V. cholerae* El Tor str. N16961, suggesting one single ancestral acquisition in the species. A noticeable exception concerns Belisama that is in two copies in two strains, one very similar to the reference and the other very divergent (56% identity in protein) suggesting multiple past events of horizontal gene transfer. This is also the case for other systems found in environmental *V. cholerae* strains. Strikingly all the homologs in the species are found in the SCI, suggesting that they have been part of integron cassettes ever since they were acquired by the species. The median genome has 10 homologous systems, but the detailed analyses of the genomes shows that most have either almost all systems or none. To understand this pattern, we represented the systems in the context of a core-genome phylogenetic tree of *V. cholerae* (Figure 5A). While the systems are scattered across the integron, their relative order is highly conserved, with some arrays presenting cassette deletions and a few presenting translocations. These few events complicate the identification of the exact order of acquisition of cassettes. Yet, it is noticeable from the phylogenetic analysis that most systems were acquired recently at the onset of the emergence of the 6th and 7th pandemic lineages (Figure 5A). This result contrasts with the analysis of previously known defence systems which were more frequent in environmental strains of *V. cholerae* (Figure S1). This suggests a change in the ecology of phage-vibrio interactions when the pandemic strains emerged.

Several systems are almost only present in *V. cholerae* (Figure 5B), usually not the most abundant ones. Interestingly, some of these showed anti-phage activity only in *E. coli* (e.g., Epona, Toutatis) and, for a given system, we could not find a clear correlation between the proportion of homologs found outside of *V. cholerae* and the specificity of the anti-phage patterns with respect to the two species (as represented in Figure 2A), *i.e.*, the systems restricted to *Vibrio* spp were not necessarily those giving null results in the *E. coli* genetic background.

We then analyzed the taxonomic distribution and genetic context of the two thirds (1698) of homologous systems that were found in other species. While our dataset had 408

genomes of other *Vibrio* spp, the systems were not very abundant in the genus (151 found). Instead, they could be found in the other Pseudomonadota (regrouped for clarity in Figure 5B). The most extreme cases were Belisama, found in 70 genera of Pseudomonadota, Nantosuelta (40 genera) and Sirona (34 genera). Our analyses already confirm that these discovered systems can be found in many other bacterial clades, even in those that were previously found without any integrons (19). For example, Cernunnos was found in *Bacteroidota* and Taranis (resembling AbiJ, initially discovered in *Lactococcus lactis*) in Bacillota and Actinomycetota.

We then investigated the genetic context of systems found outside *V. cholerae*. The ones present in integrons were usually found in other species of the *Vibrio* genus (Figure 5B). Outside integrons, the systems were identified in a variety of contexts that differed between systems. For example, several were preferentially identified in prophages, such as Sucellos (the RexAB-like system, often in *Salmonella*), Damona (*Acinetobacter*) and Cernunnos (*Escherichia*). While the unknown anti-phage systems we describe here are rarely encoded on plasmids, some of them are specifically found in chromosomes less than 20kb away from integrases, suggesting they are part of mobile genetic elements (e.g. Taranis, in *Burkholderia* and *Klebsiella*; Divona in *E. coli*) (Figure S8). Hence, these discovered defence systems occur in a variety of genetic contexts. In *Vibrio cholerae*, they are always part of the SCI, whereas in the other species they are often found in mobile genetic elements.

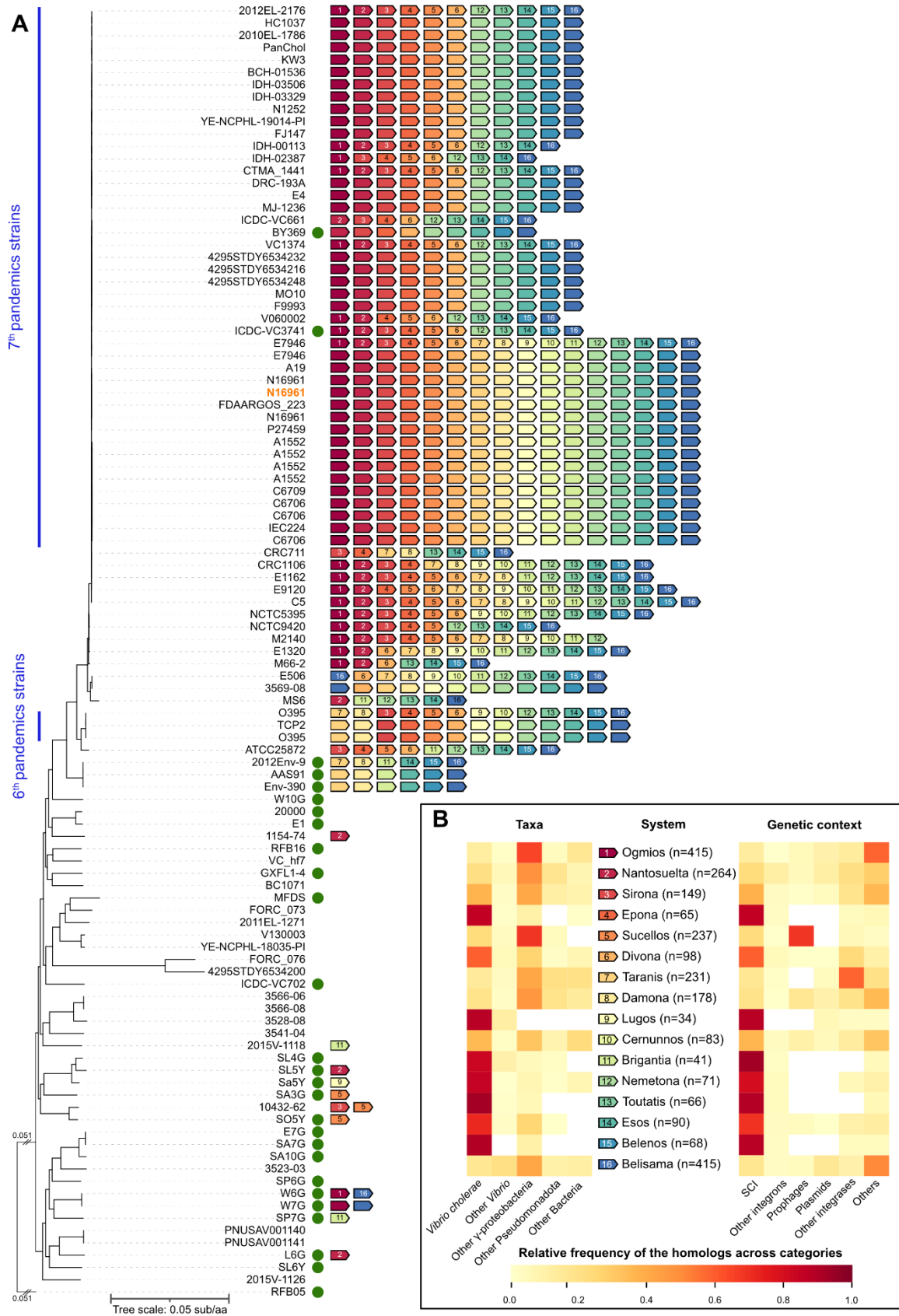


Figure 5. Analysis of the homologs of the discovered anti-phage defence systems.

A. Presence of the homologs of defence systems in *V. cholerae* plotted in the core genome tree. Defence cassettes are indicated by an arrow and ordered by their position in the SCI. The colour and number give the identity of the cassette and the relative position in the integron of the focal strain *V. cholerae* El Tor str N16961 (indicated in orange in the list of strains, see panel B). The strains from the 6th and 7th pandemic are indicated by a blue vertical line on the left. Green dot: environmental *V. cholerae*

B. Heatmaps of the relative frequency of the discovered defence systems across taxa (left) and genetic contexts (right). The scale of colours is given in the bottom of the graph. The cassettes are indicated as in panel A: within the arrow is indicated the relative order of the cassette in the focal strain *V. cholerae* El Tor str N16961 and within parenthesis is indicated the number of homologs.

DISCUSSION

We started this study to test the hypothesis that integrons are biobanks of defence systems. We confirmed that many known defence systems are in integrons. These systems are not a random sample of the types found in genomes. Instead, they tend to be among the smallest. It was previously observed that defence systems in phage satellites are smaller than those on phages, which are themselves smaller than those on other locations of the bacterial chromosome (63). This was interpreted as the result of constraints on genome size caused by phage packaging. Here, this trend is pushed to the extreme. This is probably a constraint imposed by the action of the integron integrase, because long cassettes recombine poorly (11). The low rates of mobilization of long cassettes defeat the purpose of long loci encoding defence systems being in the integron biobank. Our results reveal an additional crucial characteristics of defence systems harbored by integrons: among homologous groups of defence systems, those in integrons tend to be smaller than those in other locations of the genome. Hence, integron defence systems are small by two synergistic effects: the types of systems (those encoded in one or two genes) and the size of the genes of the defence systems (that are smaller). This suggests the existence of a process of streamlining of molecular systems to fit integron cassettes. As a result, there may be plenty of minimal versions of defence systems in integrons and these could be an untapped source of tools for biotechnology.

We were puzzled by the lack of known defence systems in the integron of our *V. cholerae* strain N16961. We therefore decided to study the large number of cassettes in the integron that were of unknown function. Since genes in integrons are often poorly or not expressed, we cloned each cassette onto a plasmid and found that ca. 18% of the tested cassettes have measurable activity against a panel of phages in *V. cholerae* or *E. coli*, meaning that at least 10% of all *Vibrio cholerae* SCI cassettes encode anti-phage systems. This is probably an underestimate. Some systems may have been missed because they provide defence against

other types of phages, might require distinct growth conditions, or may protect from other types of MGEs, e.g. plasmids (25, 64). If we take all integrons in our genome database and cluster the proteins encoded in their cassettes at low sequence similarity (30% identity) we obtain almost 25 thousand distinct families of cassettes. If 10% of them are anti-phage systems, integrons may be an extraordinary repository of anti-phage functions.

We have increased the ability to uncover defence systems by using two different genetic backgrounds and panels of phages. Most previous studies cloned defence systems in a model organism (*E. coli*, *B. subtilis*, *Pseudomonas*) and tested them against a panel of phages. We could not find a large-scale study where several systems were systematically tested both in their native host and in a model organism. Here, using the same set of systems expressed from the same plasmid backbone, we observed defence activities for different subsets of systems in the native *Vibrio* strain (9 anti-phages systems revealed using 5 phages) and in *E. coli* (9 anti-phage systems revealed using 23 phages), with little overlap between species. These intriguing differences may be caused by at least three factors. First, the matrices of defences are sparse, i.e. most systems provide resistance to a few phages. This is common in other phage-bacteria models (41, 63, 65). Using a larger number of phages might better evidence the systems that have a broader range of activity. Second, defence system expression or activity might be affected by the host genetic background. If so, unknown systems may remain to be identified in the SCI by using more diverse panels of phages and bacterial hosts. Finally, phages and *V. cholerae* strains have a long history of antagonistic co-evolution (66), and one cannot exclude that the *Vibriophages* tested in this study already developed anti-defences against some of the integron cassettes. Indeed, we have found one probable anti-defense phage protein protecting against Sucellos (Figure S7).

How do integrons favor bacterial adaptation to phage pressure? We believe this occurs in three ways. First, the integration of a novel phage-resistance cassette in a SCI facilitates the acquisition, stabilization, and expression of a potentially advantageous function under phage pressure. Second, antagonistic co-evolution between phage and bacteria, with phages eventually becoming resistant, is expected to lower the benefits associated with the phage resistance cassette. The natural dynamics of integration of novel cassettes in the SCI increasingly separates the defence system from the integron promoter, thus becoming part of a biobank of non-expressed defence systems. Third, events of recruitment of existing integron cassettes to the expressed position seem more likely (or quicker) than the acquisition of defence system by horizontal gene transfer. If the intrinsic rate of cassette shuffling is high enough, or is stimulated by the presence of a phage, then a bacterial variant where the cassette becomes

under the action of a promoter may protect the bacterium from the phage and sweep through the population. Yet, our results suggest such a scenario may not be very frequent in *Vibrio cholerae* because the shuffling rates obtained in laboratory conditions were low (17), and we found few rearrangements in the defence cassettes of natural isolates (Figure 5A). Also, systems causing cell death upon infection will not provide effective protection if they are rare in the population. We propose a second scenario that is compatible with the observed abundance of these systems in the integrons and the low rates of integron shuffling. Environmental *Vibrio cholerae* is often associated with host organisms or particulate matter where it may form metapopulations. These sub-populations may occasionally endure cassette shuffling placing a defence system near the promoter of the integron (Figure 6). If the expressed system allows infected cell survival in a diverse population, it may be immediately selected for (Figure 6B). If the system induces death of infected cells (Figure 6C and D), it will provide protection to the population (33) only if a fraction of cells containing this system remain uninfected in a homogenous population (Figure 6D). The fixation of cell death inducing mechanisms is therefore more difficult and may require population structure as provided by biofilms, processes of host infection, or colonization of particles. To test these hypotheses, we will need a better understanding of the dynamics of gene cassettes, their impact on the expression of the defence systems, and how they affect selection for anti-phage functions in the integron.

Past work on the function of integron cassettes has focused on antibiotic resistance, which is frequent in mobile integrons, but rare in SCIs. Hence, the function of the vast arrays of SCI cassettes has remained an enigma for decades. Here, we unveil a large set of functions encoded by these elements. As cassettes in SCIs can be mobilized to MIs and then spread by horizontal gene transfer, they may drive resistance to phages across Bacteria as shown by a simultaneous study by Kieffer *et al* (bioRxiv). It is possible that many of the other functions of these integrons are also unknown defence systems. The identification of SCI as a hotspot of anti-phage systems has similarities and differences with the defence islands whose study led to the discovery of so many unknown defence systems (46, 67). Both regions concentrate anti-phage systems, but defence islands seem to have originated by successive integration and excision of mobile genetic elements, suggesting that at least initially the systems were selected for the maintenance of the MGE, which may or may not favor the host (68). In contrast, SCIs are not part of MGEs and are not mobile *per se*. Their anti-phage systems are the result of the creation or exchange of cassettes between integrons, and their subsequent stabilization strongly suggests they are specifically adaptive to the host.

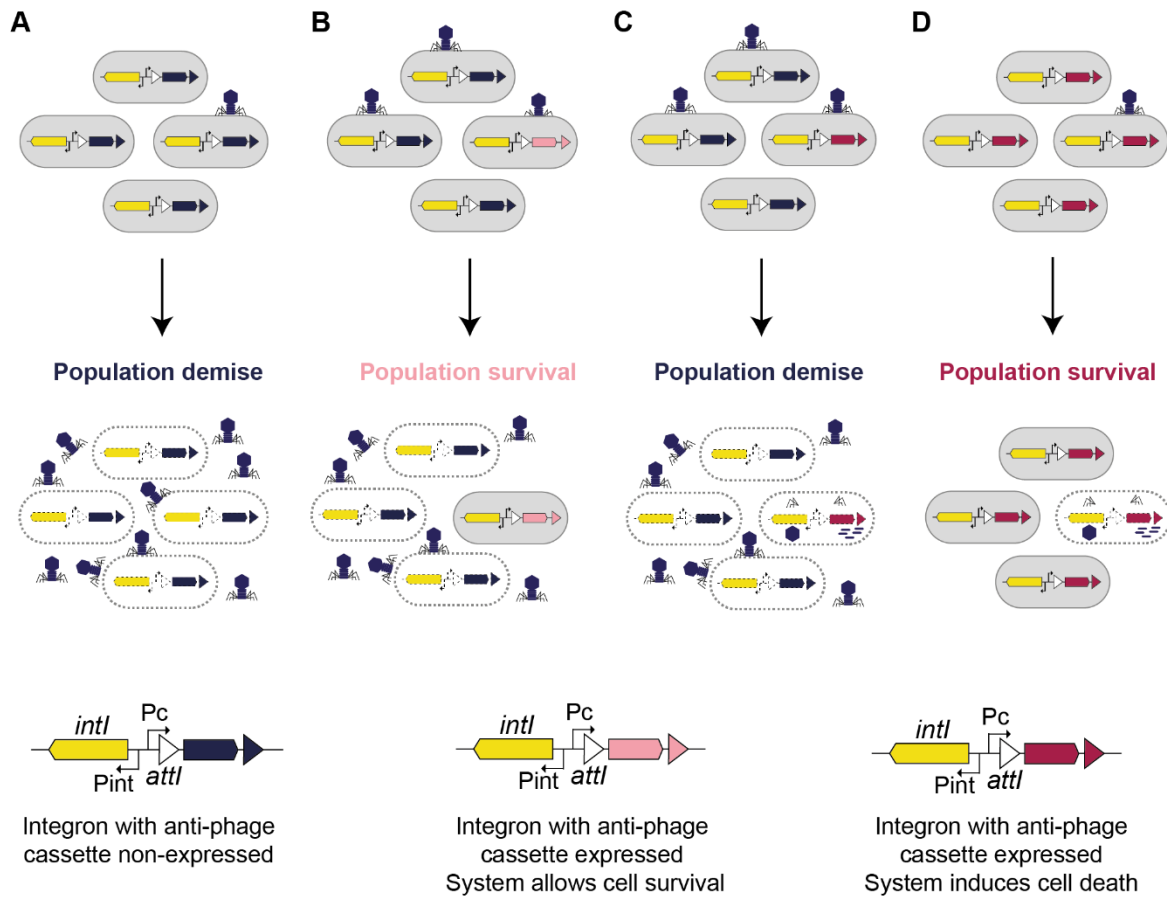


FIGURE 6: Integron cassette shuffling and bacterial adaptation to phage pressure

A. Any cell in the population expresses anti-phage cassette, the population is lysed and the phage is amplified.

B. Only a small fraction of the population expresses an anti-phage cassette that allows cell survival, the population survives.

C. Only a small fraction of the population expresses an anti-phage cassette that induces cell death, the population is lysed and the phage is amplified.

D. The whole population expresses an anti-phage cassette that induces cell death, the population survives at condition that only a part of the population was infected.

Integron with any anti-phage cassette expressed is represented at bottom left of the figure. The expressed non anti-phage cassette is represented in blue. The same integrons but with anti-phage cassettes expressed (i.e. recombined in first position of the array), are represented at the bottom right of the figure. Expressed anti-phage cassettes that allow cells to survive are shown in light pink and those that induce cell death in dark pink.

METHODS

Bioinformatics analysis

Data

Except when specifically mentioned, the genomic analyses were done with a dataset of 32,798 completely assembled genomes from the NCBI RefSeq database of reference microbial genomes (downloaded on May the 15th of 2023). The full list is available in Table S1.

Identification of integrons in complete genomes

We used the integron predictions obtained by IntegronFinder 2.0.5 (19) on the RefSeq database (default options, `-local-max` for more sensitive detection and `-gembase` to preserve annotations). The complete list of these predictions can be found in Table S3. For each genome, integrons were detected replicon by replicon. IntegronFinder provides three subtypes of integron-like elements: complete integrons (IntI plus a few cassettes flanked by *attC* sites), CALINs (Cluster of *attC* sites lacking an integrase) and loner integrases. In the current study, we considered every cassette protein belonging to integrons (67% of total) or CALIN detected by IntegronFinder.

Identification of defence systems

We searched for defence systems in all of the genomes of the database using DefenseFinder v1.2.0 (69), with the `-db-type gembase` and `-preserve-raw` options, and the DefenseFinder models database v1.2.4 (downloaded on February the 14th of 2024). This model database comprises custom HMMs (Hidden Markov Models) of the components of 152 experimentally validated defence systems that are searched for at the same time. A defence system is identified when all its mandatory components are found co-localized in the genome. The option `-db-type gembase` allows to test for this co-localization at the replicon level.

Integration of the results of IntegronFinder and DefenseFinder

The predictions of IntegronFinder and DefenseFinder were merged and analyzed jointly to pinpoint defence systems encoded in integrons. Our analysis deems cassettes as autonomous components of the integron, i.e. it assumes that systems are encoded on one single cassette. Indeed, most systems (783/789) were identified in one single integron cassette, i.e. the region between two consecutive *attC* sites. Six systems were found to be split between neighboring integron cassettes. This is unexpected, since cassette shuffling is likely to disrupt this genetic

organization and disrupt the system. Hence, we decided to discard these six systems. Similarly, we did not search systematically for components encoded in distant cassettes that could eventually produce a complete system. Likewise, we did not look for systems that could be split between an integron and another part of the genome.

In-depth analysis of defence systems in Vibrio cholerae

We made a detailed analysis of the 107 complete genomes of *Vibrio cholerae* in the database above. Strains were classified as either environmental or clinical based on the information available on RefSeq and in the literature. A pangenome, a core genome, and a phylogenetic tree were built using the serial modules of PanACoTA v1.4.1 (70). Briefly, we computed the species' pangenome using mmseqs2 v15-6f452 (71) and using a threshold of at least 80% amino acid identity and 80% coverage among genes belonging to the same pangenome family. Then, the species' persistent genome was defined as the set of pangenome families having a unique member in at least 90% of the genomes. We excluded two strains (*Vibrio cholerae* 2740-80, GCF_001683415.1; *Vibrio cholerae* strain PS4, GCF_019504425.1) that harbored less than 90% of the species' persistent genome and were hence too distant to the other strains in the species. Using MAFFT (72), we aligned individually each pangenome family belonging to the persistent genome (alignments in proteins backtranslated to codons). Finally, the resulting codon alignments were concatenated, and the phylogenetic tree was inferred with IQTree v 2.2.2.2, using the options ModelFinder and computing 1000 ultrafast bootstrap (73). The resulting tree was visualized with iTOL (74).

Identification of homologs of the discovered defence systems

We searched for homologs of the discovered defence systems in the RefSeq database using blastp (minimal coverage of the query sequence 50% and $\text{eval} < 10^{-5}$). We only made an exception for *tutA* because we realized that this small gene was often not identified in the annotation files. In this case, we used the protein of the focal strain and used tblastn to identify homologs directly in the genome nucleotide sequence (same criteria). The presence of a homologous system when the focal one had 2 genes required the identification of co-localized homologs for both genes. The dGTPase was too frequent and difficult to distinguish between homologs with different functions (34). Hence, we sampled among the hits those with $\text{eval} < 1e-73$ and coverage of the query sequence 70%. These thresholds allow to select for homologs that do not hit the current DefenseFinder model for dGTPases (Figure S8). A similar problem

was identified with the MTase, all of them hitting the DefenseFinder HMM for RM MTases, but in this case the requirement for co-localization helped categorizing the hits.

We aimed at describing the genetic context in which the homologs were found. To detect integrons and related elements, we relied on the IntegronFinder v2.0.5 predictions (same parameters as above, except for `--calin-threshold 1` to also detect SALINs (75)). We distinguished SCIs (complete integron or CALIN with >10 cassettes) from other types of elements detected by IntegronFinder, grouped into “Other integrons”. Prophages were detected with geNomad v1.5.2 (76). Replicons were classified as plasmids based on the RefSeq metadata. Integrases were identified using the PFAM profiles PF00589 for tyrosine recombinases, and the pair of profiles PF00239 and PF07508 for serine recombinases (77). All the protein profiles were searched using `hmmsearch` from the HMMer suite v.3.3.2 (hmmmer.org, default parameters). Hits were regarded as significant when their e-value was smaller than 10^{-3} and their alignment covered at least 50% of the protein profile. The complete list of these homologs can be found in Table S7.

System description

We investigated whether the newly identified systems had protein domains found in known systems using available data (Uniprot, (29)), sensitive sequence similarity detection programs (HHPred (30)), and structural similarity (AlphaFold (31) and FoldSeek (32)). All of the relevant information is available in Figure S4.

AlphaFold structure prediction

Alphafold-Multimer (53) was used to model the structure of complexes. The predictions were run via ColabFold v1.5.5 (78) using default parameters.

Bacterial strains, plasmids and primers

The different virus and bacteria strains, plasmids and primers that were used in this study are described in Table S8, S9, S10 and S11.

Media

Escherichia coli and *Vibrio cholerae* were grown in Luria Bertani lennox (LB) at 37°C. In the case of *E. coli*, antibiotics were used at the following concentrations: carbenicillin (Carb), 100 µg/ml, kanamycin (Km), 25 µg/ml, spectinomycin (Sp), 50 µg/ml. Diaminopimelic acid (DAP) was supplemented when necessary to a final concentration of 0.3 mM. To induce the P_{bad}

promoter, L-arabinose (Ara) was added to a final concentration of 2mg/ml; to repress it, glucose (Glc) was added to a final concentration of 10mg/ml. To induce the P_{tet} promoter, anhydrotetracycline (aTc) was added to a final concentration of 100 ng/ml. *V. cholerae* strains were cultivated in the same conditions and with the same antibiotic concentrations except for Sp, that were supplemented at a final concentration of 100 μ g/ml. When *V. cholerae* strains were cultivated in presence of glucose, the later concentration of Sp was increased 2-fold (200 μ g/ml). $CaCl_2$ and $MgSO_4$ were used at a final concentration of 10 mM for *V. cholerae*, and $CaCl_2$ at a final concentration of 5 mM for *E. coli*.

Strain and plasmid constructions

All PCR reactions were performed using the Phusion High-Fidelity PCR Master Mix (Thermo Scientific, Lafayette, CO, USA), and all diagnostic PCR reactions were performed using DreamTaq DNA Polymerase (Thermo Scientific, Lafayette, CO, USA). Oligonucleotides were synthesized by Eurofins Genomics. Gibson Assembly was performed using the master mix (New England Biolabs, NEB) according to the manufacturer's instructions. Genomic DNA was extracted using DNeasy® Tissue Kit (Qiagen) and Plasmid DNA using GeneJET Plasmid Miniprep kit (Thermo Scientific, Lafayette, CO, USA). Natural transformation was performed according to the procedure described in (79).

Cloning of the SCI cassettes in the pSC101 expressing vector

Among the 179 cassettes of the SCI, we excluded those of known function, those coding toxin-antitoxin genes (TA cassettes), and the second copies of duplicated cassettes bringing us down to 88 cloned cassettes. Each cassette fragment (corresponding to the 88 cassettes and *tutA*, *tutB*, *sclA* and *sclB* CDSs alone) was amplified from 8637 with corresponding .for and .rev primers. The pMP180 vector was amplified by inverse PCR with 6705 and 7086 primers. Assembly of each two fragments was achieved by performing Gibson Assembly.

Phages

ICP1, ICP2 and ICP3 phages (for the International centre for Diarrhoeal Disease Research, Bangladesh Cholera Phage) represent the three *V. cholerae*-specific, lytic (virulent) classically used phages (80). They were isolated from rice-water stool samples of cholera patients in Bangladesh (66). ICP1 and ICP3 receptors were identified as the lipopolysaccharide (LPS) O1 antigen (81) and that of ICP2 as the major outer membrane porin OmpU (82). The two other phages (24 and X29) were transmitted from the *Vibrio cholerae* CNR of Institut Pasteur. These

phages of our *vibriophage* collection encompass 5 of the 9 *vibriophage* genome clusters identified by Barman et al (83).

Phage amplification

Phages 24 and X29 were amplified using HER1051 as indicator strain and phages ICP1, ICP2 and ICP3 were amplified using MAK757. Single phages from mother stocks were used to prepare working stocks. Mother stocks were streaked into soft agar mixed with culture of the indicator strains and incubated overnight at 37°C. The next day, single plaques were cored with a cone and resuspended overnight at 4°C under agitation in 1mL of LB to let the phages diffuse in the medium. The resulting lysates were then serially diluted and mixed with bacteria at OD₆₀₀ 0.1 grown in LB + CaCl₂ + MgSO₄. Once lysed, cultures were centrifuged at 3000 g for 5 minutes and supernatants were filtered through a 0.2 μM membrane. Lysates were then titrated by drop assay and the ones with the highest titers were selected.

Coliphages were amplified on *E. coli* K-12 MG1655. Phage stocks were amplified by mixing 200 μL of an overnight culture of *E. coli* with 10 μL of phage stock solution (either pure or diluted) or a piece of frozen phage and 25 mL of warm (~50°C) LB + CaCl₂ + 0.5% agar and poured on LB + CaCl₂ + 1% agar plates. Plates on which confluent lysis was observed were used to recover the top-agar layer in 1mL of PBS and transfer it to a 50 mL conical tube. The top agar was then disrupted by vortexing until broken into small pieces. Tubes were then left to incubate 10 min at room temperature before centrifugation at 3,000 g for 5 min. Finally, the supernatant was recovered.

Phage plaque assays

Preparation of V. cholerae soft agar plates

Strains carrying each of the systems or the control plasmid were grown overnight in LB + Carb + Glc. The cultures were then diluted 1/100 in LB + Carb + Ara and grown overday up to OD₆₀₀ 0.3. Bacterial lawns were prepared by mixing 100 μL of overday culture with 4 mL of melted LB 0.5% agar + Carb + Ara + CaCl₂ + MgSO₄ and the mixture was poured onto round plates of LB 1.5% agar + Carb + Ara + CaCl₂ + MgSO₄.

Preparation of E. coli soft agar plates

Strains carrying each of the systems or the control plasmid were grown overnight in LB + Carb + Glc. Bacterial lawns were prepared by mixing 200 μL of overnight culture with 25 mL of

melted LB 0.7% agar + Carb + Ara + CaCl₂ and the mixture was poured onto squared plates of LB 1.5% agar + Carb + Ara + CaCl₂.

Serial dilutions of high-titer ($>10^8$ pfu/mL) stocks of phages were spotted on each plate and incubated overnight at 37°C. The next day, plaques were counted, and the fold resistance was measured as the number of plaques in the control plate divided by the number of plaques in the presence of each system. When plaques were too small to be counted individually, we considered the most concentrated dilution where no plaque was visible as having a single plaque.

Time course infection experiments

V. cholerae strains carrying each of the systems or the control plasmid were grown overnight in LB + Carb + Glc. The cultures were then diluted 1/100 in LB + Carb + Ara and grown for two hours. OD₆₀₀ was measured and culture concentration normalized, then diluted again 1/100 in LB + Carb + Ara + CaCl₂ + MgSO₄. Note that for Belisama, we used a lower Ara concentration (i.e. 0.01% instead of 0.2%) due to the observed gene toxicity. 180 µL of each preparation was distributed by triplicate in a 96-wells microplate and growth was monitored on a TECAN Infinite M200Pro microplate reader at 37°C with on maximum agitation with absorbance measurements (600 nm) taken at 10-min intervals. When OD_{600nm} reached mid-log phase ~0.3, cells were either kept uninfected or infected with MOI ~0.01 or MOI ~ 10 of phages. Growth was then monitored for 4 h post-infection, with the caveat that the first measurement following infection is always aberrant and was not displayed. Curves correspond to the mean of at least three technical replicates and the shade corresponds to the standard errors at each timepoint. Lysates were also collected 90 min after infection of liquid cultures at initial MOI of 10, and phage titer was quantified by plating serial dilution of the lysates on the strain containing the control plasmid as described above (i.e. Phage plaque assays).

Microscopy setup for live imaging

V. cholerae overnight cultures of the indicated strains were diluted 1/1,000 and then grown in liquid LB + Sp. When OD_{600nm} reached mid-log phase ~0.3, cells were either infected or not with MOI ~ 5 of phages and immediately immobilized on 1.4% agarose-padded slides containing the growth medium. A coverslip was placed on top of the agarose pad and sealed with a vaseline:lanolin:paraffin mix (ratio 1:1:1) to prevent pad evaporation. Slides were incubated at 37°C under an inverted wide-field microscope (Zeiss ApoTome) for time-lapse

video recording. Video frames were taken at 5 min interval for a total duration of 60 min, using a Plan Apo 63× objective (NA = 1.4) and a Hamamatsu sCMOS ORCA-Flash 4.0 v3 (Pasteur Institute Imaging Facility Imagopole). A total of 3 movies per condition–were recorded and analyzed with FIJI software (84).

***attC* sites pfold predictions**

The calculation of the probability of an *attC* site to fold into a recombinogenic structure (pfold) were performed as described in Vit *et al* (62) (*i.e.*, using the RNAfold program from ViennaRNA Package (85) (<http://rna.tbi.univie.ac.at>, category “RNAfold Server”).

Library cassette shuffling assay

A thermosensitive plasmid expressing the integrase under the control of a TetR-pTet promoter (pR669) was transformed in three SCI Inv Δ *IntIA* *V. cholerae* replicates. For each replicate, four clones were picked and grown overnight in LB + Sp at 30°C, then diluted 1/1000 in LB + Sp + aTc and grown overnight at 30°C to express the integrase and shuffle the cassette array. Shuffled cultures were diluted 1/100 in LB and grown overnight at 37°C twice to cure the plasmid. Cultures were then streaked on LB agar plates and 16 single clones were picked for each. *attIA* site region was amplified by PCR (using 5778 and 1401 primers) on these clones to detect individual insertion events and validate the shuffled status of our bank. Shuffled cultures displaying cassette insertions in at least 50% of individual clones were then pooled to limit the risk of bias due to an early insertion in of culture. To test the capability of shuffling of the 16 anti-phage cassettes in 1st position in the bacterial library population, we performed PCR using the 5778 primer associated with each cassette specific primer (308.for, 322.for, 356.for, 366.for, 367-368.for, 374.for, 396.for, 399.for, 409.for, 410.for, 419.for, 441.for, 446-447.for, 450.for, 457.for and 458.for). To control the presence of the 1st cassette (VCA0292), we used the 292.for. Each band was sequenced using the 5778 primer to confirm the specific shuffling event.

Phage training

Phage training was performed according to a method inspired by the Appelmans protocol (86). Phage training was started by mixing early exponential phase *V. cholerae* Δ SCI Δ *ddmABC* expressing or not *sucellos* with 12 different serial dilutions of phage 24 (2×10^9 PFU in the first well). The co-cultures were performed in a 96-well plate, at 180 rpm and at 37°C. After 24h incubation, all wells showing lysis were pooled along with the first non-lysed well past the

point of lysis. The pool was centrifugated to eliminate remaining bacteria. The resulting lysate was used for the next round and the process was repeated 5 times (i.e. 6 rounds in total) up to the evolving phage 24 was able to lyse at the smallest concentration. Escaper phenotype was confirmed by plaque assays and phages from 6 different lineages were isolated by picking into single plaques.

Phage DNA extraction

Wt and escaper phages were amplified as described above to obtain high titer stocks ($>10^8$ pfu/mL). Aliquots of 500 μ L were treated by TURBO DNase (Thermo Fisher Scientific) and RNase A (Qiagen) for 1 hour at 37°C to degrade cellular nucleic acids. DNase was inactivated with the addition of 20 μ L of EDTA 0.5 mM. Aliquots were then treated with 0.5 mg/mL of proteinase K (Qiagen) and SDS 0.5% for one hour at 56°C to degrade phage capsids. DNA was then purified as follows: 500 μ L ml of a phenol-chloroform-isoamylalcohol (PCI, 25:24:1) solution (Sigma) was added to the sample which was then vortexed and centrifuged for 10 minutes at 16,000 g. The upper aqueous phase (around 400 μ L) was transferred to a fresh tube containing 400 μ L of chloroform-isoamylalcohol (24:1). The sample was further vortexed and centrifuged for 10 minutes at 16, 000 g, and the upper aqueous phase (around 250 μ L) was transferred to a fresh tube containing 25 μ l of 3 M sodium acetate. The sample was mixed with 2.5 volumes of cold 100% ethanol and incubated for one hour at -20°C to precipitate DNA. After centrifugation (1 min at 16,000 g), the DNA pellet was washed with 70% cold ethanol. Finally, the pellet was air-dried and resuspended in 50 μ L of distilled water.

Phage sequencing and analysis

Phage DNA was prepared and sequenced at the P2M, Institut Pasteur, using a Nextera XT DNA library preparation kit and the NextSeq 500 System WGS. Trimming and clipping were carried out using *AlienTrimmer* v2.0 with the following parameters: “-q 15 -l 50 -p 50” Trimmed reads of *wt* phage 24 were mapped using bowtie2 (87) on the 24 reference genome to verify it. As the repeated region in *24_60* gene was difficult to assemble, it was amplified by PCR with primers 7423 and 7424, and verified by Sanger sequencing. Mutations were only found in *24_60* gene and a new reference genome was generated. This new reference was used to analyze mutations in escaper phages using bowtie2 and IGV browser (88). Mutations in *24_60* gene were verified by Sanger sequencing.

Golden Gate Assembly

The crRNA spacer was introduced into the pBA559 Cas13 expressing vector through a “Golden Gate Assembly”. The spacer was ordered as two complementary primers targeting the transcript of the gene 24_60 of the phage 24 and, with 4 bp 5' overhangs matching the BsaI-digested destination plasmid staggered ends. For the phosphorylation, 6 μ L (10 μ M) of each 7419 and 7420 primer is mixed with 2 μ L of T4 DNA Ligase Buffer (10X) and 0.4 μ L of T4 PNK for a total volume of 20 μ L and then incubated for 30 min at 37°C. To insert the guide into the pBA559 vector, we prepare the following mix: 40 ng of the pBA559 plasmid, 2 μ L of annealed oligos, 1 μ L of Cutsmart buffer, 1 μ L of BsaI enzyme, 1 μ L of ATP (1mM), 1 μ L of ligase, 2 μ L H₂O. We then incubate the mix in a thermocycler with the following steps: 3 min at 37°C (for digestion), 4 min at 16°C (for ligation) and alternate between those steps 25 times. We finish with one cycle of 5 min at 50°C and 5 min at 80°C for enzymes inactivation. At this step, the mix is ready to be transformed in a cloning *E. coli* Top10 strain (Invitrogen) and the successful sgRNA insertions are screened by sequencing using the 7421 primer.

Two-step phage editing and phage enrichment

To delete the gene 60 of the phage 24, we used the two-step phage editing and phage enrichment method developed by Adler et al, (89). First, the phages infect homology vector-containing *V. cholerae* strain at a low MOI, producing a mixed population of original and edited phages. Second, this phage population is diluted and infects a Cas13a-expressing *V. cholerae* strain targeting the original phage locus, enriching the edited phages relative to original. CRISPR-Cas13a proteins are RNA-guided RNA nucleases that bind to complementary target phage transcripts and ensure general, non-specific RNA degradation of infected cells.

Phage editing

To create genome-edited phage 24 lysate, the phage-editing strain (W303) containing the homologous recombination vector (pJET-24_60) was grown overnight in LB at 37 °C. The strain was diluted into fresh media (LB + Carb) to an OD_{600nm} of 0.04 or 0.4 in a 96-well plate. Wildtype phage 24 was added to each well to achieve an MOI of 1, 0.1 or 0.01. Infection was monitored in a TECAN shaking at 37 °C at 180 rpm, with OD_{600nm} readings every 5 min. Infection was allowed to proceed until there was a visible population crash (~7 hours). The well showing a lysis were pooled, centrifuged and filtered on 0.45 μ M membrane. These lysates comprise a mixture of homologous recombination-edited phage and wildtype phage.

Edited phage enrichment

To enrich genome-edited phage lysates, a phage counterselection strain consisting of W303 containing an ‘enrichment’ Cas13a vector (pBA559-24_60) was grown overnight in LB + Cm media at 37 °C. 100 µl of saturated overnight culture was mixed with molten LB top agar (0.7%) supplemented with Cm and poured on a LB agar plate containing Cm. Dilutions of phage 24 were then spotted and the plates incubated at 37°C. Plaques were isolated and purified for three rounds using the counterselection strain to ensure thorough removal of any remaining wildtype phages. Lysates were centrifugate to remove the remaining bacteria. These lysates comprised an enriched mixture of homologous recombination-edited phage. PCR and sequencing using 7326 and 7327 primers (located outside the homologous region) were performed on phages to confirm the 24_60 gene deletion.

Statistics and Reproducibility

All experimental data are representative of the results of at least three independent biological repeats. All replication attempts were successful. Statistical analysis of experimental data was done using GraphPad Prism 9.1.2 (GraphPad Software, Inc., CA, USA). Statistical analyses of the computational data were done in JMP (mixed models) and Python (the rest).

FUNDING

This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-UMR 3525), the Fondation pour la Recherche Médicale (FRM Grants No. EQU202103012569 and EQU201903007835), ANR Chromintevol (ANR-21-CE12-0002-01), and by the French Government’s Investissement d’Avenir program Laboratoire d’Excellence ‘Integrative Biology of Emerging Infectious Diseases’ [ANR-10-LABX-62-IBEID]. E.L. is supported by the Direction Générale de l’Armement (DGA).

ACKNOWLEDGEMENTS

We thank Florian Tesson for his technical guidance on the analysis of newly discovered anti-phage systems. We gratefully acknowledge François-Xavier Weill and Caroline Rouard from the Centre National de Référence des Vibrions et du Choléra (Institut Pasteur) from the gift of *Vibrio* phages. We gratefully acknowledge the UtechS Photonic BioImaging (Imagopole), C2RT, Institut Pasteur, supported by the French National Research Agency (France BioImaging; ANR-10-INBS-04; Investments for the Future). This work used the

computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paris. A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

AUTHOR CONTRIBUTIONS

B.D, M.B, F.D, W. S, K.D, P.A.K and C. L conceived the experimental project, designed the experiments, constructed strains and plasmids, performed the phage assays and analyzed the results. J.B performed Microscopy experiments. E.L and E.P.C.R conceived the computational project, designed the analyses, and analyzed the results. E.L made the bioinformatics analysis with the help of M.T. C.L, B.D, E.P.C.R, E. L wrote the paper with input from the other authors. C.L, D.M and E.P.C.R acquired funding. C.L, D.M, E.P.C.R and F.L.R supervised the project.

COMPETING INTEREST

None declared

REFERENCES

1. A. Chevallereau, B. J. Pons, S. van Houte, E. R. Westra, Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* **20**, 49-62 (2022).
2. H. Georjon, A. Bernheim, The highly diverse antiphage defence systems of bacteria. *Nat Rev Microbiol* **21**, 686-700 (2023).
3. A. Bernheim, R. Sorek, The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol* **18**, 113-119 (2020).
4. A. Agapov *et al.*, Multi-layered genome defences in bacteria. *Curr Opin Microbiol* **78**, 102436 (2024).
5. F. A. Hussain *et al.*, Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science* **374**, 488-492 (2021).
6. D. Mayo-Munoz, R. Pinilla-Redondo, S. Camara-Wilpert, N. Birkholz, P. C. Fineran, Inhibitors of bacterial immune systems: discovery, mechanisms and applications. *Nat Rev Genet*, (2024).
7. J. A. Escudero, C. Loot, A. Nivina, D. Mazel, The Integron: Adaptation On Demand. *Microbiology spectrum* **3**, MDNA3-0019-2014 (2015).
8. E. Richard, B. Darracq, C. Loot, D. Mazel, Unbridled Integrons: A Matter of Host Factors. *Cells* **11**, (2022).
9. G. Cambray, A. M. Guerout, D. Mazel, Integrons. *Annu Rev Genet* **44**, 141-166 (2010).
10. E. L. Fonseca, A. C. Vicente, Integron Functionality and Genome Innovation: An Update on the Subtle and Smart Strategy of Integrase and Gene Cassette Expression Regulation. *Microorganisms* **10**, (2022).
11. C. Loot *et al.*, Differences in Integron Cassette Excision Dynamics Shape a Trade-Off between Evolvability and Genetic Capacitance. *mBio* **8**, (2017).

12. E. Guerin *et al.*, The SOS response controls integron recombination. *Science* **324**, 1034 (2009).
13. H. W. Stokes, R. M. Hall, A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Molecular Microbiology* **3**, 1669-1683 (1989).
14. M. Buongiorno Pereira *et al.*, A comprehensive survey of integron-associated genes present in metagenomes. *BMC Genomics* **21**, 495 (2020).
15. T. M. Ghaly *et al.*, Functional enrichment of integrons: Facilitators of antimicrobial resistance and niche adaptation. *iScience* **26**, 108301 (2023).
16. P. H. Oliveira, M. Touchon, E. P. Rocha, The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* **42**, 10618-10631 (2014).
17. E. Richard *et al.*, Cassette recombination dynamics within chromosomal integrons are regulated by toxin-antitoxin systems. *Sci Adv* **10**, eadj3498 (2024).
18. D. Jurenas, N. Fraikin, F. Goormaghtigh, L. Van Melderen, Biology and evolution of bacterial toxin-antitoxin systems. *Nat Rev Microbiol* **20**, 335-350 (2022).
19. B. Neron *et al.*, IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria, with a Focus on Antibiotic Resistance in Klebsiella. *Microorganisms* **10**, (2022).
20. Y. Li *et al.*, Prophage encoding toxin/antitoxin system PfiT/PfiA inhibits Pf4 production in *Pseudomonas aeruginosa*. *Microb Biotechnol* **13**, 1132-1144 (2020).
21. E. Krin *et al.*, Systematic transcriptome analysis allows the identification of new type I and type II Toxin/Antitoxin systems located in the superintegron of *Vibrio cholerae*. *Res Microbiol* **174**, 103997 (2023).
22. L. Loeff, A. Walter, G. Rosalen, M. Jinek, DNA end sensing and cleavage by the Shedu anti-phage defense system. *bioRxiv*, (2023).
23. E. Krin *et al.*, Expansion of the SOS regulon of *Vibrio cholerae* through extensive transcriptome analysis and experimental validation. *BMC Genomics* **19**, 373 (2018).
24. P. Blanco *et al.*, Chromosomal Integrons are Genetically and Functionally Isolated Units of Genomes. *bioRxiv*, (2023).
25. M. Jaskolska, D. W. Adams, M. Blokesch, Two defence systems eliminate plasmids from seventh pandemic *Vibrio cholerae*. *Nature* **604**, 323-329 (2022).
26. T. A. Boyle, A. Hatoum-Aslan, Recurring and emerging themes in prokaryotic innate immunity. *Curr Opin Microbiol* **73**, 102324 (2023).
27. G. Mariano, T. R. Blower, Conserved domains can be found across distinct phage defence systems. *Mol Microbiol* **120**, 45-53 (2023).
28. F. Rousset, R. Sorek, The evolutionary success of regulated cell death in bacterial immunity. *Curr Opin Microbiol* **74**, 102312 (2023).
29. C. UniProt, UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515 (2019).
30. J. Soding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248 (2005).
31. M. Varadi *et al.*, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444 (2022).
32. M. van Kempen *et al.*, Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243-246 (2024).
33. A. Lopatina, N. Tal, R. Sorek, Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu Rev Virol* **7**, 371-384 (2020).

34. N. Tal *et al.*, Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nat Microbiol* **7**, 1200-1209 (2022).
35. D. H. Parma *et al.*, The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes & development* **6**, 497-510 (1992).
36. S. J. Labrie, J. E. Samson, S. Moineau, Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**, 317-327 (2010).
37. Y. M. Deng, M. L. Harvey, C. Q. Liu, N. W. Dunn, A novel plasmid-encoded phage abortive infection system from *Lactococcus lactis* biovar. diacetylactis. *FEMS Microbiol Lett* **146**, 149-154 (1997).
38. A. Forde, G. F. Fitzgerald, Bacteriophage defence systems in lactic acid bacteria. *Antonie Van Leeuwenhoek* **76**, 89-113 (1999).
39. F. Tesson *et al.*, A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases. *bioRxiv*, (2024).
40. V. Anantharaman, K. S. Makarova, A. M. Burroughs, E. V. Koonin, L. Aravind, Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol Direct* **8**, 15 (2013).
41. A. Fillol-Salom *et al.*, Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. *Cell* **185**, 3248-3262 e3220 (2022).
42. J. Lossouarn *et al.*, *Enterococcus faecalis* Countermeasures Defeat a Virulent Picovirinae Bacteriophage. *Viruses* **11**, (2019).
43. J. M. Bujnicki, L. Rychlewski, Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene* **267**, 183-191 (2001).
44. E. M. Kibby *et al.*, Bacterial NLR-related proteins protect against phage. *Cell* **186**, 2410-2424 e2418 (2023).
45. L. A. Gao *et al.*, Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* **377**, eabm4096 (2022).
46. S. Doron *et al.*, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, (2018).
47. S. Ryazansky, A. Kulbachinskiy, A. A. Aravin, The Expanded Universe of Prokaryotic Argonaute Proteins. *mBio* **9**, (2018).
48. E. Macdonald *et al.*, The novel anti-phage system Shield co-opts an RmuC domain to mediate phage defense across *Pseudomonas* species. *PLoS Genet* **19**, e1010784 (2023).
49. T. A. Tarkowski, D. Mooney, L. C. Thomason, F. W. Stahl, Gene products encoded in the ninR region of phage lambda participate in Red-mediated recombination. *Genes Cells* **7**, 351-363 (2002).
50. M. Sokolowska, H. Czapinska, M. Bochtler, Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res* **37**, 3799-3810 (2009).
51. S. Nair, I. Derre, T. Msadek, O. Gaillot, P. Berche, CtsR controls class III heat shock gene expression in the human pathogen *Listeria monocytogenes*. *Mol Microbiol* **35**, 800-811 (2000).
52. J. Durairaj *et al.*, Uncovering new families and folds in the natural protein universe. *Nature* **622**, 646-653 (2023).
53. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, (2021).
54. A. Millman *et al.*, Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-1561 e1512 (2020).
55. C. Souque, J. A. Escudero, R. C. MacLean, Integron activity accelerates the evolution of antibiotic resistance. *Elife* **10**, (2021).

56. O. Barraud, M. C. Ploy, Diversity of Class 1 Integron Gene Cassette Rearrangements Selected under Antibiotic Pressure. *J Bacteriol* **197**, 2171-2178 (2015).
57. D. Hocquet *et al.*, Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS Pathog* **8**, e1002778 (2012).
58. P. Blanco *et al.*, Identification of promoter activity in gene-less cassettes from Vibrionaceae superintegrons. *Nucleic Acids Res*, (2024).
59. C. Loot, D. Bikard, A. Rachlin, D. Mazel, Cellular pathways controlling integron cassette site folding. *EMBO J* **29**, 2623-2634 (2010).
60. A. Nivina, J. A. Escudero, C. Vit, D. Mazel, C. Loot, Efficiency of integron cassette insertion in correct orientation is ensured by the interplay of the three unpaired features of attC recombination sites. *Nucleic Acids Res* **44**, 7792-7803 (2016).
61. M. Bouvier, G. Demarre, D. Mazel, Integron cassette insertion: a recombination process involving a folded single strand substrate. *Embo J* **24**, 4356-4367 (2005).
62. C. Vit, C. Loot, J. A. Escudero, A. Nivina, D. Mazel, Integron Identification in Bacterial Genomes and Cassette Recombination Assays. *Methods Mol Biol* **2075**, 189-208 (2020).
63. F. Rousset *et al.*, Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740-753 e745 (2022).
64. P. D. Zongo *et al.*, An antiplasmid system drives antibiotic resistance gene integration in carbapenemase-producing *Escherichia coli* lineages. *Nature communications* **15**, 4093 (2024).
65. R. Barcia-Cruz *et al.*, Phage-inducible chromosomal minimalist islands (PICMIs), a novel family of small marine satellites of virulent phages. *Nature communications* **15**, 664 (2024).
66. K. D. Seed *et al.*, Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *mBio* **2**, e00334-00310 (2011).
67. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol* **193**, 6039-6056 (2011).
68. E. P. C. Rocha, D. Bikard, Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? *PLoS Biol* **20**, e3001514 (2022).
69. F. Tesson *et al.*, Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nature communications* **13**, 2561 (2022).
70. A. Perrin, E. P. C. Rocha, PanACoTA: a modular tool for massive microbial comparative genomics. *NAR Genom Bioinform* **3**, lqaa106 (2021).
71. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).
72. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).
73. B. Q. Minh *et al.*, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534 (2020).
74. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021).
75. C. Loot *et al.*, Integron cassettes integrate into bacterial genomes via widespread non-classical attG sites. *Nat Microbiol* **9**, 228-240 (2024).
76. A. P. Camargo *et al.*, Identification of mobile genetic elements with geNomad. *Nat Biotechnol*, (2023).
77. J. Mistry *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419 (2021).

78. M. Mirdita *et al.*, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).
79. R. L. Marvig, M. Blokesch, Natural transformation of *Vibrio cholerae* as a tool--optimizing the procedure. *BMC Microbiol* **10**, 155 (2010).
80. M. Yen, L. S. Cairns, A. Camilli, A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models. *Nature communications* **8**, 14187 (2017).
81. K. D. Seed *et al.*, Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in *Vibrio cholerae* O1. *PLoS Pathog* **8**, e1002917 (2012).
82. K. D. Seed *et al.*, Evolutionary consequences of intra-patient phage predation on microbial populations. *Elife* **3**, e03497 (2014).
83. R. K. Barman, A. K. Chakrabarti, S. Dutta, Screening of Potential *Vibrio cholerae* Bacteriophages for Cholera Therapy: A Comparative Genomic Approach. *Frontiers in microbiology* **13**, 803933 (2022).
84. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).
85. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* **6**, 26 (2011).
86. B. H. Burrowes, I. J. Molineux, J. A. Fralick, Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans Protocol. *Viruses* **11**, (2019).
87. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
88. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192 (2013).
89. B. A. Adler *et al.*, Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. *Nat Microbiol* **7**, 1967-1979 (2022).
90. S. G. Bhandare *et al.*, Complete Genome Sequences of *Vibrio cholerae*-Specific Bacteriophages 24 and X29. *Genome Announc* **5**, (2017).
91. E. Maffei *et al.*, Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. *PLoS Biol* **19**, e3001424 (2021).
92. M. Galtier *et al.*, Bacteriophages Targeting Adherent Invasive *Escherichia coli* Strains as a Promising New Treatment for Crohn's Disease. *J Crohns Colitis* **11**, 840-847 (2017).
93. F. Le Roux, J. Binesse, D. Saulnier, D. Mazel, Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene *vsm* by use of a novel counterselectable suicide vector. *Appl Environ Microbiol* **73**, 777-784 (2007).
94. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000).
95. F. de Lemos Martins, F. Fournes, M. V. Mazzuoli, D. Mazel, M. E. Val, *Vibrio cholerae* chromosome 2 copy number is controlled by the methylation-independent binding of its monomeric initiator to the chromosome 1 *crtS* site. *Nucleic Acids Res* **46**, 10145-10156 (2018).
96. C. Vit *et al.*, Cassette recruitment in the chromosomal Integron of *Vibrio cholerae*. *Nucleic Acids Res* **49**, 5654-5670 (2021).

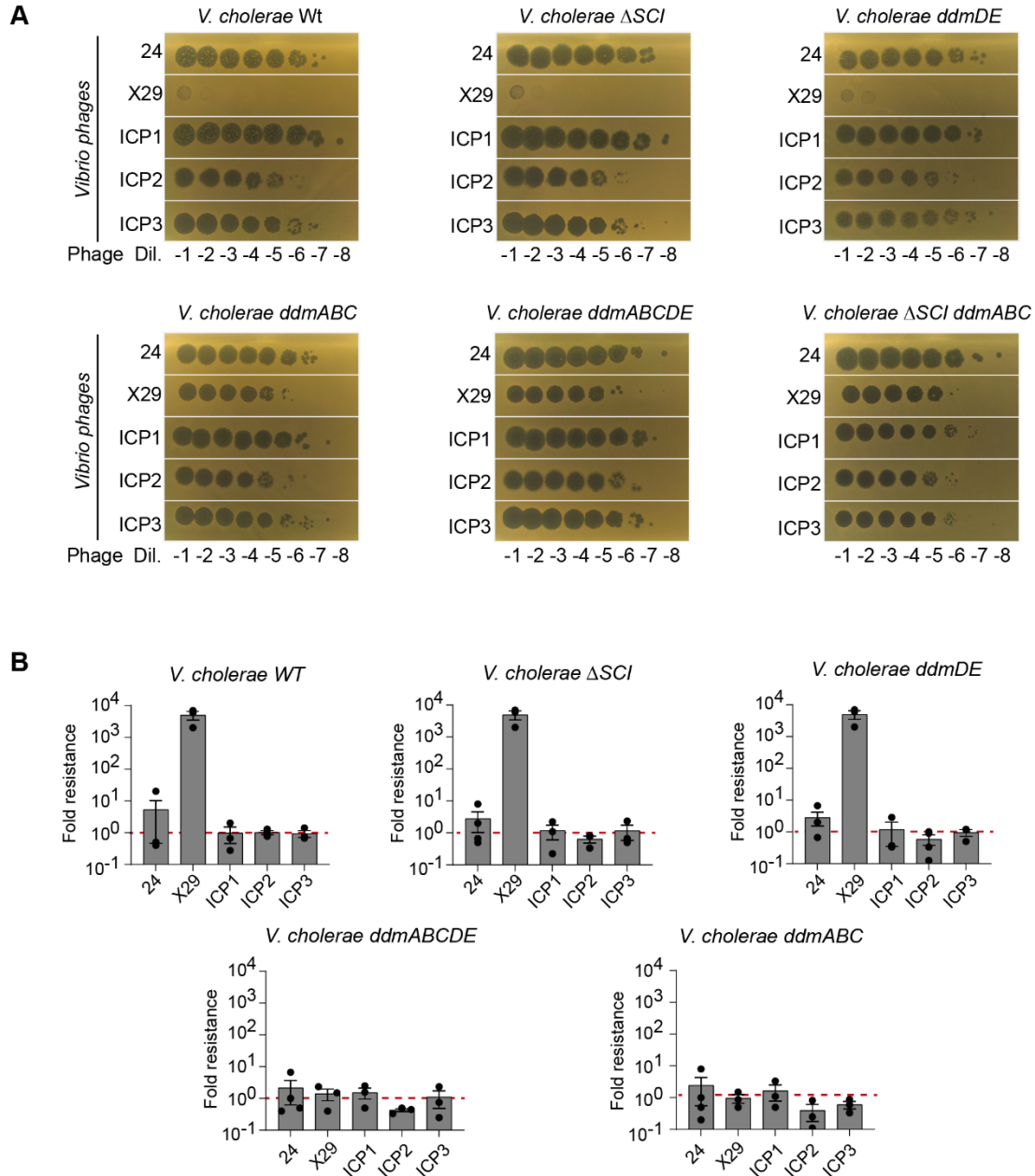


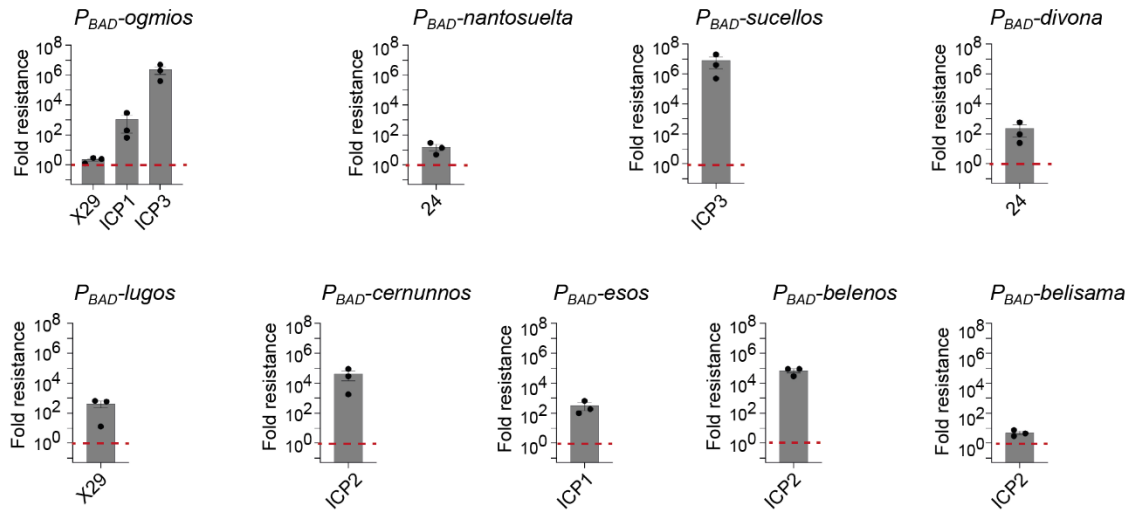
Figure S2: Detailed fold resistance of *V. cholerae* mutant strains against phages

A. 10-fold serial dilutions of a high titer lysate of phages spotted on *V. cholerae* strains. Representative images from a single replicate out of three independent replicates are shown.

B. Plaque-forming units were measured for each phage on the indicated *V. cholerae* strains and on the ΔSCI *ddmABC* cells. Fold-resistance was measured as the ratio between these two values. The bar charts show the mean fold resistance for three independent replicates ($n=3$, individual plots). Error bars shown the standard error of mean (SEM).

A

Phage resistance in V. cholerae



B

Phage resistance in E. coli

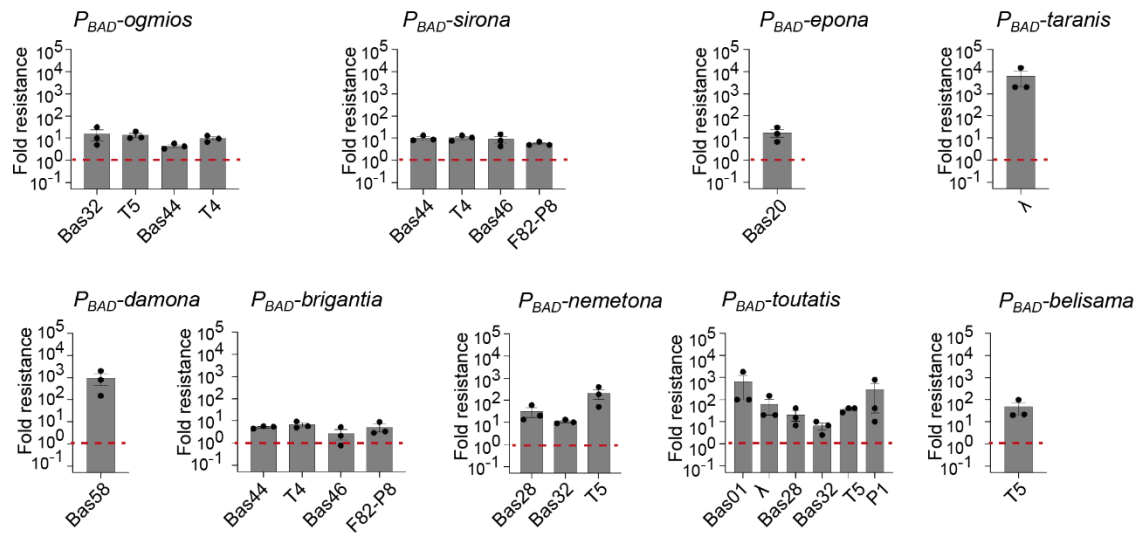


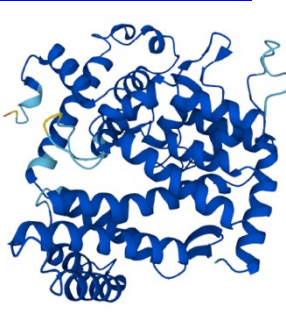
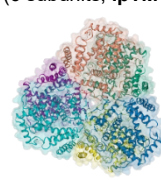
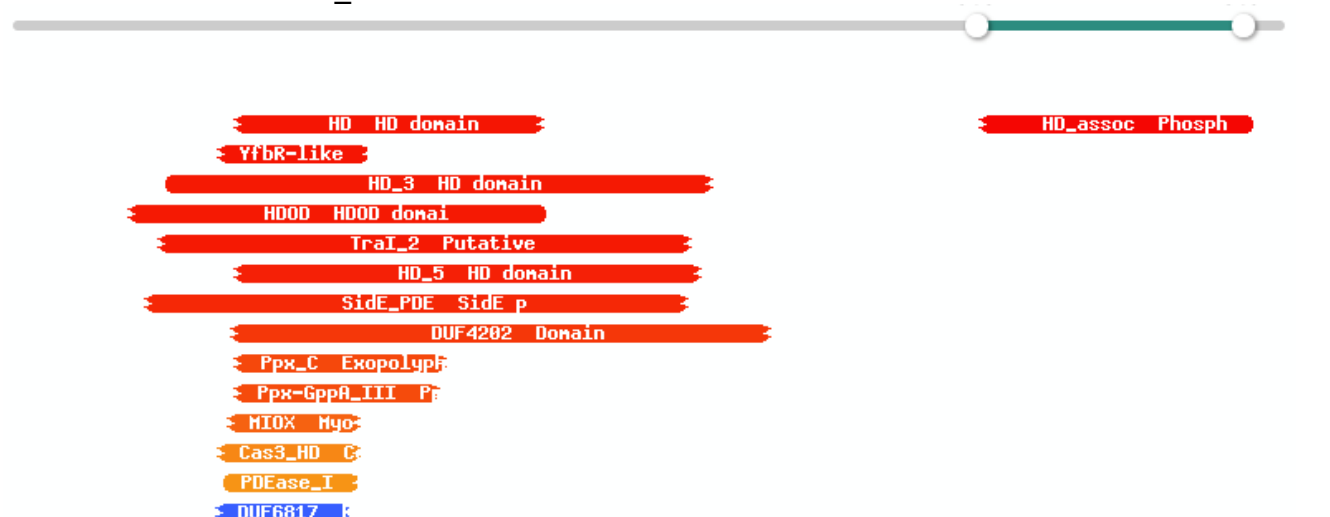
Figure S3: Detailed fold resistance of all defence systems against phages, Related to Figure 2

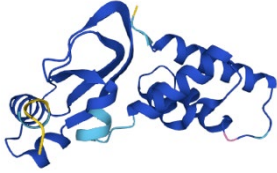
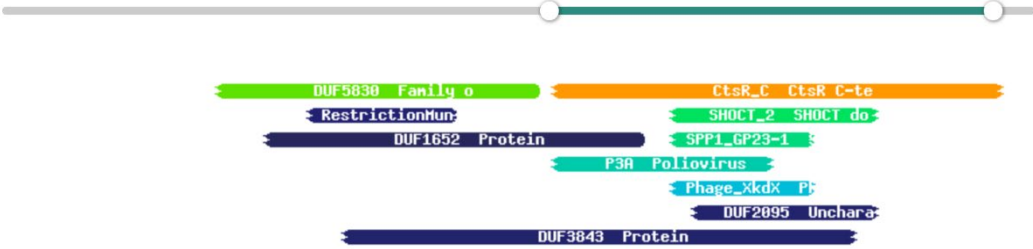
Plaque-forming units were measured for each phage on *V. cholerae* (A) or *E. coli* (B) cells harboring either a control plasmid or a defence system. Fold-resistance was measured as the ratio between these two values. The bar charts show the mean fold resistance for three independent replicates (n=3, individual plots). Error bars shown the standard error of mean (SEM).

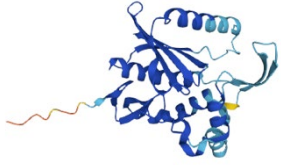

Figure S4: Detailed description of the anti-phage systems found in the *V. cholerae* SCI.

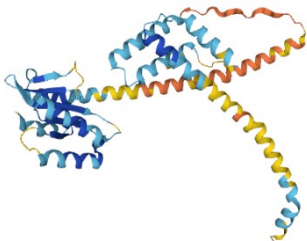
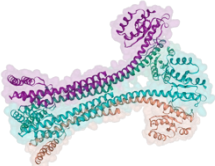
All the available data gathered is summarized in an individual sheet by system. The first sheet is a template sheet that explains where each data comes from.

System name	Uniprot annotation Gene identifier in <i>V. cholerae</i> El Tor str. N16961 / Uniprot ID	Length Gene length in aa
AlphaFold2 structure available at EBI Result of the best AlphaFold2 multimer cofold prediction (nb of subunits / ipTM)	AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”, except when less than 3 are uncharacterized) AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”, except when less than 3 are uncharacterized)	
Results of Foldseek on the AFDB-Proteome database Results of Foldseek on the AFDB-SwissProt database Results of Foldseek on the PDB100 database		
HHPred hits vs PFAM-A_v36		
InterPro/PFAM hits according to Uniprot		
Miscellaneous information found in literature		
Selected references corresponding to the other features above		

<h1>Ogmios</h1>	Deoxyguanosinetriphosphate triphosphohydrolase-like protein 2 VCA0308 / Uniprot ID Q9KMM3	Length 402 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (6 subunits, ipTM = 0.77)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="507 241 1465 474"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A1M6EJV6-F1</td> <td>dGTPase</td> <td><i>Aquimarina spongiae</i></td> <td>447</td> <td>95.19</td> </tr> <tr> <td>AF-A0A6L6UET7-F1</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase</td> <td><i>Winogradskyella endarachnes</i></td> <td>445</td> <td>95.06</td> </tr> <tr> <td>AF-G8R1E7-F1</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase</td> <td><i>Owenweeksia hongkongensis</i></td> <td>458</td> <td>95.06</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="507 524 1465 748"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A381NSQ7-F1</td> <td>HDC domain-containing protein</td> <td>Unknown (marine metagenome)</td> <td>439</td> <td>95.12</td> </tr> <tr> <td>AF-A0A024FFC9-F1</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase</td> <td><i>Winogradskyella sp. PG-2</i></td> <td>444</td> <td>94.94</td> </tr> <tr> <td>AF-W0EHQ4-F1</td> <td>HD domain-containing protein</td> <td><i>Desulfitobacterium metallireducens</i></td> <td>468</td> <td>94.94</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A1M6EJV6-F1	dGTPase	<i>Aquimarina spongiae</i>	447	95.19	AF-A0A6L6UET7-F1	Deoxyguanosinetriphosphate triphosphohydrolase	<i>Winogradskyella endarachnes</i>	445	95.06	AF-G8R1E7-F1	Deoxyguanosinetriphosphate triphosphohydrolase	<i>Owenweeksia hongkongensis</i>	458	95.06	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A381NSQ7-F1	HDC domain-containing protein	Unknown (marine metagenome)	439	95.12	AF-A0A024FFC9-F1	Deoxyguanosinetriphosphate triphosphohydrolase	<i>Winogradskyella sp. PG-2</i>	444	94.94	AF-W0EHQ4-F1	HD domain-containing protein	<i>Desulfitobacterium metallireducens</i>	468	94.94
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A1M6EJV6-F1	dGTPase	<i>Aquimarina spongiae</i>	447	95.19																																						
AF-A0A6L6UET7-F1	Deoxyguanosinetriphosphate triphosphohydrolase	<i>Winogradskyella endarachnes</i>	445	95.06																																						
AF-G8R1E7-F1	Deoxyguanosinetriphosphate triphosphohydrolase	<i>Owenweeksia hongkongensis</i>	458	95.06																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A381NSQ7-F1	HDC domain-containing protein	Unknown (marine metagenome)	439	95.12																																						
AF-A0A024FFC9-F1	Deoxyguanosinetriphosphate triphosphohydrolase	<i>Winogradskyella sp. PG-2</i>	444	94.94																																						
AF-W0EHQ4-F1	HD domain-containing protein	<i>Desulfitobacterium metallireducens</i>	468	94.94																																						
<p>Foldseek (AFDB-Proteome)</p> <table border="1" data-bbox="129 790 1465 902"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-K0ESW1-F1-model_v4</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase-like ...</td> <td><i>Nocardia brasiliensis</i> ATCC 700358</td> <td>0.99</td> <td>25.9</td> <td>0.821</td> <td>11-392</td> <td></td> </tr> <tr> <td>AF-P9WNY7-F1-model_v4</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase-like ...</td> <td><i>Mycobacterium tuberculosis</i> H37Rv</td> <td>0.99</td> <td>25.6</td> <td>0.816</td> <td>11-395</td> <td></td> </tr> <tr> <td>AF-Q9CCG3-F1-model_v4</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase-like ...</td> <td><i>Mycobacterium leprae</i> TN</td> <td>0.99</td> <td>24.4</td> <td>0.815</td> <td>11-393</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-K0ESW1-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Nocardia brasiliensis</i> ATCC 700358	0.99	25.9	0.821	11-392		AF-P9WNY7-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Mycobacterium tuberculosis</i> H37Rv	0.99	25.6	0.816	11-395		AF-Q9CCG3-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Mycobacterium leprae</i> TN	0.99	24.4	0.815	11-393									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-K0ESW1-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Nocardia brasiliensis</i> ATCC 700358	0.99	25.9	0.821	11-392																																				
AF-P9WNY7-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Mycobacterium tuberculosis</i> H37Rv	0.99	25.6	0.816	11-395																																				
AF-Q9CCG3-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Mycobacterium leprae</i> TN	0.99	24.4	0.815	11-393																																				
<p>Foldseek (AFDB-SwissProt)</p> <table border="1" data-bbox="129 936 1465 1048"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q9KMM3-F1-model_v4</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase-like ...</td> <td><i>Vibrio cholerae</i> O1 biovar El Tor str...</td> <td>1.00</td> <td>100</td> <td>1.000</td> <td>1-402</td> <td></td> </tr> <tr> <td>AF-Q4FM11-F1-model_v4</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase-like ...</td> <td><i>Candidatus Pelagibacter ubique</i> HT...</td> <td>0.98</td> <td>25.3</td> <td>0.827</td> <td>3-350</td> <td></td> </tr> <tr> <td>AF-Q9RNN5-F1-model_v4</td> <td>Deoxyguanosinetriphosphate triphosphohydrolase-like ...</td> <td><i>Zymomonas mobilis</i> subsp. mobilis...</td> <td>0.98</td> <td>24.8</td> <td>0.826</td> <td>7-397</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-Q9KMM3-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	1.00	100	1.000	1-402		AF-Q4FM11-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Candidatus Pelagibacter ubique</i> HT...	0.98	25.3	0.827	3-350		AF-Q9RNN5-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Zymomonas mobilis</i> subsp. mobilis...	0.98	24.8	0.826	7-397									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-Q9KMM3-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	1.00	100	1.000	1-402																																				
AF-Q4FM11-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Candidatus Pelagibacter ubique</i> HT...	0.98	25.3	0.827	3-350																																				
AF-Q9RNN5-F1-model_v4	Deoxyguanosinetriphosphate triphosphohydrolase-like ...	<i>Zymomonas mobilis</i> subsp. mobilis...	0.98	24.8	0.826	7-397																																				
<p>Foldseek (PDB100)</p> <table border="1" data-bbox="129 1081 1465 1193"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>7lwz-assembly1_B</td> <td>Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...</td> <td><i>Vibrio cholerae</i> O1 biovar El Tor str...</td> <td>0.98</td> <td>28.6</td> <td>0.801</td> <td>15-392</td> <td></td> </tr> <tr> <td>7lwz-assembly1_F</td> <td>Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...</td> <td><i>Vibrio cholerae</i> O1 biovar El Tor str...</td> <td>0.97</td> <td>29</td> <td>0.795</td> <td>15-392</td> <td></td> </tr> <tr> <td>7lwz-assembly1_D</td> <td>Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...</td> <td><i>Vibrio cholerae</i> O1 biovar El Tor str...</td> <td>0.98</td> <td>28.7</td> <td>0.793</td> <td>15-391</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	7lwz-assembly1_B	Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	0.98	28.6	0.801	15-392		7lwz-assembly1_F	Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	0.97	29	0.795	15-392		7lwz-assembly1_D	Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	0.98	28.7	0.793	15-391									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
7lwz-assembly1_B	Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	0.98	28.6	0.801	15-392																																				
7lwz-assembly1_F	Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	0.97	29	0.795	15-392																																				
7lwz-assembly1_D	Apo Structure of <i>Vibrio cholerae</i> dGTPase protein VC19...	<i>Vibrio cholerae</i> O1 biovar El Tor str...	0.98	28.7	0.793	15-391																																				
<p>HHPred hits vs PFAM-A_v36</p>  <p>HD HD domain</p> <p>Yfbr-like</p> <p>HD_3 HD domain</p> <p>HOOD HOOD domai</p> <p>TraI_2 Putative</p> <p>HD_5 HD domain</p> <p>SidE_PDE SidE p</p> <p>DUF4202 Domain</p> <p>Ppx_C Exopolyp</p> <p>Ppx-GppA_III P</p> <p>MIOX Myo</p> <p>Cas3_HD C</p> <p>PDEase_I</p> <p>DUF6817</p>																																										
<p>InterPro/PFAM hits (UniProt)</p> <p>IPR006261 (dNTPase); IPR023023 (dNTPase_2); IPR003607 (HD/PDEase_dom); IPR006674 (HD_domain); IPR026875 (PHydrolase_assoc_dom); PF01966 (HD); PF13286 (HD_assoc)</p>																																										
<p>Miscellaneous information</p> <p>Significant similarity with previously identified defensive dGTPases.</p>																																										
<p>Selected references</p> <p>Tal N, Millman A, Stokar-Avihail A, et al. Bacteria deplete deoxynucleotides to defend against bacteriophage infection. <i>Nat Microbiol.</i> 2022;7(8):1200-1209. doi:10.1038/s41564-022-01158-0 https://defensefinder.mdmlab.fr/wiki/defense-systems/dgtpase (DefenseFinder wiki about dGTPases)</p>																																										

Nantosuelta	Uncharacterized protein VCA0322 / Uniprot ID Q9KML5	Length 150 aa																																								
<p>AlphaFold structure</p>  <p>Cofold: None</p>	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 277 1466 479"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A6G6NIZ1-F1</td> <td>Uncharacterized protein</td> <td><i>Aeromonas veronii</i></td> <td>148</td> <td>94.06</td> </tr> <tr> <td>AF-A0A7U8WNF3-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio cholerae</i> MAK 757</td> <td>150</td> <td>93.38</td> </tr> <tr> <td>AF-A0A241PVU6-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio anguillarum</i></td> <td>148</td> <td>93.31</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 533 1466 734"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A3N0IUV5-F1</td> <td>Uncharacterized protein</td> <td><i>Eggerthella sinensis</i></td> <td>152</td> <td>95.12</td> </tr> <tr> <td>AF-A0A653I3Z7-F1</td> <td>Uncharacterized protein</td> <td><i>Limnobacter</i> sp. 130</td> <td>155</td> <td>93.94</td> </tr> <tr> <td>AF-A0A822MNB0-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio crassostreae</i></td> <td>148</td> <td>93.81</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A6G6NIZ1-F1	Uncharacterized protein	<i>Aeromonas veronii</i>	148	94.06	AF-A0A7U8WNF3-F1	Uncharacterized protein	<i>Vibrio cholerae</i> MAK 757	150	93.38	AF-A0A241PVU6-F1	Uncharacterized protein	<i>Vibrio anguillarum</i>	148	93.31	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A3N0IUV5-F1	Uncharacterized protein	<i>Eggerthella sinensis</i>	152	95.12	AF-A0A653I3Z7-F1	Uncharacterized protein	<i>Limnobacter</i> sp. 130	155	93.94	AF-A0A822MNB0-F1	Uncharacterized protein	<i>Vibrio crassostreae</i>	148	93.81
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A6G6NIZ1-F1	Uncharacterized protein	<i>Aeromonas veronii</i>	148	94.06																																						
AF-A0A7U8WNF3-F1	Uncharacterized protein	<i>Vibrio cholerae</i> MAK 757	150	93.38																																						
AF-A0A241PVU6-F1	Uncharacterized protein	<i>Vibrio anguillarum</i>	148	93.31																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A3N0IUV5-F1	Uncharacterized protein	<i>Eggerthella sinensis</i>	152	95.12																																						
AF-A0A653I3Z7-F1	Uncharacterized protein	<i>Limnobacter</i> sp. 130	155	93.94																																						
AF-A0A822MNB0-F1	Uncharacterized protein	<i>Vibrio crassostreae</i>	148	93.81																																						
Foldseek (AFDB-Proteome)																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q8CYR4-F1-model_v4</td> <td>Uncharacterized protein</td> <td><i>Streptococcus pneumoniae</i> R6</td> <td>0.91</td> <td>9.4</td> <td>0.499</td> <td>1-150</td> <td></td> </tr> <tr> <td>AF-A0A0R0JG8-F1-model_v4</td> <td>Uncharacterized protein</td> <td>Glycine max</td> <td>0.09</td> <td>5.5</td> <td>0.411</td> <td>1-150</td> <td></td> </tr> <tr> <td>AF-A0A0H3GP32-E1-model_v4</td> <td>Uncharacterized protein</td> <td><i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> ...</td> <td>0.12</td> <td>2.6</td> <td>0.393</td> <td>53-150</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	AF-Q8CYR4-F1-model_v4	Uncharacterized protein	<i>Streptococcus pneumoniae</i> R6	0.91	9.4	0.499	1-150		AF-A0A0R0JG8-F1-model_v4	Uncharacterized protein	Glycine max	0.09	5.5	0.411	1-150		AF-A0A0H3GP32-E1-model_v4	Uncharacterized protein	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> ...	0.12	2.6	0.393	53-150									
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																			
AF-Q8CYR4-F1-model_v4	Uncharacterized protein	<i>Streptococcus pneumoniae</i> R6	0.91	9.4	0.499	1-150																																				
AF-A0A0R0JG8-F1-model_v4	Uncharacterized protein	Glycine max	0.09	5.5	0.411	1-150																																				
AF-A0A0H3GP32-E1-model_v4	Uncharacterized protein	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> ...	0.12	2.6	0.393	53-150																																				
Foldseek (AFDB-SwissProt)																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-P4S919-F1-model_v4</td> <td>Uncharacterized protein YqbC</td> <td><i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168</td> <td>0.12</td> <td>5.1</td> <td>0.291</td> <td>32-150</td> <td></td> </tr> <tr> <td>AF-Q8B9J2-F1-model_v4</td> <td>Malonyl CoA-acyl carrier protein transacylase</td> <td><i>Bacillus subtilis</i></td> <td>0.18</td> <td>1.6</td> <td>0.285</td> <td>16-150</td> <td></td> </tr> <tr> <td>AF-B5FE04-F1-model_v4</td> <td>Elongation factor P-like protein</td> <td><i>Aliivibrio fischeri</i> MU11</td> <td>0.09</td> <td>1</td> <td>0.284</td> <td>1-148</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	AF-P4S919-F1-model_v4	Uncharacterized protein YqbC	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	0.12	5.1	0.291	32-150		AF-Q8B9J2-F1-model_v4	Malonyl CoA-acyl carrier protein transacylase	<i>Bacillus subtilis</i>	0.18	1.6	0.285	16-150		AF-B5FE04-F1-model_v4	Elongation factor P-like protein	<i>Aliivibrio fischeri</i> MU11	0.09	1	0.284	1-148									
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																			
AF-P4S919-F1-model_v4	Uncharacterized protein YqbC	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	0.12	5.1	0.291	32-150																																				
AF-Q8B9J2-F1-model_v4	Malonyl CoA-acyl carrier protein transacylase	<i>Bacillus subtilis</i>	0.18	1.6	0.285	16-150																																				
AF-B5FE04-F1-model_v4	Elongation factor P-like protein	<i>Aliivibrio fischeri</i> MU11	0.09	1	0.284	1-148																																				
Foldseek (PDB100)																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>3kup-assembly2_D</td> <td>Crystal Structure of the CBX3 Chromo Shadow Domain</td> <td><i>Homo sapiens</i></td> <td>0.09</td> <td>3.4</td> <td>0.473</td> <td>8-150</td> <td></td> </tr> <tr> <td>3p7i-assembly1_B</td> <td>Drosophila HP1a chromo shadow domain</td> <td><i>Drosophila melanogaster</i></td> <td>0.09</td> <td>3.5</td> <td>0.455</td> <td>7-143</td> <td></td> </tr> <tr> <td>3vmt-assembly2_B</td> <td>Crystal structure of Staphylococcus aureus membrane-bound tr...</td> <td><i>Staphylococcus aureus</i> subsp. <i>aureus</i> M...</td> <td>0.15</td> <td>4.7</td> <td>0.329</td> <td>15-150</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	3kup-assembly2_D	Crystal Structure of the CBX3 Chromo Shadow Domain	<i>Homo sapiens</i>	0.09	3.4	0.473	8-150		3p7i-assembly1_B	Drosophila HP1a chromo shadow domain	<i>Drosophila melanogaster</i>	0.09	3.5	0.455	7-143		3vmt-assembly2_B	Crystal structure of Staphylococcus aureus membrane-bound tr...	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> M...	0.15	4.7	0.329	15-150									
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																			
3kup-assembly2_D	Crystal Structure of the CBX3 Chromo Shadow Domain	<i>Homo sapiens</i>	0.09	3.4	0.473	8-150																																				
3p7i-assembly1_B	Drosophila HP1a chromo shadow domain	<i>Drosophila melanogaster</i>	0.09	3.5	0.455	7-143																																				
3vmt-assembly2_B	Crystal structure of Staphylococcus aureus membrane-bound tr...	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> M...	0.15	4.7	0.329	15-150																																				
HHPred hits vs PFAM-A_v36																																										
 <p>Diagram showing HHPred hits vs PFAM-A_v36. The main protein is represented by a green bar. Below it, several domain hits are shown as colored bars with their names: DUF5830 Family o (green), RestrictionMun (blue), DUF1652 Protein (dark blue), DUF3843 Protein (dark blue), CtsR_C CtsR C-te (orange), SHOCT_2 SHOCT do (green), SPP1_GP23-1 (green), P3R Poliovirus (teal), Phage_XkdX Pf (teal), and DUF2095 Unchara (dark blue).</p>																																										
InterPro/PFAM hits (UniProt)																																										
None																																										
Miscellaneous information																																										
<p>Nantosuelta shows some similarity with the C-terminal dimerization domain of CtsR (PF17727), a protein that controls heat shock gene expression in <i>Listeria monocytogenes</i> (Nair et al., 2002)</p>																																										
Selected references																																										
<p>Nair S, Derré I, Msadek T, Gaillot O, Berche P. CtsR controls class III heat shock gene expression in the human pathogen <i>Listeria monocytogenes</i>. <i>Mol Microbiol.</i> 2000;35(4):800-811. doi:10.1046/j.1365-2958.2000.01752.x</p>																																										

<h1>Sirona</h1>	Uncharacterized protein VCA0356 / Uniprot ID Q9KMJ2	Length 235 aa					
<p>AlphaFold structure</p>  <p>Cofold: None</p>	AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)						
	AFDB acc.	Description	Species	Length	pLDDT		
	AF-A0A0G0XTC8-F1	HNH_5 domain-containing protein	<i>Candidatus Woesebacteria bacterium</i>	203	93.44		
	AF-A0A495AVZ4-F1	Uncharacterized protein	<i>Vogesella indigofera</i>	243	94.38		
AF-A0A0E3Z328-F1	Uncharacterized protein	<i>Pseudoxanthomonas suwonensis</i>	245	94.31			
AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)							
AFDB acc.	Description	Species	Length	pLDDT			
AF-A0A2M7V7W9-F1	HNH_5 domain-containing protein	<i>Candidatus Magasanikbacteria bacterium</i>	203	94.31			
AF-A0A3A4T1H2-F1	HNH_5 domain-containing protein	<i>Desulfobacteraceae bacterium</i>	216	91.94			
AF-K2D3N0-F1	HNH family endonuclease	Unknown (uncultured bacterium)	226	86.56			
Foldseek (AFDB-Proteome)							
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
AF-Q2HYQ4-F1-model_v4	Probable transcriptional regulator	<i>Pseudomonas aeruginosa</i> PAO1	0.09	3.4	0.375	43-250	≡
AF-Q32FD1-F1-model_v4	Transcriptional regulator SlyA	<i>Shigella dysenteriae</i> Sd197	0.09	3.3	0.371	42-201	≡
AF-Q12049-F1-model_v4	Protein THP3	<i>Saccharomyces cerevisiae</i> S288C	0.12	1.3	0.269	7-224	≡
Foldseek (AFDB-SwissProt)							
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
AF-Q8D2L8-F1-model_v4	Transcriptional regulator SlyA	<i>Wigglesworthia glossinidia</i> endosymbiont...	0.10	2.7	0.379	42-201	≡
AF-B5XWN4-F1-model_v4	Transcriptional regulator SlyA	<i>Klebsiella pneumoniae</i> 342	0.10	2.8	0.379	42-200	≡
AF-P69889-F1-model_v4	Transcriptional regulator SlyA	<i>Yersinia pseudotuberculosis</i> IP 32953	0.09	2.2	0.378	42-201	≡
Foldseek (PDB100)							
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
4rqr-assembly1_A	Crystal Structure of Putative MarR Family Transcriptional Regul...	<i>Acinetobacter baylyi</i> ADP1	0.11	3.4	0.400	43-203	≡
1z6b-assembly4_B	Crystal structure of Plasmodium falciparum FabZ at 2.1 A	<i>Plasmodium falciparum</i>	0.09	3.1	0.398	10-225	≡
4rou-assembly1_B-2	Crystal Structure of Putative MarR Family Transcriptional Regul...	<i>Acinetobacter baylyi</i> ADP1	0.11	3.3	0.395	43-203	≡
HHPred hits vs PFAM-A_v36							
							
InterPro/PFAM hits (UniProt): None							
Miscellaneous information							
The type II restriction endonuclease hit (RE_Alw26IDE) only covers 10% of the PFAM domain. The distantly related λ's NinG (PF05766) or Rap (recombination adept with plasmid) is involved in homologous recombination catalysed by Escherichia coli's RecBCD pathway (Tarkowski et al., 2002).							
Selected references							
Tarkowski TA, Mooney D, Thomason LC, Stahl FW. Gene products encoded in the ninR region of phage lambda participate in Red-mediated recombination. <i>Genes Cells</i> . 2002;7(4):351-363. doi:10.1046/j.1365-2443.2002.00531.x							
Wu CC, Lin JLJ, Yuan HS. Structures, Mechanisms, and Functions of His-Me Finger Nucleases. <i>Trends Biochem Sci</i> . 2020;45(11):935-946. doi:10.1016/j.tibs.2020.07.002							

<h1>Epona</h1>	Restriction endonuclease type IV Mrr domain-containing protein VCA0366 / Uniprot ID Q9KMI4	Length 311 aa																				
<p>AlphaFold structure</p>  <p>Cofold (4 subunits, <i>ipTM</i> = 0.24)</p> 	AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)																					
	<table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A2D5P7V7-F1</td> <td>Mrr_cat domain-containing protein</td> <td><i>Arcobacter spp</i></td> <td>320</td> <td>83.25</td> </tr> <tr> <td>AF-A0A111C9B2-F1</td> <td>Restriction endonuclease</td> <td><i>Pseudomonas simiae</i></td> <td>312</td> <td>81.62</td> </tr> <tr> <td>AF-A0A2E4Q244-F1</td> <td>Mrr_cat domain-containing protein</td> <td><i>Cobetia sp</i></td> <td>295</td> <td>81.44</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A2D5P7V7-F1	Mrr_cat domain-containing protein	<i>Arcobacter spp</i>	320	83.25	AF-A0A111C9B2-F1	Restriction endonuclease	<i>Pseudomonas simiae</i>	312	81.62	AF-A0A2E4Q244-F1	Mrr_cat domain-containing protein	<i>Cobetia sp</i>	295	81.44
AFDB acc.	Description	Species	Length	pLDDT																		
AF-A0A2D5P7V7-F1	Mrr_cat domain-containing protein	<i>Arcobacter spp</i>	320	83.25																		
AF-A0A111C9B2-F1	Restriction endonuclease	<i>Pseudomonas simiae</i>	312	81.62																		
AF-A0A2E4Q244-F1	Mrr_cat domain-containing protein	<i>Cobetia sp</i>	295	81.44																		
	AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)																					
	<table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A7X2M0U1-F1</td> <td>Mrr_cat domain-containing protein</td> <td><i>Bacillus lacus</i></td> <td>309</td> <td>84.75</td> </tr> <tr> <td>AF-A0A521XC13-F1</td> <td>Mrr_cat domain-containing protein</td> <td><i>Bacteriodes bacterium</i></td> <td>317</td> <td>83.69</td> </tr> <tr> <td>AF-F3LFX7-F1</td> <td>Mrr_cat domain-containing protein</td> <td><i>gamma proteobacterium IMCC1989</i></td> <td>320</td> <td>83.06</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A7X2M0U1-F1	Mrr_cat domain-containing protein	<i>Bacillus lacus</i>	309	84.75	AF-A0A521XC13-F1	Mrr_cat domain-containing protein	<i>Bacteriodes bacterium</i>	317	83.69	AF-F3LFX7-F1	Mrr_cat domain-containing protein	<i>gamma proteobacterium IMCC1989</i>	320	83.06
AFDB acc.	Description	Species	Length	pLDDT																		
AF-A0A7X2M0U1-F1	Mrr_cat domain-containing protein	<i>Bacillus lacus</i>	309	84.75																		
AF-A0A521XC13-F1	Mrr_cat domain-containing protein	<i>Bacteriodes bacterium</i>	317	83.69																		
AF-F3LFX7-F1	Mrr_cat domain-containing protein	<i>gamma proteobacterium IMCC1989</i>	320	83.06																		

Foldseek (AFDB-Proteome)

Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
AF-G6TKX1-F1-model_v4	t-SNARE coiled-coil homology domain-containing protein	<i>Glycine max</i>	0.09	3.8	0.328	54-230	
AF-A0A0P0W8V3-F1-model_v4	Oso4g0280200 protein	<i>Oryza sativa Japonica Group</i>	0.11	1.7	0.227	25-300	

Foldseek (AFDB-SwissProt)

Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
AF-Q9SELS-F1-model_v4	Vesicle transport v-SNARE 12	<i>Arabidopsis thaliana</i>	0.10	4.5	0.358	91-210	

Foldseek (PDB100)

Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
3mvd-assembly1_A	Truncated Human ATP-Citrate Lyase with Citrate Bound	<i>Homo sapiens</i>	0.14	2.2	0.323	35-311	
8e40-assembly1_A	Full-length APOBEC3G in complex with HIV-1 Vif, CBF-beta, a...	<i>Macaca mulatta</i>	0.10	2.7	0.299	42-308	
5te1-assembly1_A	C20S, C293G Mutant N-terminal Human ATP Citrate Lyase Bo...	<i>Homo sapiens</i>	0.15	2.3	0.278	25-311	

HHPred hits vs PFAM-A_v36



InterPro/PFAM hits (UniProt)

[IPR007560](#) (Restrict endonuc IV Mrr); [PF04471](#) (Mrr_cat)

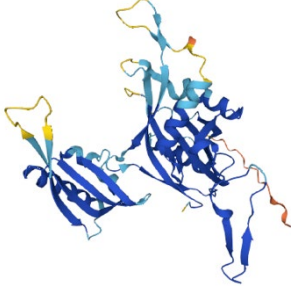
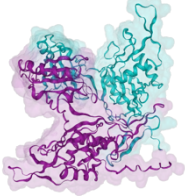
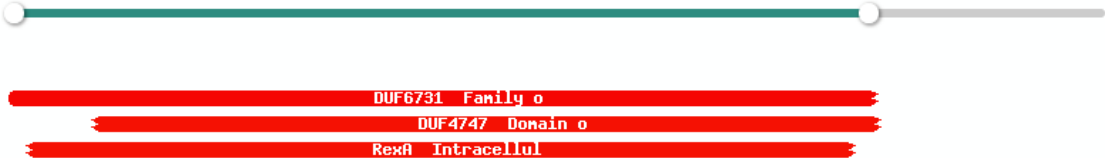
Miscellaneous information: 2 transmembrane domains

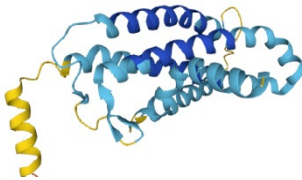
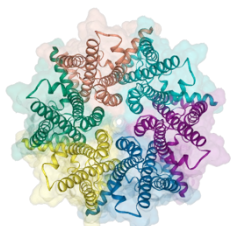


The Mrr_cat domain is part of the PDDEXK endonuclease clan CL0326, a catalytic domain found in type II and type IV restriction enzymes (Bujnicki et al., 2001) and type IV restriction enzymes. It can also be found in other recently described defence systems' effectors such as Lamassu, Avs and NLR-related systems.

Selected references

Bujnicki JM, Rychlewski L. Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene*. 2001;267(2):183-191. doi:10.1016/s0378-1119(01)00405-x

<h1>Sucellos (SciA)</h1>	Uncharacterized protein VCA0367 / Uniprot ID Q9KMI3	Length 340 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.81)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 264 1461 468"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A3B7LV14-F1</td> <td>DUF4747 family protein</td> <td><i>Acinetobacter chinensis</i></td> <td>292</td> <td>92</td> </tr> <tr> <td>AF-A0A376MST8-F1</td> <td>DUF4747 family protein</td> <td><i>Escherichia coli</i></td> <td>323</td> <td>88.25</td> </tr> <tr> <td>AF-A0A859GMZ5-F1</td> <td>DUF4747 family protein</td> <td><i>Vibrio cholerae</i></td> <td>327</td> <td>88.19</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 517 1461 721"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A389LZ03-F1</td> <td>Recombination-associated protein RdgC</td> <td>Unknown (γ-proteobacteria)</td> <td>293</td> <td>94</td> </tr> <tr> <td>AF-A0A506FFS8-F1</td> <td>DUF4747 family protein</td> <td><i>Acinetobacter baumannii</i></td> <td>292</td> <td>93.06</td> </tr> <tr> <td>AF-A0A849UIK3-F1</td> <td>DUF4747 family protein</td> <td><i>Methylotenera sp</i></td> <td>291</td> <td>92.69</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A3B7LV14-F1	DUF4747 family protein	<i>Acinetobacter chinensis</i>	292	92	AF-A0A376MST8-F1	DUF4747 family protein	<i>Escherichia coli</i>	323	88.25	AF-A0A859GMZ5-F1	DUF4747 family protein	<i>Vibrio cholerae</i>	327	88.19	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A389LZ03-F1	Recombination-associated protein RdgC	Unknown (γ -proteobacteria)	293	94	AF-A0A506FFS8-F1	DUF4747 family protein	<i>Acinetobacter baumannii</i>	292	93.06	AF-A0A849UIK3-F1	DUF4747 family protein	<i>Methylotenera sp</i>	291	92.69
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A3B7LV14-F1	DUF4747 family protein	<i>Acinetobacter chinensis</i>	292	92																																						
AF-A0A376MST8-F1	DUF4747 family protein	<i>Escherichia coli</i>	323	88.25																																						
AF-A0A859GMZ5-F1	DUF4747 family protein	<i>Vibrio cholerae</i>	327	88.19																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A389LZ03-F1	Recombination-associated protein RdgC	Unknown (γ -proteobacteria)	293	94																																						
AF-A0A506FFS8-F1	DUF4747 family protein	<i>Acinetobacter baumannii</i>	292	93.06																																						
AF-A0A849UIK3-F1	DUF4747 family protein	<i>Methylotenera sp</i>	291	92.69																																						
<p>Foldseek (AFDB-Proteome)</p> <table border="1" data-bbox="134 860 1461 936"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0H3GLU8-F1-model_v4</td> <td>Uncharacterized protein</td> <td><i>Klebsiella pneumoniae subsp. pneumoniae</i></td> <td>0.60</td> <td>5.5</td> <td>0.533</td> <td>1-321</td> <td></td> </tr> <tr> <td>AF-Q803R3-F1-model_v4</td> <td>Tyrosine-protein phosphatase non-receptor type</td> <td><i>Danio rerio</i></td> <td>0.09</td> <td>2.8</td> <td>0.267</td> <td>1-256</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-A0A0H3GLU8-F1-model_v4	Uncharacterized protein	<i>Klebsiella pneumoniae subsp. pneumoniae</i>	0.60	5.5	0.533	1-321		AF-Q803R3-F1-model_v4	Tyrosine-protein phosphatase non-receptor type	<i>Danio rerio</i>	0.09	2.8	0.267	1-256																	
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-A0A0H3GLU8-F1-model_v4	Uncharacterized protein	<i>Klebsiella pneumoniae subsp. pneumoniae</i>	0.60	5.5	0.533	1-321																																				
AF-Q803R3-F1-model_v4	Tyrosine-protein phosphatase non-receptor type	<i>Danio rerio</i>	0.09	2.8	0.267	1-256																																				
<p>Foldseek (AFDB-SwissProt): No hit</p>																																										
<p>Foldseek (PDB100): No hit</p>																																										
<p>HHPred hits vs PFAM-A_v36</p> 																																										
<p>InterPro/PFAM hits (UniProt): None</p>																																										
<p>Miscellaneous information</p> <p>Homologous with RexAB from phage λ (Parma et al., 1992) based on structure similarity. RexA senses phage infection and triggers the membrane-anchored protein RexB, an ion channel which destroys the membrane potential resulting in cell death (Labrie et al., 2010). SciA is not matched by the PFAM domain of RexA with default cutoffs. Several prophage-encoded RexAB-like defense systems were identified in Actinomycetota (Montgomery et al., 2019; Gentile et al., 2019; Mageeney et al., 2020) and are also not detected by DefenseFinder. DUF6731 is also found in PIC1-encoded defense systems (Fillol-Salom et al., 2022)</p>																																										
<p>Selected references</p> <p>Parma DH, Snyder M, Sobolevski S, Nawroz M, Brody E, Gold L. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. <i>Genes Dev.</i> 1992;6(3):497-510. doi:10.1101/gad.6.3.497</p> <p>Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. <i>Nat Rev Microbiol.</i> 2010;8(5):317-327. doi:10.1038/nrmicro2315</p> <p>Montgomery MT, Guerrero Bustamante CA, Dedrick RM, Jacobs-Sera D, Hatfull GF. Yet More Evidence of Collusion: a New Viral Defense System Encoded by <i>Gordonia</i> Phage CarolAnn. <i>mBio.</i> 2019;10(2):e02417-18. Published 2019 Mar 19. doi:10.1128/mBio.02417-18</p> <p>Gentile GM, Wetzel KS, Dedrick RM, et al. More Evidence of Collusion: a New Prophage-Mediated Viral Defense System Encoded by Mycobacteriophage Sbash. <i>mBio.</i> 2019;10(2):e00196-19. Published 2019 Mar 19. doi:10.1128/mBio.00196-19</p> <p>Mageeney CM, Mohammed HT, Dies M, et al. <i>Mycobacterium</i> Phage Butters-Encoded Proteins Contribute to Host Defense against Viral Attack. <i>mSystems.</i> 2020;5(5):e00534-20. Published 2020 Oct 6. doi:10.1128/mSystems.00534-20</p> <p>Fillol-Salom A, Rostøl JT, Ojiogu AD, et al. Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. <i>Cell.</i> 2022;185(17):3248-3262.e20. doi:10.1016/j.cell.2022.07.014</p> <p>https://defensefinder.mdmlab.fr/wiki/defense-systems/rexab (DefenseFinder wiki about RexAB)</p>																																										

<h1>Sucellos</h1> <p>(SciB)</p>	Uncharacterized protein VCA0368 / Uniprot ID Q9KMI2	Length 204 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (6 subunits, ipTM = 0.89)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-Q6D1I9-F1</td> <td>Putative membrane protein</td> <td><i>Pectobacterium atrosepticum</i></td> <td>207</td> <td>84.5</td> </tr> <tr> <td>AF-A0A379S1Q5-F1</td> <td>Phage membrane protein</td> <td><i>Salmonella enterica subsp. arizonae</i></td> <td>199</td> <td>75.31</td> </tr> <tr> <td>AF-M7CV74-F1</td> <td>Phage membrane protein</td> <td><i>Marinobacter santoriniensis</i></td> <td>210</td> <td>69.81</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-F6AMC0-F1</td> <td>Phage membrane protein</td> <td><i>Delftia sp. (strain Cs1-4)</i></td> <td>204</td> <td>75.38</td> </tr> <tr> <td>AF-M5DNA8-F1</td> <td>Phage membrane protein</td> <td><i>Thalassolituus oleivorans MIL-1</i></td> <td>202</td> <td>72.62</td> </tr> <tr> <td>AF-A0A6J4FA48-F1</td> <td>Phage membrane protein</td> <td><i>Magnetospirillum sp. LM-5</i></td> <td>203</td> <td>67.56</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-Q6D1I9-F1	Putative membrane protein	<i>Pectobacterium atrosepticum</i>	207	84.5	AF-A0A379S1Q5-F1	Phage membrane protein	<i>Salmonella enterica subsp. arizonae</i>	199	75.31	AF-M7CV74-F1	Phage membrane protein	<i>Marinobacter santoriniensis</i>	210	69.81	AFDB acc.	Description	Species	Length	pLDDT	AF-F6AMC0-F1	Phage membrane protein	<i>Delftia sp. (strain Cs1-4)</i>	204	75.38	AF-M5DNA8-F1	Phage membrane protein	<i>Thalassolituus oleivorans MIL-1</i>	202	72.62	AF-A0A6J4FA48-F1	Phage membrane protein	<i>Magnetospirillum sp. LM-5</i>	203	67.56
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-Q6D1I9-F1	Putative membrane protein	<i>Pectobacterium atrosepticum</i>	207	84.5																																						
AF-A0A379S1Q5-F1	Phage membrane protein	<i>Salmonella enterica subsp. arizonae</i>	199	75.31																																						
AF-M7CV74-F1	Phage membrane protein	<i>Marinobacter santoriniensis</i>	210	69.81																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-F6AMC0-F1	Phage membrane protein	<i>Delftia sp. (strain Cs1-4)</i>	204	75.38																																						
AF-M5DNA8-F1	Phage membrane protein	<i>Thalassolituus oleivorans MIL-1</i>	202	72.62																																						
AF-A0A6J4FA48-F1	Phage membrane protein	<i>Magnetospirillum sp. LM-5</i>	203	67.56																																						

Foldseek (AFDB-Proteome)

Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
AF-A0A5K4EK17-F1-model_v4	Uncharacterized protein	Schistosoma mansoni	0.28	2.5	0.474	49-164	
AF-C6T1G0-F1-model_v4	CASP-like protein 7	Glycine max	0.33	7	0.473	52-193	
AF-K7VVC0-F1-model_v4	CASP-like protein	Zea mays	0.33	7.6	0.471	53-193	

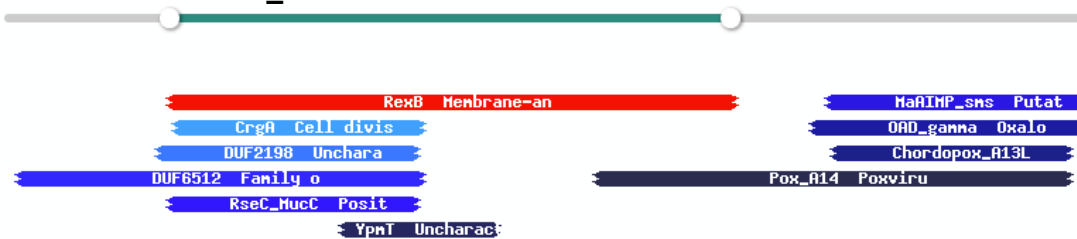
Foldseek (AFDB-SwissProt)

Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
AF-Q82195-F1-model_v4	Phenylalanine ammonia-lyase	Agaricus bisporus	0.25	6.6	0.500	51-189	
AF-P0D159-F1-model_v4	Casparian strip membrane protein 1	Lactuca sativa	0.30	8.9	0.475	52-193	
AF-P0D158-F1-model_v4	CASP-like protein 2	Lotus japonicus	0.33	4.4	0.473	52-193	

Foldseek (PDB100)

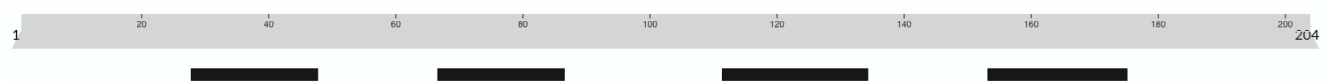
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment
6mux-assembly1_f	The structure of the Plasmodium falciparum 20S protea...	Plasmodium falciparum 3D7	0.47	3.5	0.526	34-186	
8dt0-assembly1_A	Scaffolding protein functional sites using deep learning	synthetic construct	0.30	3.8	0.523	35-187	
8dt0-assembly2_B	Scaffolding protein functional sites using deep learning	synthetic construct	0.30	3.8	0.519	35-186	

HHPred hits vs PFAM-A_v36



InterPro/PFAM hits (UniProt): None

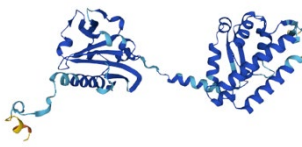
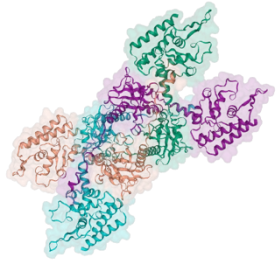

Miscellaneous information: 4 transmembrane domains


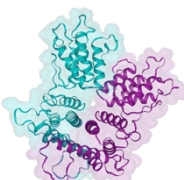
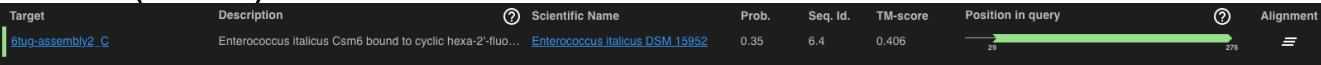
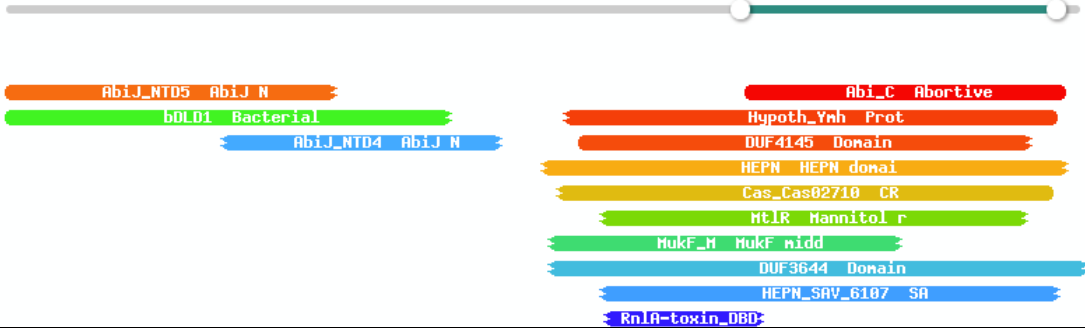


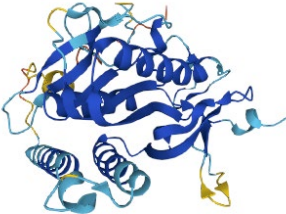
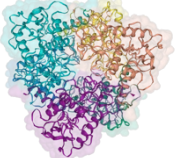
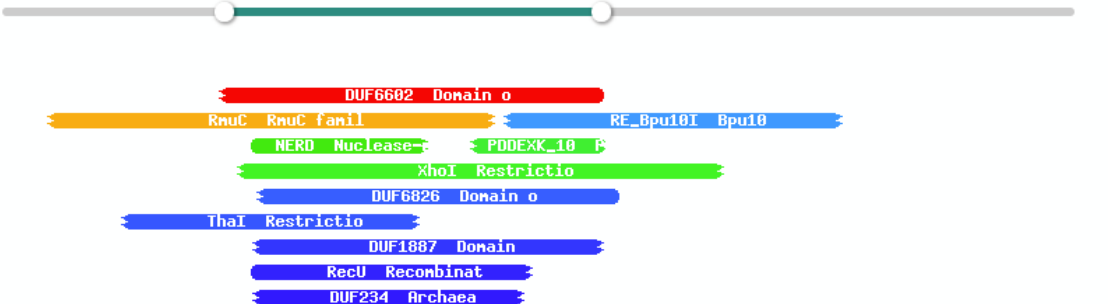
Cf SciA's miscellaneous information about RexAB

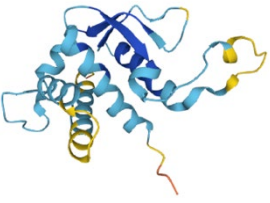
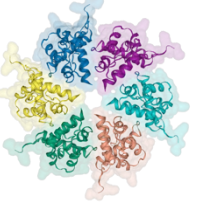
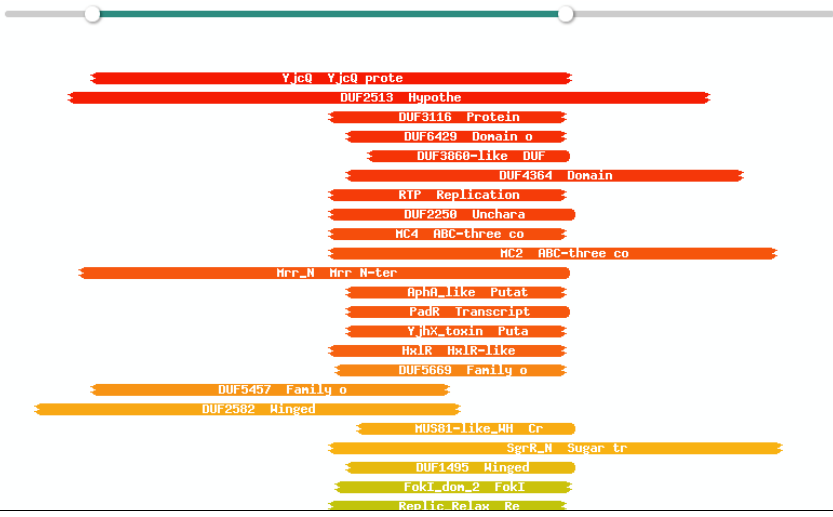
Selected references

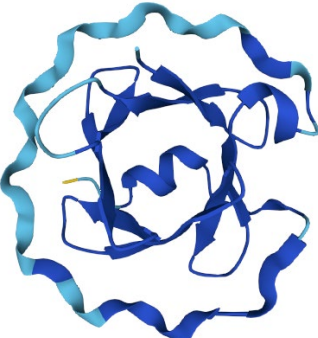

Cf SciA's references

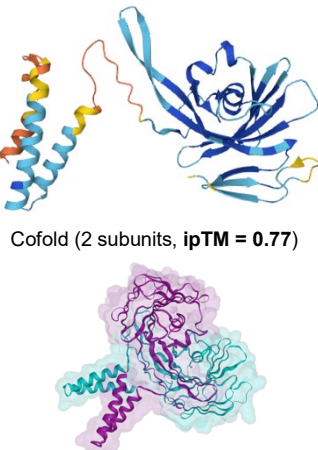


<h1>Divona</h1>	DUF4365 domain-containing protein VCA0374 / Uniprot ID Q9KMH9	Length 345 aa																																																								
<p>AlphaFold structure</p>  <p>Cofold (4 subunits, ipTM = 0.84)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A844PJ97-F1</td> <td>DUF1817 domain-containing protein</td> <td><i>Aliivibrio fischeri</i></td> <td>329</td> <td>92.25</td> </tr> <tr> <td>AF-A0A8B0IVC8-F1</td> <td>DUF4365 domain-containing protein</td> <td><i>Vibrio cholerae (strain MO10)</i></td> <td>335</td> <td>92.12</td> </tr> <tr> <td>AF-A0A559J5N6-F1</td> <td>DUF4365 and DUF1817 domain-containing protein</td> <td><i>Cohnella sp. G13</i></td> <td>333</td> <td>91.88</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A6L6PLJ5-F1</td> <td>DUF1817 domain-containing protein</td> <td><i>Duganella radidis</i></td> <td>339</td> <td>93.75</td> </tr> <tr> <td>AF-A0A2X1GV75-F1</td> <td>DUF4365 domain-containing protein</td> <td><i>Burkholderia cepacia</i></td> <td>338</td> <td>92.75</td> </tr> <tr> <td>AF-A0A2N1HLB0-F1</td> <td>DUF4365 domain-containing protein</td> <td><i>Colwellia sp. Bg11-28</i></td> <td>324</td> <td>92.69</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A844PJ97-F1	DUF1817 domain-containing protein	<i>Aliivibrio fischeri</i>	329	92.25	AF-A0A8B0IVC8-F1	DUF4365 domain-containing protein	<i>Vibrio cholerae (strain MO10)</i>	335	92.12	AF-A0A559J5N6-F1	DUF4365 and DUF1817 domain-containing protein	<i>Cohnella sp. G13</i>	333	91.88	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A6L6PLJ5-F1	DUF1817 domain-containing protein	<i>Duganella radidis</i>	339	93.75	AF-A0A2X1GV75-F1	DUF4365 domain-containing protein	<i>Burkholderia cepacia</i>	338	92.75	AF-A0A2N1HLB0-F1	DUF4365 domain-containing protein	<i>Colwellia sp. Bg11-28</i>	324	92.69																
AFDB acc.	Description	Species	Length	pLDDT																																																						
AF-A0A844PJ97-F1	DUF1817 domain-containing protein	<i>Aliivibrio fischeri</i>	329	92.25																																																						
AF-A0A8B0IVC8-F1	DUF4365 domain-containing protein	<i>Vibrio cholerae (strain MO10)</i>	335	92.12																																																						
AF-A0A559J5N6-F1	DUF4365 and DUF1817 domain-containing protein	<i>Cohnella sp. G13</i>	333	91.88																																																						
AFDB acc.	Description	Species	Length	pLDDT																																																						
AF-A0A6L6PLJ5-F1	DUF1817 domain-containing protein	<i>Duganella radidis</i>	339	93.75																																																						
AF-A0A2X1GV75-F1	DUF4365 domain-containing protein	<i>Burkholderia cepacia</i>	338	92.75																																																						
AF-A0A2N1HLB0-F1	DUF4365 domain-containing protein	<i>Colwellia sp. Bg11-28</i>	324	92.69																																																						
<p>Foldseek (AFDB-Proteome)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-B4FZ14-F1-model_v4</td> <td>Protein CHLORORESPIRATORY REDUCTION 6 chloro...</td> <td><i>Zea mays</i></td> <td>0.16</td> <td>14.8</td> <td>0.433</td> <td>185-341</td> <td></td> </tr> <tr> <td>AF-O82258-F1-model_v4</td> <td>Protein CHLORORESPIRATORY REDUCTION 6, chlor...</td> <td><i>Arabidopsis thaliana</i></td> <td>0.15</td> <td>15</td> <td>0.431</td> <td>185-341</td> <td></td> </tr> <tr> <td>AF-Q0PC90-F1-model_v4</td> <td>Uncharacterized protein</td> <td><i>Campylobacter jejuni subsp. jejuni...</i></td> <td>0.10</td> <td>7.7</td> <td>0.281</td> <td>10-185</td> <td></td> </tr> </tbody> </table> <p>Foldseek (AFDB-SwissProt)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-O82258-F1-model_v4</td> <td>Protein CHLORORESPIRATORY REDUCTION 6, chlor...</td> <td><i>Arabidopsis thaliana</i></td> <td>0.15</td> <td>15</td> <td>0.431</td> <td>185-341</td> <td></td> </tr> <tr> <td>AF-Q5WGB6-F1-model_v4</td> <td>Holliday junction resolvase RecU</td> <td><i>Alkalihalobacillus clausii KSM-K16</i></td> <td>0.09</td> <td>5.5</td> <td>0.389</td> <td>5-165</td> <td></td> </tr> </tbody> </table> <p>Foldseek (PDB100)</p> <p>No hit</p> <p>HHPred hits vs PFAM-A_v36</p> 			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-B4FZ14-F1-model_v4	Protein CHLORORESPIRATORY REDUCTION 6 chloro...	<i>Zea mays</i>	0.16	14.8	0.433	185-341		AF-O82258-F1-model_v4	Protein CHLORORESPIRATORY REDUCTION 6, chlor...	<i>Arabidopsis thaliana</i>	0.15	15	0.431	185-341		AF-Q0PC90-F1-model_v4	Uncharacterized protein	<i>Campylobacter jejuni subsp. jejuni...</i>	0.10	7.7	0.281	10-185		Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-O82258-F1-model_v4	Protein CHLORORESPIRATORY REDUCTION 6, chlor...	<i>Arabidopsis thaliana</i>	0.15	15	0.431	185-341		AF-Q5WGB6-F1-model_v4	Holliday junction resolvase RecU	<i>Alkalihalobacillus clausii KSM-K16</i>	0.09	5.5	0.389	5-165	
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																																			
AF-B4FZ14-F1-model_v4	Protein CHLORORESPIRATORY REDUCTION 6 chloro...	<i>Zea mays</i>	0.16	14.8	0.433	185-341																																																				
AF-O82258-F1-model_v4	Protein CHLORORESPIRATORY REDUCTION 6, chlor...	<i>Arabidopsis thaliana</i>	0.15	15	0.431	185-341																																																				
AF-Q0PC90-F1-model_v4	Uncharacterized protein	<i>Campylobacter jejuni subsp. jejuni...</i>	0.10	7.7	0.281	10-185																																																				
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																																			
AF-O82258-F1-model_v4	Protein CHLORORESPIRATORY REDUCTION 6, chlor...	<i>Arabidopsis thaliana</i>	0.15	15	0.431	185-341																																																				
AF-Q5WGB6-F1-model_v4	Holliday junction resolvase RecU	<i>Alkalihalobacillus clausii KSM-K16</i>	0.09	5.5	0.389	5-165																																																				
<p>InterPro/PFAM hits (UniProt)</p> <p>IPR014946 (CRR6); IPR025375 (DUF4365); PF08847 (Crr6); PF14280 (DUF4365)</p>																																																										
<p>Miscellaneous information</p> <p>The domain DUF4365 is found in RecB-like nucleases of the Mrr subfamily (Ryazansky et al., 2018) and in pAgo (sometimes mentioned as a PDDEXK domain (Lu et al., 2023))</p>																																																										
<p>Selected references</p> <p>Lu X, Xiao J, Wang L, Zhu B, Huang F. The nuclease-associated short prokaryotic Argonaute system nonspecifically degrades DNA upon activation by target recognition. <i>Nucleic Acids Res.</i> 2024;52(2):844-855. doi:10.1093/nar/gkad1145</p> <p>Ryazansky S, Kulbachinskiy A, Aravin AA. The Expanded Universe of Prokaryotic Argonaute Proteins. <i>mBio.</i> 2018;9(6):e01935-18. Published 2018 Dec 18. doi:10.1128/mBio.01935-18</p> <p>https://defensefinder.mdmlab.fr/wiki/defense-systems/pago (DefenseFinder wiki about pAgo)</p>																																																										

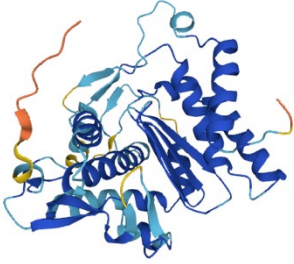

<h1>Taranis</h1>	Abortive infection protein-like C-terminal domain-containing protein VCA0396 / Uniprot ID Q9KMG2	Length 276 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.63)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="512 241 1461 443"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-G4IA54-F1</td> <td>Abi_C domain-containing protein</td> <td><i>Mycolicibacterium rhodesiae JS60</i></td> <td>277</td> <td>88.5</td> </tr> <tr> <td>AF-A0A5S4V200-F1</td> <td>Abortive infection family protein</td> <td><i>Agromyces sp. NEAU-184</i></td> <td>274</td> <td>87.75</td> </tr> <tr> <td>AF-A0A1A0VD76-F1</td> <td>Abi_C domain-containing protein</td> <td><i>Mycolicibacterium peregrinum</i></td> <td>274</td> <td>87.38</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="512 497 1461 698"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A098BNY7-F1</td> <td>Abi_C domain-containing protein</td> <td><i>Rhodococcus ruber</i></td> <td>273</td> <td>91.12</td> </tr> <tr> <td>AF-A0A150HCU4-F1</td> <td>Abi_C domain-containing protein</td> <td><i>Brevibacterium ravensturgense</i></td> <td>268</td> <td>88.56</td> </tr> <tr> <td>AF-A0A1X2F575-F1</td> <td>Abi_C domain-containing protein</td> <td><i>Mycolicibacterium wolinskyi</i></td> <td>274</td> <td>88.44</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-G4IA54-F1	Abi_C domain-containing protein	<i>Mycolicibacterium rhodesiae JS60</i>	277	88.5	AF-A0A5S4V200-F1	Abortive infection family protein	<i>Agromyces sp. NEAU-184</i>	274	87.75	AF-A0A1A0VD76-F1	Abi_C domain-containing protein	<i>Mycolicibacterium peregrinum</i>	274	87.38	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A098BNY7-F1	Abi_C domain-containing protein	<i>Rhodococcus ruber</i>	273	91.12	AF-A0A150HCU4-F1	Abi_C domain-containing protein	<i>Brevibacterium ravensturgense</i>	268	88.56	AF-A0A1X2F575-F1	Abi_C domain-containing protein	<i>Mycolicibacterium wolinskyi</i>	274	88.44
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-G4IA54-F1	Abi_C domain-containing protein	<i>Mycolicibacterium rhodesiae JS60</i>	277	88.5																																						
AF-A0A5S4V200-F1	Abortive infection family protein	<i>Agromyces sp. NEAU-184</i>	274	87.75																																						
AF-A0A1A0VD76-F1	Abi_C domain-containing protein	<i>Mycolicibacterium peregrinum</i>	274	87.38																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A098BNY7-F1	Abi_C domain-containing protein	<i>Rhodococcus ruber</i>	273	91.12																																						
AF-A0A150HCU4-F1	Abi_C domain-containing protein	<i>Brevibacterium ravensturgense</i>	268	88.56																																						
AF-A0A1X2F575-F1	Abi_C domain-containing protein	<i>Mycolicibacterium wolinskyi</i>	274	88.44																																						
<p>Foldseek (AFDB-Proteome) No hit</p> <p>Foldseek (AFDB-SwissProt) No hit</p> <p>Foldseek (PDB100)</p> 																																										
<p>HHPred hits vs PFAM-A_v36</p> 																																										
<p>InterPro/PFAM hits (UniProt) IPR026001 (Abi-like_C); PF14355 (Abi_C)</p>																																										
<p>Miscellaneous information</p> <p>The PFAM domain PF14335 (Abi_C) corresponds to the C-terminus of the <i>Lactococcus lactis</i> abortive infection protein Abi-859, a system later renamed AbiJ (Forde et al., 1999). While the C-terminus of AbiJ is conserved and belongs to the superfamily of HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domains, its N-terminus can vary (Anantharaman et al., 2013). The N-terminus of Taranis is homologous to AbiJ_NTD5 (PF18865). Of note, several systems containing an Abi_C or an AbiJ_NTD domain were recently identified in phage-inducible chromosomal islands (PICIs) of <i>Klebsiella</i> (Fillol-Salom et al., 2022).</p>																																										
<p>Selected references</p> <p>Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. <i>Biol Direct.</i> 2013;8:15. Published 2013 Jun 15. doi:10.1186/1745-6150-8-15</p> <p>Fillol-Salom A, Rostøl JT, Ojiogu AD, et al. Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. <i>Cell.</i> 2022;185(17):3248-3262.e20. doi:10.1016/j.cell.2022.07.014</p> <p>Forde A, Fitzgerald GF. Bacteriophage defence systems in lactic acid bacteria. <i>Antonie Van Leeuwenhoek.</i> 1999;76(1-4):89-113.</p> <p>Garcia-Doval C, Schwede F, Berk C, et al. Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. <i>Nat Commun.</i> 2020;11(1):1596. Published 2020 Mar 27. doi:10.1038/s41467-020-15334-5 https://defensefinder.mdmlab.fr/wiki/defense-systems/abij (DefenseFinder wiki about AbiJ)</p>																																										

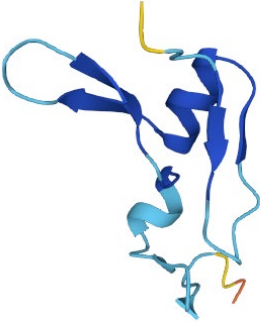
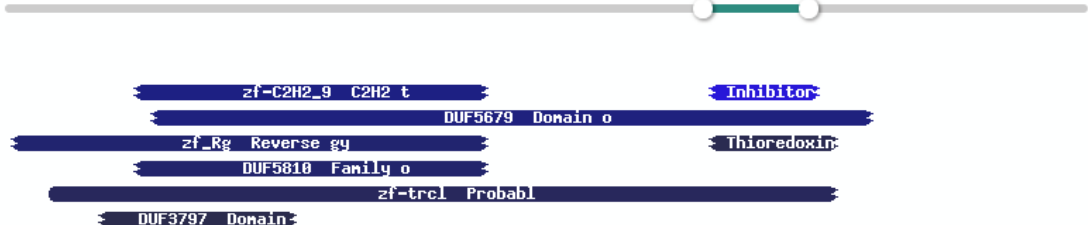
<h1>Damona</h1>	DUF6602 domain-containing protein VCA0399 / Uniprot ID Q9KMF9	Length 291 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (6 subunits, ipTM = 0.81)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A7V2QXG1-F1</td> <td>Uncharacterized protein</td> <td><i>Dehalococcoidia spp</i></td> <td>269</td> <td>91.69</td> </tr> <tr> <td>AF-A0A6G1X0H5-F1</td> <td>Uncharacterized protein</td> <td><i>Hadesarchaea spp</i></td> <td>292</td> <td>87.5</td> </tr> <tr> <td>AF-A0A0K9UJB3-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio cholerae 2740-80</i></td> <td>280</td> <td>86.31</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A4S4BUT7-F1</td> <td>Uncharacterized protein</td> <td><i>Bacillus sp. DSL-17</i></td> <td>250</td> <td>94.12</td> </tr> <tr> <td>AF-A0A7V1NT95-F1</td> <td>Uncharacterized protein</td> <td><i>Anaerolineae spp</i></td> <td>271</td> <td>92.25</td> </tr> <tr> <td>AF-A0A1Q6U7H6-F1</td> <td>Uncharacterized protein</td> <td><i>Azospirillum sp. 47_25</i></td> <td>286</td> <td>91.12</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A7V2QXG1-F1	Uncharacterized protein	<i>Dehalococcoidia spp</i>	269	91.69	AF-A0A6G1X0H5-F1	Uncharacterized protein	<i>Hadesarchaea spp</i>	292	87.5	AF-A0A0K9UJB3-F1	Uncharacterized protein	<i>Vibrio cholerae 2740-80</i>	280	86.31	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A4S4BUT7-F1	Uncharacterized protein	<i>Bacillus sp. DSL-17</i>	250	94.12	AF-A0A7V1NT95-F1	Uncharacterized protein	<i>Anaerolineae spp</i>	271	92.25	AF-A0A1Q6U7H6-F1	Uncharacterized protein	<i>Azospirillum sp. 47_25</i>	286	91.12
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A7V2QXG1-F1	Uncharacterized protein	<i>Dehalococcoidia spp</i>	269	91.69																																						
AF-A0A6G1X0H5-F1	Uncharacterized protein	<i>Hadesarchaea spp</i>	292	87.5																																						
AF-A0A0K9UJB3-F1	Uncharacterized protein	<i>Vibrio cholerae 2740-80</i>	280	86.31																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A4S4BUT7-F1	Uncharacterized protein	<i>Bacillus sp. DSL-17</i>	250	94.12																																						
AF-A0A7V1NT95-F1	Uncharacterized protein	<i>Anaerolineae spp</i>	271	92.25																																						
AF-A0A1Q6U7H6-F1	Uncharacterized protein	<i>Azospirillum sp. 47_25</i>	286	91.12																																						
<p>Foldseek (AFDB-Proteome)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-P44733-F1-model_v4</td> <td>DNA recombination protein RmuC homolog</td> <td><i>Haemophilus influenzae Rd_KW20</i></td> <td>0.30</td> <td>5.5</td> <td>0.380</td> <td>18-278</td> <td></td> </tr> <tr> <td>AF-Q9HWA3-F1-model_v4</td> <td>TadG</td> <td><i>Pseudomonas aeruginosa PAO1</i></td> <td>0.10</td> <td>2.7</td> <td>0.254</td> <td>16-281</td> <td></td> </tr> <tr> <td>AF-Q8IPN2-F1-model_v4</td> <td>5-hydroxytryptamine (Serotonin) receptor 2A, isoform B</td> <td><i>Drosophila melanogaster</i></td> <td>0.09</td> <td>1.5</td> <td>0.213</td> <td>42-291</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-P44733-F1-model_v4	DNA recombination protein RmuC homolog	<i>Haemophilus influenzae Rd_KW20</i>	0.30	5.5	0.380	18-278		AF-Q9HWA3-F1-model_v4	TadG	<i>Pseudomonas aeruginosa PAO1</i>	0.10	2.7	0.254	16-281		AF-Q8IPN2-F1-model_v4	5-hydroxytryptamine (Serotonin) receptor 2A, isoform B	<i>Drosophila melanogaster</i>	0.09	1.5	0.213	42-291									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-P44733-F1-model_v4	DNA recombination protein RmuC homolog	<i>Haemophilus influenzae Rd_KW20</i>	0.30	5.5	0.380	18-278																																				
AF-Q9HWA3-F1-model_v4	TadG	<i>Pseudomonas aeruginosa PAO1</i>	0.10	2.7	0.254	16-281																																				
AF-Q8IPN2-F1-model_v4	5-hydroxytryptamine (Serotonin) receptor 2A, isoform B	<i>Drosophila melanogaster</i>	0.09	1.5	0.213	42-291																																				
<p>Foldseek (AFDB-SwissProt)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-P0DTF8-F1-model_v4</td> <td>Endodeoxyribonuclease NucC</td> <td><i>Pseudomonas aeruginosa</i></td> <td>0.91</td> <td>15.1</td> <td>0.747</td> <td>31-291</td> <td></td> </tr> <tr> <td>AF-D7Y2H5-F1-model_v4</td> <td>Endodeoxyribonuclease NucC</td> <td><i>Escherichia coli MS 115-1</i></td> <td>0.91</td> <td>16.2</td> <td>0.747</td> <td>31-291</td> <td></td> </tr> <tr> <td>AF-P0DUD6-F1-model_v4</td> <td>CRISPR-associated endodeoxyribonuclease NucC</td> <td><i>Vibrio metoecus</i></td> <td>0.91</td> <td>15.7</td> <td>0.740</td> <td>31-291</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-P0DTF8-F1-model_v4	Endodeoxyribonuclease NucC	<i>Pseudomonas aeruginosa</i>	0.91	15.1	0.747	31-291		AF-D7Y2H5-F1-model_v4	Endodeoxyribonuclease NucC	<i>Escherichia coli MS 115-1</i>	0.91	16.2	0.747	31-291		AF-P0DUD6-F1-model_v4	CRISPR-associated endodeoxyribonuclease NucC	<i>Vibrio metoecus</i>	0.91	15.7	0.740	31-291									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-P0DTF8-F1-model_v4	Endodeoxyribonuclease NucC	<i>Pseudomonas aeruginosa</i>	0.91	15.1	0.747	31-291																																				
AF-D7Y2H5-F1-model_v4	Endodeoxyribonuclease NucC	<i>Escherichia coli MS 115-1</i>	0.91	16.2	0.747	31-291																																				
AF-P0DUD6-F1-model_v4	CRISPR-associated endodeoxyribonuclease NucC	<i>Vibrio metoecus</i>	0.91	15.7	0.740	31-291																																				
<p>Foldseek (PDB100)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>6q1h-assembly1_B</td> <td>Structure of <i>P. aeruginosa</i> ATCC27853 NucC, cAAA-bo...</td> <td><i>Pseudomonas aeruginosa</i></td> <td>0.91</td> <td>14.4</td> <td>0.749</td> <td>31-291</td> <td></td> </tr> <tr> <td>6p7c-assembly1_C</td> <td>Structure of <i>E. coli</i> MS115-1 NucC, cAAA-bound form</td> <td><i>Escherichia coli MS 115-1</i></td> <td>0.91</td> <td>15.9</td> <td>0.748</td> <td>31-291</td> <td></td> </tr> <tr> <td>6uxf-assembly1_A</td> <td>Structure of <i>V. metoecus</i> NucC, hexamer form</td> <td><i>Vibrio metoecus</i></td> <td>0.89</td> <td>16</td> <td>0.721</td> <td>31-291</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	6q1h-assembly1_B	Structure of <i>P. aeruginosa</i> ATCC27853 NucC, cAAA-bo...	<i>Pseudomonas aeruginosa</i>	0.91	14.4	0.749	31-291		6p7c-assembly1_C	Structure of <i>E. coli</i> MS115-1 NucC, cAAA-bound form	<i>Escherichia coli MS 115-1</i>	0.91	15.9	0.748	31-291		6uxf-assembly1_A	Structure of <i>V. metoecus</i> NucC, hexamer form	<i>Vibrio metoecus</i>	0.89	16	0.721	31-291									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
6q1h-assembly1_B	Structure of <i>P. aeruginosa</i> ATCC27853 NucC, cAAA-bo...	<i>Pseudomonas aeruginosa</i>	0.91	14.4	0.749	31-291																																				
6p7c-assembly1_C	Structure of <i>E. coli</i> MS115-1 NucC, cAAA-bound form	<i>Escherichia coli MS 115-1</i>	0.91	15.9	0.748	31-291																																				
6uxf-assembly1_A	Structure of <i>V. metoecus</i> NucC, hexamer form	<i>Vibrio metoecus</i>	0.89	16	0.721	31-291																																				
<p>HHPred hits vs PFAM-A_v36</p> 																																										
<p>InterPro/PFAM hits (UniProt) IPR046537 (DUF6602); PF20247 (DUF6602)</p>																																										
<p>Miscellaneous information RmuC (PF02646) is part of the PDDEXK endonuclease clan (CL0326), a key domain in the defense system Shield (Macdonald et al., 2023), where it has non-specific (endo)nuclease activity and can degrade phage, plasmid and chromosomal DNA. While Damona is not part of a system recognized by DefenseFinder, it has some similarity to HMMs of DefenseFinder: a PDDEXK Cas protein (Cas_PD_DExK) and a CBASS endonuclease (CBASS_Endonuc_small), two defense systems relying on nuclease activity. NucC (found in the Foldseek hits) is the effector DNase of a CBASS defence system (Lau et al., 2020).</p>																																										
<p>Selected references Lau RK, Ye Q, Birkholz EA, et al. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. <i>Mol Cell</i>. 2020;77(4):723-733.e6. doi:10.1016/j.molcel.2019.12.010 Macdonald E, Wright R, Connolly JPR, et al. The novel anti-phage system Shield co-opts an RmuC domain to mediate phage defense across <i>Pseudomonas</i> species. <i>PLoS Genet</i>. 2023;19(6):e1010784. Published 2023 Jun 5. doi:10.1371/journal.pgen.1010784 https://defensefinder.mdmlab.fr/wiki/defense-systems/cbass (DefenseFinder wiki about CBASS)</p>																																										

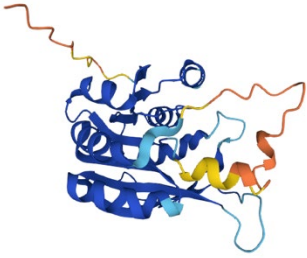
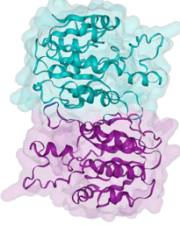
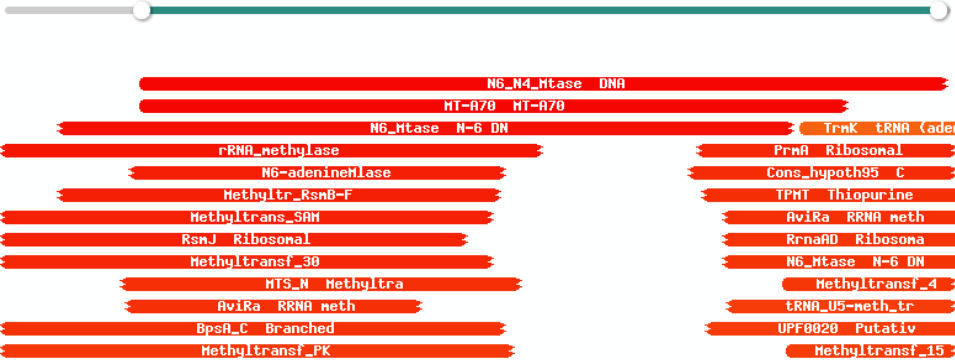
<h1>Lugos</h1>	Uncharacterized protein VCA0409 / Uniprot ID Q9KMF1	Length 152 aa																																																																																																
<p>AlphaFold structure</p>  <p>Cofold (6 subunits, ipTM = 0.82)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A690UNM6-F1</td> <td>Uncharacterized protein</td> <td><i>Campylobacter jejuni</i></td> <td>133</td> <td>90.31</td> </tr> <tr> <td>AF-A0A825JNV4-F1</td> <td>Uncharacterized protein</td> <td><i>Campylobacter coli</i></td> <td>133</td> <td>90.12</td> </tr> <tr> <td>AF-A0A2A5M3Z1-F1</td> <td>Uncharacterized protein</td> <td><i>Campylobacter sp. BCW_8712</i></td> <td>133</td> <td>89.94</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A661D5D2-F1</td> <td>Uncharacterized protein</td> <td>Unknown (γ-proteobacteria)</td> <td>124</td> <td>93.38</td> </tr> <tr> <td>AF-A0A2P8F058-F1</td> <td>Uncharacterized protein</td> <td><i>Marinobacterium halophilum</i></td> <td>126</td> <td>92.06</td> </tr> <tr> <td>AF-A0A7M1LGZ6-F1</td> <td>Uncharacterized protein</td> <td><i>Campylobacter corcagiensis</i></td> <td>126</td> <td>91.94</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A690UNM6-F1	Uncharacterized protein	<i>Campylobacter jejuni</i>	133	90.31	AF-A0A825JNV4-F1	Uncharacterized protein	<i>Campylobacter coli</i>	133	90.12	AF-A0A2A5M3Z1-F1	Uncharacterized protein	<i>Campylobacter sp. BCW_8712</i>	133	89.94	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A661D5D2-F1	Uncharacterized protein	Unknown (γ -proteobacteria)	124	93.38	AF-A0A2P8F058-F1	Uncharacterized protein	<i>Marinobacterium halophilum</i>	126	92.06	AF-A0A7M1LGZ6-F1	Uncharacterized protein	<i>Campylobacter corcagiensis</i>	126	91.94																																																								
	AFDB acc.	Description	Species	Length	pLDDT																																																																																													
AF-A0A690UNM6-F1	Uncharacterized protein	<i>Campylobacter jejuni</i>	133	90.31																																																																																														
AF-A0A825JNV4-F1	Uncharacterized protein	<i>Campylobacter coli</i>	133	90.12																																																																																														
AF-A0A2A5M3Z1-F1	Uncharacterized protein	<i>Campylobacter sp. BCW_8712</i>	133	89.94																																																																																														
AFDB acc.	Description	Species	Length	pLDDT																																																																																														
AF-A0A661D5D2-F1	Uncharacterized protein	Unknown (γ -proteobacteria)	124	93.38																																																																																														
AF-A0A2P8F058-F1	Uncharacterized protein	<i>Marinobacterium halophilum</i>	126	92.06																																																																																														
AF-A0A7M1LGZ6-F1	Uncharacterized protein	<i>Campylobacter corcagiensis</i>	126	91.94																																																																																														
<p>Foldseek (AFDB-Proteome)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q9I014-F1-model_v4</td> <td>Uncharacterized protein</td> <td><i>Pseudomonas aeruginosa</i> PAO1</td> <td>0.93</td> <td>11.8</td> <td>0.707</td> <td>4-152</td> <td></td> </tr> <tr> <td>AF-Q57874-F1-model_v4</td> <td>Uncharacterized protein MJ0432</td> <td><i>Methanocaldococcus jannaschii</i> DS...</td> <td>0.20</td> <td>12.8</td> <td>0.497</td> <td>41-112</td> <td></td> </tr> <tr> <td>AF-IBX7F9-F1-model_v4</td> <td>Transcriptional regulator RV3488</td> <td><i>Mycobacterium tuberculosis</i> H37Rv</td> <td>0.23</td> <td>8.4</td> <td>0.483</td> <td>5-135</td> <td></td> </tr> </tbody> </table> <p>Foldseek (AFDB-SwissProt)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q57874-F1-model_v4</td> <td>Uncharacterized protein MJ0432</td> <td><i>Methanocaldococcus jannaschii</i> DS...</td> <td>0.20</td> <td>12.8</td> <td>0.497</td> <td>41-112</td> <td></td> </tr> <tr> <td>AF-IBX7F9-F1-model_v4</td> <td>Transcriptional regulator RV3488</td> <td><i>Mycobacterium tuberculosis</i> H37Rv</td> <td>0.23</td> <td>8.4</td> <td>0.483</td> <td>5-135</td> <td></td> </tr> <tr> <td>AF-P0C1U8-F1-model_v4</td> <td>Transcriptional regulator SarA</td> <td><i>Staphylococcus aureus</i></td> <td>0.21</td> <td>7.6</td> <td>0.437</td> <td>6-111</td> <td></td> </tr> </tbody> </table> <p>Foldseek (PDB100)</p> <table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>6apw-assembly1_B</td> <td>Structural basis of cohesin ring opening</td> <td><i>Saccharomyces cerevisiae</i> S288C</td> <td>0.21</td> <td>7</td> <td>0.537</td> <td>6-143</td> <td></td> </tr> <tr> <td>5h67-assembly1_C</td> <td>Crystal structure of the Bacillus subtilis SMC head dom...</td> <td><i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 1...</td> <td>0.14</td> <td>6.8</td> <td>0.512</td> <td>9-95</td> <td></td> </tr> <tr> <td>7q34-assembly2_C</td> <td>Crystal structure of the multidrug binding transcriptional ...</td> <td><i>Lactococcus lactis</i> subsp. <i>lactis</i></td> <td>0.25</td> <td>11.3</td> <td>0.488</td> <td>6-112</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	AF-Q9I014-F1-model_v4	Uncharacterized protein	<i>Pseudomonas aeruginosa</i> PAO1	0.93	11.8	0.707	4-152		AF-Q57874-F1-model_v4	Uncharacterized protein MJ0432	<i>Methanocaldococcus jannaschii</i> DS...	0.20	12.8	0.497	41-112		AF-IBX7F9-F1-model_v4	Transcriptional regulator RV3488	<i>Mycobacterium tuberculosis</i> H37Rv	0.23	8.4	0.483	5-135		Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	AF-Q57874-F1-model_v4	Uncharacterized protein MJ0432	<i>Methanocaldococcus jannaschii</i> DS...	0.20	12.8	0.497	41-112		AF-IBX7F9-F1-model_v4	Transcriptional regulator RV3488	<i>Mycobacterium tuberculosis</i> H37Rv	0.23	8.4	0.483	5-135		AF-P0C1U8-F1-model_v4	Transcriptional regulator SarA	<i>Staphylococcus aureus</i>	0.21	7.6	0.437	6-111		Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	6apw-assembly1_B	Structural basis of cohesin ring opening	<i>Saccharomyces cerevisiae</i> S288C	0.21	7	0.537	6-143		5h67-assembly1_C	Crystal structure of the Bacillus subtilis SMC head dom...	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 1...	0.14	6.8	0.512	9-95		7q34-assembly2_C	Crystal structure of the multidrug binding transcriptional ...	<i>Lactococcus lactis</i> subsp. <i>lactis</i>	0.25	11.3	0.488	6-112	
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																																																																											
AF-Q9I014-F1-model_v4	Uncharacterized protein	<i>Pseudomonas aeruginosa</i> PAO1	0.93	11.8	0.707	4-152																																																																																												
AF-Q57874-F1-model_v4	Uncharacterized protein MJ0432	<i>Methanocaldococcus jannaschii</i> DS...	0.20	12.8	0.497	41-112																																																																																												
AF-IBX7F9-F1-model_v4	Transcriptional regulator RV3488	<i>Mycobacterium tuberculosis</i> H37Rv	0.23	8.4	0.483	5-135																																																																																												
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																																																																											
AF-Q57874-F1-model_v4	Uncharacterized protein MJ0432	<i>Methanocaldococcus jannaschii</i> DS...	0.20	12.8	0.497	41-112																																																																																												
AF-IBX7F9-F1-model_v4	Transcriptional regulator RV3488	<i>Mycobacterium tuberculosis</i> H37Rv	0.23	8.4	0.483	5-135																																																																																												
AF-P0C1U8-F1-model_v4	Transcriptional regulator SarA	<i>Staphylococcus aureus</i>	0.21	7.6	0.437	6-111																																																																																												
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																																																																											
6apw-assembly1_B	Structural basis of cohesin ring opening	<i>Saccharomyces cerevisiae</i> S288C	0.21	7	0.537	6-143																																																																																												
5h67-assembly1_C	Crystal structure of the Bacillus subtilis SMC head dom...	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 1...	0.14	6.8	0.512	9-95																																																																																												
7q34-assembly2_C	Crystal structure of the multidrug binding transcriptional ...	<i>Lactococcus lactis</i> subsp. <i>lactis</i>	0.25	11.3	0.488	6-112																																																																																												
<p>HHPred hits vs PFAM-A_v36</p>  <p>YjcQ YjcQ_prot</p> <p>DUF2513 Hypothe</p> <p>DUF3116 Protein</p> <p>DUF6429 Domain o</p> <p>DUF3860-like DUF</p> <p>DUF4364 Domain</p> <p>RTP Replication</p> <p>DUF2250 Unchara</p> <p>HC4 ABC-three co</p> <p>HC2 ABC-three co</p> <p>Mrr_N Mrr N-ter</p> <p>AphA_like Putat</p> <p>PadR Transcript</p> <p>YjxX_toxin Puta</p> <p>HxIR HxIR-like</p> <p>DUF5669 Family o</p> <p>DUF9457 Family o</p> <p>DUF2582 Winged</p> <p>MUS81-like_MH Cr</p> <p>SgrR_N Sugar tr</p> <p>DUF1495 Winged</p> <p>FokI_dom_2 FokI</p> <p>Replic_Relax Re</p>																																																																																																		
<p>InterPro/PFAM hits (UniProt): None</p>																																																																																																		
<p>Miscellaneous information</p> <p>The PFAM domain of YjcQ (PF09639) belongs to the helix-turn-helix clan of DNA-binding motifs. YjcQ is a protein of unknown function found in <i>B. subtilis</i> (Serizawa et al., 2004). The Mrr_N domain can be found fused to the Mrr family REase.</p>																																																																																																		
<p>Selected references</p> <p>Serizawa M, Yamamoto H, Yamaguchi H, et al. Systematic analysis of SigD-regulated genes in <i>Bacillus subtilis</i> by DNA microarray and Northern blotting analyses. <i>Gene</i>. 2004;329:125-136. doi:10.1016/j.gene.2003.12.024</p>																																																																																																		

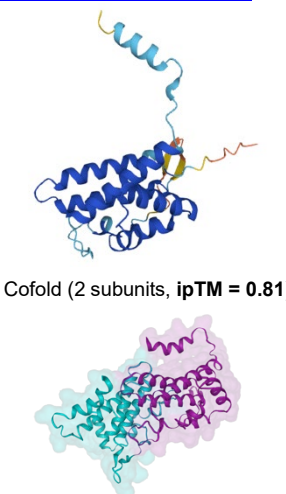
<h1>Cernunnos</h1>	4-fold beta flower domain-containing protein VCA0410 / Uniprot ID Q9KMF0	Length 120 aa																																								
<p>AlphaFold structure</p>  <p>Cofold: None</p>	AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”) <table border="1" data-bbox="507 241 1460 443"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A329E589-F1</td> <td>YD repeat-containing protein</td> <td><i>Vibrio diazotrophicus</i></td> <td>120</td> <td>92.38</td> </tr> <tr> <td>AF-A0A4Q0UC55-F1</td> <td>Collagen-like protein</td> <td><i>Arcobacter skirrowii</i> <i>CCUG 10374</i></td> <td>118</td> <td>89.56</td> </tr> <tr> <td>AF-A0A837HGE0-F1</td> <td>Uncharacterized protein</td> <td><i>Microgenomates</i> group bacterium</td> <td>122</td> <td>96.81</td> </tr> </tbody> </table> AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”) <table border="1" data-bbox="507 492 1460 721"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0G0A503-F1</td> <td>Uncharacterized protein</td> <td><i>Candidatus</i> <i>Woesebacteria bacterium</i></td> <td>122</td> <td>96.88</td> </tr> <tr> <td>AF-A0A7S8W6W9-F1</td> <td>Uncharacterized protein</td> <td><i>Acinetobacter sp.</i> <i>YH12138</i></td> <td>120</td> <td>96.44</td> </tr> <tr> <td>AF-A0A356J188-F1</td> <td>Uncharacterized protein</td> <td><i>Stenotrophomonas</i> <i>spp</i></td> <td>120</td> <td>96.38</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A329E589-F1	YD repeat-containing protein	<i>Vibrio diazotrophicus</i>	120	92.38	AF-A0A4Q0UC55-F1	Collagen-like protein	<i>Arcobacter skirrowii</i> <i>CCUG 10374</i>	118	89.56	AF-A0A837HGE0-F1	Uncharacterized protein	<i>Microgenomates</i> group bacterium	122	96.81	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A0G0A503-F1	Uncharacterized protein	<i>Candidatus</i> <i>Woesebacteria bacterium</i>	122	96.88	AF-A0A7S8W6W9-F1	Uncharacterized protein	<i>Acinetobacter sp.</i> <i>YH12138</i>	120	96.44	AF-A0A356J188-F1	Uncharacterized protein	<i>Stenotrophomonas</i> <i>spp</i>	120	96.38
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A329E589-F1	YD repeat-containing protein	<i>Vibrio diazotrophicus</i>	120	92.38																																						
AF-A0A4Q0UC55-F1	Collagen-like protein	<i>Arcobacter skirrowii</i> <i>CCUG 10374</i>	118	89.56																																						
AF-A0A837HGE0-F1	Uncharacterized protein	<i>Microgenomates</i> group bacterium	122	96.81																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A0G0A503-F1	Uncharacterized protein	<i>Candidatus</i> <i>Woesebacteria bacterium</i>	122	96.88																																						
AF-A0A7S8W6W9-F1	Uncharacterized protein	<i>Acinetobacter sp.</i> <i>YH12138</i>	120	96.44																																						
AF-A0A356J188-F1	Uncharacterized protein	<i>Stenotrophomonas</i> <i>spp</i>	120	96.38																																						
<p>Foldseek (AFDB-Proteome) No hit</p> <p>Foldseek (AFDB-SwissProt) No hit</p> <p>Foldseek (PDB100) No hit</p>																																										
<p>HHPred hits vs PFAM-A_v36</p>  <p> Bflower_4-fold b Bflower_2_5-fold Bflower_2_5-fold Bflower_3_6-fold Scramblase Scra DUF3659 Protein LOR LURP-one-re DUF5944 Family o MD40_3 MD domain Scramblase Scra PRC PRC-barrel Cytochrom_D1 Cy FAM194 FAM194 p DUF5046 Domain o DUF2171 Uncha NPFI_C Nucleosi SH3_7 SH3 domai DUF5052 Domain o UPF0257 Unchara eIF2A Eukaryoti </p>																																										
<p>InterPro/PFAM hits (UniProt) IPR048911 (Bflower); IPR006530 (YD); PF21784 (Bflower)</p>																																										
<p>Miscellaneous information The 4-fold beta flower domain (PF21784) is a recently described structural outlier when clustering AlphaFold's predictions (no / rare homologs in the PDB, cf Durairaj et al., 2024 for further details).</p>																																										
<p>Selected references Durairaj J, Waterhouse AM, Mets T, et al. Uncovering new families and folds in the natural protein universe. <i>Nature</i>. 2023;622(7983):646-653. doi:10.1038/s41586-023-06622-3</p>																																										

<h1>Brigantia</h1>	Uncharacterized protein VCA0419 / Uniprot ID Q9KME2	Length 261 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.77)</p>	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="507 241 1460 443"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A1B7XMS2-F1</td> <td>Uncharacterized protein</td> <td><i>Halodesulfovibrio spirochaetisodalis</i></td> <td>220</td> <td>84.88</td> </tr> <tr> <td>AF-A0A1N6FR19-F1</td> <td>Uncharacterized protein</td> <td><i>Halodesulfovibrio marinisediminis</i></td> <td>222</td> <td>84.19</td> </tr> <tr> <td>AF-U4E4A0-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio nigripulchritudo FTn2</i></td> <td>235</td> <td>84.06</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="507 495 1460 694"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A7C8HYU4-F1</td> <td>Uncharacterized protein</td> <td><i>Halodesulfovibrio sp. MK-HDV</i></td> <td>220</td> <td>86.5</td> </tr> <tr> <td>AF-A0A7T9CMS6-F1</td> <td>Uncharacterized protein</td> <td><i>Holophagales spp</i></td> <td>213</td> <td>85</td> </tr> <tr> <td>AF-A0A0M0HIC9-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio nereis</i></td> <td>235</td> <td>84.5</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A1B7XMS2-F1	Uncharacterized protein	<i>Halodesulfovibrio spirochaetisodalis</i>	220	84.88	AF-A0A1N6FR19-F1	Uncharacterized protein	<i>Halodesulfovibrio marinisediminis</i>	222	84.19	AF-U4E4A0-F1	Uncharacterized protein	<i>Vibrio nigripulchritudo FTn2</i>	235	84.06	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A7C8HYU4-F1	Uncharacterized protein	<i>Halodesulfovibrio sp. MK-HDV</i>	220	86.5	AF-A0A7T9CMS6-F1	Uncharacterized protein	<i>Holophagales spp</i>	213	85	AF-A0A0M0HIC9-F1	Uncharacterized protein	<i>Vibrio nereis</i>	235	84.5
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A1B7XMS2-F1	Uncharacterized protein	<i>Halodesulfovibrio spirochaetisodalis</i>	220	84.88																																						
AF-A0A1N6FR19-F1	Uncharacterized protein	<i>Halodesulfovibrio marinisediminis</i>	222	84.19																																						
AF-U4E4A0-F1	Uncharacterized protein	<i>Vibrio nigripulchritudo FTn2</i>	235	84.06																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A7C8HYU4-F1	Uncharacterized protein	<i>Halodesulfovibrio sp. MK-HDV</i>	220	86.5																																						
AF-A0A7T9CMS6-F1	Uncharacterized protein	<i>Holophagales spp</i>	213	85																																						
AF-A0A0M0HIC9-F1	Uncharacterized protein	<i>Vibrio nereis</i>	235	84.5																																						
<p>Foldseek (AFDB-Proteome)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0R0IP82-F1-model_v4</td> <td>FYVE-type domain-containing protein</td> <td><i>Glycine max</i></td> <td>0.10</td> <td>1.5</td> <td>0.196</td> <td>20-254</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-A0A0R0IP82-F1-model_v4	FYVE-type domain-containing protein	<i>Glycine max</i>	0.10	1.5	0.196	20-254																									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-A0A0R0IP82-F1-model_v4	FYVE-type domain-containing protein	<i>Glycine max</i>	0.10	1.5	0.196	20-254																																				
<p>Foldseek (AFDB-SwissProt)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-O6NGP1-F1-model_v4</td> <td>Polyribonucleotide nucleotidyltransferase</td> <td><i>Corynebacterium diphtheriae</i> NCTC...</td> <td>0.10</td> <td>2.7</td> <td>0.230</td> <td>3-251</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-O6NGP1-F1-model_v4	Polyribonucleotide nucleotidyltransferase	<i>Corynebacterium diphtheriae</i> NCTC...	0.10	2.7	0.230	3-251																									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-O6NGP1-F1-model_v4	Polyribonucleotide nucleotidyltransferase	<i>Corynebacterium diphtheriae</i> NCTC...	0.10	2.7	0.230	3-251																																				
<p>Foldseek (PDB100)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>5ygp-assembly1_A</td> <td>Crystal structure of a domain-swapped dimer of the sec...</td> <td><i>Saccharomyces cerevisiae</i> S288C</td> <td>0.10</td> <td>6.7</td> <td>0.383</td> <td>80-257</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	5ygp-assembly1_A	Crystal structure of a domain-swapped dimer of the sec...	<i>Saccharomyces cerevisiae</i> S288C	0.10	6.7	0.383	80-257																									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
5ygp-assembly1_A	Crystal structure of a domain-swapped dimer of the sec...	<i>Saccharomyces cerevisiae</i> S288C	0.10	6.7	0.383	80-257																																				
<p>HHPred hits vs PFAM-A_v36</p>  <p>YtpI YtpI-like</p>																																										
<p>InterPro/PFAM hits (UniProt): None</p>																																										
<p>Miscellaneous information: 1 transmembrane domain</p> 																																										
<p>Selected references</p>																																										

<h1>Nemetona</h1>	HNH endonuclease VCA0441 / Uniprot ID Q9KMC3	Length 309 aa																																								
<p>AlphaFold structure</p>  <p>Cofold: None</p>	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 241 1461 443"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A5C8V7Q5-F1</td> <td>HNH endonuclease</td> <td><i>Flagellimonas hymeniacidonis</i></td> <td>306</td> <td>91.19</td> </tr> <tr> <td>AF-A0A552X4I3-F1</td> <td>Uncharacterized protein</td> <td><i>Aliidiomarina sp. IM1326</i></td> <td>297</td> <td>91.06</td> </tr> <tr> <td>AF-A0A0U2WZK4-F1</td> <td>Uncharacterized protein</td> <td><i>Pseudoalteromonas translucida</i></td> <td>297</td> <td>90.94</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 495 1461 696"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A090VK89-F1</td> <td>Uncharacterized protein</td> <td><i>Algibacter lectus</i></td> <td>303</td> <td>92.44</td> </tr> <tr> <td>AF-A0A1F9B2K4-F1</td> <td>Uncharacterized protein</td> <td>Unknown (δ-proteobacteria)</td> <td>258</td> <td>89.12</td> </tr> <tr> <td>AF-A0A228GKZ8-F1</td> <td>Uncharacterized protein</td> <td><i>Burkholderia sp. AU15512</i></td> <td>284</td> <td>87.38</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A5C8V7Q5-F1	HNH endonuclease	<i>Flagellimonas hymeniacidonis</i>	306	91.19	AF-A0A552X4I3-F1	Uncharacterized protein	<i>Aliidiomarina sp. IM1326</i>	297	91.06	AF-A0A0U2WZK4-F1	Uncharacterized protein	<i>Pseudoalteromonas translucida</i>	297	90.94	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A090VK89-F1	Uncharacterized protein	<i>Algibacter lectus</i>	303	92.44	AF-A0A1F9B2K4-F1	Uncharacterized protein	Unknown (δ -proteobacteria)	258	89.12	AF-A0A228GKZ8-F1	Uncharacterized protein	<i>Burkholderia sp. AU15512</i>	284	87.38
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A5C8V7Q5-F1	HNH endonuclease	<i>Flagellimonas hymeniacidonis</i>	306	91.19																																						
AF-A0A552X4I3-F1	Uncharacterized protein	<i>Aliidiomarina sp. IM1326</i>	297	91.06																																						
AF-A0A0U2WZK4-F1	Uncharacterized protein	<i>Pseudoalteromonas translucida</i>	297	90.94																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A090VK89-F1	Uncharacterized protein	<i>Algibacter lectus</i>	303	92.44																																						
AF-A0A1F9B2K4-F1	Uncharacterized protein	Unknown (δ -proteobacteria)	258	89.12																																						
AF-A0A228GKZ8-F1	Uncharacterized protein	<i>Burkholderia sp. AU15512</i>	284	87.38																																						
<p>Foldseek (AFDB-Proteome)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q3UCV8-F1-model_v4</td> <td>Ubiquitin thioesterase otulin</td> <td>Mus musculus</td> <td>0.10</td> <td>2.6</td> <td>0.299</td> <td></td> <td></td> </tr> <tr> <td>AF-A0A3P7FA37-F1-model_v4</td> <td>Uncharacterized protein</td> <td>Wuchereria bancrofti</td> <td>0.11</td> <td>1.6</td> <td>0.229</td> <td></td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	AF-Q3UCV8-F1-model_v4	Ubiquitin thioesterase otulin	Mus musculus	0.10	2.6	0.299			AF-A0A3P7FA37-F1-model_v4	Uncharacterized protein	Wuchereria bancrofti	0.11	1.6	0.229																		
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																			
AF-Q3UCV8-F1-model_v4	Ubiquitin thioesterase otulin	Mus musculus	0.10	2.6	0.299																																					
AF-A0A3P7FA37-F1-model_v4	Uncharacterized protein	Wuchereria bancrofti	0.11	1.6	0.229																																					
<p>Foldseek (AFDB-SwissProt)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q3UCV8-F1-model_v4</td> <td>Ubiquitin thioesterase otulin</td> <td>Mus musculus</td> <td>0.10</td> <td>2.6</td> <td>0.299</td> <td></td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	AF-Q3UCV8-F1-model_v4	Ubiquitin thioesterase otulin	Mus musculus	0.10	2.6	0.299																										
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																			
AF-Q3UCV8-F1-model_v4	Ubiquitin thioesterase otulin	Mus musculus	0.10	2.6	0.299																																					
<p>Foldseek (PDB100)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>4a0a-assembly1_B</td> <td>Structure of hsDDB1-drDDB2 bound to a 16 bp CPD-du...</td> <td>Danio rerio</td> <td>0.09</td> <td>1.2</td> <td>0.289</td> <td></td> <td></td> </tr> <tr> <td>8tb-assembly1_B</td> <td>Cryo-EM structure of the PP2A.B55-ARPP19 complex</td> <td>Homo sapiens</td> <td>0.09</td> <td>1.4</td> <td>0.275</td> <td></td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment	4a0a-assembly1_B	Structure of hsDDB1-drDDB2 bound to a 16 bp CPD-du...	Danio rerio	0.09	1.2	0.289			8tb-assembly1_B	Cryo-EM structure of the PP2A.B55-ARPP19 complex	Homo sapiens	0.09	1.4	0.275																		
Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment																																			
4a0a-assembly1_B	Structure of hsDDB1-drDDB2 bound to a 16 bp CPD-du...	Danio rerio	0.09	1.2	0.289																																					
8tb-assembly1_B	Cryo-EM structure of the PP2A.B55-ARPP19 complex	Homo sapiens	0.09	1.4	0.275																																					
<p>HHPred hits vs PFAM-A_v36</p>																																										
																																										
<p>InterPro/PFAM hits (UniProt): None</p>																																										
<p>Miscellaneous information</p>																																										
<p>Nemetona is annotated as an HNH endonuclease on Uniprot (Uniprot entry: Q9KMC3), following the deep learning method ProtNLM (https://www.uniprot.org/help/ProtNLM). Like Sirona, Nemetona has also weak similarity to the PFAM domain of a type II restriction endonuclease (PF09665, Re_Alw26IDE) and NinG.</p>																																										
<p>Selected references</p>																																										
<p>Tarkowski TA, Mooney D, Thomason LC, Stahl FW. Gene products encoded in the ninR region of phage lambda participate in Red-mediated recombination. <i>Genes Cells</i>. 2002;7(4):351-363. doi:10.1046/j.1365-2443.2002.00531.x</p> <p>Wu CC, Lin JLJ, Yuan HS. Structures, Mechanisms, and Functions of His-Me Finger Nucleases. <i>Trends Biochem Sci</i>. 2020;45(11):935-946. doi:10.1016/j.tibs.2020.07.002</p>																																										

<h1>Toutatis</h1> <p>(TutA)</p>	Haemagglutinin VCA0446 / Uniprot ID H9L4P8	Length 63 aa																																								
<p>AlphaFold structure</p>  <p>Cofold: None</p>	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="512 264 1460 421"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A3L8PSA4-F1</td> <td>Hemagglutinin</td> <td><i>Parashevanella curva</i></td> <td>60</td> <td>86.56</td> </tr> <tr> <td>AF-A0A833W14-F1</td> <td>Hemagglutinin</td> <td><i>Plesiomonas shigelloides</i></td> <td>60</td> <td>86.5</td> </tr> <tr> <td>AF-A0A2N0WM80-F1</td> <td>Hemagglutinin</td> <td><i>Alteromonadales spp</i></td> <td>60</td> <td>85.38</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="512 470 1460 645"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0X1KZP2-F1</td> <td>Hemagglutinin</td> <td><i>Vibrio cholerae MO10</i></td> <td>63</td> <td>86.75</td> </tr> <tr> <td>AF-A0A329TQZ5-F1</td> <td>Hemagglutinin</td> <td><i>Faecalibacterium prausnitzii</i></td> <td>61</td> <td>83.5</td> </tr> <tr> <td>AF-A0A730GQI1-F1</td> <td>Hemagglutinin</td> <td><i>Salmonella newport</i></td> <td>60</td> <td>82.06</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A3L8PSA4-F1	Hemagglutinin	<i>Parashevanella curva</i>	60	86.56	AF-A0A833W14-F1	Hemagglutinin	<i>Plesiomonas shigelloides</i>	60	86.5	AF-A0A2N0WM80-F1	Hemagglutinin	<i>Alteromonadales spp</i>	60	85.38	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A0X1KZP2-F1	Hemagglutinin	<i>Vibrio cholerae MO10</i>	63	86.75	AF-A0A329TQZ5-F1	Hemagglutinin	<i>Faecalibacterium prausnitzii</i>	61	83.5	AF-A0A730GQI1-F1	Hemagglutinin	<i>Salmonella newport</i>	60	82.06
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A3L8PSA4-F1	Hemagglutinin	<i>Parashevanella curva</i>	60	86.56																																						
AF-A0A833W14-F1	Hemagglutinin	<i>Plesiomonas shigelloides</i>	60	86.5																																						
AF-A0A2N0WM80-F1	Hemagglutinin	<i>Alteromonadales spp</i>	60	85.38																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A0X1KZP2-F1	Hemagglutinin	<i>Vibrio cholerae MO10</i>	63	86.75																																						
AF-A0A329TQZ5-F1	Hemagglutinin	<i>Faecalibacterium prausnitzii</i>	61	83.5																																						
AF-A0A730GQI1-F1	Hemagglutinin	<i>Salmonella newport</i>	60	82.06																																						
<p>Foldseek (AFDB-Proteome)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0H4J5E3-F1-model_v4</td> <td>Uncharacterized protein</td> <td>Neisseria gonorrhoeae FA 1090</td> <td>0.47</td> <td>16.3</td> <td>0.438</td> <td>4-52</td> <td></td> </tr> <tr> <td>AF-K7L115-F1-model_v4</td> <td>TRASH domain-containing protein</td> <td>Glycine max</td> <td>0.16</td> <td>6.1</td> <td>0.310</td> <td>3-62</td> <td></td> </tr> <tr> <td>AF-Q6P6G7-F1-model_v4</td> <td>Probable ribosome biogenesis protein RLP24</td> <td>Rattus norvegicus</td> <td>0.16</td> <td>5</td> <td>0.286</td> <td>4-69</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-A0A0H4J5E3-F1-model_v4	Uncharacterized protein	Neisseria gonorrhoeae FA 1090	0.47	16.3	0.438	4-52		AF-K7L115-F1-model_v4	TRASH domain-containing protein	Glycine max	0.16	6.1	0.310	3-62		AF-Q6P6G7-F1-model_v4	Probable ribosome biogenesis protein RLP24	Rattus norvegicus	0.16	5	0.286	4-69									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-A0A0H4J5E3-F1-model_v4	Uncharacterized protein	Neisseria gonorrhoeae FA 1090	0.47	16.3	0.438	4-52																																				
AF-K7L115-F1-model_v4	TRASH domain-containing protein	Glycine max	0.16	6.1	0.310	3-62																																				
AF-Q6P6G7-F1-model_v4	Probable ribosome biogenesis protein RLP24	Rattus norvegicus	0.16	5	0.286	4-69																																				
<p>Foldseek (AFDB-SwissProt)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-P61124-F1-model_v4</td> <td>50S ribosomal protein L24e</td> <td>Pyrococcus horikoshii OT3</td> <td>0.14</td> <td>10</td> <td>0.356</td> <td>4-69</td> <td></td> </tr> <tr> <td>AF-C5A1V7-F1-model_v4</td> <td>50S ribosomal protein L24e</td> <td>Thermopoccus gammatolerans E-13</td> <td>0.14</td> <td>8.4</td> <td>0.351</td> <td>4-69</td> <td></td> </tr> <tr> <td>AF-P38665-F1-model_v4</td> <td>60S ribosomal protein L24</td> <td>Kluyveromyces lactis NRRL Y-1140</td> <td>0.15</td> <td>0</td> <td>0.286</td> <td>3-69</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-P61124-F1-model_v4	50S ribosomal protein L24e	Pyrococcus horikoshii OT3	0.14	10	0.356	4-69		AF-C5A1V7-F1-model_v4	50S ribosomal protein L24e	Thermopoccus gammatolerans E-13	0.14	8.4	0.351	4-69		AF-P38665-F1-model_v4	60S ribosomal protein L24	Kluyveromyces lactis NRRL Y-1140	0.15	0	0.286	3-69									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-P61124-F1-model_v4	50S ribosomal protein L24e	Pyrococcus horikoshii OT3	0.14	10	0.356	4-69																																				
AF-C5A1V7-F1-model_v4	50S ribosomal protein L24e	Thermopoccus gammatolerans E-13	0.14	8.4	0.351	4-69																																				
AF-P38665-F1-model_v4	60S ribosomal protein L24	Kluyveromyces lactis NRRL Y-1140	0.15	0	0.286	3-69																																				
<p>Foldseek (PDB100)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>8q9i-assembly1_Z</td> <td>Structure of Candida albicans 80S ribosome in complex...</td> <td>Candida albicans</td> <td>0.16</td> <td>0</td> <td>0.377</td> <td>4-69</td> <td></td> </tr> <tr> <td>2x7n-assembly1_D</td> <td>Mechanism of eIF6s anti-association activity</td> <td>Saccharomyces cerevisiae</td> <td>0.15</td> <td>6.5</td> <td>0.377</td> <td>3-60</td> <td></td> </tr> <tr> <td>8eub-assembly1_AW</td> <td>Hypopseudouridylylated yeast 80S bound with Taura syn...</td> <td>Saccharomyces cerevisiae</td> <td>0.16</td> <td>0</td> <td>0.377</td> <td>4-69</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	8q9i-assembly1_Z	Structure of Candida albicans 80S ribosome in complex...	Candida albicans	0.16	0	0.377	4-69		2x7n-assembly1_D	Mechanism of eIF6s anti-association activity	Saccharomyces cerevisiae	0.15	6.5	0.377	3-60		8eub-assembly1_AW	Hypopseudouridylylated yeast 80S bound with Taura syn...	Saccharomyces cerevisiae	0.16	0	0.377	4-69									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
8q9i-assembly1_Z	Structure of Candida albicans 80S ribosome in complex...	Candida albicans	0.16	0	0.377	4-69																																				
2x7n-assembly1_D	Mechanism of eIF6s anti-association activity	Saccharomyces cerevisiae	0.15	6.5	0.377	3-60																																				
8eub-assembly1_AW	Hypopseudouridylylated yeast 80S bound with Taura syn...	Saccharomyces cerevisiae	0.16	0	0.377	4-69																																				
<p>HHPred hits vs PFAM-A_v36</p>																																										
 <p>zf-C2H2_9 C2H2 t Inhibitor</p> <p>DUF5679 Domain o</p> <p>zf_Rg Reverse gy Thioredoxin</p> <p>DUF5810 Family o</p> <p>zf-trc1 Probabl</p> <p>DUF3797 Domain</p>																																										
<p>InterPro/PFAM hits (UniProt): None</p>																																										
<p>Miscellaneous information</p>																																										
<p>TutA, while homolog to one domain of the restriction endonuclease Hpy99I involved in cleavage site recognition (Sokolowska et al., 2009) does not contain catalytic residues (data not shown).</p>																																										
<p>Selected references</p>																																										
<p>Sokolowska M, Czapinska H, Bochtler M. Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA. <i>Nucleic Acids Res.</i> 2009;37(11):3799-3810. doi:10.1093/nar/gkp228</p>																																										

<h1>Toutatis</h1> <p>(TutB)</p>	Methyltransferase VCA0447 / Uniprot ID H9L4S1	Length 230 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.88)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="507 264 1465 495"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A1F6LI02-F1</td> <td>Methyltransferase</td> <td><i>Candidatus Magasanikbacteria spp</i></td> <td>255</td> <td>95.75</td> </tr> <tr> <td>AF-A0A349PHI4-F1</td> <td>Methyltransferase</td> <td><i>Candidatus Moranbacteria spp</i></td> <td>255</td> <td>95.31</td> </tr> <tr> <td>AF-A0A2I0QLZ0-F1</td> <td>Modification methylase</td> <td><i>Candidatus Altiarchaeales spp</i></td> <td>259</td> <td>95.31</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="507 544 1465 723"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A7V1LSO5-F1</td> <td>Methyltransferase</td> <td><i>Ignavibacteria spp</i></td> <td>260</td> <td>95.12</td> </tr> <tr> <td>AF-A0A1F6JLU0-F1</td> <td>Methyltransferase</td> <td><i>Candidatus Levybacteria spp</i></td> <td>234</td> <td>95.12</td> </tr> <tr> <td>AF-A0A2I0QLA2-F1</td> <td>Modification methylase</td> <td><i>Candidatus Altiarchaeales spp</i></td> <td>259</td> <td>95.06</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A1F6LI02-F1	Methyltransferase	<i>Candidatus Magasanikbacteria spp</i>	255	95.75	AF-A0A349PHI4-F1	Methyltransferase	<i>Candidatus Moranbacteria spp</i>	255	95.31	AF-A0A2I0QLZ0-F1	Modification methylase	<i>Candidatus Altiarchaeales spp</i>	259	95.31	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A7V1LSO5-F1	Methyltransferase	<i>Ignavibacteria spp</i>	260	95.12	AF-A0A1F6JLU0-F1	Methyltransferase	<i>Candidatus Levybacteria spp</i>	234	95.12	AF-A0A2I0QLA2-F1	Modification methylase	<i>Candidatus Altiarchaeales spp</i>	259	95.06
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A1F6LI02-F1	Methyltransferase	<i>Candidatus Magasanikbacteria spp</i>	255	95.75																																						
AF-A0A349PHI4-F1	Methyltransferase	<i>Candidatus Moranbacteria spp</i>	255	95.31																																						
AF-A0A2I0QLZ0-F1	Modification methylase	<i>Candidatus Altiarchaeales spp</i>	259	95.31																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A7V1LSO5-F1	Methyltransferase	<i>Ignavibacteria spp</i>	260	95.12																																						
AF-A0A1F6JLU0-F1	Methyltransferase	<i>Candidatus Levybacteria spp</i>	234	95.12																																						
AF-A0A2I0QLA2-F1	Modification methylase	<i>Candidatus Altiarchaeales spp</i>	259	95.06																																						
<p>Foldseek (AFDB-Proteome)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-Q93GQ3-F1-model_v4</td> <td>Methyltransferase</td> <td><i>Salmonella enterica subsp. enterica...</i></td> <td>0.99</td> <td>27.8</td> <td>0.829</td> <td>14-230</td> <td></td> </tr> <tr> <td>AF-A0A0H3GZ55-F1-model_v4</td> <td>Methyltransferase</td> <td><i>Klebsiella pneumoniae subsp. pneu...</i></td> <td>0.99</td> <td>25.2</td> <td>0.819</td> <td>14-230</td> <td></td> </tr> <tr> <td>AF-O24891-F1-model_v4</td> <td>Methyltransferase</td> <td><i>Helicobacter pylori 26695</i></td> <td>0.99</td> <td>23.1</td> <td>0.808</td> <td>14-230</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-Q93GQ3-F1-model_v4	Methyltransferase	<i>Salmonella enterica subsp. enterica...</i>	0.99	27.8	0.829	14-230		AF-A0A0H3GZ55-F1-model_v4	Methyltransferase	<i>Klebsiella pneumoniae subsp. pneu...</i>	0.99	25.2	0.819	14-230		AF-O24891-F1-model_v4	Methyltransferase	<i>Helicobacter pylori 26695</i>	0.99	23.1	0.808	14-230									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-Q93GQ3-F1-model_v4	Methyltransferase	<i>Salmonella enterica subsp. enterica...</i>	0.99	27.8	0.829	14-230																																				
AF-A0A0H3GZ55-F1-model_v4	Methyltransferase	<i>Klebsiella pneumoniae subsp. pneu...</i>	0.99	25.2	0.819	14-230																																				
AF-O24891-F1-model_v4	Methyltransferase	<i>Helicobacter pylori 26695</i>	0.99	23.1	0.808	14-230																																				
<p>Foldseek (AFDB-SwissProt)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-P00473-F1-model_v4</td> <td>Modification methylase Hhall</td> <td><i>Haemophilus parahaemolyticus</i></td> <td>0.99</td> <td>22.2</td> <td>0.831</td> <td>5-230</td> <td></td> </tr> <tr> <td>AF-Q8S4X2-F1-model_v4</td> <td>Putative methylase YubD</td> <td><i>Escherichia coli K-12</i></td> <td>0.99</td> <td>28.7</td> <td>0.826</td> <td>14-230</td> <td></td> </tr> <tr> <td>AF-P50178-F1-model_v4</td> <td>Modification methylase LiaDCHIB</td> <td><i>Lactococcus cremoris</i></td> <td>0.99</td> <td>20</td> <td>0.800</td> <td>4-230</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-P00473-F1-model_v4	Modification methylase Hhall	<i>Haemophilus parahaemolyticus</i>	0.99	22.2	0.831	5-230		AF-Q8S4X2-F1-model_v4	Putative methylase YubD	<i>Escherichia coli K-12</i>	0.99	28.7	0.826	14-230		AF-P50178-F1-model_v4	Modification methylase LiaDCHIB	<i>Lactococcus cremoris</i>	0.99	20	0.800	4-230									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-P00473-F1-model_v4	Modification methylase Hhall	<i>Haemophilus parahaemolyticus</i>	0.99	22.2	0.831	5-230																																				
AF-Q8S4X2-F1-model_v4	Putative methylase YubD	<i>Escherichia coli K-12</i>	0.99	28.7	0.826	14-230																																				
AF-P50178-F1-model_v4	Modification methylase LiaDCHIB	<i>Lactococcus cremoris</i>	0.99	20	0.800	4-230																																				
<p>Foldseek (PDB100)</p>																																										
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>8s9m-assembly1_A-2</td> <td>DNA cytosine-N4 methyltransferase (residues 79-324) f...</td> <td><i>Adineta vaga</i></td> <td>0.99</td> <td>21.1</td> <td>0.805</td> <td>12-230</td> <td></td> </tr> <tr> <td>5hjl-assembly2_E</td> <td>crystal structure of M1.HpyAVI-SAM complex</td> <td><i>Helicobacter pylori 26695</i></td> <td>0.96</td> <td>23.4</td> <td>0.797</td> <td>14-230</td> <td></td> </tr> <tr> <td>5hjl-assembly4_G</td> <td>crystal structure of M1.HpyAVI-SAM complex</td> <td><i>Helicobacter pylori 26695</i></td> <td>0.96</td> <td>24.7</td> <td>0.795</td> <td>14-230</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	8s9m-assembly1_A-2	DNA cytosine-N4 methyltransferase (residues 79-324) f...	<i>Adineta vaga</i>	0.99	21.1	0.805	12-230		5hjl-assembly2_E	crystal structure of M1.HpyAVI-SAM complex	<i>Helicobacter pylori 26695</i>	0.96	23.4	0.797	14-230		5hjl-assembly4_G	crystal structure of M1.HpyAVI-SAM complex	<i>Helicobacter pylori 26695</i>	0.96	24.7	0.795	14-230									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
8s9m-assembly1_A-2	DNA cytosine-N4 methyltransferase (residues 79-324) f...	<i>Adineta vaga</i>	0.99	21.1	0.805	12-230																																				
5hjl-assembly2_E	crystal structure of M1.HpyAVI-SAM complex	<i>Helicobacter pylori 26695</i>	0.96	23.4	0.797	14-230																																				
5hjl-assembly4_G	crystal structure of M1.HpyAVI-SAM complex	<i>Helicobacter pylori 26695</i>	0.96	24.7	0.795	14-230																																				
<p>HHPred hits vs PFAM-A_v36</p>																																										
																																										
<p>InterPro/PFAM hits (UniProt) IPR002941(DNA_methylase_N4/N6); IPR001091(RM_Methyltransferase) ; IPR029063 (SAM-dependent MTases sf) ; PF01555 (N6_N4_Mtase)</p>																																										
<p>Miscellaneous information TutB is annotated as an N6_N4 methyltransferase and hits the DefenseFinder HMM corresponding to a type II methyltransferase.</p>																																										
<p>Selected references Mruk I, Kobayashi I. To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. <i>Nucleic Acids Res.</i> 2014;42(1):70-86. doi:10.1093/nar/gkt711 https://defensefinder.mdmlab.fr/wiki/defense-systems/rm (DefenseFinder wiki about RMs)</p>																																										

<h1>Esos</h1>	HEPN domain-containing protein VCA0450 / Uniprot ID H9L4T1	Length 176 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.81)</p>	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-Q08796-F1</td> <td>DNA for RVC repeated sequence</td> <td><i>Vibrio cholerae</i></td> <td>176</td> <td>85.56</td> </tr> <tr> <td>AF-A0A7Y9VDV2-F1</td> <td>HEPN domain-containing protein</td> <td><i>Pseudomonas sp. H1F10D</i></td> <td>162</td> <td>82.06</td> </tr> <tr> <td>AF-A0A7Y0VRK5-F1</td> <td>Uncharacterized protein</td> <td><i>Pseudomonas sp. WS 5413</i></td> <td>166</td> <td>88.31</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A6N4DAB7-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio vulnificus</i></td> <td>162</td> <td>89.25</td> </tr> <tr> <td>AF-A0A7Y1MPS7-F1</td> <td>Uncharacterized protein</td> <td><i>Pseudomonas gessardii</i></td> <td>168</td> <td>89.19</td> </tr> <tr> <td>AF-A0A7G8YDR9-F1</td> <td>Uncharacterized protein</td> <td><i>Pseudomonas protegens</i></td> <td>162</td> <td>88.44</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-Q08796-F1	DNA for RVC repeated sequence	<i>Vibrio cholerae</i>	176	85.56	AF-A0A7Y9VDV2-F1	HEPN domain-containing protein	<i>Pseudomonas sp. H1F10D</i>	162	82.06	AF-A0A7Y0VRK5-F1	Uncharacterized protein	<i>Pseudomonas sp. WS 5413</i>	166	88.31	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A6N4DAB7-F1	Uncharacterized protein	<i>Vibrio vulnificus</i>	162	89.25	AF-A0A7Y1MPS7-F1	Uncharacterized protein	<i>Pseudomonas gessardii</i>	168	89.19	AF-A0A7G8YDR9-F1	Uncharacterized protein	<i>Pseudomonas protegens</i>	162	88.44
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-Q08796-F1	DNA for RVC repeated sequence	<i>Vibrio cholerae</i>	176	85.56																																						
AF-A0A7Y9VDV2-F1	HEPN domain-containing protein	<i>Pseudomonas sp. H1F10D</i>	162	82.06																																						
AF-A0A7Y0VRK5-F1	Uncharacterized protein	<i>Pseudomonas sp. WS 5413</i>	166	88.31																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A6N4DAB7-F1	Uncharacterized protein	<i>Vibrio vulnificus</i>	162	89.25																																						
AF-A0A7Y1MPS7-F1	Uncharacterized protein	<i>Pseudomonas gessardii</i>	168	89.19																																						
AF-A0A7G8YDR9-F1	Uncharacterized protein	<i>Pseudomonas protegens</i>	162	88.44																																						

Foldseek (AFDB-Proteome)

Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment
AF-Q2FZX1-F1-model_v4	Uncharacterized protein	Staphylococcus aureus subsp. aure...	0.54	4.9	0.564		
AF-A5Z2V6-F1-model_v4	Uncharacterized protein	Caenorhabditis elegans	0.28	9.2	0.448		
AF-C0NH73-F1-model_v4	HHE domain-containing protein	Histoplasma capsulatum G186AR	0.21	3	0.400		

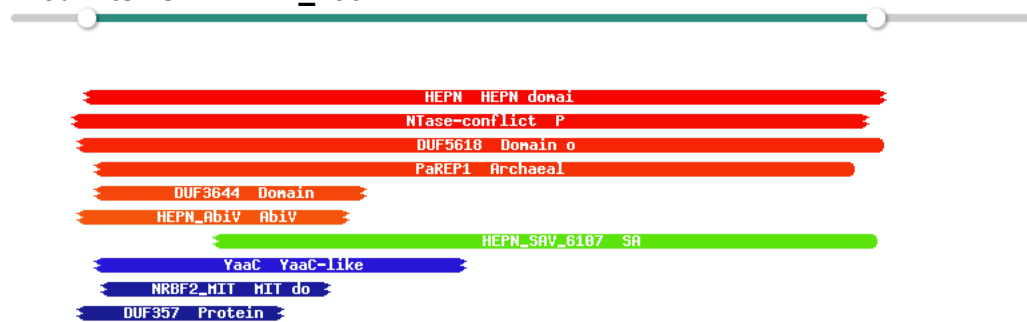
Foldseek (AFDB-SwissProt)

Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment
AF-P00153-F1-model_v4	Cytochrome c'	Rhodocyclus tenuis	0.20	4	0.444		
AF-P75925-F1-model_v4	Cytochrome b561 homolog 2	Escherichia coli K-12	0.25	4.5	0.417		
AF-Q8RGY9-F1-model_v4	Na(+)/H(+) antiporter subunit G	Alkalihalophilus pseudofirmus OF4	0.09	2.7	0.351		

Foldseek (PDB100)

Target	Description	Scientific Name	Prob.	Seq. id.	TM-score	Position in query	Alignment
4npf-assembly1_A	Crystal structure of HEPN domain protein	Agrobacterium fabrum str. C58	0.63	11.9	0.604		
7bxo-assembly1_D	Crystal structure of the toxin-antitoxin with AMP-PNP	Shewanella oneidensis MR-1	0.44	5.6	0.564		
7bxo-assembly1_E	Crystal structure of the toxin-antitoxin with AMP-PNP	Shewanella oneidensis MR-1	0.41	5.8	0.548		

HHPred hits vs PFAM-A_v36



InterPro/PFAM hits (UniProt): None

Miscellaneous information

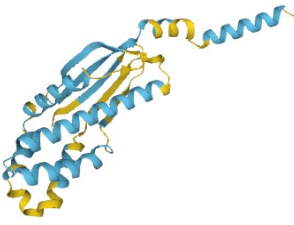
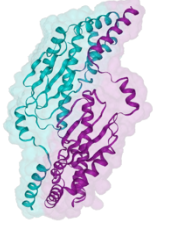
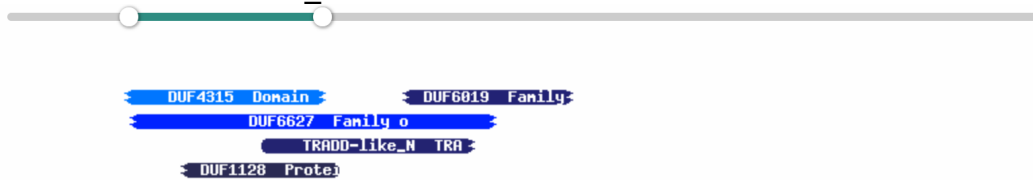
HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domains are present in several defense systems, especially in the Abi family (Anantharaman et al., 2013). Of note, several systems containing a HEPN domain were recently identified in phage-inducible chromosomal islands (PICIs) of *Klebsiella* (Fillol-Salom et al., 2022) as well as in P4-like satellites of *E.coli* (Rousset et al., 2022).

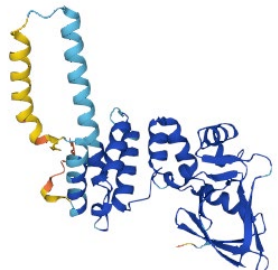
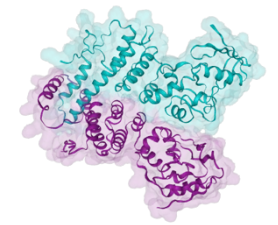

Selected references

Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol Direct.* 2013;8:15. Published 2013 Jun 15. doi:10.1186/1745-6150-8-15

Fillol-Salom A, Rostøl JT, Ojiogu AD, et al. Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. *Cell.* 2022;185(17):3248-3262.e20. doi:10.1016/j.cell.2022.07.014

Rousset F, Depardieu F, Miele S, et al. Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe.* 2022;30(5):740-753.e5. doi:10.1016/j.chom.2022.02.018

<h1>Belenos</h1>	Uncharacterized protein VCA0457 / Uniprot ID Q9KMB6	Length 204 aa																																
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.84)</p> 	AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)																																	
	<table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0X1KZR1-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio cholerae MO10</i></td> <td>204</td> <td>73.38</td> </tr> <tr> <td>AF-Q9KMB6-F1</td> <td>Uncharacterized protein</td> <td><i>Vibrio cholerae serotype O1</i></td> <td>204</td> <td>72.5</td> </tr> <tr> <td>AF-A0A4R3XIZ5-F1</td> <td>Uncharacterized protein</td> <td><i>Methylomonas methanica</i></td> <td>204</td> <td>71.38</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A0X1KZR1-F1	Uncharacterized protein	<i>Vibrio cholerae MO10</i>	204	73.38	AF-Q9KMB6-F1	Uncharacterized protein	<i>Vibrio cholerae serotype O1</i>	204	72.5	AF-A0A4R3XIZ5-F1	Uncharacterized protein	<i>Methylomonas methanica</i>	204	71.38												
AFDB acc.	Description	Species	Length	pLDDT																														
AF-A0A0X1KZR1-F1	Uncharacterized protein	<i>Vibrio cholerae MO10</i>	204	73.38																														
AF-Q9KMB6-F1	Uncharacterized protein	<i>Vibrio cholerae serotype O1</i>	204	72.5																														
AF-A0A4R3XIZ5-F1	Uncharacterized protein	<i>Methylomonas methanica</i>	204	71.38																														
	AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)																																	
	<table border="1"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A2W1J6J0-F1</td> <td>Uncharacterized protein</td> <td><i>Acaryochloris sp. RCC1774</i></td> <td>189</td> <td>82.44</td> </tr> <tr> <td>AF-A0A2G2Q9Y0-F1</td> <td>Uncharacterized protein</td> <td><i>Methylophaga spp</i></td> <td>207</td> <td>76.81</td> </tr> <tr> <td>AF-A0A2V2R8V1-F1</td> <td>Uncharacterized protein</td> <td><i>Acidobacteria spp</i></td> <td>186</td> <td>75.44</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A2W1J6J0-F1	Uncharacterized protein	<i>Acaryochloris sp. RCC1774</i>	189	82.44	AF-A0A2G2Q9Y0-F1	Uncharacterized protein	<i>Methylophaga spp</i>	207	76.81	AF-A0A2V2R8V1-F1	Uncharacterized protein	<i>Acidobacteria spp</i>	186	75.44												
AFDB acc.	Description	Species	Length	pLDDT																														
AF-A0A2W1J6J0-F1	Uncharacterized protein	<i>Acaryochloris sp. RCC1774</i>	189	82.44																														
AF-A0A2G2Q9Y0-F1	Uncharacterized protein	<i>Methylophaga spp</i>	207	76.81																														
AF-A0A2V2R8V1-F1	Uncharacterized protein	<i>Acidobacteria spp</i>	186	75.44																														
Foldseek (AFDB-Proteome)																																		
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-A0A0N4UB57-F1-model_v4</td> <td>Uncharacterized protein</td> <td>Dracunculus medinensis</td> <td>0.11</td> <td>4.2</td> <td>0.393</td> <td></td> <td></td> </tr> <tr> <td>AF-A0A1E1JKH1-F1-model_v4</td> <td>Uncharacterized protein</td> <td>Caenorhabditis elegans</td> <td>0.09</td> <td>3</td> <td>0.384</td> <td></td> <td></td> </tr> <tr> <td>AF-P0A0V0-F1-model_v4</td> <td>LPP20 lipoprotein</td> <td>Helicobacter pylori 26695</td> <td>0.14</td> <td>3.6</td> <td>0.384</td> <td></td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-A0A0N4UB57-F1-model_v4	Uncharacterized protein	Dracunculus medinensis	0.11	4.2	0.393			AF-A0A1E1JKH1-F1-model_v4	Uncharacterized protein	Caenorhabditis elegans	0.09	3	0.384			AF-P0A0V0-F1-model_v4	LPP20 lipoprotein	Helicobacter pylori 26695	0.14	3.6	0.384		
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																											
AF-A0A0N4UB57-F1-model_v4	Uncharacterized protein	Dracunculus medinensis	0.11	4.2	0.393																													
AF-A0A1E1JKH1-F1-model_v4	Uncharacterized protein	Caenorhabditis elegans	0.09	3	0.384																													
AF-P0A0V0-F1-model_v4	LPP20 lipoprotein	Helicobacter pylori 26695	0.14	3.6	0.384																													
Foldseek (AFDB-SwissProt)																																		
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-P36586-F1-model_v4</td> <td>GTP-binding protein ypt5</td> <td>Schizosaccharomyces pombe 972h-</td> <td>0.10</td> <td>3.2</td> <td>0.313</td> <td></td> <td></td> </tr> <tr> <td>AF-B4HR78-F1-model_v4</td> <td>Protein kintoun</td> <td>Drosophila sechellia</td> <td>0.10</td> <td>3.1</td> <td>0.206</td> <td></td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-P36586-F1-model_v4	GTP-binding protein ypt5	Schizosaccharomyces pombe 972h-	0.10	3.2	0.313			AF-B4HR78-F1-model_v4	Protein kintoun	Drosophila sechellia	0.10	3.1	0.206										
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																											
AF-P36586-F1-model_v4	GTP-binding protein ypt5	Schizosaccharomyces pombe 972h-	0.10	3.2	0.313																													
AF-B4HR78-F1-model_v4	Protein kintoun	Drosophila sechellia	0.10	3.1	0.206																													
Foldseek (PDB100)																																		
<table border="1"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>5ok8-assembly1_A</td> <td>Crystal structure of protein Lpp20 (HP1456) from Helico...</td> <td>Helicobacter pylori J99</td> <td>0.12</td> <td>2.6</td> <td>0.427</td> <td></td> <td></td> </tr> <tr> <td>2av5-assembly1_A</td> <td>Crystal structure of Pyrococcus furiosus Pop5, an archa...</td> <td>Pyrococcus furiosus</td> <td>0.09</td> <td>1.7</td> <td>0.389</td> <td></td> <td></td> </tr> <tr> <td>1re5-assembly1_D</td> <td>Crystal structure of 3-carboxy-cis-cis-muconate lactonizi...</td> <td>Pseudomonas putida KT2440</td> <td>0.35</td> <td>2.4</td> <td>0.357</td> <td></td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	5ok8-assembly1_A	Crystal structure of protein Lpp20 (HP1456) from Helico...	Helicobacter pylori J99	0.12	2.6	0.427			2av5-assembly1_A	Crystal structure of Pyrococcus furiosus Pop5, an archa...	Pyrococcus furiosus	0.09	1.7	0.389			1re5-assembly1_D	Crystal structure of 3-carboxy-cis-cis-muconate lactonizi...	Pseudomonas putida KT2440	0.35	2.4	0.357		
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																											
5ok8-assembly1_A	Crystal structure of protein Lpp20 (HP1456) from Helico...	Helicobacter pylori J99	0.12	2.6	0.427																													
2av5-assembly1_A	Crystal structure of Pyrococcus furiosus Pop5, an archa...	Pyrococcus furiosus	0.09	1.7	0.389																													
1re5-assembly1_D	Crystal structure of 3-carboxy-cis-cis-muconate lactonizi...	Pseudomonas putida KT2440	0.35	2.4	0.357																													
HHPred hits vs PFAM-A_v36																																		
 <p>DUF4315 Domain DUF6019 Family</p> <p>DUF6627 Family o</p> <p>TRADD-like_N TRA</p> <p>DUF1128 Prote</p>																																		
InterPro/PFAM hits (UniProt): None																																		
Miscellaneous information																																		
Selected references																																		

<h1>Belisama</h1>	DUF2806 domain-containing protein VCA0458 / Uniprot ID Q9KMB5	Length 285 aa																																								
<p>AlphaFold structure</p>  <p>Cofold (2 subunits, ipTM = 0.60)</p> 	<p>AFDB50/MMseqs2 (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 241 1468 470"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A7T9JX00-F1</td> <td>DUF2806 domain-containing protein</td> <td><i>Candidatus Peregrinibacteria bacterium</i></td> <td>288</td> <td>91.38</td> </tr> <tr> <td>AF-A0A7U1AIZ1-F1</td> <td>DUF2806 domain-containing protein</td> <td><i>Tamlana sp. s12</i></td> <td>285</td> <td>91.19</td> </tr> <tr> <td>AF-A0A1C0TNP1-F1</td> <td>TIGR03899 family protein</td> <td><i>Pseudoalteromonas luteoviolacea</i></td> <td>288</td> <td>90.38</td> </tr> </tbody> </table> <p>AFDB/Foldseek (Top 3 pLDDT not “uncharacterized”)</p> <table border="1" data-bbox="509 526 1468 728"> <thead> <tr> <th>AFDB acc.</th> <th>Description</th> <th>Species</th> <th>Length</th> <th>pLDDT</th> </tr> </thead> <tbody> <tr> <td>AF-A0A8A7F9H1-F1</td> <td>TIGR03899 family protein</td> <td><i>Pseudoalteromonas viridis</i></td> <td>293</td> <td>90.5</td> </tr> <tr> <td>AF-K1ZJM5-F1</td> <td>Membrane-fusion protein</td> <td>Unknown (uncultured bacterium)</td> <td>287</td> <td>90.44</td> </tr> <tr> <td>AF-A0A7V1CWC6-F1</td> <td>TIGR03899 family protein</td> <td><i>Pseudoalteromonas prydzensis</i></td> <td>295</td> <td>90.38</td> </tr> </tbody> </table>		AFDB acc.	Description	Species	Length	pLDDT	AF-A0A7T9JX00-F1	DUF2806 domain-containing protein	<i>Candidatus Peregrinibacteria bacterium</i>	288	91.38	AF-A0A7U1AIZ1-F1	DUF2806 domain-containing protein	<i>Tamlana sp. s12</i>	285	91.19	AF-A0A1C0TNP1-F1	TIGR03899 family protein	<i>Pseudoalteromonas luteoviolacea</i>	288	90.38	AFDB acc.	Description	Species	Length	pLDDT	AF-A0A8A7F9H1-F1	TIGR03899 family protein	<i>Pseudoalteromonas viridis</i>	293	90.5	AF-K1ZJM5-F1	Membrane-fusion protein	Unknown (uncultured bacterium)	287	90.44	AF-A0A7V1CWC6-F1	TIGR03899 family protein	<i>Pseudoalteromonas prydzensis</i>	295	90.38
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A7T9JX00-F1	DUF2806 domain-containing protein	<i>Candidatus Peregrinibacteria bacterium</i>	288	91.38																																						
AF-A0A7U1AIZ1-F1	DUF2806 domain-containing protein	<i>Tamlana sp. s12</i>	285	91.19																																						
AF-A0A1C0TNP1-F1	TIGR03899 family protein	<i>Pseudoalteromonas luteoviolacea</i>	288	90.38																																						
AFDB acc.	Description	Species	Length	pLDDT																																						
AF-A0A8A7F9H1-F1	TIGR03899 family protein	<i>Pseudoalteromonas viridis</i>	293	90.5																																						
AF-K1ZJM5-F1	Membrane-fusion protein	Unknown (uncultured bacterium)	287	90.44																																						
AF-A0A7V1CWC6-F1	TIGR03899 family protein	<i>Pseudoalteromonas prydzensis</i>	295	90.38																																						
<p>Foldseek (AFDB-Proteome)</p> <table border="1" data-bbox="127 869 1466 929"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>AF-A0A132Z9X2-F1-model_v4</td> <td>Uncharacterized protein</td> <td><i>Enterococcus faecalis</i></td> <td>0.30</td> <td>6.4</td> <td>0.511</td> <td>15-238</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	AF-A0A132Z9X2-F1-model_v4	Uncharacterized protein	<i>Enterococcus faecalis</i>	0.30	6.4	0.511	15-238																									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
AF-A0A132Z9X2-F1-model_v4	Uncharacterized protein	<i>Enterococcus faecalis</i>	0.30	6.4	0.511	15-238																																				
<p>Foldseek (AFDB-SwissProt)</p>																																										
<p>No hit</p>																																										
<p>Foldseek (PDB100)</p> <table border="1" data-bbox="127 1093 1466 1153"> <thead> <tr> <th>Target</th> <th>Description</th> <th>Scientific Name</th> <th>Prob.</th> <th>Seq. Id.</th> <th>TM-score</th> <th>Position in query</th> <th>Alignment</th> </tr> </thead> <tbody> <tr> <td>7upp-assembly1_A</td> <td>Crystal structure of designed heterotrimeric assembly D...</td> <td><i>synthetic construct</i></td> <td>0.12</td> <td>3.5</td> <td>0.329</td> <td>10-284</td> <td></td> </tr> </tbody> </table>			Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment	7upp-assembly1_A	Crystal structure of designed heterotrimeric assembly D...	<i>synthetic construct</i>	0.12	3.5	0.329	10-284																									
Target	Description	Scientific Name	Prob.	Seq. Id.	TM-score	Position in query	Alignment																																			
7upp-assembly1_A	Crystal structure of designed heterotrimeric assembly D...	<i>synthetic construct</i>	0.12	3.5	0.329	10-284																																				
<p>HHPred hits vs PFAM-A_v36</p> 																																										
<p>InterPro/PFAM hits (UniProt) IPR021254 (DUF2806) ; PF10987 (DUF2806)</p>																																										
<p>Miscellaneous information</p> <p>It is not part of the systems detected by DefenseFinder but does hit the PFAM HMM of Abiα (PF14337, formerly DUF4393). HHPred finds similarity between Abiα and DUF2806 (PF10987, found in Belisama) mentioned by the discoverers of Abiα. Abiα is an Abi system mostly found in prophages (Lossouarn et al., 2019).</p>																																										
<p>Selected references</p> <p>Lossouarn J, Briet A, Moncaut E, et al. Enterococcus faecalis Countermeasures Defeat a Virulent Picovirinae Bacteriophage. <i>Viruses</i>. 2019;11(1):48. Published 2019 Jan 10. doi:10.3390/v11010048</p>																																										

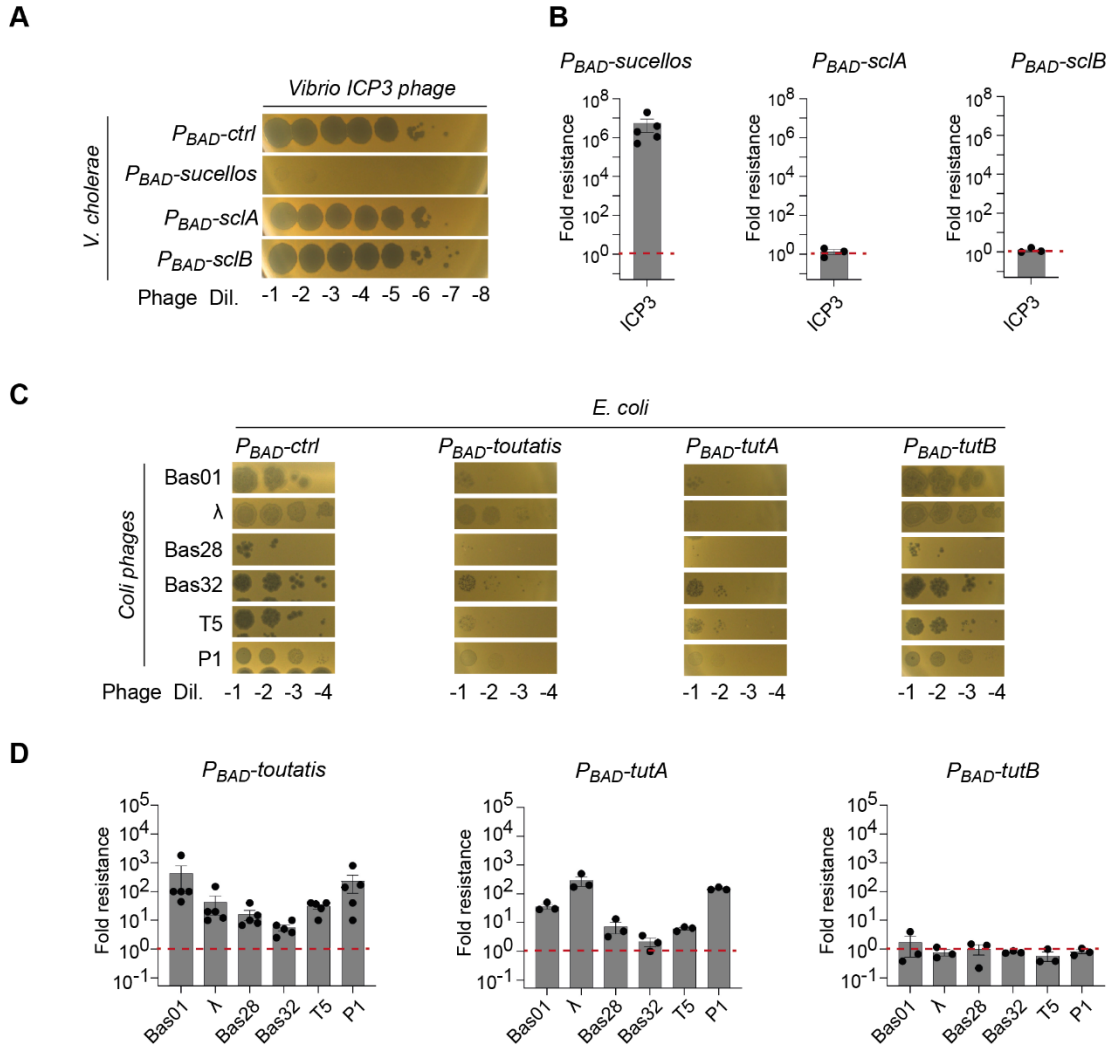


Figure S5: Detailed fold resistance of Sucellos and Toutatis defence systems

A and C. 10-fold serial dilutions of a high titer lysate of phages spotted on *V. cholerae* (**A**) and *E. coli* (**C**) strains harboring a plasmid expressing the defence systems or not ($P_{BAD-ctrl}$). Representative images from a single replicate out of independent replicates are shown.

B and D. Plaque-forming units were measured for each phage on *V. cholerae* (**B**) or *E. coli* (**D**) strains harboring the control plasmid or the defence systems. Fold-resistance was measured as the ratio between these two values. The bar charts show the mean fold resistance for at least three independent replicates ($n \geq 3$, individual plots). Error bars shown the standard error of mean (SEM).

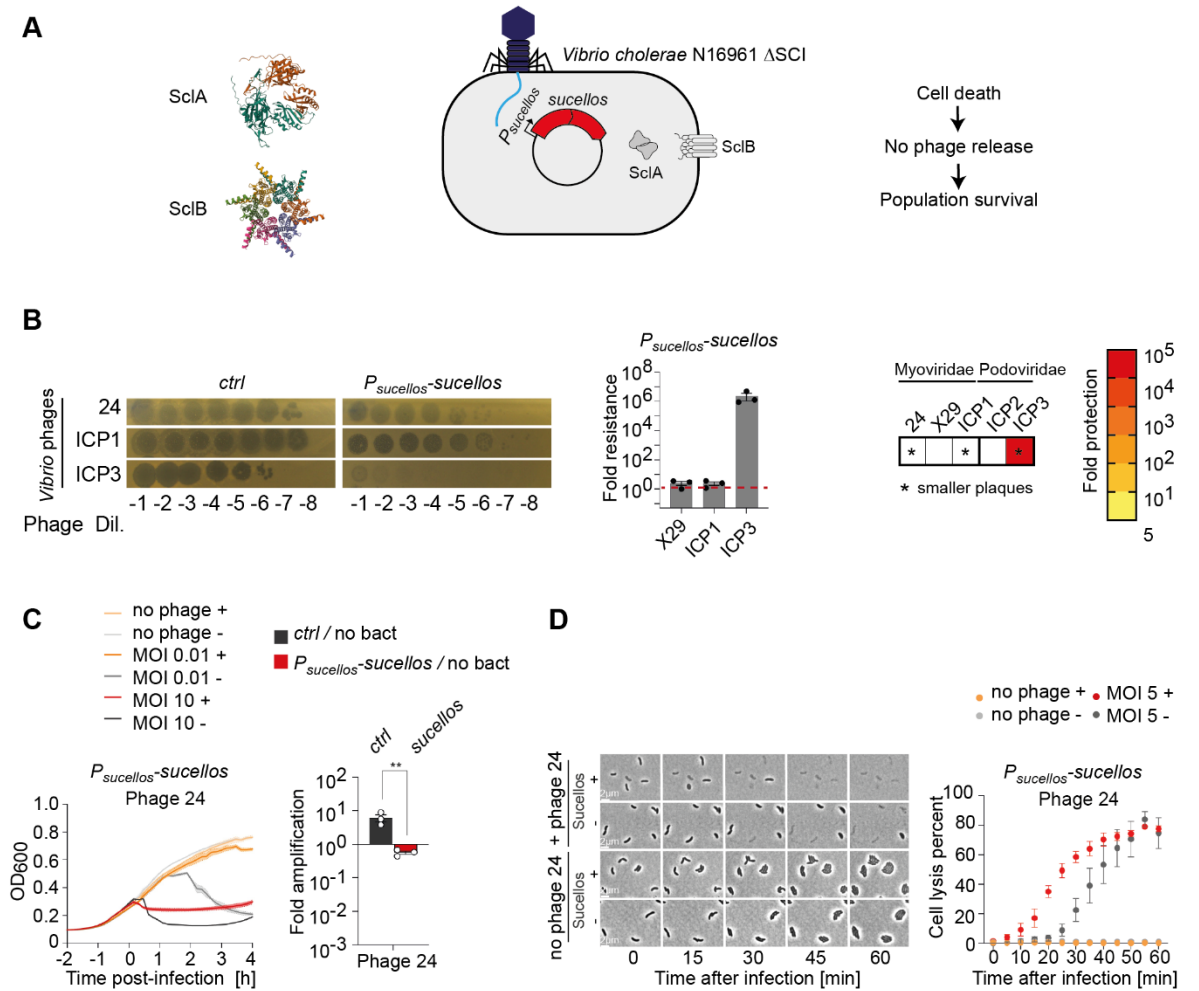


Figure S6: Analysis of the Sucellos system

A. AlphaFold structure prediction of the SclA and SclB proteins and model for the defence activity.

B. 10-fold serial dilutions of a high titer lysate of phages spotted on *V. cholerae* strains. Representative images from a single replicate out of three independent replicates are shown (top panel). Plaque-forming units were measured for each phage on the *V. cholerae* strains harbouring a plasmid expressing *sucellos* (*P_{sucellos-sucellos}*) or not (*ctrl*). Fold-resistance was measured as the ratio between these two values. The bar charts show the mean fold resistance for three independent replicates ($n=3$, individual plots). Error bars shown the standard error of mean (SEM) (bottom right panel). Heatmap of the fold resistance phenotype representing the mean fold resistance of the three independent replicates. Phages are indicated and classed according to morphology. * stands for a marked reduction in plaque size.

C. As in Figure 4A and B.

D. Analysis using phase contrast microscopy at 0 min and up to 60 min after infection with phage 24 at an MOI of 5. Cell lysis percent was analyzed every 5 min. Each dot represents the mean of three independent experiments and error bars show the standard error of mean (SEM). The top row shows *V. cholerae* cells containing the anti-phage system and the bottom one without the anti-phage system. Just after phage adding, phage-mediated cell lysis is observed in cells in which the system is present. Representative images from a single replicate out of three independent replicates are shown.

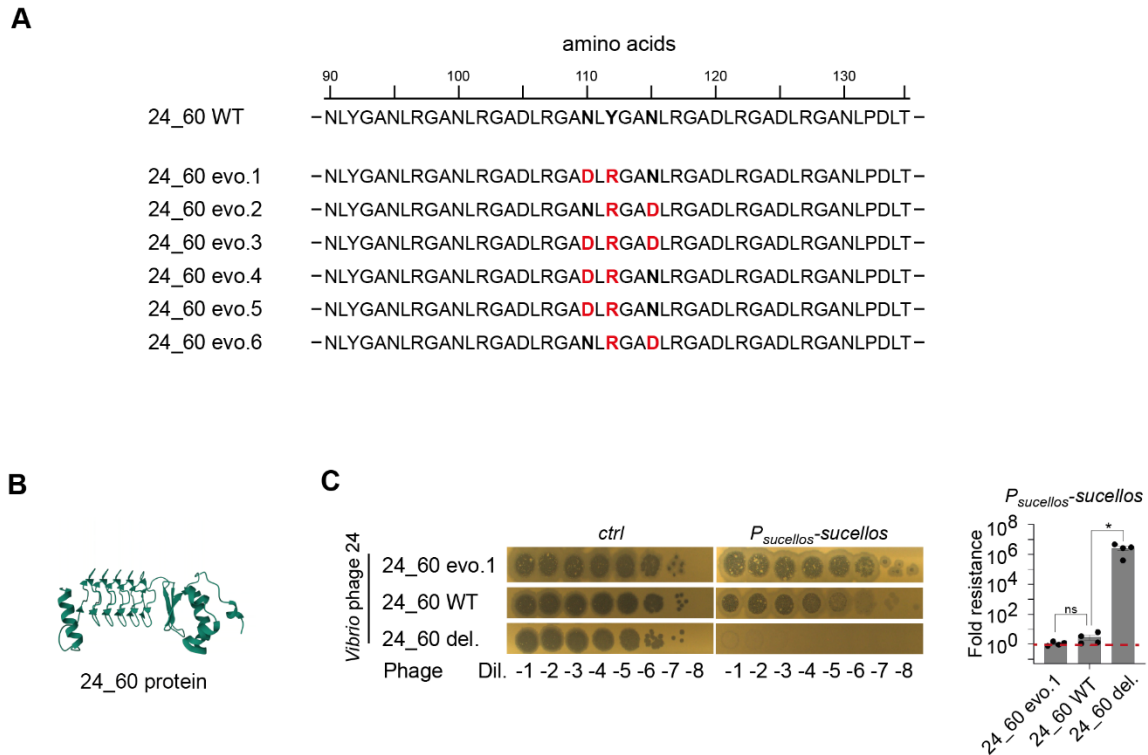


Figure S7: Analysis of the function of the 24_60 phage protein

A. The 24_60 amino acid sequences of the *wt* phage 24 and the 6 escapers phages are shown with mutated residues in red.

B. AlphaFold structure prediction of the 24_60 protein

C. 10-fold serial dilutions of a high titer lysate of phages 24 spotted on *V. cholerae* strains. Representative images from a single replicate out of three independent replicates are shown (left panel). Plaque-forming units were measured for each phage 24 on the *V. cholerae* strains harbouring a plasmid expressing *sucellos* (*P_{sucellos}-sucellos*) or not (*P_{sucellos}-ctrl*). Fold-resistance was measured as the ratio between these two values. The bar charts show the mean fold resistance for four independent replicates (n=4, individual plots). Error bars shown the standard error of mean (SEM) (right panel). Statistical comparisons (Student's t-test, two-tailed) are as follow: ns, not significant; *Pvalue < 0.05.

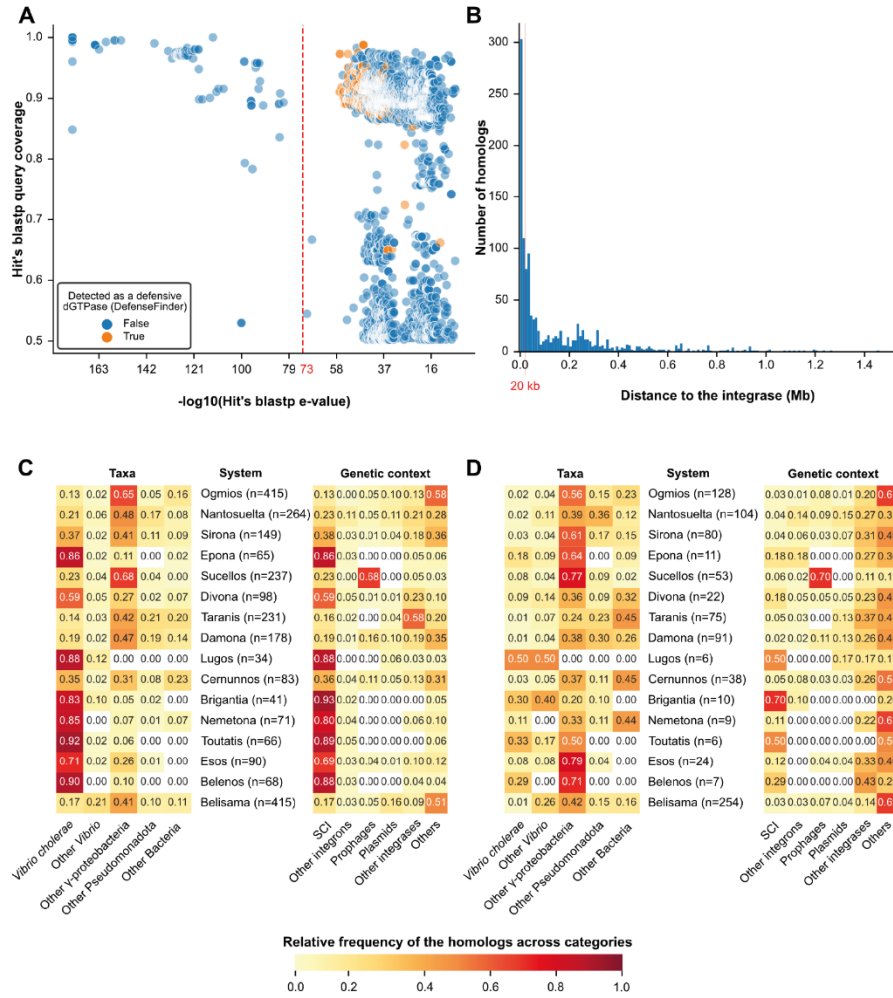


Figure S8. Detailed analysis of the frequency and locations of the newly discovered anti-phage defence systems

A. Scatter plot of the homologs of Ogmios identified with blastp. All the hits with a query coverage greater than 0.5 and an e-value smaller than 10^{-5} (initial thresholds) are displayed. The homologs classified as a defensive dGTPase by DefenseFinder are indicated in orange. The e-value threshold specifically selected for Ogmios homologs is indicated as a red vertical line (10^{-7.3}).

B. Histogram of the smallest distance of a system homolog to an integrase. The distance has been computed for the 1175 homologs that were not found to be located in integrons, prophages nor plasmids. A vertical red line indicates a *distance of less than 20 kb, retained as the threshold for being close to an integrase.

C. Detailed version of Figure 5B. Heatmaps of the relative frequency of the novel defence systems across taxa (left) and genetic contexts (right). A number indicates the rounded value of the frequency in each square. The scale of colours is given in the bottom of the graph. The number of systems homologs is indicated after the systems' names.

D. Similar analysis as **C**, except that the homologs were clustered with an identity threshold of 95% (coverage threshold of 80%) to remove redundancy. The number of unique systems homologs (at these thresholds) is indicated after the systems' names.